


Major Depressive Disorder Classification Based on Different Convolutional Neural Network Models: Deep Learning Approach

Clinical EEG and Neuroscience
1–14
© EEG and Clinical Neuroscience
Society (ECNS) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1550059420916634
journals.sagepub.com/home/eeg


Caglar Uyulan¹ , Türker Tekin Ergüzel² , Huseyin Unubol^{3,4}, Merve Cebi^{3,4}, Gokben Hizli Sayar^{3,4}, Mehdi Nezhadas⁵, and Nevzat Tarhan^{3,4}

Abstract

The human brain is characterized by complex structural, functional connections that integrate unique cognitive characteristics. There is a fundamental hurdle for the evaluation of both structural and functional connections of the brain and the effects in the diagnosis and treatment of neurodegenerative diseases. Currently, there is no clinically specific diagnostic biomarker capable of confirming the diagnosis of major depressive disorder (MDD). Therefore, exploring translational biomarkers of mood disorders based on deep learning (DL) has valuable potential with its recently underlined promising outcomes. In this article, an electroencephalography (EEG)-based diagnosis model for MDD is built through advanced computational neuroscience methodology coupled with a deep convolutional neural network (CNN) approach. EEG recordings are analyzed by modeling 3 different deep CNN structure, namely, ResNet-50, MobileNet, Inception-v3, in order to dichotomize MDD patients and healthy controls. EEG data are collected for 4 main frequency bands (Δ , θ , α , and β , accompanying spatial resolution with location information by collecting data from 19 electrodes. Following the pre-processing step, different DL architectures were employed to underline discrimination performance by comparing classification accuracies. The classification performance of models based on location data, MobileNet architecture generated 89.33% and 92.66% classification accuracy. As to the frequency bands, delta frequency band outperformed compared to other bands with 90.22% predictive accuracy and area under curve (AUC) value of 0.9 for ResNet-50 architecture. The main contribution of the study is the delineation of distinctive spatial and temporal features using various DL architectures to dichotomize 46 MDD subjects from 46 healthy subjects. Exploring translational biomarkers of mood disorders based on DL perspective is the main focus of this study and, though it is challenging, with its promising potential to improve our understanding of the psychiatric disorders, computational methods are highly worthy for the diagnosis process and valuable in terms of both speed and accuracy compared with classical approaches.

Keywords

convolutional neural network, deep learning, major depressive disorder, EEG signal processing, classification

Received May 14, 2019; revised February 23, 2020; accepted March 6, 2020.

Introduction

Major depressive disorder (MDD) is a kind of mental illness influencing the general behavioral, emotional and thinking abilities negatively. Depression symptoms comprise feeling sadness, loss of interest and concentration, having a physical dysfunction, insomnia, feeling guilty, making decisions difficulty, thoughts of suicide, and so forth. Depression varies as mild, moderate, or severe related to the severity of the symptoms. It can be treatable through psychotherapy, medical countermeasures, or electroconvulsive therapy. However, the precondition of treating is that the necessity of a well-diagnosing process. There exist obstructions of diagnosing MDD. Symptoms must last at least 2 weeks; the symptoms of depression are confused with other medical conditions (eg, thyroid problems, brain tumour, or vitamin deficiency). Depression-related conditions are peripartum or seasonal depression, persistent depressive disorder, premenstrual dysphoric

disorder, disruptive mood dysregulation disorder, and bipolar disorders.^{1–4}

According to the literature research, in depressed people, the quantity of neurotransmitter release and concentration of

¹Department of Mechatronics, Faculty of Engineering, Bulent Ecevit University, Zonguldak, Turkey

²Department of Software Engineering, Faculty of Engineering and Natural Sciences, Uskudar University, Istanbul, Turkey

³Department of Psychology, Faculty of Humanities and Social Sciences, Uskudar University, Istanbul, Turkey

⁴NP Istanbul Brain Hospital, Istanbul, Turkey

⁵Yildiz Technical University, Istanbul, Turkey

Corresponding Author:

Türker Tekin Ergüzel, Department of Software Engineering, Faculty of Engineering and Natural Sciences, Uskudar University, Altunizade Mah. Haluk Turksoy Sk. No:14, Uskudar, Istanbul, 34662, Turkey.
Email: turker.erguzel@uskudar.edu.tr

receptors in the synapse is lower as compared with healthy ones. The neuroimaging tools, which record the activity of the brain, have biomarker value in affective disorders. EEG, which is one of these devices, is more preferred than other tools because it is noninvasive, economical, practical, and easy to operate. The accuracy, robustness, and reliability of the EEG-related methodology combined with deep learning (DL) has proven with many research in brain-computer interface (BCI), especially in motor imagery task classification,⁵⁻¹³ epileptic seizure prediction and detection,¹⁴⁻²³ drivers fatigue prediction,^{24,25} emotion and affective state classification,²⁶⁻³² sleep stage detection,³³⁻³⁶ prognosis in rapid eye movement behavior disorder,³⁷ EEG-based diagnosis of various neurodegenerative diseases, including attention deficit/hyperactivity disorder,³⁸ schizophrenia,^{39,40} Creutzfeldt-Jacob disease,⁴¹ Parkinson's disease,⁴² Alzheimer's disease,⁴³ mild cognitive impairment,⁴⁴ predicting transcranial direct current stimulation treatment outcomes of patients with MDD has been studied in the recent literature.⁴⁵

The performance of a machine learning methodology based on the pretreatment EEG for MDD prognosis of treatment with a selective reuptake inhibitor medication,⁴⁶ and based on the prediction of antidepressant treatment response using EEG were investigated.⁴⁷ DL is a way to extract useful information or pattern from data in an automated manner by using multilayered neural networks and optimization. The practical implementation can be done through the libraries in Python, TensorFlow, PyTorch, Caffe. The hard part of the methodology is to collect quality data and ask pragmatical questions about real problems. Nowadays, DL research continues with increasing acceleration due to the digitization of information and the ability to access data in a distributive fashion, substantial-scale executable graphics processing unit (GPU) hardware technology and robust algorithms and tools that enable a person to solve the higher level of theoretical problems as fast as possible. In recent years, many breakthroughs like AlphaGo, self-driving cars, speech detection, face recognition, text-to-speech generation, handwriting transcription, medical diagnosis, digital assistants, social recommendations are mainly originated from DL methods.⁴⁸⁻⁵¹ DL covers many different algorithms such as long- and short-term memory, convolutional neural network (CNNs), recurrent neural networks, generative adversarial networks, deep Boltzmann machine, deep belief networks, and so forth. CNNs have proved their superiority on computer vision, medical image analysis, image classification, semantic segmentation.⁵²⁻⁵⁸ Besides, recent studies underline the impact and remarkable contribution of DL applications in neuroscience using neuroimaging data. EEG signal has complex, chaotic, and non-stationary nature. Thus, the differences between healthy and depressed persons cannot be observed through graphical representations in the time domain, and intrinsically a computer-aided methodology may reveal the varieties by using pattern analysis.

A data mining methodology was developed to classify EEGs of 53 MDD patients and 43 healthy volunteers (HVs). Linear discriminant analysis (LDA) mapped the features into a new

feature space, and genetic algorithm (GA) identified the most significant features. After that, a predictive model through a decision tree algorithm was built to reveal rules and hidden patterns based on the reduced and mapped features. Application of LDA and GA markedly reduced the total number of utilized features, and the model demonstrated average classification accuracy (MDD vs HV) of 80%. The best results from model testing with additional test EEG recordings from 9 MDD patients and 35 HV individuals showed an accuracy of 80% with an average sensitivity of 70%, a specificity of 76%, and a positive and negative predictive value of 74% and 75%, respectively.⁵⁹ An experiment was conducted using EEG data collected 37 subjects from a 128 channel HydroCel Geodesic Sensor Net. MDD patients and healthy controls were discriminated applying support vector machine (SVM), Bayesian Network, Random Forest, Logistic Regression, k-nearest neighbor (KNN). Optimal performance was achieved utilizing a combination of feature selection method GreedyStepwise based on correlation features selection and KNN classifier for the beta frequency band. Accuracies achieved 92% and 98%, and area under the curve (AUC) achieved 0.957 and 0.997, respectively.⁶⁰

In this article, the contribution of the DL approach to classifying MDD subjects and healthy subjects were underlined. Raw EEG data collected from electrode channels were filtered into its 4 main frequency components in time domain, and the electrodes were spatially mapped as left or right hemisphere according to their locations. The preprocessed data, which corresponds to the frequency and spatial information, are converted to a spatiotemporal mesh, and the matrix was expressed as grayscale image data for each subject. Later the mesh was fed into CNN to construct an adaptively learning model.

