

Article

Deep Transfer Learning for Parkinson's Disease Monitoring by Image-Based Representation of Resting-State EEG Using Directional Connectivity

Emad Arasteh ¹, Ailar Mahdizadeh ¹, Maryam S. Mirian ², Soojin Lee ² and Martin J. McKeown ^{2,3,*} 

¹ Electrical and Computer Engineering Department, University of Tehran, Tehran 14395-515, Iran; emad.arasteh@ut.ac.ir (E.A.); ailar.mahdizadeh@ut.ac.ir (A.M.)

² Pacific Parkinson's Research Centre, Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, BC V6T 2B5, Canada; maryam.mirian@ubc.ca (M.S.M.); soojin.lee@ndcn.ox.ac.uk (S.L.)

³ Faculty of Medicine (Neurology), University of British Columbia, Vancouver, BC V6T 2B5, Canada

* Correspondence: martin.mckeown@ubc.ca

Abstract: Parkinson's disease (PD) is characterized by abnormal brain oscillations that can change rapidly. Tracking neural alternations with high temporal resolution electrophysiological monitoring methods such as EEG can lead to valuable information about alterations observed in PD. Concomitantly, there have been advances in the high-accuracy performance of deep neural networks (DNNs) using few-patient data. In this study, we propose a method to transform resting-state EEG data into a deep latent space to classify PD subjects from healthy cases. We first used a general orthogonalized directed coherence (gOPDC) method to compute directional connectivity (DC) between all pairwise EEG channels in four frequency bands (Theta, Alpha, Beta, and Gamma) and then converted the DC maps into 2D images. We then used the VGG-16 architecture (trained on the ImageNet dataset) as our pre-trained model, enlisted weights of convolutional layers as initial weights, and fine-tuned all layer weights with our data. After training, the classification achieved 99.62% accuracy, 100% precision, 99.17% recall, 0.9958 F1 score, and 0.9958 AUC averaged for 10 random repetitions of training/evaluating on the proposed deep transfer learning (DTL) network. Using the latent features learned by the network and employing LASSO regression, we found that latent features (as opposed to the raw DC values) were significantly correlated with five clinical indices routinely measured: left and right finger tapping, left and right tremor, and body bradykinesia. Our results demonstrate the power of transfer learning and latent space derivation for the development of oscillatory biomarkers in PD.

Keywords: rsEEG; gOPDC; MVAR; directional connectivity; UPDRS; deep transfer learning; EEG neuromarkers



Citation: Arasteh, E.; Mahdizadeh, A.; Mirian, M.S.; Lee, S.; McKeown, M.J. Deep Transfer Learning for Parkinson's Disease Monitoring by Image-Based Representation of Resting-State EEG Using Directional Connectivity. *Algorithms* **2022**, *15*, 5. <https://doi.org/10.3390/a15010005>

Academic Editor: Maryam Ravan

Received: 15 November 2021

Accepted: 21 December 2021

Published: 24 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder affecting seniors [1]. The motor features of PD include bradykinesia, posture, and rest tremors, while non-motor complications can include constipation, postural hypotension, depression, and dementia. Pathologically, PD is characterized primarily by the degeneration of dopaminergic cells in the substantia nigra in the midbrain [2], leading to dopamine depletion in the striatum [3]. This neurochemical alteration impairs neuronal processing in the basal ganglia [4], resulting in prominent 8–30 Hz oscillations in local field potential recordings.

Non-invasive neuroimaging modalities (e.g., fMRI, MEG, EEG) have assisted in detecting altered brain networks in PD. As fMRI has limited temporal resolution (~sec) [5], it cannot directly detect features such as pathological oscillations seen at higher frequencies (e.g., 8–30 Hz). On the other hand, fMRI is capable of assessing changes in functional connectivity between macroscopic brain regions. Prior studies have demonstrated that

changes in striatal dopamine profoundly affect cortico-striatal connectivity, with decreases in connectivity between the posterior putamen and cortical somatosensory and motor regions [5]. Resting-state fMRI studies have also shown the disruption of connectivity in cortico-striatal networks [6]. The improved temporal resolution of EEG and MEG allows for the assessment of oscillations and has also demonstrated alterations in connectivity in PD at much more rapid time scales [7]. Thus, researchers are starting to explore ways to utilize rsEEG in diagnosing neurodegenerative diseases such as PD [8].

Several methods already exist to analyze rsEEG signals. These include spectral and complexity analysis [9], energy and entropy features followed by a decision tree classifier [10], and time-frequency properties of components from blind source separation [11]. In many of the aforementioned approaches (e.g., [12]), the EEG signals are divided into short (e.g., 2-s) epochs, which may fail to capture low-frequency dynamics of potential diagnostic utility. Here, we used a full 60 s of data to avoid this complication. Moreover, authors in this paper used time-frequency representations in which the channels' location are not considered as we consider here.

In [13], the authors have developed a classification approach using extracted features with the aim of emotion classification in HC and advanced (Hoehn–Yahr scale 4–5 stage) PD subjects, in contrast to the mildly affected people assessed here (Hoehn–Yahr 1–2).

Many previous approaches do not fully consider the spatial, temporal, and spectral properties of EEG signals while looking at the interaction between EEG channels may also be fruitful [14]. For example, in [15], a 1D-CNN was developed to classify features (as opposed to utilizing the raw data). Based on the assumption that the spatial location of the channels may benefit classification, EEG signals can be converted into multispectral images for subsequent classification [16]. In a related study, a feature-fusion multispectral image method (FMIM) was proposed to analyze multi-channel EEG signals and proved superior to conventional multispectral image methods [17].

Here, our aim is to consider channels connectivity as a key factor for Parkinson's classification, which would provide us with the exact interaction of channels instead of considering only their relative locations.

We propose using images based on the directional connectivity (DC) between EEG channels for the classification of the EEG from PD subjects. DC approaches model how neural activity in one region modulates neural activity in another location (but not necessarily vice versa). One effective way to compute DC is to employ multivariate autoregressive (MVAR) modeling [18]. This utilizes a linear combination of previous values of one channel to find its causal influences on another one, akin to Granger causality [19,20].

In this paper, we demonstrate that a latent space found by a deep transfer learning method to discriminate between HC and PD proves superior to standard methods such as sparse discriminant analyses in associating the EEG and clinical features such as UP-DRS3 [18]. This implies that a non-linear combination of features is more likely to prove more valuable in creating PD-specific biomarkers.

2. Materials and Methods

2.1. Overview

Here, we explore an MVAR-based DC measure to calculate connection strength between brain channels based on the gOPDC method (see Methods, below). DC is computed from EEG data for four bands: Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–45 Hz). We arrange DC matrices as 2D image-based representations of EEG signals and train a deep transfer learning model based on these acquired images. Then, we extract latent features of this model and demonstrate that these are more closely aligned with three important clinical indices of tremor, finger tap performance, and body bradykinesia compared to the original DC values, with each clinical feature having a unique DC signature. The general schematic overview of the procedures carried out in this research is depicted in Figure 1.

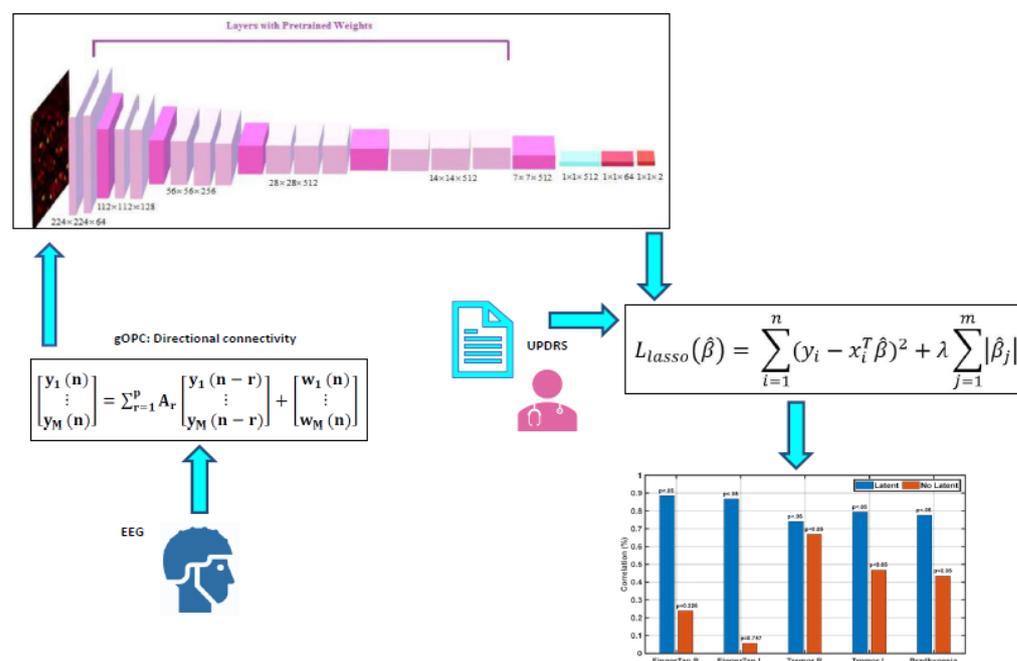


Figure 1. General schematic overview of the proposed approach.