Materials and Methods

Participants

This research was conducted at the Neuropsychiatry Istanbul Hospital to determine the ability of Quantitative EEG (QEEG) to classify MDD subjects and healthy controls. This retrospective study was formally approved by the local medical research ethics committee. Participants first visited a psychiatrist to determine whether they met the inclusion criteria. All subjects had abstained from psychotropic medications for ≥ 2 weeks before enrollment. Subjects with nonpsychotic depressive disorder, as defined by the International Statistical Classification of Diseases and Related Health Problems criteria, and determined using the 17-item Hamilton Depression Rating Scale (HAM-D; score >14) were eligible. Diagnosis of recurrent MDD and absence of any other mental disorder was confirmed by using Composite International Diagnostic Interview (CIDI): CIDI is a standardized, fully-structured interview developed by the World Health Organization (WHO) that provides current and lifetime ICD-10 (International Classification of Diseases, 10th Revision) and DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition) diagnoses.⁶¹ It is widely used as a research diagnostic interview and has been adapted to

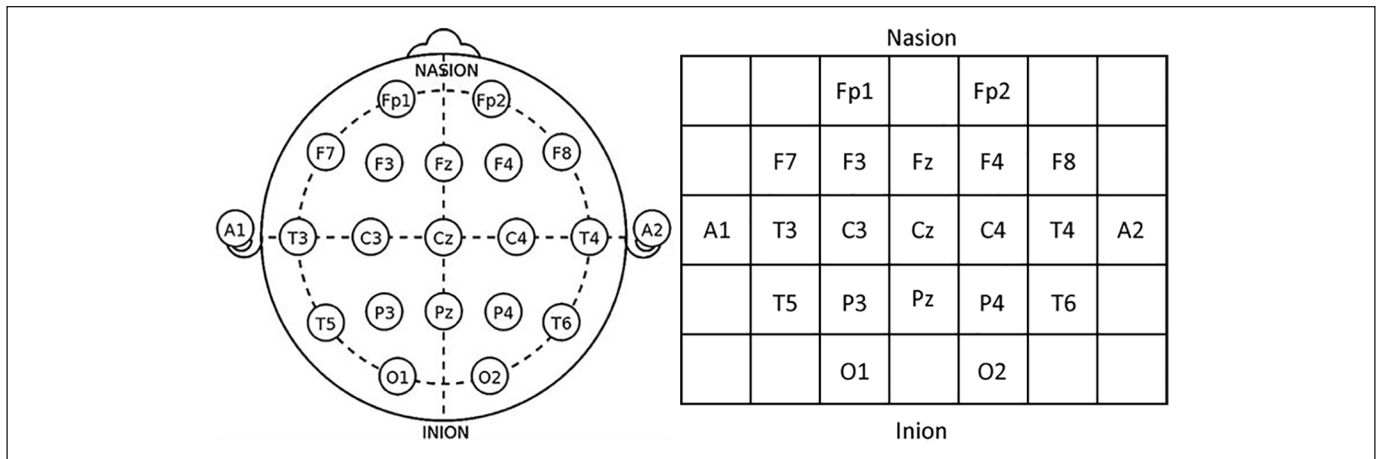


Figure 1. Electrode placement and mesh matrix representing the 10-20 placement system for a 19-electrode cap.

Turkish population.⁶² A total of 46 patients aged between 22 and 51 years (average age 39.3 years) with MDD, resistant to medication treatment, completed the protocols and were examined for the study. On the day before treatment a baseline clinical assessment was conducted by a psychiatrist using the HAM-D. During the study, patients were assessed clinically, neuropsychologically, and QEEG. Laboratory studies (complete blood count, chemistry, and thyroid-stimulating hormone), urine toxicology screening, and electrocardiography were performed at screening. Patients with organic brain disorders, pacemakers, psychotic symptoms, dementia, delirium, substance-related disorders, cluster A or B axis II disorders, electroconvulsive therapy in the prior 6 months, history of craniotomy, skull fracture, seizures, significant neurologic illness, suicidal intent, planning, or attempts were ineligible. The control group included 46 healthy subjects aged between 20 and 45 years (average age 37.1 years) who have no history of neurological or psychiatric disease and have not taken any drugs.

Data Acquisition and Preprocessing

EEG data used in the experiment collected from a total of 46 control and 46 depressed subjects. The EEG electrodes mounted in the neuro-headset, and the mesh matrix representing the 10-20 placement system for a 19 electrode cap is depicted in Figure 1.

EEG data were collected through “Neuroscan/ Scan LT” neuro-headset, which saves EEG signals from 19 channels of the headset, and the electrodes were marked at Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2, respectively. Samples were collected from 19 channels at a sample rate of 125 Hz. Signal dynamic range is ± 130 mV (22-bit analog to digital converter) with a bandwidth (3 dB down) DC to 262 Hz. The computer interface is provided through a universal serial bus. The notch filter of 50 Hz was applied to attenuate power line interference. The artifacts caused by muscle activity, eye movements, eye blink, and transmission line were eliminated from the EEG signal using the methodology inspired by

Cai et al.⁶³ Three band-stop filters whose cutoff frequencies are 3 Hz, 35 Hz, and 50 Hz were applied to the raw EEG signals.⁶⁴ Then the fast independent component analysis (FastICA) was performed evaluating the residual artefacts as an independent component for avoiding the time-consuming manual selection of artifacts. Independent component analysis (ICA) does not necessitate a specific montage, and only postulates the statistical independence and linear codependence of channels.⁶⁵

The bipolar EEG signal of each subject was acquired from the left hemisphere through Fp1, F7, F3, T3, C3, T5, P3, and O1 electrodes and from the right hemisphere via Fp2, F8, F4, T4, C4, T6, P4, and O2 electrodes while the subjects were awake but relaxed with eyes closed, for a duration 200 seconds after beginning of resting. The frequency components delta, theta, alpha, and beta were obtained by using 12th-order Butterworth bandpass filter. The lower and upper 3 dB cutoff frequencies are 8 to 12.5 Hz for the α band, 13 to 30 Hz for the β band, 4.5 to 7.5 Hz for the θ band, and 2 to 4 Hz for the Δ band. The first and second stopband attenuations were 0.0001, passband ripple was 0.0575, and density factor was selected as 20. Parks-McClellan optimal FIR (finite impulse response) filter algorithm, whose goal is to minimize the error in the passbands and stopbands by applying the Chebyshev approximation, was used in the implementation of the filters and the evaluation of the filter coefficients.^{64,65}

Image Construction From EEG Time Series

EEG signals are constituted from multiple time series measured from various spatial locations over the cortex. The spatial dimensions of the EEG signal representing brain regions are matched with electrode positions. The electrode measurements are transformed into a 2-dimensional (2D) mesh-dependent grayscale image data to preserve the internal structure of the data in the spatiotemporal coupled domain. The sequence of images derived from successively time windows represents the transient changes in brain activity. The location of electrodes is projected from a 3D space into a 2D

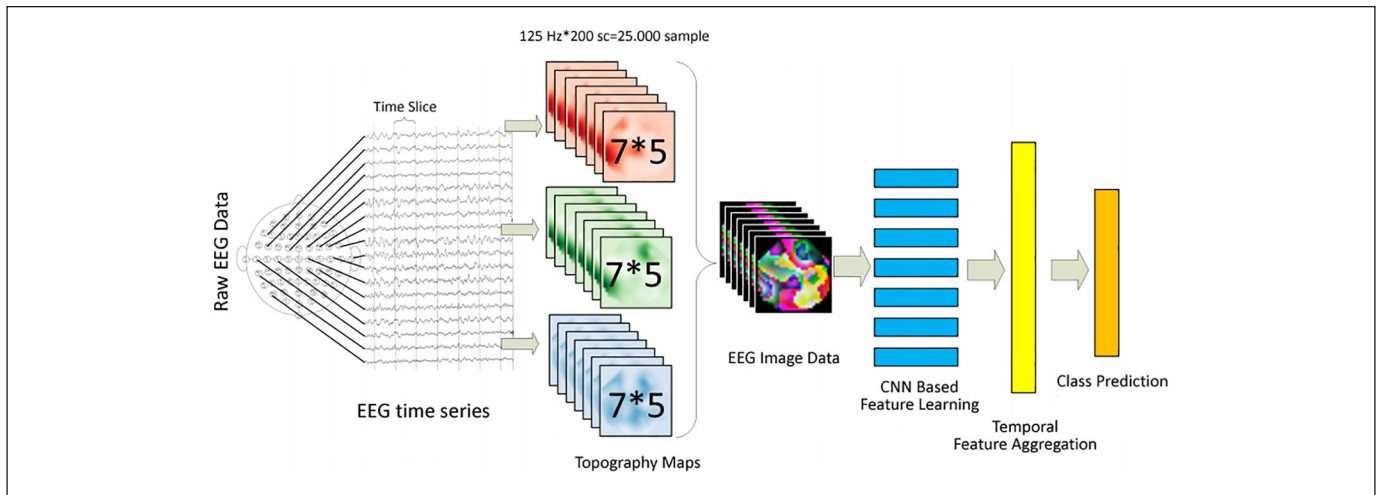


Figure 2. Flowchart of the procedure.

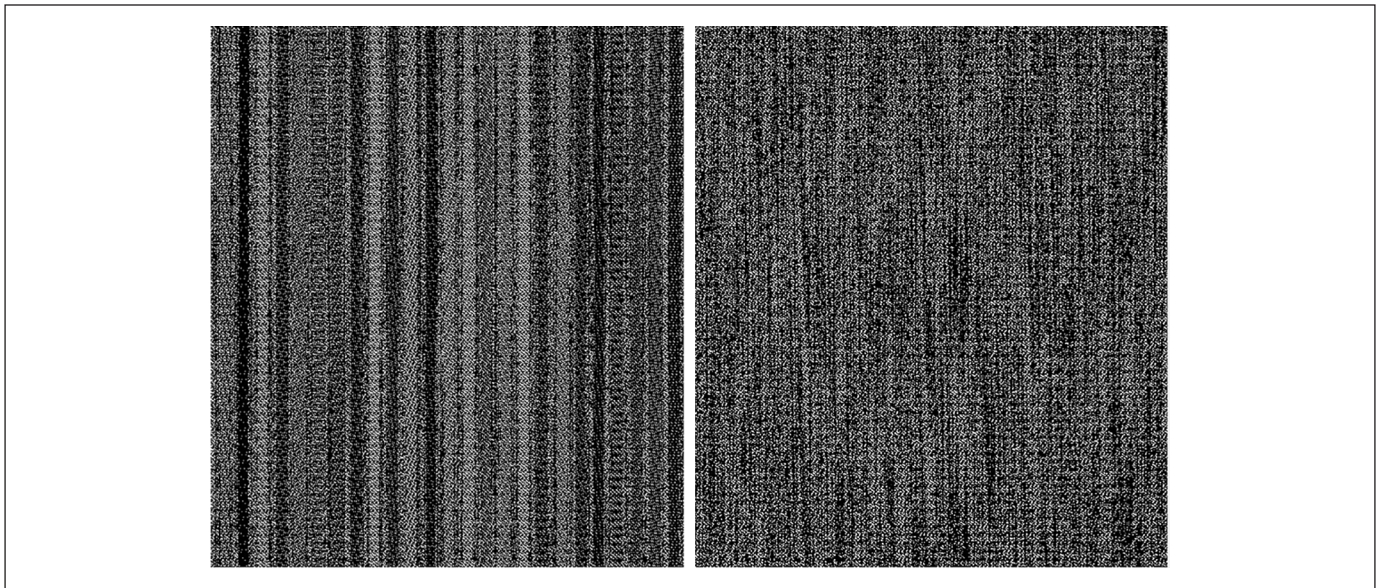


Figure 3. Grayscale image of raw EEG data for a randomly selected healthy subjects and those with major depressive disorder with $7 * 5 * 200 * 125$ resolution.

surface by keeping the relative distance among adjacent electrodes constant via azimuthal equidistant projection or polar projection method.⁶⁶⁻⁶⁹ This procedure is repeated for each frequency band of interest and left and right hemispheres, and topographically preserved $7 * 5 * 200 * 125$ -sized spatiotemporal mesh data are built, then fed into CNNs to reveal the inherent structure of EEG data. The $7 * 5$ mesh matrix represents the spatial information in a single time window as depicted in Figure 1. The temporal data size is obtained by multiplication of the sampling frequency (125 Hz.) with the duration of EEG recordings (200 seconds). CNNs deal with variations in space and time domains due to their ability to learn 2D demonstration of the data. The architecture of the proposed approach is given in Figure 2.

Where EEG image dataset is given as input to the CNN-based model and is stored as grayscale as given in Figure 3. Each subject is represented with $7 * 5$ (for 10-20 system of electrode placement) $* 200$ seconds recording duration $* 125$ Hz sampling frequency data. Besides, each pixel is expressed as a grayscale image that is the normalized amplitude value between 0 and 1.

Each data segmentation corresponding to the subjects has $200 * 125$ (25 000) samples ($* 200$ seconds recording duration $* 125$ Hz sampling frequency). Therefore, the shape of the tensor for each subject is structured as (25 000, 7, 5) with a 3D dimension. The total sample volume of the aggregated data is 25 000 (temporal samples in each electrode) $* 92$ (total subject number) $* 19$ (electrode number) = 43 700 000.

Methodology

DL utilizes many layers stacked together to process big data, leads to recognize an object, detect speech, translate languages with high accuracy, and so forth, is a subfield of machine learning approach with high computational capabilities. DL architecture has grid-like topology convolved with hidden layers, exploits with more massive datasets, learns automatically from raw data, brings more accuracy as compared to the traditional machine learning.^{70,71} Various quantitative EEG techniques are used in order to classify subjects. Absolute power, relative power, and cordance are promising markers with their predictive and discriminative capabilities. Since biomarkers underline specific markers searching significant patterns of the signals is a valuable process to classify discrete groups. So, the use of DL approach is another valuable perspective focusing on the structure of raw data rather than a processed biomarker.⁷²⁻⁷⁵

CNNs are similar to neural networks, but the most significant difference is the convolution operation and does not require a priori assumptions on the parameters for distinction. The matrix multiplication in the conventional neural network layers is done by a matrix of parameters with a separate parameter defining the reciprocal action among each input unit and each output unit. On the contrary, CNNs contain sparse interactions or connectivity, whose kernel have made smaller than the input. Through the help of this attribute, small and meaningful features in the data can be revealed.⁷⁶ Moreover, the complicated interactions among many variables can be efficiently represented by CNNs. Convolutional layer helps networks to detect image-based features like edges by convolving filters whose weights and biases are trained in a specific manner through the images. Apart from convolution layer, there exist other functional layers like rectified linear unit (ReLU) layer, pooling layer and fully connected layer. ReLU layer is an activation layer, which is applied in the element of images. The primary purpose of this layer is to eliminate the nonlinear features of images by converting negative values to zero.⁷⁷ Pooling layer reduces the computational cost of CNNs by downsampling features in the spatial dimension. For example, a maximum pooling operation extracts the maximum output within a rectangular neighborhood. Another pooling operation evaluates the average or weighted average of a rectangular neighborhood, or \mathcal{L}^2 -norm of a rectangular neighbourhood.⁷⁸ Fully connected layer works as a conventional neural network, and the probability calculation for each class is in the class prediction block that uses softmax function in our study.

ResNet-50 Convolutional Architecture

A residual learning framework was presented to ease the training of deep neural networks. It can be deduced from the empirical evidence that this kind of networks can gain accuracy from remarkably growth depth. ResNet works as feature extraction architecture for classification regression semantic segmentation, and so on. Through this architecture, a 28% relative

improvement was obtained on the Common Objects in Context (COCO) object detection dataset. ResNet won ImageNet competition on the tasks ImageNet detection and localization, and COCO detection and segmentation in 2015. Deep networks integrate the various level of features enriched by the number of stacked layers. Plain networks are not capable of calculating the gradient correctly; for this reason, after reaching the depth point in the convolutional layer, the training and testing error rises. Residual networks perform better than plain networks without having identity connections by overcoming the gradient vanishing/exploding problem. While deeper networks start converging via gradient descent with backpropagation, a degradation problem occurs with the network depth increasing. Incrementing the number of layers cause to higher training error. Identity mapping layers are added to the shallower model, and the other layers are copied from the learned shallower model to solve gradient vanishing existed in plain networks. As a result, the training error in the deeper model is restricted to its shallower counterpart. If $Q(x)$ is considered as a mapping, which comprises a few stacked layers with x referring to the first input set of these layers, one can assume that a residual function $R(x) = Q(x) - x$ is approximated asymptotically. The original function becomes $R(x) + x$. When the added layers are built as identity mappings, the training error of the deeper model does not exceed the training error of its shallower counterpart. Through the help of this residual learning, the weights of the multiple nonlinear layers approach to zero when the identity mappings are optimal. A building block is given as in Equation 1:

$$y = R(x, \{\psi_i\}) + \psi_s x \quad (1)$$

where x and y are the input and output vectors of the layers, respectively. $R(x, \{\psi_i\})$ corresponds to the learned residual mapping on convolutional or fully connected layers. ψ_s is the shortcut connections to match the dimensions with $R(\cdot)$. The element-wise addition is applied on 2 feature maps. 3×3 -sized filter is mostly used in the convolutional layers. The filter number in the layers is the same for the same output feature map size. When the feature map size becomes halve, the number of filters become doubled to keep the time complexity per layer. Downsampling is performed directly to convolutional layers, whose stride number is 2. The identity shortcut connections are embedded if the input and output have the same dimensions. When the dimensions increase, the identity mapping process is still applied with extra zero entries padding for increasing dimensions or the shortcut projection is utilized to equalize the dimensions.⁷⁹

ResNet has a significant advantage in solving the gradients vanish problem and makes the deep network trainable even if there exist outnumbered layers. Each sequence learning component comprises inputs, convolution layers, residual units, and additional layers, respectively. Residual functions consist of multiple convolution layers and batch normalization layers also with several ReLU transition function. This structure allows faster training, higher learning rate, and ease the

Table 1. MobileNet Body Architecture.⁷¹

Type/Stride	Filter Shape	Input Size
Conv2/s2	3*3*3*32	224*224*3
Conv dw/s1	3*3*32 dw	112*112*32
Conv/s1	1*1*32*64	112*112*32
Conv dw/s2	3*3*64 dw	112*112*64
Conv/s1	1*1*64*128	56*56*64
Conv dw/s1	3*3*128 dw	56*56*64
Conv/s1	1*1*128*128	56*56*128
Conv dw/s2	3*3*128 dw	56*56*128
Conv/s1	1*1*128*256	28*28*56
Conv dw/s1	3*3*256 dw	28*28*56
Conv/s1	1*1*256*256	28*28*256
Conv dw/s2	3*3*256 dw	28*28*256
Conv /s1	1*1*256*512	14*14*256
5*(Conv dw/s1)	3*3*512 dw	14*14*512
(Conv/s1)	1*1*512*512	14*14*512
Conv dw/s2	3*3*512*dw	14*14*512
Conv/s1	1*1*512*1024	7*7*512
Conv dw/s2	3*3*1024 dw	7*7*1024
Conv/s1	1*1*1024*1024	7*7*1024
Avg pool/s1	Pool 7*7	7*7*1024
Fc/s1	1024*1000	1*1*1024
Softmax/s1	Classifier	1*1*1000

Abbreviation: dw, depthwise.

network weight initialization, regularisation, and also improves the network performance. The task images in a sequence are built taking into account their time order, and thus the temporal dependency is captured.^{80,81}

MobileNet Convolutional Architecture

MobileNets was firstly developed for mobile and embedded vision applications. They are based on a streamlined architecture that uses depthwise separable convolutions. MobileNets yield small networks that match the application-based latency and size and optimize for latency without considering speed. A standard convolutional layer takes as input a $D_F \times D_F \times M$ feature map F and yields a $D_G \times D_G \times N$ feature map G . D_F is the spatial width and height of a square input feature map, M is the input depth, D_G is the spatial width and height of a square output feature map, and N is the output depth. The convolution kernel K has the size $D_K \times D_K \times M \times N$. D_K is the spatial width and height of the kernel.

The output feature map for standard convolution with stride one and padding is evaluated as in Equation 2:

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} * F_{k+i-1,l+j-1,m} \quad (2)$$

MobileNets use depthwise separable convolutions, which are composed of 3×3 -sized depthwise convolutions and 1×1 -sized pointwise convolutions, to separate the interaction among the number of output channels and the kernel size by applying a single filter to each input channel (input depth).