2.2. Participants

The EEG data used in this paper are the same as [21], where a classical machine learning algorithm, sparse discriminant analyses [22], was used to discriminate between PD and control subjects. More detailed information is described in [21]. Moreover, 15 PD subjects (7 males; age: 67.3 ± 6.5 years) and 18 HC (9 males; age: 67.6 ± 8.9 years) subjects participated in the experiment. Because, in this study, we were utilizing the raw data, we were stringent in excluding subjects that had excessive muscle artifact in the EEG. Four HC and five PD subjects were excluded. Table 1 shows the characteristics of the subjects. The Hoehn and Yahr stages [23] range from 0—no disease to 1—Unilateral involvement only, 2—Bilateral involvement without impairment of balance, 3—Mild to moderate involvement with some postural instability, 4—Severe disability, and 5—Wheelchair-bound or bedridden. All participants provided written, informed consent. The study procedure was approved by the Clinical Research Ethics Board at the University of British Columbia (UBC).

Table 1. Demographic and clinical characteristics of the study.

	PD (n = 15)	HC (n = 18)
Age (year)	67.3 ± 6.5	67.6 ± 8.9
Sex (male/female)	7/8	9/9
Disease Duration (years), mean (std)	7.4 (4.3)	-
UPDRS II, mean (std)	14.8 (8.1)	-
UPDRS III, mean (std)	23.3 (9.1)	-
Hoehn and Yahr scale, mean	1.3 (1–2)	-

2.3. Study Protocols

The participants were seated in front of a computer screen, continuously gazing at a fixed target while EEG was recorded for 60 s.

The PD participants stopped taking levodopa medication and dopamine agonists for at least 12 h and 18 h, respectively, before the experiments. More detailed information is available in [21].

2.4. Recording and Pre-Processing EEG Signals

Data were acquired by 27 scalp electrodes, positioned according to the international 10–20 placement standard [21], as in Figure S1. Pre-processing steps are mainly based on what was reported in [21]. As explained in [21,24], the noise-removed EEG signals were filtered into four EEG frequency bands, Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–45 Hz), by a two-way finite impulse response (FIR) filter (the eegfilt function in EEGLAB: data sampling rate of 1000 Hz, filter order = 3*(sampling rate/ low edge frequency) with 3 minimum 3 low-edge frequency cycle and minimum filter order of 15-fractional width of transition zones: 0.25–1 frame per epoch).

2.5. Directional Connectivity

Functional connectivity algorithms typically use a linear mapping between two connections, leading to symmetrical connectivity matrices between EEG channels [25]. However, there is no guarantee that there will be abnormal connections in both directions equally between two brain regions [26]. To compute directional connectivity, we can assume that an asymmetrical linear operator H has been multiplied with zero-mean sources W that form the measured signals. Then, assuming the lack of instantaneous feedback between the received signals, the power spectral matrix of these signals can be decomposed, having components of H (a transfer function) and diagonal covariance matrix of W , which is Σ as shown in Equation (1) [25], where the superscript H means the Hermitian operator:

$$S(f) = H \Sigma H^H \tag{1}$$

To compute this asymmetrical H matrix, we suppose a multivariate autoregressive (MVAR) model for the received signals to find a correlation pattern between values of a “cause” source to later values of its “effect” (similar to that in Granger causality). In this way, a time series $y(n)$ with L samples can be modeled by a causal MVAR of order p :

$$\begin{bmatrix} y_1(n) \\ \vdots \\ y_M(n) \end{bmatrix} = \sum_{r=1}^p A_r \begin{bmatrix} y_1(n-r) \\ \vdots \\ y_M(n-r) \end{bmatrix} + \begin{bmatrix} w_1(n) \\ \vdots \\ w_M(n) \end{bmatrix} \tag{2}$$

where $w = [w_1 \ \dots \ w_M]^T$ are the assumed white noise vector with zero mean and a diagonal covariance matrix with nonzero elements equal to $\langle ww^T \rangle = \text{diag}\{\lambda_{kk}^2\}$. The A_r matrix elements are equal to the linear delay between two channels at a specific delay. For instance, m -th row and n -th column element, a_{mn}^r , is the r -th delay between channel m (cause) and channel n (effect) [27]. There can be two time-variant and time-invariant cases that the latter one forces A_r elements’ values to be dependent on the time.

Taking the Z-transform of both sides of Equation (2), the time-varying case A_r can be computed by an algebraic equation (as shown in Equation (3), where Z [25] means the Z-transform of a function or matrix). The frequency components of H can be computed by utilizing the correspondence of $z = e^{j2\pi f}$ between the Fourier and Z transforms. The $A(n, f)$ matrix is related to the power component in each time and frequency bin:

$$HZ\{W\} = \left(\sum_{r=1}^p A_r(n)z^{-r} \right) Z\{W\} + Z\{W\} \Rightarrow A(n, f) = I - \sum_{r=1}^p A_r(n)z^{-r} \Big|_{z=e^{j2\pi f}}; H(z, n) = A^{-1}(n, f) \tag{3}$$

Partial directed coherence (PDC) is a technique to compute directional connectivity in the frequency domain:

$$|\tilde{\pi}_{kl}(n, f)| \triangleq \frac{|A_{kl}(n, f)|}{\sqrt{\sum_k |A_{kl}(n, f)|^2}} \tag{4}$$

Equation (4) implies that the PDC from channel l to k is the ratio of in the k -th series due to the past of the l -th series. This method, based on a frequency picture of Granger causality, utilizes directional connection to compute the strength of the causal link between channels. In this manner, the null-hypothesis for Granger causality is no connection from l -th to k -th region ($H_0: \tilde{\pi}_{kl}(n, f) = 0$) [28].

Unfortunately, PDC as typically written has two major disadvantages. First, it is not scale-invariant, i.e., if one of the signals is amplified by a constant value and its spectral representation is unchanged, then the value of PDC is reduced [29]. To overcome this issue, each power component in Equation (4) can be divided by the related element in the cross-covariance matrix to obtain an estimate for general partial directed coherence (gPDC). Another drawback of the PDC is its sensitivity to volume conduction effects. Volume conduction is a serious problem for directional connectivity computations of EEG signals [30] due to the relatively poor spatial resolution of scalp EEG. A single source on the brain, through volume conduction, can reach adjacent channels leading to erroneous inferences about connectivity [18].

In [31], a method was proposed to orthogonalize the power envelope of the received signals to suppress volume conduction effects. By orthogonalizing power signals at each channel, adjacent channels do not share the trivial co-variability made by common sources. Reference [18] proves that the aforementioned orthogonalizing procedure can be reflected in the imaginary part of cross-spectral density between two channels using the MVAR modeling of the received signals. This method (using the time scaling technique) is called general orthogonal partial directed coherence (gOPDC). If we assume that at most K independent sources affect each EEG channel and the sources add linearly, such a relationship based on Equation (1) in the frequency domain is:

$$Y_i(f) = \sum_{k=1}^K v_{ik} S_k(f) \tag{5}$$

Equation (5) can be written in its matrix form:

$$Y(f) = VS(f) \tag{6}$$

$Y(f) \in C^M$ is a multi-channel EEG collected datum in the frequency domain, $S(f) \in C^M$ is the multivariate source signal in the frequency domain, and $V \in \mathbb{R}^{M \times M}$ includes all source weights.

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1k} \\ \vdots & \ddots & \vdots \\ v_{M1} & \cdots & v_{Mk} \end{bmatrix} \tag{7}$$

The cross-spectral density function $C_{ij}(f)$ between $Y_i(f)$ and $Y_j(f)$ is defined by Equation (8):

$$C_{ij}(f) = \langle Y_i(f) Y_j^*(f) \rangle = \sum_{k=1}^K V_{ik} V_{jk} \langle |S_k|^2 \rangle \tag{8}$$

The MVAR model in Equation (8) aids in checking the directional connectivity in the cross-spectral density. The innovative idea behind [18] is that instead of having two orthogonal signals, they search for orthogonal MVAR representations. The imaginary part of $Imag\{C_{ij}(f)\}$, in effect, subtracts the volume conduction effects from the received signal. It can be shown that this value can be computed by Equation (9) [27]. Σ_w^{-1} is the inverse covariance matrix of l -th column of $A(n, f)$, and $a_l(n, f)$ is the l -th column of $A(n, f)$. The inverse covariance value is used to ensure scale invariance:

$$\Psi_{kl}(n, f) = \frac{|Real\{A_{kl}(n, f)\}|}{\sqrt{a_l^H(n, f) a_l(n, f)}} \times \frac{|Imag\{A_{kl}(n, f)\}|}{\sqrt{a_l^H(n, f) a_l(n, f)}} \text{ if } k \neq l \tag{9}$$

Analogous to the definition of gPDC, OPDC can be extended to gOPDC, $\tilde{\Psi}_{kl}(n, f)$ by taking the effect of time series scaling into consideration:

$$\tilde{\Psi}_{kl}(n, f) = \frac{1}{\lambda_{kk}^2} \times \frac{|Real\{A_{kl}(n, f)\}|}{\sqrt{a_l^H(n, f) \Sigma_w^{-1} a_l(n, f)}} \times \frac{|Imag\{A_{kl}(n, f)\}|}{\sqrt{a_l^H(n, f) \Sigma_w^{-1} a_l(n, f)}} \text{ if } k \neq l \quad (10)$$

We computed DC from EEG signals during rest using the model in Equation (1) and the formula of Equation (10). Each element of DC matrix is the mean value of time components and frequency bins in the regarded band derived from Equation (10). After computation of gOPDC via Equation (10), we analyzed the resulting directional connectivity to see how strong the connections were. Directional connectivity (DC) in this paper is computed by averaging over time and frequency bins' (in each of the 4 bands) component values of any computed gOPDC connection. Because the number of variables is 2808 ($4 \times 2 \times \binom{27}{2} - 4$ frequency bands, 2 directions of DC, and $\binom{27}{2}$ related to all possible existing connections based on choosing 2 combinations of 27 numbers of EEG channels) (based on 4 frequency bands and 2 directions), the DC matrix is 33×2808 , including 2808 for each of 18 HC and 15 PD subjects. The optimum orders of the autoregressive model were chosen based on the minimum value of the Bayesian Information Criterion (BIC).