They also store both batch norm and ReLU nonlinearities in each layers except to the final fully connected layer, which is fed into a softmax layer for classification.

Depthwise convolution with one filter per input channel is constituted as in Equation 3:

$$\tilde{G}_{k,l,m} = \sum_{i,j} \tilde{K}_{i,j,m} * F_{k+i-1,l+j-1,m} \quad (3)$$

where \tilde{K} is the depthwise convolutional kernel whose size is $D_K \times D_K \times M$.

The m th channel in F is subjected to the m th filter in \tilde{K} to obtain the m th channel of the filtered output feature map \tilde{G} . The cost of the depthwise convolution is evaluated as $D_k \cdot D_k \cdot M \cdot D_F \cdot D_F$. Downsampling process is realized with strided convolution in the depthwise convolutions. The spatial resolution is reduced to 1 by a final average pooling before fully connected layer.⁷¹

The cost of the depthwise separable convolution is obtained as $D_k \cdot D_k \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$.

The convolution is constructed by applying filtering and combining process, and the computation cost is reduced as in Equation 4:

$$\frac{D_k \cdot D_k \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_k \cdot D_k \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_k^2} \quad (4)$$

To construct smaller and obtain less computational burden, a width multiplier parameter (ζ) should be adopted. The width multiplier thins a network uniformly layer by layer. Then, the computational cost of the convolution becomes $D_k \cdot D_k \cdot \zeta M \cdot D_F \cdot D_F + \zeta M \cdot \zeta N \cdot D_F \cdot D_F$.

To increase the speed and reduce computational effort, the resolution factor ($1 > \rho > 0$) is another parameter. This parameter is added to the inputs of the images and hidden layers as multipliers. After this process, the computational cost of the convolution is updated as $D_k \cdot D_k \cdot \zeta M \cdot \rho D_F \cdot \rho D_F + \zeta M \cdot \zeta N \cdot \rho D_F \cdot \rho D_F$. The structure of the MobileNet used in this article is given in Table 1.

The utilization of the fast downsampling strategy attenuates the computational burden of the layers before the smallest spatial dimensions. The depthwise separable convolutions improve the representation capability of generated features. When the number of channels is increased, this action contributes to larger information capacity, which is crucial of affecting the network performance under minimal computational resources. Evaluations of the actual inference time show that MobileNets is faster as compared with the ShuffleNet on an ARM-based device under the same complexity.⁸² Depth multipliers increase the number of feature maps corresponding to the input image channels, and width multiplier balances the accuracy and computational effort.⁸³

Inception-v3 Convolutional Architecture

Inception-v3 is a kind of transfer learning-type network, which allows the user to retrain the final layer of an existing model;

Table 2. Inception-v3 Body Architecture.

Type/Stride	Filter Shape	Input Size
Conv/s2	3*3	299*299*3
Conv/s1	3*3	149*149*32
Conv/s1 padded	3*3	147*147*32
Pool/s2	3*3	147*147*64
Conv/s1	3*3	73*73*64
Conv/s2	3*3	71*71*80
Conv/s1	3*3	35*35*192
3 × Inception		35*35*288

Diagram of a single Inception module. The 'Base' block feeds into four parallel paths: a 1x1 conv, a 3x3 conv, a 3x3 conv followed by a 1x1 conv, and a 3x3 conv followed by a 1x1 conv followed by a 3x3 conv. These paths are then concatenated into a 'Filter Concat' block.

8*8*1280

Diagram of a 2x Inception module. The 'Base' block feeds into four parallel paths: a 1x1 conv, a 1x3 conv, a 3x1 conv, and a 3x3 conv. These paths are then concatenated into a 'Filter Concat' block.

Pool	8*8	8*8*2048
Linear	Logits	1*1*2048
Softmax	Classifier	1*1*1000

thus, the training time and the size of the dataset required is reduced. Inception-v3 CNN concept was specialized in factorization ideas. The same network efficiency can be sustained even if reducing the number of connections or parameters by factorizing convolutions. For example, if one utilizes 2 layers of 3×3 filters rather than 1 layer of a 5×5 filter, the number of parameters is diminished by 28%. The factorisation procedure can be made through asymmetric convolutions (3×1 and 1×3 filters vs 3×3 filters with a 33% reduction ratio). In Inception-v3, a regularizer is replaced for auxiliary classifier role. The efficient grid size reduction is applied by concatenating the sets of feature maps.⁸⁴ Inception networks provide better solutions to calculation cost and gradient vanishing. In order to get accurate results in these networks, both local information and global information are accessed using various kernel dimensions. It is possible to reach global information by choosing a large kernel and local information by selecting a small kernel. By this way, the network is expanding rather than deepening. The classification accuracy was enhanced by integrating

recurrent layers to the Inception network.⁸⁵ As the models getting deeper, the dimensions become narrower. Therefore, information loss occurs. To avoid loss of information, the model Inception-v1 is expanded to Inception-v3. The structure of the Inception-v3 utilized in this article is given in Table 2.

A 48-layer Inception-v3 model was proposed, which is an improved version of the GoogleNet architecture in 2015. Unlike the AlexNet and VGG architectures, some of the calculations, for example, 1×1 , 3×3 , and 5×5 convolutions and max-pooling, are calculated simultaneously and chained. This concatenation is termed “inception layer.” The improved version, the 5×5 convolutions are replaced by 2 successive 3×3 convolutions. Feature extraction part using a CNN, and classification part using fully connected and softmax layers. The pre-trained Inception-v3 model succeeds in state-of-the-art accuracy for distinguishing universal things using 1000 classes, like “Zebra,” “Computer,” and “Dishes.” The model extracts features from input images in the first part and classifies them based on those features in the second part. Inception-V3 are also utilized in super-resolution, object-detection, segmentation, object tracking human pose-estimation, and video classification. For instance, ImageNet is the most potent model owned by the Google brain team. Transfer learning concept in Inception-V3 extracts bottleneck 2048-dimensional vector features from each frame.⁸⁶⁻⁸⁹

Training, Testing, and Validation Process

Our implementation is derived from the publicly available “Google Colaboratory,” which is a free cloud service giving an opportunity to AI (artificial intelligence) developers to apply their DL-based algorithms. Our datasets are trained on “Google Colaboratory,” including several significant modifications, which allows evaluations on multiple GPUs. The GPU model is Tesla k80 supporting all Python codes and DL libraries. It is easy to upload data from “Google Drive Application” to train the model. Multi-GPU training exploits data parallelism and is carried out by splitting each batch of training images into several GPU batches, processed in parallel on each GPU. The gradient of the full batch is obtained by averaging the computed GPU batch gradients. Gradient computation is synchronous across the GPUs, so the result is precisely the same as when training on a single GPU. In this article, all datasets are split into 5 sections and trained with 5-fold cross-validation (CV) method.

The models obtained from the training phase are tested section-by-section, and finally, the average accuracy is obtained. The systematic hyperparameter optimization is conducted in the process. The most important part of the unsupervised hyperparameter optimization comprises 2 problems. The first one is the implementation of the resulting metrics as a single score and the second is to find the best metric from the hyperparameter optimization process. Keras is used for constructing the classifier model, and Talos is utilized for hyperparameter optimization. First, the initial parameter boundaries such as

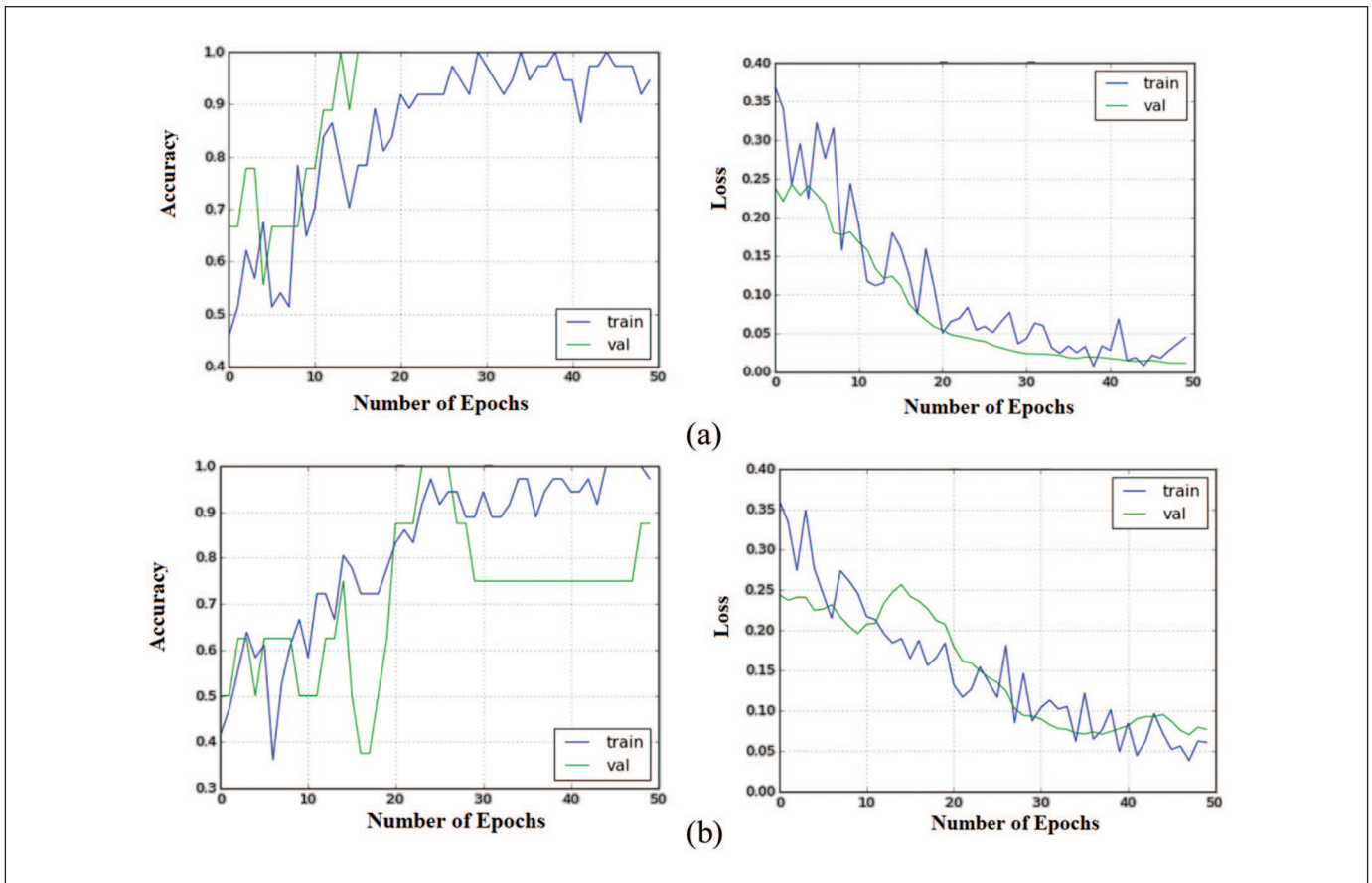


Figure 4. (a) MobileNet right hemisphere train and validation accuracies. (b) ResNet-50 delta band train and validation accuracies and loss functions.

learning rate, hidden layers, batch size, epochs, dropout, weight regularizer, optimizer, losses, activation, last activation, are chosen. Then, each permutation existed in the parameter space is successively tried in the train and validation processes. Finally, the best parameter set, which forces both train and validation loss close to zero, is selected. The best combination is found as SGD (stochastic gradient descent) optimizer with a 0.0002 initial learning rate and 0.9 momentum value, weight decay value = 0.0005, batch size = 64, epochs = 200, loss function = cross entropy.

Results

The train and validation accuracies and loss functions for right hemisphere obtained from MobileNet model are given in Figure 4a, and ResNet-50 Delta band validation accuracy plots through the training process and the loss functions are given in Figure 4b.

The averaged classification results for each epoch under a 5-fold CV are given in Tables 3, 4, 5, and 6 where the shaded rows and columns are representing the outperforming models and features.

In addition to classification accuracies of the models (Table 7), Gini coefficients that are recently being used in classification

problems were also calculated. Gini coefficient can be straight away generated from the AUC value, which is the ratio of the area between the ROC (receiver operating characteristic) curve and the diagonal line and the area of the above triangle in the ROC curve. $Gini = (2 * AUC - 1)$ is the equation used to express the Gini coefficient in terms of AUC value. The Gini values were calculated as 0.8 for ResNet-50, while that value is 0.76 for MobileNet, and 0.64 for Inception-v3. While the performance of FFT (fast Fourier transform) approach is comparatively low with its Gini value 0.59. Here, 0.82 points out a quite satisfying score in terms of model performance. Apart from the classification performance, a graphical representation of comparative performances of the models was also plotted as ROC curves as given in Figure 5.

FFT was used to calculate absolute and relative power in delta, theta, alpha, and beta bands. Following the FFT transform, cordance values are calculated. Cordance is a method of EEG spectra to yield values that have stronger correlation with regional cerebral perfusion. This correlation provides a physiological basis for interpreting this measure. The algorithm for cordance calculation yields 2 indicators for each electrode site in each frequency band: a categorical value (concordant or discordant state) and a numerical value. EEG cordance method combines both absolute and relative EEG power, and negative values of this measure (discordance) specifically in

Table 3. Classification Accuracy of Deep Learning Architectures for Left and Right Hemispheres.

Deep Learning Architecture	Hemispheric Classification Accuracy (%)	
	Left	Right
ResNet-50	75.55	87.64
Inception-v3	67.88	72.60
MobileNet	89.33	92.66

Table 4. Classification Accuracy of Deep Learning Architectures for Delta, Theta, Alpha, Beta Bands.

Deep Learning Architecture	Frequency Band Based Classification Accuracy (%)			
	Delta	Theta	Alpha	Beta
ResNet-50	90.22	73.67	66.64	71.25
Inception-v3	82.11	62.60	62.41	63.59
MobileNet	88.19	80.25	63.58	70.45

Table 5. Confusion Matrix of ResNet-50 For Delta Frequency Band.

Confusion Matrix	Predicted Label		
	No. of Samples	MDD	Healthy Controls
True label	MDD	41	4
	Healthy controls	5	42

Abbreviation: MDD, major depressive disorder.

Table 6. AUC Values of Deep Learning Architectures for Delta, Theta, Alpha and Beta Frequency Bands.

Deep Learning Architecture	Frequency Band Based Classification Accuracy (%)			
	Delta	Theta	Alpha	Beta
ResNet-50	0.9	0.69	0.68	0.7
Inception-v3	0.85	0.6	0.62	0.63
MobileNet	0.88	0.76	0.62	0.67

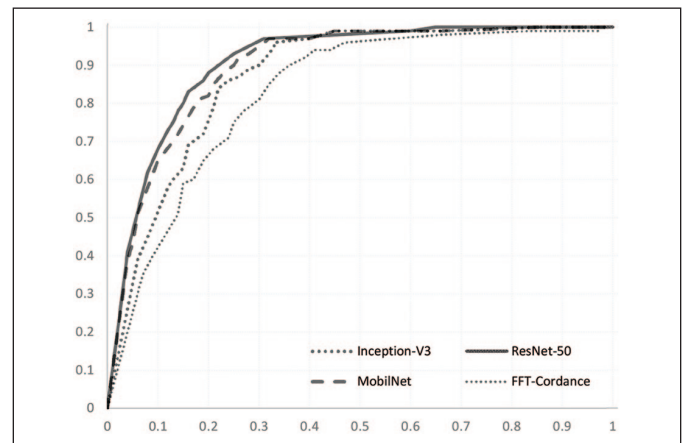
Abbreviation: AUC, area under the curve.

theta is believed to reflect low perfusion or metabolism, whereas positive values (concordance) are thought to reflect high perfusion or metabolism. Concordance combines complementary information from absolute and relative power of EEG spectra to yield values having stronger correlation with regional cerebral perfusion compared with stand-alone measures. The algorithm normalizes power across both electrode sites and frequency bands in 3 steps: First, absolute power values are reattributed to each individual electrode by averaging power from all bipolar electrode pairs sharing that

Table 7. Classification Performance of the Models for Testing Data.

Model	Frequency Band–Based Classification Accuracy (%)			
	Delta	Theta	Alpha	Beta
ResNet-50	89.47	84.21	78.94	78.94
Inception-V3	84.21	73.68	68.42	68.42
MobileNet	78.94	78.94	73.68	73.68
FFT-Cordance	73.68	73.68	68.42	68.42

Abbreviation: FFT, fast Fourier transform.

**Figure 5.** Receiver operating characteristic (ROC) curves of 3 different deep learning architectures and a fast Fourier transform (FFT)-based model.

electrode. It has been under lined that a stronger relation between QEEG measures and perfusion of underlying brain is provided by electrode referencing on the basis of power averaging. The relative power values are then calculated dividing total power values for each electrode site in each frequency band to reattributed absolute power values. Following the former step, the maximum absolute and relative power values (A_{maxf} , R_{maxf}) in each frequency band (f) are determined for each individual EEG recording to obtain normalized absolute ($ANORM(s,f)$) and normalized relative ($RNORM(s,f)$) power values. This normalization process places absolute and relative power values into a combined unit (producing values between 0 and 1). Finally, the cordance values at each electrode site (s) for each frequency band (f) are calculated by adding the $ANORM$ to $RNORM$, after half-maximal values are subtracted: $CORDANCE_{(s,f)} = (A_{Norm(s,f)} - 0.5) + (R_{Norm(s,f)} - 0.5)$.

The performance of DL-and FFT-based biomarker models are compared. As it is seen from the figure, the competitive performance of the models is noteworthy. Besides, ResNet-50 outperforms to MobilNet and Inception both in terms of Gini scores and classification accuracy. The performances of

classifiers with aforementioned methods were also compared in terms of their classification for spatial resolution of the data. As it is listed in Table 3, MobilNet outperforms compared with ResNet-50 and Inception to emphasize the contribution of different approaches for various perspectives. The performance of FFT model was satisfactory but not competitive compared with DL approaches.

Discussion and Conclusions

The DL approach based on CNN was implemented in this article to diagnose MDD. The feature extraction and feature selection processes are not necessitated for classification in the proposed model. The highest spatially dependent classification accuracies are found via MobileNet architecture with 89.33% and 92.66% for left and right hemisphere, respectively. The highest frequency-dependent accuracies are achieved through ResNet-50 architecture with 90.22% in the delta band and through MobileNet architecture with 80.25% in the theta band, respectively. The predictive accuracies supported with 0.9 and 0.76 AUC values for delta and theta band. The pathophysiology of MDD resides inadequately understood; however, the utilization of noninvasive neuroimaging techniques has dramatically enhanced our means to understand this disorder. The results of the present DL study revealed that the higher average delta amplitude in MDD compared with that of healthy control subjects might be a biomarker for MDD.