2.6. Image-Based Representation of DC

1. In order to increase our sample size, for each of the 4 frequency bands, we created 2D matrices (27×27) as a heatmap since there were 27 channels, and there were 26 channels that can be directionally connected to each channel (connectivity of a channel with itself is ignored). In order to have a square matrix for sizing of the heatmap, we created a 27×27 matrix and set the main diagonal components of the matrix to zero.
2. We normalized each matrix individually. In fact, we had normalized each matrix of DC (27×27) using the min/max approach.
3. We represented the results by computing heat maps, i.e., 2D representations of the values in a data matrix, in which colors represent their variability in intensity. Figure 2, as an example, shows two sample heat maps from two PD and HC cases. Larger values were represented by lighter pixels and smaller values by darker pixels.

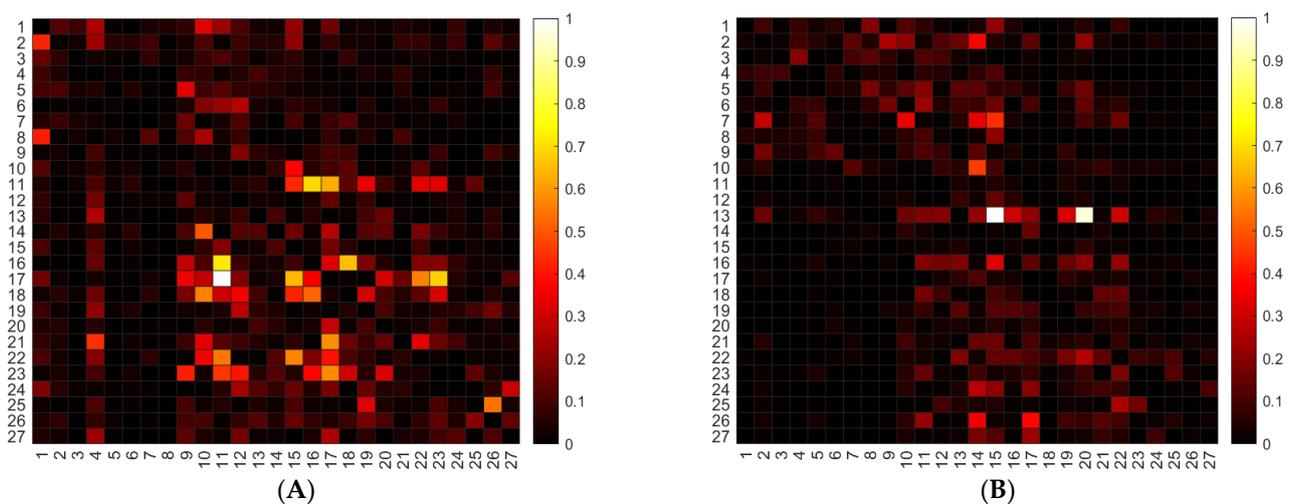


Figure 2. (A) Subject 11 DC of healthy class in the alpha band. (B) Subject 10 (Total Unified Parkinson’s Disease Rating Scale (UPDRS) = 22) DC of Parkinson class in the alpha band.

4. We resized the heatmap images to 224×224 in order to make them fit the architecture of the VGG-16 architecture.

After we computed heat maps of all bands, Theta, Alpha, Beta, and Gamma, we obtained 18 HC subjects' 2D arranged DC matrices (18 subjects \times 4 bands = 72 heatmaps) and 15 PD subjects' 2D arranged DC matrices (15 subjects \times 4 bands = 60 heatmaps). These were then used as the alternative compact representations of the EEG dataset for the binary classification task via deep transfer learning.

2.7. Transfer Learning

Traditionally, the goal of transfer learning is to solve the problem of domain gap when the knowledge gained on training data (source) needs to be re-used on different test (target) data. This aids in the problem of insufficient training data [32]. The primary layers of a DCNN (Deep Convolutional Neural Network), trained on a large dataset, can extract general features, as shown by [33,34]. Therefore, by recruiting deep transfer learning on a big dataset, we can fine-tune a pre-trained model on our small-size dataset and solve a binary classification task to PD/HC much more efficiently.

1. We used the VGG-16 architecture [35] trained on the ImageNet dataset as our pre-trained model, and we used only weights of convolutional layers of the pre-trained model.
2. To better fit our classification task, we modified the base model of VGG-16 in this order: (A) Implementing a fully connected layer ($1 \times 1 \times 512$) after the last max-pooling layer ($7 \times 7 \times 512$). (B) Developing a fully connected layer ($1 \times 1 \times 64$) with activation function of "Relu" to better map reduced features to the last layer. (C) Applying the fully connected layer ($1 \times 1 \times 2$) with the activation function of "Softmax" for the classification task consists of two categories, Parkinson's disease and healthy controls.
3. We tried two techniques of fine-tuning: The network, which we call the Totally trained model, is represented in Figure 3, while the other alternative is represented in the Appendix A just for further information.
4. To implement the first technique, we used weights of convolutional layers as initial weights and updated all layers' weights with our data. The network architecture is shown in Figure 3. Tables 2 and 3 represent the details of this network. In addition, Figure A1 shows the best performance of the proposed model over the epochs.

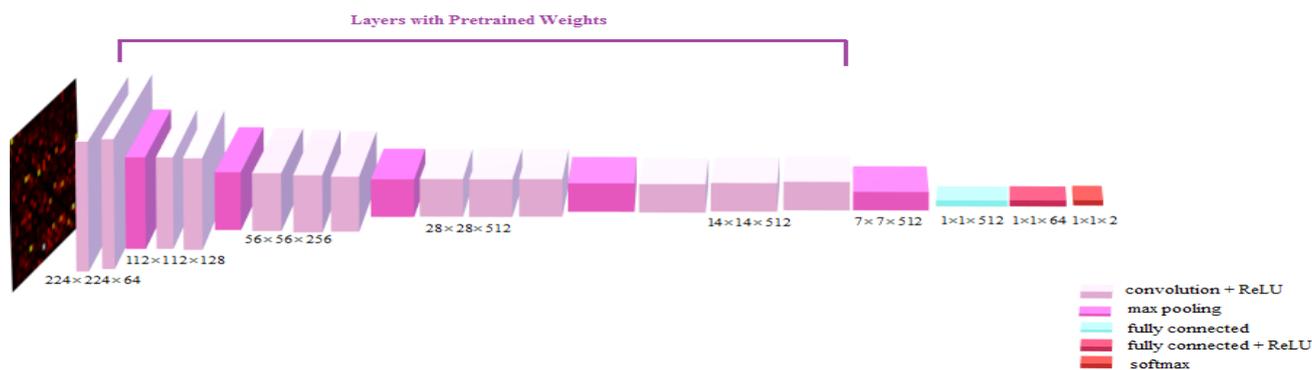


Figure 3. Totally trained network architecture.

We compared the best performance of both techniques (see the Appendix A for the alternative one) to choose the best fine-tuning strategy. As it is shown in Table 4, The totally-trained model surpassed the limited-trained model (expanded upon in the Appendix A). Our implementation was performed in Python using Keras with Tensorflow Backend and underwent 43 s of training on Google Colab. The test data set was 25% of data that had been randomly selected.

Table 2. Totally trained model's characteristics.

	First Model
Optimizer	SGD
Learning rate	0.01
Decay	0.001
Batch size	8
Loss function	Binary cross entropy

Table 3. Totally trained model's summary.

Layer	Output Shape	Param
Functional VGG16	(None, 7, 7, 512)	14,714,688
Max Pooling	(None, 1, 1, 512)	0
Flatten	(None, 512)	0
Dense	(None, 64)	32,832
Dense	(None, 2)	130
Total params: 14,747,650		
Trainable params: 14,747,650		
Non-trainable params: 0		

Table 4. Comparing the performance of fine-tuning strategies.

	Totally-Trained Model	Limited-Trained Model
Train accuracy	1.00	1.00
Test accuracy	1.00	0.871
Train loss	0.0004	0.013
Test loss	0.053	0.274

2.8. Performance Evaluation

We used accuracy, precision, recall, and F1 score concepts defined by Equations (11)–(14), respectively:

$$\text{Accuracy}(\%) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100 \quad (11)$$

$$\text{Precision}(\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \quad (12)$$

$$\text{Recall}(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \quad (13)$$

$$\text{F1 score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (14)$$

2.9. Feature Extraction from the Network

To investigate the clinical relevance of the features learned by the model, we saved the weights of the best model, fed every data sample in, and extracted the features of the second last layer. This resulted in a vector of (1×64) for each data point for each subject.

In [21], after the computation of phase-locking value (PLV) as a functional connectivity metric, sparse discriminant analysis [36] was used to find different main connections between HC and PD cases. This reference suggested that PLV variability lowers after the development of PD, and there is a significant negative correlation between PLV variability and disease severity. Here, we used the same EEG data as [21] and acquired new features

from the above DC approach by transfer learning. We then used the latent space features learned by the network to determine if the latent space is associated with disease severity.

We use the aforementioned (1×64) vector as 64 new latent features for DC of each subject. The UPDRS of healthy subjects were considered zero to incorporate both HC and PD connectivity values in the correlation computation [37]. To investigate whether these extracted features differed from raw connectivity values in predicting clinical scores, we fitted separate LASSO regression models [38] to both the extracted latent features and the raw DC values [39]. We then computed the correlation between the actual clinical scores and the scores estimated by the fitted models.

For the clinical scores, we checked for 5 sub-scores including both symmetric and asymmetric symptoms to compare latent and non-latent space results: Right and left “Finger tapping performance” and “Tremor” may present asymmetrically, while “Body Bradykinesia” is midline and is not associated with a specific side.

The LASSO regression models were computed for 30 runs separately for both latent and non-latent cases. In each run, out of 33 observations (33 subjects’ data), 25 random observations were assigned to the train and the rest 8 to the test set. In addition, as there were five statistical tests (the null hypothesis is that the resultant correlation is not significantly different from zero), p -values were corrected by the FDR method to deal with multiple comparisons. The corrected p -values less than 0.05 were considered significant. After completing 30 runs of the LASSO, the average of 30 best beta coefficients (coefficients that resulted in the least MSE at each run) were considered as the LASSO coefficients for the final LASSO model. Then, these vectors were chosen as the final LASSO coefficients of the models and were applied to the latent space and non-latent space data to estimate each clinical score.

3. Results

3.1. PD Classification Performance

To compare the proposed method against when only the four last layers were fine-tuned (Appendix A) and a prior DL model on the same data set [8], we evaluated it by feeding in randomly selected test data.

We tested the model 10 times (on randomly chosen train/test partitions). The training accuracy was constantly 100%, but the test accuracy varied in the range of 91 to 100%.

The average values of performance metrics are reported in Table 5. We extracted the last layer’s scores assigned to each class (HC/PD) for 10-fold repetition as predicted classes, compared them to the actual classes, and computed the True Negative, True Positive, False Negative, and False Positive declared in Figure 4.

In [15], a CNN model is suggested comprised of four consecutive blocks which are comprised of a 1D convolutional layer, max pooling layer, and activation layer, and the activation function is Relu, finally followed by three fully connected layers with Softmax activation function. The learning rate equals to 0.001, and the optimizer is Adam.

Table 5. Performance comparison of models.

	Proposed	CRNN [8]	SVM-RBF [8]	[15]
Accuracy (%)	99.6	99.2	95.4	95.4
Precision (%)	100	98.9	96.3	95.2
Recall (%)	99.17	99.4	94.3	95.5
F1 score	0.995	0.992	0.953	0.953
AUC (area under the ROC curve)	0.995	0.992	0.954	0.954

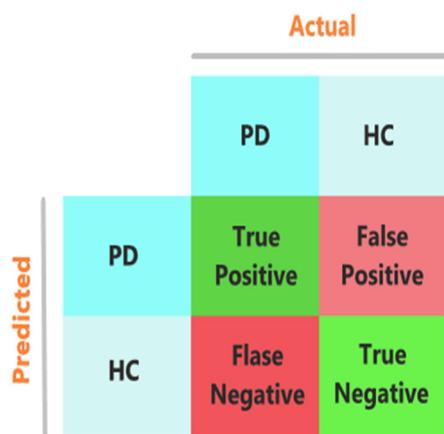


Figure 4. The confusion matrix: TP, TN, FP, and FN.

Previous work on the same data set included applications on a variety of machine learning models, including “KNN”, “SVM-L”, “SVM-P2”, “SVM-P3”, “SVM-RBF”, and “RF” for 779 extracted time-series features, including autocorrelation, quantiles, and number of zero-crossings [8]. In addition, they used univariate statistical analyses (ANOVA) and identified significant features, with a significance level of 0.05. Then, using these significant features, they trained the models. The best performance for these models is reported as “SVM-RBF” in Table 5. The proposed approach achieved 99.62% accuracy, 100% precision, 99.17% recall, 0.995 F1 score, and 0.995 AUC.

3.2. Clinical Relevance of Deep-Transfer-Learning-Based Classification

Figure 5 illustrates the correlation and the related *p*-values found by the fitted LASSO models on the two latent space and non-latent space cases. The values of the coefficients are plotted in A5 (Appendix A).

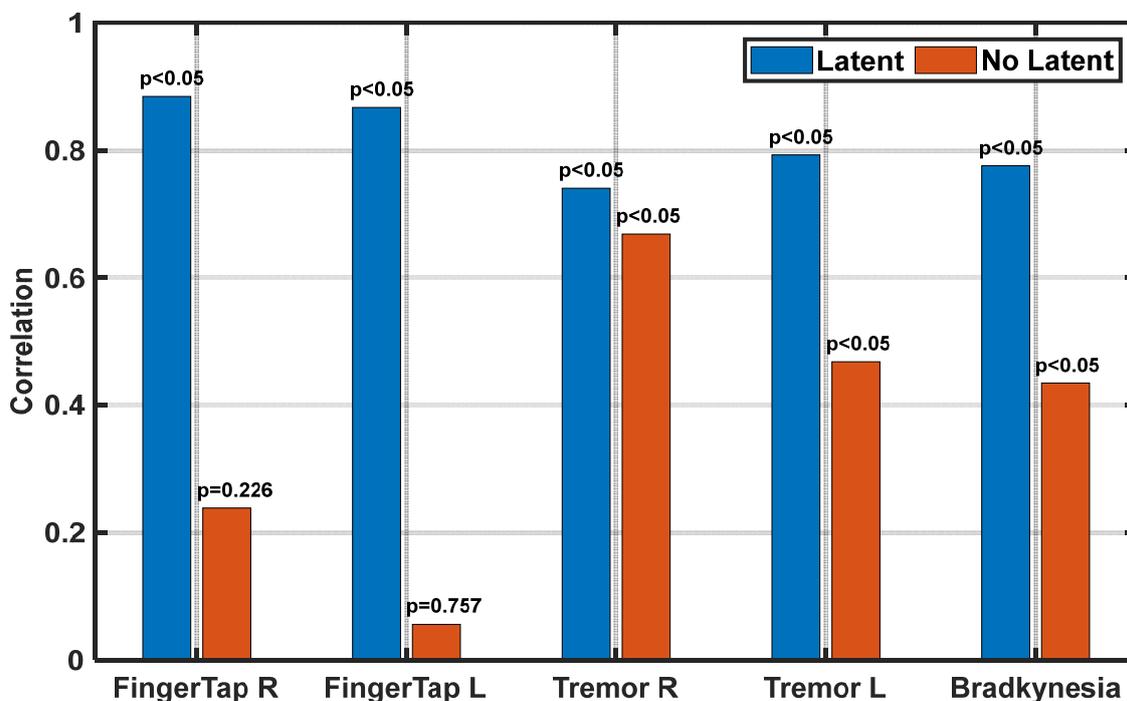


Figure 5. Correlations and the related *p*-values for the fitted LASSO models on two latent space and non-latent space cases to regress the clinical scores.

To quantitatively compare the correlation scores found by LASSO models in the latent and non-latent spaces, we use t-statistics on the Fisher z-transform of these correlation values to see whether latent space correlation values among the 30 runs were significantly higher than the non-latent space.

The t-test outcomes are shown in Table 6 and demonstrate the superiority of the latent space in predicting clinical aspects of the disease.

Table 6. t-Statistics values between the correlation of latent and non-latent space values with each clinical data sample for the 30 runs.

Clinical Test	Right Finger Tap	Left Finger Tap	Right Tremor	Left Tremor	Bradykinesia
<i>t</i> -test	6.0533	10.15	7.63	5.07	4.12
<i>p</i> -value	$<1 \times 10^{-7}$	$<2 \times 10^{-14}$	$<3 \times 10^{-10}$	$<5 \times 10^{-6}$	$<2 \times 10^{-4}$

Since it is difficult to visualize the heat map that would maximize each feature, we found the DC image (heat map) of which subject, in what health status (HC or PD), and in which frequency band led to the highest value in each feature (Table 7). The 10 highest values of each DC matrix (leading to maximum values in each feature) are shown in an arrow type plot in Figure 6.