Delta power is related to reduced alertness and sleeps in the human brain.⁹⁰ EEG delta oscillations also determined as a correlate of basic homeostatic and motivational processes.⁹¹ However, delta oscillations are also recognized when the body is at rest but awake.⁹² A recent simultaneous fMRI-EEG study ascertained a clear correlation between frontal delta power and activity in the default mode network (DMN).⁹³ DMN is an intrinsic connectivity network which shows a high degree of interconnectedness even in the absence of a task. The research showed that MDD patients demonstrate increased connectivity within the DMN which is active in a resting state.^{94,95} Activity in the DMN is attenuated when the brain is task-oriented, and the network is more active when the brain is at rest. Results previously reported in the literature, denote correlations between the DMN and the delta frequencies.⁹⁶ The DMN includes the precuneus/posterior cingulate cortex, the medial prefrontal cortex, and medial, lateral, and inferior parietal cortex.⁹⁷ Numerous studies have shown that the DMN is related to self-referential processes. Patients with MDD frequently tend to heavily introspect and reflect upon their self, which has been described as “increased self-focus.”⁹⁸ The increased self-focus of depressive patients suggested being related to increased connectivity within the DMN.⁹⁹

Moreover, an increased “dominance” of the DMN in depression patients was found to be associated with the severity of self-focused rumination.¹⁰⁰ The extent to which the differences in delta amplitude may be a potentially relevant clinical tool warrants further

research. Prospectively, the utilization of DL may prove valuable biomarkers in psychiatry. Studies focusing on frequency-based biomarkers like absolute and relative power,¹⁰¹⁻¹⁰⁴ cordance,¹⁰⁵⁻¹⁰⁷ and coherence,¹⁰⁸⁻¹¹¹ generate promising outcomes and underline the frequency based potential of the marker by using processed but there is no clinically specific diagnostic biomarker capable of confirming the diagnosis of MDD. Therefore, exploring translational biomarkers of mood disorders based on DL has valuable potential with its recently underlined promising outcomes.

The proposed methodology can be adapted to the computer-assisted diagnosis of depression to validate the diagnosis made by clinicians. The predictive performance is enhanced by using a more extensive set of EEG data.

Author Contributions

CU contributed to design, contributed to analysis of data, drafted manuscript, critically revised manuscript, gave final approval, and agrees to be accountable for all aspects of work ensuring integrity and accuracy. TTE contributed to design, contributed to analysis of data, drafted manuscript, critically revised manuscript, gave final approval, and agrees to be accountable for all aspects of work ensuring integrity and accuracy. HU contributed to conception and design, contributed to interpretation of data, drafted manuscript, critically revised manuscript, gave final approval, and agrees to be accountable for all aspects of work ensuring integrity and accuracy. MC contributed to design, contributed to acquisition of data, drafted manuscript, critically revised manuscript, gave final approval, and agrees to be accountable for all aspects of work ensuring integrity and accuracy. GHS contributed to design, contributed to interpretation of data, drafted manuscript, critically revised manuscript, gave final approval, and agrees to be accountable for all aspects of work ensuring integrity and accuracy. MN contributed to design, contributed to acquisition of data, drafted manuscript, critically revised manuscript, gave final approval, and agrees to be accountable for all aspects of work ensuring integrity and accuracy. NT contributed to design, contributed to interpretation of data, drafted manuscript, critically revised manuscript, gave final approval, and agrees to be accountable for all aspects of work ensuring integrity and accuracy.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding


The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

ORCID iDs

Çaglar Uyulan  <https://orcid.org/0000-0002-6423-6720>

Türker Tekin Ergüzel  <https://orcid.org/0000-0001-8438-6542>

References

- Zarate CA Jr, Mathews D, Ibrahim L, et al. A randomized trial of a low-trapping nonselective N-methyl-D-aspartate channel blocker in major depression. *Biol Psychiatry*. 2013;74:257-264. doi:10.1016/j.biopsych.2012.10.019
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62:593-602. doi:10.1001/archpsyc.62.6.593
- Ballenger J. A randomized trial of an N-methyl-d-aspartate antagonist in treatment-resistant major depression. In: Talbott JA, Ballenger JC, Buckley PF, Frances RJ, Markowitz JC, Sarles RM, eds. *Yearbook of Psychiatry and Applied Mental Health*. St. Louis, MO: Mosby; 2008:204-205. doi:10.1016/s0084-3970(08)70782-5
- Lally N, Nugent AC, Luckenbaugh DA, Ameli R, Roiser JP, Zarate CA. Anti-anhedonic effect of ketamine and its neural correlates in treatment-resistant bipolar depression. *Transl Psychiatry*. 2014;4:e469. doi:10.1038/tp.2014.105
- Paranjape PN, Dhabu MM, Deshpande PS, Kekre AM. Cross-correlation aided ensemble of classifiers for BCI oriented EEG study. *IEEE Access*. 2019;7:11985-11996. doi:10.1109/access.2019.2892492
- Zhang J, Yan C, Gong X. Deep convolutional neural network for decoding motor imagery based brain-computer interface. Paper presented at: IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC); October 22-25, 2017; Xiamen, China. doi:10.1109/icspcc.2017.8242581
- Kumar S, Sharma A, Mamun K, Tsunoda T. A deep learning approach for motor imagery EEG signal classification. Paper presented at: 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE); December 5-6, 2016; Nadi, Fiji. doi:10.1109/apwc-on-cse.2016.017
- Sakhavi S, Guan C, Yan S. Parallel convolutional-linear neural network for motor imagery classification. Paper presented at: 23rd European Signal Processing Conference (EUSIPCO); August 31 to September 4, 2015; Nice, France. doi:10.1109/eusipco.2015.7362882
- Shen Y, Lu H, Jia J. Classification of motor imagery EEG signals with deep learning models. In: Sun Y, Lu H, Zhang L, Yang J, Huang H, eds. *IScIDE 2017: Intelligence Science and Big Data Engineering. Lecture Notes in Computer Science*. Vol 10559. Cham, Switzerland: Springer; 2017:181-190. doi:10.1007/978-3-319-67777-4_16
- Dose H, Moller JS, Puthusserypady S, Iversen HK. A deep learning MI—EEG classification model for BCIs. Paper presented at: 26th European Signal Processing Conference (EUSIPCO); September 3-7, 2018; Rome, Italy. doi:10.23919/eusipco.2018.8553332
- Lu N, Li T, Ren X, Miao H. A deep learning scheme for motor imagery classification based on restricted Boltzmann machines. *IEEE Trans Neural Syst Rehabil Eng*. 2017;25:566-576. doi:10.1109/tnsre.2016.2601240
- Chiarelli AM, Croce P, Merla A, Zappasodi F. Deep learning for hybrid EEG-fNIRS brain-computer interface: application to motor imagery classification. *J Neural Eng*. 2018;5:036028. doi:10.1088/1741-2552/aaaf82
- Yang H, Sakhavi S, Ang KK, Guan C. On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification. *Conf Proc IEEE Eng Med Biol Soc*. 2015;2015:2620-2623. doi:10.1109/embc.2015.7318929
- Golmohammadi M, Ziyabari S, Shah V, Obeid I, Picone J. Deep architectures for spatio-temporal modeling: automated seizure detection in scalp EEGs. Paper presented at: 17th IEEE International Conference on Machine Learning and Applications (ICMLA); December 17-20, 2018; Orlando, FL. doi:10.1109/ICMLA.2018.00118
- Hosseini MP, Pompili D, Elisevich K, Soltanian-Zadeh H. Optimized deep learning for EEG big data and seizure prediction BCI via Internet of things. *IEEE Transact Big Data*. 2017;3:392-404. doi:10.1109/tbdata.2017.2769670
- Tsiouris KM, Pezoulas VC, Zervakis M, Konitsiotis S, Koutsouris DD, Fotiadis DI. A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput Biol Med*. 2018;99:24-37. doi:10.1016/j.combiomed.2018.05.019
- Liang J, Lu R, Zhang C, Wang F. Predicting seizures from electroencephalography recordings: a knowledge transfer strategy. Paper presented at: IEEE International Conference on Healthcare Informatics (ICHI); October 4-7, 2016; Chicago, IL. doi:10.1109/ichi.2016.27
- Yuan Y, Xun G, Ma F, et al. A novel channel-aware attention framework for multi-channel EEG seizure detection via multi-view deep learning. Paper presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); March 4-7, 2018; Las Vegas, NV. doi:10.1109/bhi.2018.8333405
- Birjandtalab J, Heydarzadeh M, Nourani M. Automated EEG-based epileptic seizure detection using deep neural networks. Paper presented at: IEEE International Conference on Healthcare Informatics (ICHI); August 23-26, 2017; Park City, UT. doi:10.1109/ichi.2017.55
- Jingwei L, Yin C, Weidong Z. Deep learning EEG response representation for brain-computer interface. Paper presented at: 34th Chinese Control Conference (CCC); July 28-30, 2015; Hangzhou, China. doi:10.1109/chicc.2015.7260182
- Cao Y, Guo Y, Yu H, Yu X. Epileptic seizure auto-detection using deep learning method. Paper presented at: 4th International Conference on Systems and Informatics (ICSAI); November 11-13, 2017; Hangzhou, China. doi:10.1109/icsai.2017.8248445
- Alkanhal I, Kumar BVKV, Savvides M. Automatic seizure detection via an optimized image-based deep feature learning. Paper presented at: 17th IEEE International Conference on Machine Learning and Applications (ICMLA); December 17-20, 2018; Orlando, FL. doi:10.1109/icmla.2018.00086
- Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of a seizure using EEG signals. *Comput Biol Med*. 2018;100:270-278. doi:10.1016/j.combiomed.2017.09.017
- Hajinoroozi M, Zhang JM, Huang Y. Drivers fatigue prediction by deep covariance learning from EEG. Paper presented at: IEEE International Conference on Systems, Man, and Cybernetics (SMC); October 5-8, 2017; Banff, AB. doi:10.1109/smc.2017.8122609
- Hajinoroozi M, Mao Z, Huang Y. Prediction of drivers drowsy and alert states from EEG signals with deep learning. Paper presented at: IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP); December 13-16, 2015; Cancun, Mexico. doi:10.1109/camsap.2015.7383844