Table 7. A related subject, health status (PD/HC), and the band of the DC heat map leading to the maximum value among 64 features for each clinical test.

Clinical Test	Right Finger Tap	Left Finger Tap	Right Tremor	Left Tremor	Bradykinesia
Subject	2	11	2	9	2
Health Status	PD	PD	PD	PD	HC
Frequency Sub-Band	Beta	Beta	Beta	Gamma	Alpha

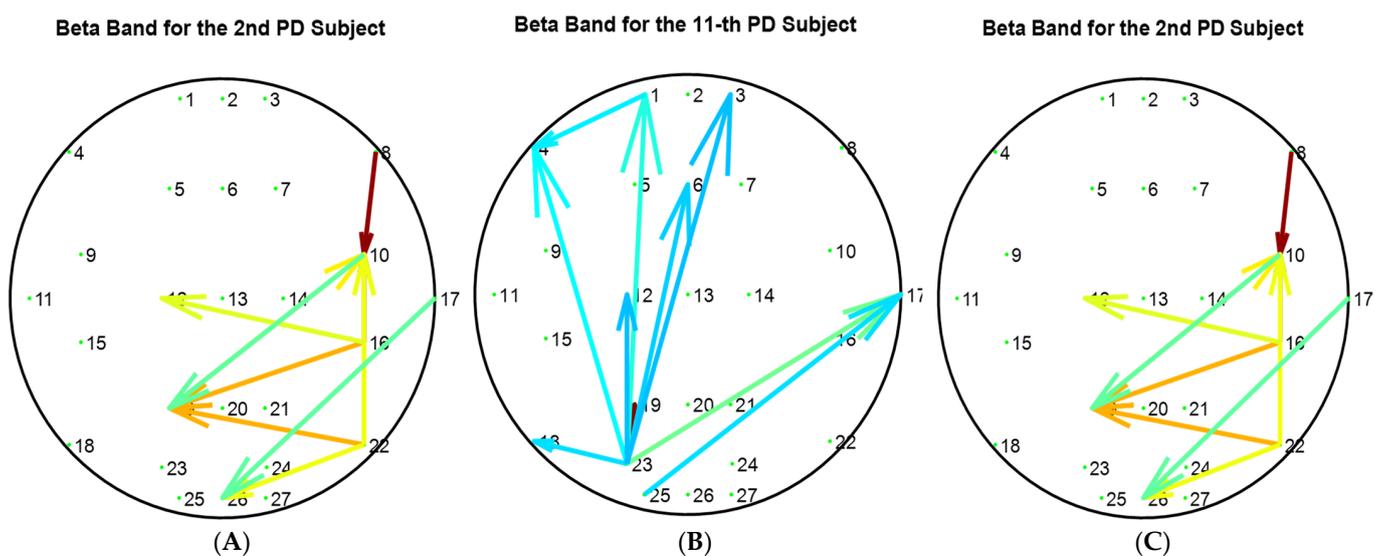


Figure 6. Cont.

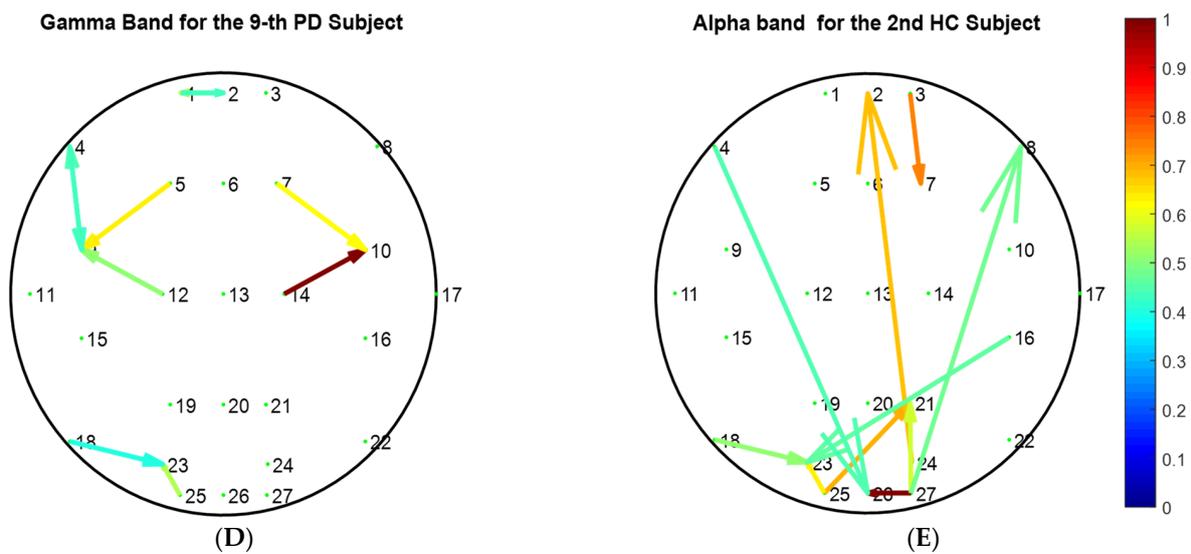


Figure 6. Ten highest normalized directional connections’ values that cause maximum value in feature related to the maximum LASSO coefficient for latent space: (A) Right finger tap test; (B) Left finger tap test; (C) Right tremor test; (D) Left tremor test; (E) Bradykinesia.

To establish the most influential EEG bands in non-latent space, we drew the main neural oscillations related to the right and left finger tap (the two cases mostly worse than the latent-space data) for non-latent data (Figure 7).

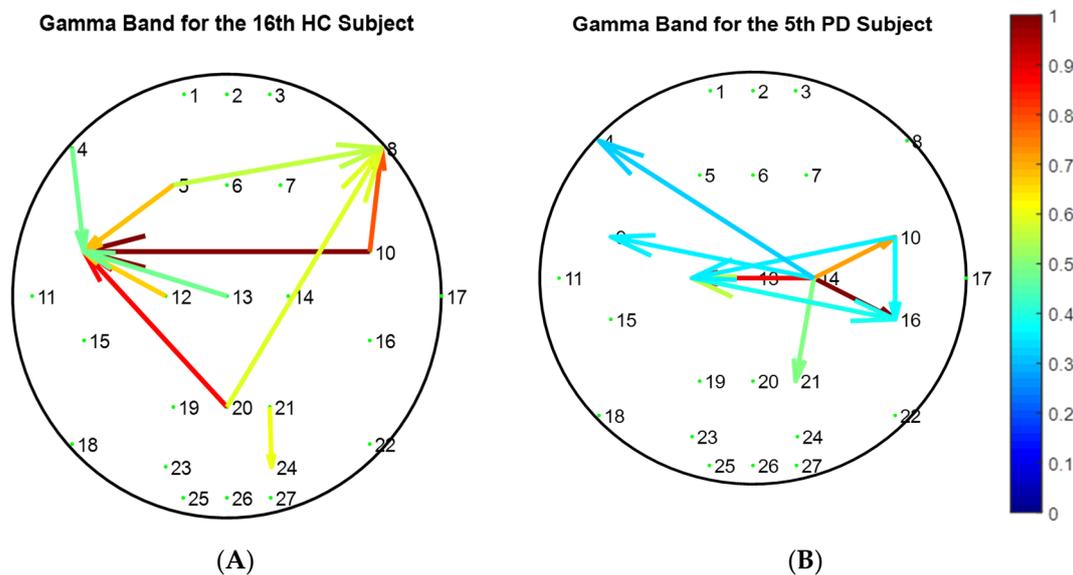


Figure 7. 10 highest normalized directional connections’ values that cause maximum value in feature related to the maximum LASSO coefficient for non-latent space: (A) Right finger tap test; (B) Left finger tap test.

4. Discussion

We found that DC measures are useful to discriminate PD subjects from controls. Importantly, we found that the latent space obtained by TL was more closely aligned with clinical indices related to movement, tremor, and body bradykinesia. It is perhaps unsurprising that the latent space, which will likely contain non-linear combinations of DC would more closely align with clinical performance, as presumably there will be a complex non-linear mapping between EEG connectivity measures and crude clinical performance

metrics. What is perhaps surprising is that TL, from a network trained on natural images (as the VGG-16 network used here is), results in a latent space that is more closely aligned to clinical indices.

We are acutely aware that our study may potentially be under-powered. However, we note that (1) as is typical with patient populations, the number of subjects tends to be quite low, and as a result of the challenges associated with obtaining recordings in these populations, (2) we have spent particular effort in ensuring that overfitting has not taken place, and (3) we are using transfer learning, which is a key aspect of the novelty of the proposed approach. The whole point of transfer learning is to allow for smaller data sets to be employed to allow for informed descriptions of disease states, and (4) we augmented the data by a factor of four by generating independent heat maps for each frequency subband.

The goal of using LASSO was to (1) select a sparse set of features and then (2) solve the regression problem to linearly map these sparse features to the target clinical scores. The beta coefficients are found by jointly minimizing the mean square error (MSE) between original and modeled values and applying an L1 norm constraint on the coefficients to promote sparsity. In Figure 5 we show that the LASSO method using the latent space values more closely predicts the clinical scores compared to the DC values used to develop the latent space. This proves that there exists a linear mapping from extracted features (of latent space) to the five clinical scores considered in this paper.

It is interesting that the network was agnostic to the particular band used, yet still provided excellent classification performance. This implies that the presence of effective connectivity between pairs of channels may be more informative for disease discrimination than the particular frequencies involved. Note that we were still able, post-hoc, to infer which frequency bands were most informative by deducting which particular heat maps were most informative in the final latent space.

It is also notable that many of the DC connections in the latest space appear ipsilateral to the hand in the finger tapping and tremor tasks (Figure 6). While mirror movements have been suggested as a reason for ipsilateral activation in PD subjects [40], it is important to note that the features we isolate best discriminate between PD and HC subjects. PD subjects tend to demonstrate more bilateral activation with unilateral movement than HC subjects [41]. Thus, the connections we found that best discriminate between PD and HC subjects will tend to be ipsilateral to the movement. We note that most of the discriminating features are in the beta-band (Table 7). This is hardly surprising, given the prominent role of altered beta-band oscillations in PD pathophysiology [42].

The proposed method had better results in terms of the accuracy, precision, AUC, and F1 score compared to when only the last four layers were fine-tuned (A2), a prior DL model on the same data set [8], and another deep learning approach for PD diagnosis [36], with approximately 10% of the run time of [8]. This likely reflects the relatively small data set, so that the other DL approaches could not fully capture the discrimination subspace given the data available. The transfer learning approach, adopted here, would appear to make discrimination tasks on smaller data sets more feasible.

Finally, we note that the maximum values of a feature originated from different connections, different subjects, and different bands. This suggests that the features we found were not based on overfitting a small subset of subjects.

5. Conclusions

Computing DC in the EEG provided us with an innovative technique to generate image representation of the 1D signal. Transfer learning achieved excellent classification performance and provided a latent space that is more closely related to clinical indices. Our results suggest transfer learning applied to an image classifier incorporating heatmaps may provide a way to assist in the classification of small-sample data sets.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/a15010005/s1>, Figure S1: Placement of 27 EEG electrodes. (A) 10–20 system configuration. Ref. [18] (B) Numbering of recorded channels.

Author Contributions: E.A. and A.M.: software, methodology, formal analysis; A.M., E.A. and M.S.M.: visualization, validation, draft preparation; E.A., A.M., M.S.M. and M.J.M.: conceptualization, writing—review and editing; M.J.M.: funding acquisition; S.L.: data collection. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the John Nichol Chair in Parkinson’s Research (M.J.M.).

Institutional Review Board Statement: The studies involving human participants were reviewed and approved by Clinical Research Ethics Board at the University of British Columbia (UBC) (approval code: H09-02016).

Informed Consent Statement: The patients/participants provided their written informed consent to participate in this study.

Data Availability Statement: De-identified data will be made available via written request to the corresponding author. In this paper, we used Matlab, Python. All codes and modules are accessible via GitHub repository, which also includes a Readme file on how to run the codes step by step to generate the results [43].

Acknowledgments: The authors would like to thank Ramy Hussain for his invaluable time and efforts in consulting with us about the network design.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The best performance of the proposed model, classifying PD and HC subjects, is shown in Figure A1.

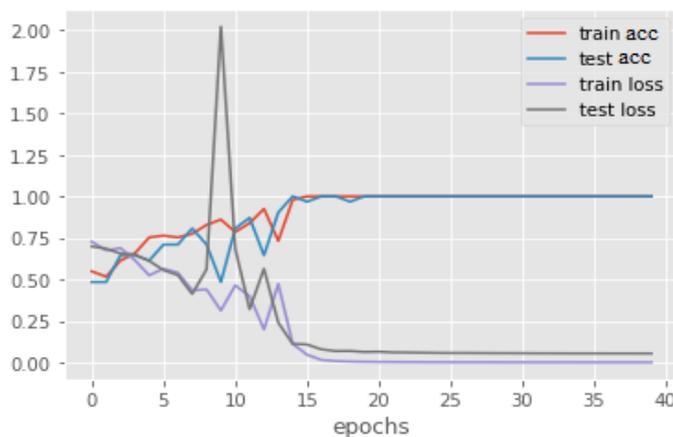


Figure A1. The Best Performance of the totally trained network for PD and HC classification.

In an alternative technique, we froze the weights of convolutional layers of the pre-trained model and updated only the last four layers with our data is shown in Figure A2. In addition, the details of the network are given in Tables A1 and A2 and the best performance of the network is shown in Figure A3.

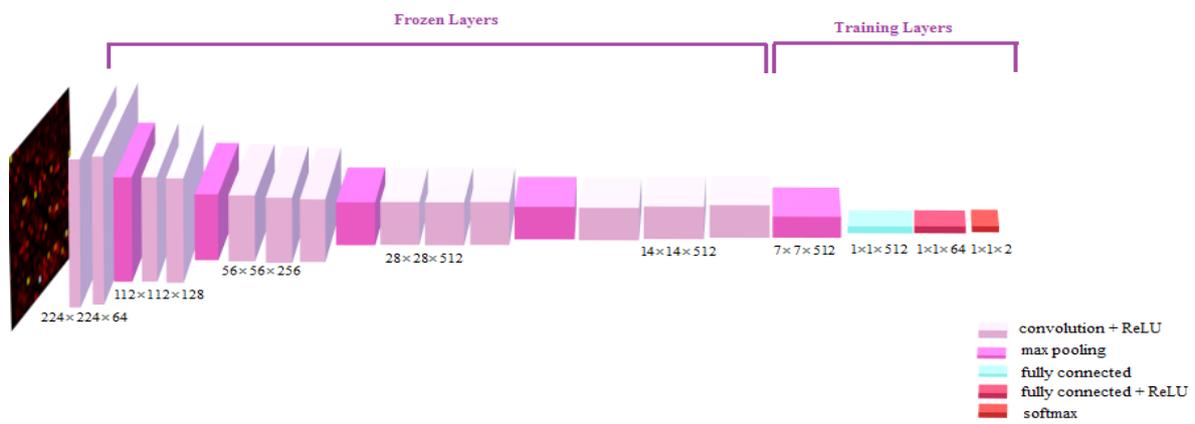


Figure A2. Limited-trained network architecture.

Table A1. Limited-trained model’s characteristics.

	First Model
Optimizer	Adam
Learning rate	0.01
Decay	0.001
Batch size	8
Loss function	Binary cross entropy

Table A2. Limited-trained model’s summary.

Layer	Output Shape	Param
Functional VGG16	(None, 7, 7, 512)	14,714,688
Max Pooling	(None, 1, 1, 512)	0
Flatten	(None, 512)	0
Dense	(None, 64)	32,832
Dense	(None, 2)	130
Total params: 14,747,650		
Trainable params: 32,962		
Non-trainable params: 14,714,688		

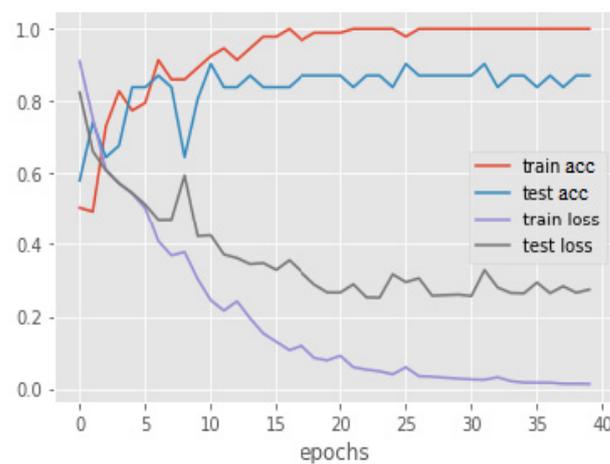


Figure A3. Best performance of the limited-trained network for PD and HC classification.