26. Jirayucharoensak S, Pan-Ngum S, Israsena P. EEG-based emotion recognition using deep learning network with principal component-based covariate shift adaptation. *ScientificWorldJournal*. 2014;2014:627892. doi:10.1155/2014/627892
27. Liao S, Wu C, Huang H, Cheng W, Liu Y. Major depression detection from EEG Signals using kernel eigen-filter-bank common spatial patterns. *Sensors (Basel)*. 2017;17:E1385. doi:10.3390/s17061385
28. Xu H, Plataniotis KN. Affective states classification using EEG and semi-supervised deep learning approaches. Paper presented at: IEEE 18th International Workshop on Multimedia Signal Processing (MMSP); September 21-23, 2016; Montreal, Quebec, Canada. doi:10.1109/mmisp.2016.7813351
29. Zheng WL, Zhu JY, Peng Y, Lu BL. EEG-based emotion classification using deep belief networks. Paper presented at: IEEE International Conference on Multimedia and Expo (ICME); July 14-18, 2014; Chengdu, China. doi:10.1109/icme.2014.6890166
30. Jia X, Li K, Li X, Zhang A. A novel semi-supervised deep learning framework for affective state recognition on EEG signals. Paper presented at: IEEE International Conference on Bioinformatics and Bioengineering; November 10-12, 2014; Boca Raton, FL. doi:10.1109/bibe.2014.26
31. Li X, Hu B, Sun S, Cai H. EEG-based mild depressive detection using feature selection methods and classifiers. *Comput Methods Programs Biomed*. 2016;136:151-161. doi:10.1016/j.cmpb.2016.08.010
32. Mei H, Xu X. EEG-based emotion classification using a convolutional neural network. Paper presented at: International Conference on Security, Pattern Analysis, and Cybernetics (SPAC); December 15-17, 2017; Shenzhen, China. doi:10.1109/spac.2017.8304263
33. Chambon S, Thorey V, Arnal PJ, Mignot E, Gramfort A. A deep learning architecture to detect events in EEG signals during sleep. Paper presented at: IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP); September 17-20, 2018; Aalborg, Denmark. doi:10.1109/mlsp.2018.8517067
34. Tan D, Zhao R, Sun J, Qin W. Sleep spindle detection using deep learning: a validation study based on crowdsourcing. Paper presented at: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 25-29, 2015; Milan, Italy. doi:10.1109/embc.2015.7318980
35. Chambon S, Galtier MN, Arnal PJ, Wainrib G, Gramfort A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Trans Neural Syst Rehabil Eng*. 2018;26:758-769. doi:10.1109/tnsre.2018.2813138
36. Manzano M, Guillén A, Rojas I, Herrera LJ. Deep learning using EEG data in time and frequency domains for sleep stage classification. In: Rojas I, Joya G, Catala A, eds. *IWANN 2017: Advances in Computational Intelligence. Lecture Notes in Computer Science*. Vol 10305. Cham, Switzerland: Springer; 2017:132-141. doi:10.1007/978-3-319-59153-7_12
37. Ruffini G, Soria DI, Dubreuil L, et al. Deep learning with EEG spectrograms in rapid eye movement behavior disorder [published online May 18, 2018]. *Front Neurol*. doi:10.1101/240267
38. Kuang D, He L. Classification on ADHD with deep learning. Paper presented at: International Conference on Cloud Computing and Big Data; November 12-14, 2014; Wuhan, China. doi:10.1109/ccbd.2014.42
39. Patel P, Aggarwal P, Gupta A. Classification of schizophrenia versus normal subjects using deep learning. Paper presented at: ICVGIP16: Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing; December 2016; Guwahati, India. doi:10.1145/3009977.3010050
40. Li Y, Li T. Deep learning classification study of first-episode treatment-naïve schizophrenia using brain cortical area and cognitive features. Paper presented at: Conference on Cognitive Computational Neuroscience; September 5-8, 2018; Philadelphia, PA. doi:10.32470/ccn.2018.1187-0
41. Morabito FC, Campolo M, Mammone N, et al. Deep learning representation from electroencephalography of early-stage Creutzfeldt-Jakob disease and features for differentiation from rapidly progressive dementia. *Int J Neural Syst*. 2017;27:1650039. doi:10.1142/s0129065716500398
42. Oh SL, Hagiwara Y, Raghavendra U, et al. A deep learning approach for Parkinson's disease diagnosis from EEG signals [published online August 30, 2018]. *Neural Comput Appl*. doi:10.1007/s00521-018-3689-5
43. Morabito FC, Campolo M, Labate D, et al. A longitudinal EEG study of Alzheimers disease progression based on a complex network approach. *Int J Neural Syst*. 2015;25:1550005. doi:10.1142/s0129065715500057
44. Mammone N, Bonanno L, Salvo SD, et al. Permutation disalignment index as an indirect, EEG-based, measure of brain connectivity in MCI and AD patients. *Int J Neural Syst*. 2017;27:1750020. doi:10.1142/s0129065717500204
45. Al-Kaysi AM, Al-Ani A, Boonstra TW. A multichannel deep belief network for the classification of EEG data. In: Arik S, Huang T, Lai W, Liu Q, eds. *ICONIP 2015: Neural Information Processing. Lecture Notes in Computer Science*. Vol 9492. Cham, Switzerland: Springer; 2015:38-45. doi:10.1007/978-3-319-26561-2_5
46. Khodayari-Rostamabad A, Reilly JP, Hasey GM, Bruin HD, Maccrimmon DJ. A machine learning approach using EEG data to predict response to SSRI treatment for major depressive disorder. *Clin Neurophysiol*. 2013;124:1975-1985. doi:10.1016/j.clinph.2013.04.010
47. Jaworska N, Salle SD, Ibrahim M, Blier P, Knott V. Leveraging machine learning approaches for predicting antidepressant treatment response using electroencephalography (EEG) and clinical data. *Front Psychiatry*. 2019;9:768. doi:10.3389/fpsy.2018.00768
48. Granter SR, Beck AH, Papke DJ Jr. AlphaGo, deep learning, and the future of the human microscopist. *Arch Pathol Lab Med*. 2017;141:619-621. doi:10.5858/arpa.2016-0471-ED
49. Surden H, Williams M. How self-driving cars work. *SSRN Electron J*. <https://ssrn.com/abstract=2784465>. Published May 28, 2016. Accessed March 20, 2020. doi:10.2139/ssrn.2784465
50. Triantafyllidou D, Tefas A. Face detection based on deep convolutional neural networks exploiting incremental facial part learning. Paper presented at: 23rd International Conference on Pattern Recognition (ICPR); December 4-8, 2016; Cancun, Mexico. doi:10.1109/icpr.2016.7900186
51. Kuhnle A, Copestake A. Deep learning evaluation using deep linguistic processing. Paper presented at: Proceedings of the Workshop on Generalization in the Age of Deep Learning; June 2018; New Orleans, LA. doi:10.18653/v1/w18-1003
52. Benuwa BB, Zhan YZ, Ghansah B, Wornyo DK, Kataka FB. A review of deep machine learning. *Int J Eng Res Africa*. 2016;24:124-136. doi:10.4028/www.scientific.net/jera.24.124

53. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press; 2016.
54. Al-Kaysi AM, Al-Ani A, Loo CK, et al. Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification. *J Affect Disord*. 2017;208:597-603. doi:10.1016/j.jad.2016.10.021
55. Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. *Neurosci Biobehav Rev*. 2017;74(pt A):58-75. doi:10.1016/j.neubiorev.2017.01.002
56. Acharya UR, Sudarshan VK, Adeli H, et al. A novel depression diagnosis index using nonlinear features in EEG signals. *Eur Neurol*. 2015;74:79-83. doi:10.1159/000438457
57. Hosseiniard B, Moradi MH, Rostami R. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput Methods Programs Biomed*. 2013;109:339-345. doi:10.1016/j.cmpb.2012.10.008
58. Akar SA, Kara S, Agambayev S, Bilgiç V. Nonlinear analysis of EEGs of patients with major depression during different emotional states. *Comput Biol Med*. 2015;67:49-60. doi:10.1016/j.combiomed.2015.09.019
59. Mohammadi M, Al-Azab F, Raahemi B, et al. Data mining EEG signals in depression for their diagnostic value. *BMC Med Inform Decis Mak*. 2015;15:108. doi:10.1186/s12911-015-0227-6
60. Li X, Song D, Zhang P, Yu G, Hou Y, Hu B. Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network. Paper presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 15-18, 2016; Shenzhen, China. doi:10.1109/bibm.2016.7822545
61. Mumtaz W, Xia L, Ali SS, Yasin MA, Hussain M, Malik AS. Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). *Biomed Signal Proc Control*. 2017;31:108-115. doi:10.1016/j.bspc.2016.07.006
62. Liao CY, Chen RC, Tai SK. Emotion stress detection using EEG signal and deep learning technologies. Paper presented at: IEEE International Conference on Applied System Invention (ICASI); April 13-17, 2018; Chiba, Japan. doi:10.1109/icas.2018.8394414
63. Cai H, Sha X, Han X, Wei S, Hu B. Pervasive EEG diagnosis of depression using Deep Belief Network with three-electrodes EEG collector. Paper presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 15-18; 2016; Shenzhen, China. doi:10.1109/bibm.2016.7822696
64. Parks TW, Burrus CS. *Digital Filter Design*. New York, NY: Wiley; 1987.
65. Rabiner LR, McClellan JH, Parks TW. FIR digital filter design techniques using weighted Chebyshev approximation. *Proc IEEE*. 1975;63:595-610.
66. Bashivan P, Rish I, Yeasin M, Codella N. Learning representations from EEG with deep recurrent-convolutional neural networks. Conference Paper at ICLR. <https://arxiv.org/abs/1511.06448>. Published 2016. Accessed March 20, 2020.
67. Dose H, Möller JS, Iversen HK, Puthusserypady S. An end-to-end deep learning approach to MI-EEG signal classification for BCIs. *Expert Syst Appl*. 2018;114:532-542. doi:10.1016/j.eswa.2018.08.031
68. Stober S, Sternin A, Owen AM, Grahn JA. Deep feature learning for EEG recordings. Conference Paper for ICLR. <https://arxiv.org/abs/1511.04306>. Published 2016. Accessed March 20, 2020.
69. Wen T, Zhang Z. Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. *IEEE Access*. 2018;6:25399-25410. doi:10.1109/access.2018.2833746
70. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV. doi:10.1109/cvpr.2016.90
71. Howard AG, Zhu M, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. <https://arxiv.org/abs/1704.04861>. Published 2017. Accessed March 20, 2020.
72. Erguzel TT, Tas C, Cebi M. A wrapper-based approach for feature selection and classification of major depressive disorder—bipolar disorders. *Comput Biol Med*. 2015;64:127-137. doi:10.1016/j.combiomed.2015.06.021
73. Erguzel TT, Ozekes S, Gultekin S, Tarhan N, Sayar GH, Bayram A. Neural network based response prediction of rTMS in major depressive disorder using QEEG cordance. *Psychiatry Investig*. 2015;12:61-65. doi:10.4306/pi.2015.12.1.6189.
74. Erguzel TT, Ozekes S, Gultekin S, Tarhan N. Ant colony optimization based feature selection method for QEEG data classification. *Psychiatry Investig*. 2014;11:243-250. doi:10.4306/pi.2014.11.3.243
75. Erguzel TT, Sayar GH, Tarhan N. Artificial intelligence approach to classify unipolar and bipolar disorders. *Neural Comput Appl*. 2016;27:1607-1616. doi:10.1007/s00521-015-1959-z
76. Zhang, Q. Convolutional neural networks. Paper presented at: 3rd International Conference on Electromechanical Control Technology and Transportation; January 19-21, 2018; Chongqing, China. doi:10.5220/0006972204340439
77. Ide H, Kurita T. Improvement of learning for CNN with ReLU activation by sparse regularization. Paper presented at: International Joint Conference on Neural Networks (IJCNN); May 14-19, 2017; Anchorage, AK. doi:10.1109/ijcnn.2017.7966185
78. Nagi J, Ducatelle F, Di Caro GA, et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition. Paper presented at: IEEE International Conference on Signal and Image Processing Applications (ICSIPA); November 16-18, 2011; Kuala Lumpur, Malaysia.
79. Ioannidou A, Chatzilari E, Nikolopoulos S, Kompatsiaris I. Deep learning advances in computer vision with 3D data. *ACM Comput Surveys*. 2017;50:20. doi:10.1145/3042064
80. Zhai D, Liu A, Chen S, Li Z, Zhang X. SeqST-ResNet: a sequential spatial temporal ResNet for task prediction in spatial crowdsourcing. In: Li G, Yang J, Gama J, Natwichai J, Tong Y, eds. *DASFAA 2019: Database Systems for Advanced Applications. Lecture Notes in Computer Science*. Vol 11446. Cham, Switzerland: Springer; 2019:260-275. doi:10.1007/978-3-030-18576-3_16
81. Chen Z, Xie Z, Zhang W, Xu X. ResNet and model fusion for automatic spoofing detection. Paper presented at: Interspeech; August 20-24, 2017; Stockholm, Sweden. doi:10.21437/interspeech.2017-1085
82. Qin Z, Zhang Z, Chen X, Wang C, Peng Y. Fd-MobileNet: improved Mobilenet with a fast downsampling strategy. Paper presented at: 25th IEEE International Conference on Image

- Processing (ICIP); October 7-10, 2018; Athens, Greece. doi:10.1109/icip.2018.8451355
83. Chen HY, Su CY. An enhanced hybrid MobileNet. Paper presented at: 9th International Conference on Awareness Science and Technology (ICAST); September 19-21, 2018; Fukuoka, Japan. doi:10.1109/icawst.2018.8517177
 84. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV. doi:10.1109/cvpr.2016.308
 85. Alom MdZ, Hasan M, Yakopcic C, Taha TM. Inception recurrent convolutional neural network for object recognition. <https://arxiv.org/abs/1704.07709>. Published 2017. Accessed March 20, 2020.
 86. Sinha R, Clarke J. When technology meets technology: retrained "Inception V3" classifier for NGS based pathogen detection. Paper presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); November 13-16, 2017; Kansas City, MO. doi:10.1109/bibm.2017.8217942
 87. Mednikov Y, Nehemia S, Zheng B, Benzaquen O, Lederman D. Transfer representation learning using Inception-V3 for the detection of masses in mammography. Paper presented at: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 18-21, 2018; Honolulu, HI. doi:10.1109/embc.2018.8512750
 88. Sai VB, Rao GN, Ramya M, Sree YS, Anuradha T. Classification of skin cancer images using TensorFlow and Inception v3. *Int J Eng Technol*. 2018;7:717-721. doi:10.14419/ijet.v7i2.7.10930
 89. Cheng K, Khokhar MS, Rao Y, Tahir R. Multi-camera background and scene activity modelling based on Spearman correlation analysis and Inception-V3 Network. Paper presented at: IEEE 35th International Conference on Data Engineering Workshops (ICDEW); April 8-12, 2019; Macao, China. doi:10.1109/icdew.2019.00058
 90. Hlinka J, Alexakis C, Diukova A, Liddle PF, Auer DP. Slow EEG pattern predicts reduced intrinsic functional connectivity in the default mode network: an inter-subject analysis. *Neuroimage*. 2010;53:239-246.
 91. Knyazev GG, Savostyanov AN, Bocharov AV, et al. Task-positive and task-negative networks in major depressive disorder: a combined fMRI and EEG study. *J Affect Disord*. 2018;235:211-219.
 92. Chen AC, Feng W, Zhao H, Yin Y, Wang P. EEG default mode network in the human brain: spectral regional field powers. *Neuroimage*. 2008;41:561-574.
 93. Neuner I, Arrubla J, Werner CJ, et al. The default mode network and EEG regional spectral power: a simultaneous fMRI-EEG study. *PLoS One*. 2014;9:e88214.
 94. Knyazev GG. EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neurosci Biobehav Rev*. 2012;36:677-695.
 95. Gusnard DA, Raichle ME. Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci*. 2001;2:685-694.
 96. Hamilton JP, Farmer M, Fogelman P, Gotlib IH. Depressive rumination, the default-mode network, and the dark matter of clinical neuroscience. *Biol Psychiatry*. 2015;78:224-230.
 97. Raichle ME. The brain's default mode network. *Annu Rev Neurosci*. 2015;38:433-447.
 98. Grimm S, Ernst J, Boesiger P, et al. Increased self-focus in major depressive disorder is related to neural abnormalities in subcortical-cortical midline structures. *Hum Brain Mapp*. 2009;30:2617-2627.
 99. Posner J, Hellerstein DJ, Gat I, et al. Antidepressants normalize the default mode network in patients with dysthymia. *JAMA Psychiatry*. 2013;70:373-382.
 100. Hamilton JP, Furman DJ, Chang C, Thomason ME, Dennis E, Gotlib IH. Default-mode and task-positive network activity in major depressive disorder: implications for adaptive and maladaptive rumination. *Biol Psychiatry*. 2011;70:327-333.
 101. Zoon HF, Veth CP, Arns M, et al. EEG alpha power as an intermediate measure between brain-derived neurotrophic factor Val66Met and depression severity in patients with major depressive disorder. *J Clin Neurophysiol*. 2013;30:261-267. doi:10.1097/WNP.0b013e3182933d6e108.
 102. Jeong HG, Ko YH, Han C, Kim YK, Joe SH. Distinguishing quantitative electroencephalogram findings between adjustment disorder and major depressive disorder. *Psychiatry Investig*. 2013;10:62-68. doi:10.4306/pi.2013.10.1.62
 103. Kam JW, Bolbecker AR, O'Donnell BF, Hetrick WP, Brenner CA. Resting state EEG power and coherence abnormalities in bipolar disorder and schizophrenia. *J Psychiatr Res*. 2013;47:1893-1901.
 104. Knott V, Mahoney C, Kennedy S, Evans K. EEG power, frequency, asymmetry and coherence in male depression. *Psychiatry Res*. 2001;106:123-140.
 105. Bares M, Brunovsky M, Novak T, et al. The change of prefrontal QEEG theta cordance as a predictor of response to bupropion treatment in patients who had failed to respond to previous antidepressant treatments. *Eur Neuropsychopharmacol*. 2010;20:459-466.
 106. Leuchter AF, Cook IA, Lufkin RB, et al. Cordance: a new method for assessment of cerebral perfusion and metabolism using quantitative electroencephalography. *Neuroimage*. 1994;1:208-219.
 107. Bares M, Brunovsky M, Kopecek M, et al. Changes in QEEG prefrontal cordance as a predictor of response to antidepressants in patients with treatment resistant depressive disorder: a pilot study. *J Psychiatr Res*. 2007;41:319-325.
 108. Tas C, Cebi M, Tan O, Hizli-Sayar G, Tarhan N, Brown EC. EEG power, cordance and coherence differences between unipolar and bipolar depression. *J Affect Disord*. 2015;172:185-190.
 109. Bowyer SM. Coherence a measure of the brain networks: past and present. *Neuropsychiatr Electrophysiol*. 2016;2:1 doi:10.1186/s40810-015-0015-7.
 110. Guevara MA, Corsi-Cabrera MC. EEG coherence or EEG correlation?, *Int J Psychophysiol*. 1996;23:145-153. doi:10.1016/S0167-8760(96)00038-4
 111. Srinivasan R, Winter WR, Ding J, Nunez PL. EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics, *J Neurosci Methods*. 2007;166:41-52.