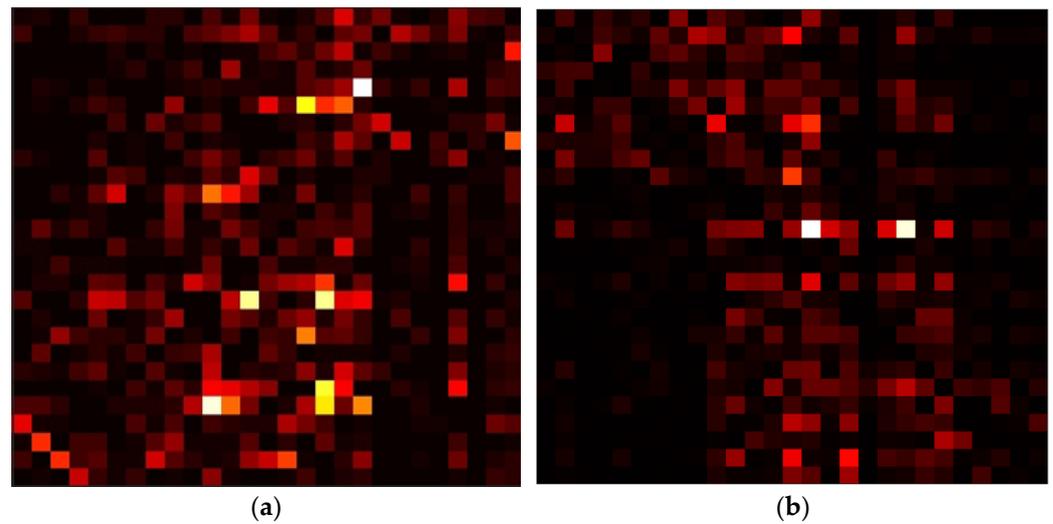


Figure A4. (a) Subject 11's DC of healthy class in alpha band, (b) Subject 10's (Total UPDRS = 22) DC of Parkinson class in alpha band.

Figure A5 depicts the mean value of the beta after 30 runs of LASSO.

If any element of the average beta vector is nonzero in at least 70% of the 30 runs, then that element is used in the final set of coefficients otherwise it is discarded. In fact, LASSO coefficients in different runs can be variable, and this causes unstable results. So, we heuristically found that stability over 70% of the runs leads to more reliable outcomes.

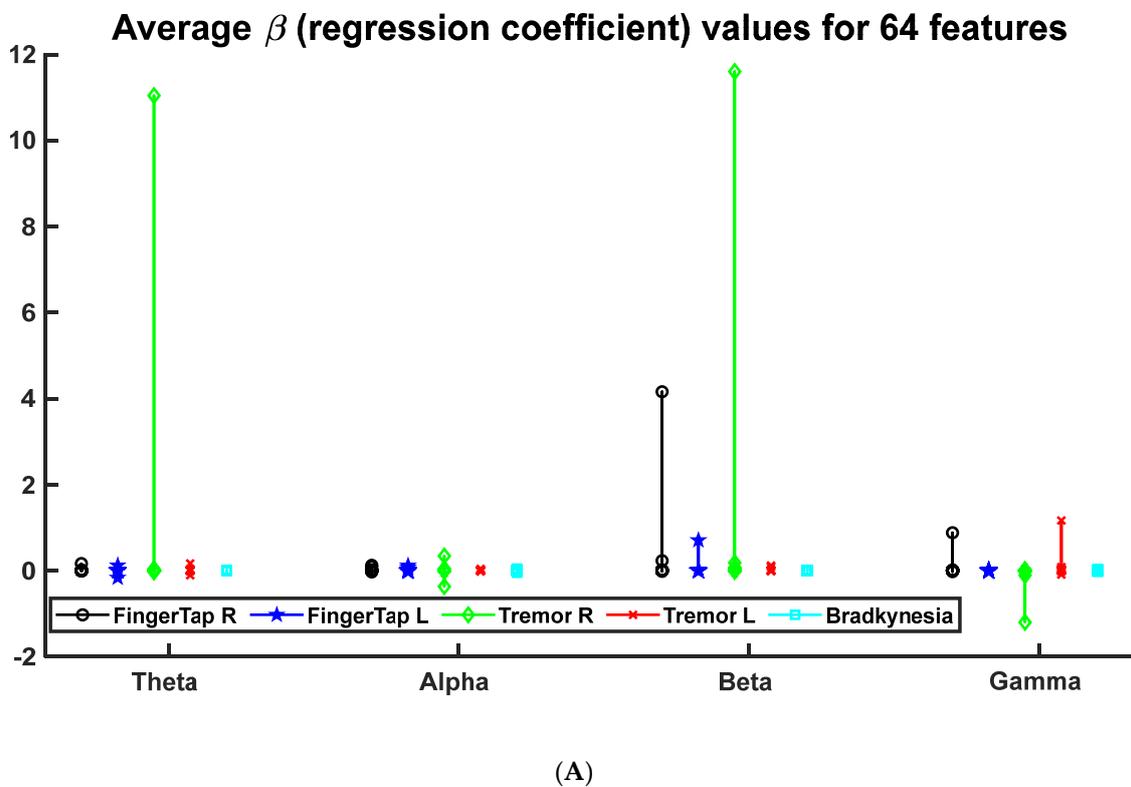


Figure A5. *Cont.*

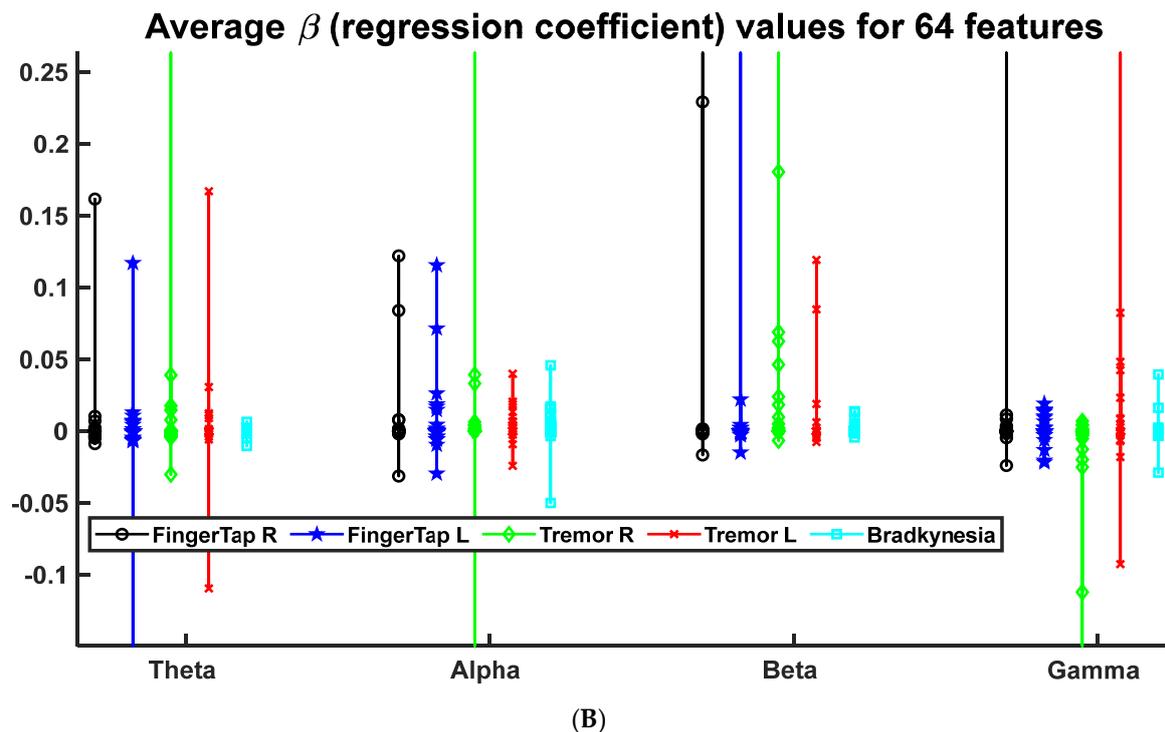


Figure A5. Average values of the LASSO coefficients: (A) All the values; (B) Zoomed graph of (A) for the near zero values.

References

- Scandalis, T.A.; Bosak, A.; Berliner, J.C.; Helman, L.L.; Wells, M.R. Resistance training and gait function in patients with Parkinson's disease. *Am. J. Phys. Med. Rehabil.* **2001**, *80*, 38–43. [[CrossRef](#)]
- Braak, H.; Del Tredici, K.; Rüb, U.; De Vos, R.A.; Steur, E.N.J.; Braak, E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **2003**, *24*, 197–211. [[CrossRef](#)]
- Brooks, D.J.; Piccini, P. Imaging in Parkinson's disease: The role of monoamines in behavior. *Biol. Psychiatry* **2006**, *59*, 908–918. [[CrossRef](#)] [[PubMed](#)]
- Rivlin-Etzion, M.; Marmor, O.; Heimer, G.; Raz, A.; Nini, A.; Bergman, H. Basal ganglia oscillations and pathophysiology of movement disorders. *Curr. Opin. Neurobiol.* **2006**, *16*, 629–637. [[CrossRef](#)]
- Helmich, R.C.; Derikx, L.C.; Bakker, M.; Scheeringa, R.; Bloem, B.R.; Toni, I. Spatial remapping of cortico-striatal connectivity in Parkinson's disease. *Cereb. Cortex* **2010**, *20*, 1175–1186. [[CrossRef](#)]
- Seibert, T.M.; Murphy, E.A.; Kaestner, E.J.; Brewer, J.B. Interregional correlations in Parkinson disease and Parkinson-related dementia with resting functional MR imaging. *Radiology* **2012**, *263*, 226–234. [[CrossRef](#)]
- Han, C.-X.; Wang, J.; Yi, G.-S.; Che, Y.-Q. Investigation of EEG abnormalities in the early stage of Parkinson's disease. *Cogn. Neurodynamics* **2013**, *7*, 351–359. [[CrossRef](#)]
- Lee, S.; Hussein, R.; Ward, R.; Wang, Z.J.; McKeown, M.J. A convolutional-recurrent neural network approach to resting-state EEG classification in Parkinson's disease. *J. Neurosci. Methods* **2021**, *361*, 109282. [[CrossRef](#)] [[PubMed](#)]
- McBride, J.C.; Zhao, X.; Munro, N.B.; Smith, C.D.; Jicha, G.A.; Hively, L.; Broster, L.S.; Schmitt, F.A.; Kryscio, R.J.; Jiang, Y. Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease. *Comput. Methods Programs Biomed.* **2014**, *114*, 153–163. [[CrossRef](#)]
- Baker, M.; Akrofi, K.; Schiffer, R.; O'Boyle, M.W. EEG patterns in mild cognitive impairment (MCI) patients. *Open Neuroimag. J.* **2008**, *2*, 52. [[CrossRef](#)]
- Vialatte, F.; Cichocki, A.; Dreyfus, G.; Musha, T.; Rutkowski, T.M.; Gervais, R. Blind source separation and sparse bump modelling of time frequency representation of EEG signals: New tools for early detection of Alzheimer's disease. In Proceedings of the 2005 IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT, USA, 28 September 2005.
- Khare, S.K.; Bajaj, V.; Acharya, U.R. PDCNNet: An automatic framework for the detection of Parkinson's Disease using EEG signals. *IEEE Sens. J.* **2021**, *21*, 17017–17024. [[CrossRef](#)]
- Krishna, N.M.; Sekaran, K.; Vamsi, A.V.N.; Ghantasala, G.P.; Chandana, P.; Kadry, S.; Blažauskas, T.; Damaševičius, R. An efficient mixture model approach in brain-machine interface systems for extracting the psychological status of mentally impaired persons using EEG signals. *IEEE Access* **2019**, *7*, 77905–77914. [[CrossRef](#)]

14. Akrofi, K.; Pal, R.; Baker, M.C.; Nutter, B.S.; Schiffer, R.W. Classification of Alzheimer's disease and mild cognitive impairment by pattern recognition of EEG power and coherence. In Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 14–19 March 2010.
15. Oh, S.L.; Hagiwara, Y.; Raghavendra, U.; Yuvaraj, R.; Arunkumar, N.; Murugappan, M.; Acharya, U.R. A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Comput. Appl.* **2020**, *32*, 10927–10933. [[CrossRef](#)]
16. Kwak, Y.; Kong, K.; Song, W.-J.; Min, B.-K.; Kim, S.-E. Multilevel feature fusion with 3d convolutional neural network for eeg-based workload estimation. *IEEE Access* **2020**, *8*, 16009–16021. [[CrossRef](#)]
17. Wen, D.; Li, P.; Li, X.; Wei, Z.; Zhou, Y.; Pei, H.; Li, F.; Bian, Z.; Wang, L.; Yin, S. The feature extraction of resting-state EEG signal from amnesic mild cognitive impairment with type 2 diabetes mellitus based on feature-fusion multispectral image method. *Neural Netw.* **2020**, *124*, 373–382. [[CrossRef](#)] [[PubMed](#)]
18. Omidvarnia, A.H.; Azemi, G.; Boashash, B.; O'Toole, J.M.; Colditz, P.; Vanhatalo, S. Orthogonalized partial directed coherence for functional connectivity analysis of newborn EEG. In Proceedings of the International Conference on Neural Information Processing, Doha, Qatar, 12–15 November 2012; Springer: Berlin/Heidelberg, Germany, 2012.
19. Granger, C.W. Investigating causal relations by econometric models and cross-spectral methods. *Econom. J. Econom. Soc.* **1969**, *37*, 424–438. [[CrossRef](#)]
20. Astolfi, L.; Cincotti, F.; Mattia, D.; Fallani, F.D.V.; Tocci, A.; Colosimo, A.; Salinari, S.; Marciani, M.G.; Hesse, W.; Witte, H. Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 902–913. [[CrossRef](#)] [[PubMed](#)]
21. Lee, S.; Liu, A.; Wang, Z.J.; McKeown, M.J. Abnormal phase coupling in Parkinson's disease and normalization effects of subthreshold vestibular stimulation. *Front. Hum. Neurosci.* **2019**, *13*, 118. [[CrossRef](#)]
22. Clemmensen, L.; Hastie, T.; Witten, D.; Ersbøll, B. Sparse discriminant analysis. *Technometrics* **2011**, *53*, 406–413. [[CrossRef](#)]
23. Hoehn, M.M.; Yahr, M.D. Parkinsonism: Onset, progression and mortality. *Neurology* **2001**, *57*, S11–S26.
24. Lee, S.; Liu, A.; McKeown, M.J. Current perspectives on galvanic vestibular stimulation in the treatment of Parkinson's disease. *Expert Rev. Neurother.* **2021**, *21*, 405–418. [[CrossRef](#)]
25. Baccala, L.A.; Sameshima, K.; Ballester, G.; do Valle, A.C.; Timo-Iaria, C. Studying the interaction between brain structures via directed coherence and Granger causality. *Appl. Signal Process.* **1998**, *5*, 40. [[CrossRef](#)]
26. Tropini, G.; Chiang, J.; Wang, Z.J.; Ty, E.; McKeown, M.J. Altered directional connectivity in Parkinson's disease during performance of a visually guided task. *Neuroimage* **2011**, *56*, 2144–2156. [[CrossRef](#)] [[PubMed](#)]
27. Omidvarnia, A.; Azemi, G.; Boashash, B.; O'Toole, J.M.; Colditz, P.B.; Vanhatalo, S. Measuring time-varying information flow in scalp EEG signals: Orthogonalized partial directed coherence. *IEEE Trans. Biomed. Eng.* **2013**, *61*, 680–693. [[CrossRef](#)]
28. Baccalá, L.A.; Sameshima, K. Partial directed coherence: A new concept in neural structure determination. *Biol. Cybern.* **2001**, *84*, 463–474. [[CrossRef](#)]
29. Baccala, L.A.; Sameshima, K.; Takahashi, D.Y. Generalized partial directed coherence. In Proceedings of the 2007 15th International Conference on Digital Signal Processing, Cardiff, UK, 1–4 July 2007.
30. Brunner, C.; Billinger, M.; Seeber, M.; Mullen, T.R.; Makeig, S. Volume conduction influences scalp-based connectivity estimates. *Front. Comput. Neurosci.* **2016**, *10*, 121. [[CrossRef](#)]
31. Hipp, J.F.; Hawellek, D.J.; Corbetta, M.; Siegel, M.; Engel, A.K. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat. Neurosci.* **2012**, *15*, 884–890. [[CrossRef](#)]
32. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A survey on deep transfer learning. In Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; Springer: Berlin/Heidelberg, Germany, 2018.
33. Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; Darrell, T. Decaf: A deep convolutional activation feature for generic visual recognition. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014.
34. Zeiler, M.D.; Fergus, R. Stochastic pooling for regularization of deep convolutional neural networks. In Proceedings of the 1st International Conference on Learning Representations, Scottsdale, AZ, USA, 2–4 May 2013.
35. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the The 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
36. Drysdale, A.T.; Grosenick, L.; Downar, J.; Dunlop, K.; Mansouri, F.; Meng, Y.; Fetcho, R.N.; Zebley, B.; Oathes, D.J.; Etkin, A. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **2017**, *23*, 28–38. [[CrossRef](#)] [[PubMed](#)]
37. MacAskill, M.R.; Graham, C.F.; Pitcher, T.L.; Myall, D.J.; Livingston, L.; van Stockum, S.; Dalrymple-Alford, J.C.; Anderson, T.J. The influence of motor and cognitive impairment upon visually-guided saccades in Parkinson's disease. *Neuropsychologia* **2012**, *50*, 3338–3347. [[CrossRef](#)]
38. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
39. Kazemi, A.; Mirian, M.; Lee, S.; McKeown, M. Galvanic Vestibular Stimulation Effects on EEG Biomarkers of Motor Vigor in Parkinson's Disease. *Front. Neurol.* **2021**, *12*, 759149. [[CrossRef](#)] [[PubMed](#)]
40. Li, J.Y.; Espay, A.J.; Gunraj, C.A.; Pal, P.K.; Cunic, D.I.; Lang, A.E.; Chen, R. Interhemispheric and ipsilateral connections in Parkinson's disease: Relation to mirror movements. *Mov. Disord.* **2007**, *22*, 813–821. [[CrossRef](#)] [[PubMed](#)]
41. Wu, T.; Hou, Y.; Hallett, M.; Zhang, J.; Chan, P. Lateralization of brain activity pattern during unilateral movement in Parkinson's disease. *Hum. Brain Mapp.* **2015**, *36*, 1878–1891. [[CrossRef](#)]

-
42. Little, S.; Brown, P. The functional role of beta oscillations in Parkinson's disease. *Parkinsonism Relat. Disord.* **2014**, *20*, S44–S48. [[CrossRef](#)]
 43. DTL-GOPDC. Available online: <https://github.com/DTL-GOPDC> (accessed on 19 December 2021).