Parkinson's disease diagnosis through electroencephalographic signal processing and neural network classification

Junfeng Zou[1,3,†], Yufei Zhao[2,4,5,†]

[1]Department of Leeds, Southwest Jiaotong University, Chengdu, 611756, China
[2]Faculty of Engineering, University of Ottawa, Ottawa, K1N 6N4, Canada

[3]el202jz@leeds.ac.uk
[4]yzhao312@uottawa.ca
[5]Corresponding author
[†]Junfeng Zou and Yufei Zhao contributed equally to this work and should be considered co-first authors.

Abstract. Parkinson's disease (PD) is the second most prevalent neurological disorder, following Alzheimer's. Despite this, there is currently no successful treatment for PD. Therefore, early detection of Parkinson's disease is crucial for preventing its progression. To address this, a computer-aided diagnosis system has been implemented to identify any abnormalities. Significant research has been conducted using speech and gait analysis. However, there is growing interest in using electroencephalographic (EEG) signals to diagnose Parkinson's disease at an early stage. This paper aims to use EEG to capture neural correlates of dysfunction in PD patients and compare with the normal ones to determine whether a person has PD. The method is to preprocess the EEG dataset using MATLAB and EEGLAB and to analyze and classify the preprocessed data using MLP neural networks, which has good expressiveness and adaptability. Our dataset contains 25 sets of data with 11 healthy people and 14 Parkinson's Disease patients. Experiments show that the model has an average test accuracy of 96.8% and average test loss of 12.8%.

Keywords: Parkinson's Disease, EEG Preprocessing, MLP neural network.

1. Introduction

Parkinson's disease primarily occurs due to the loss of dopaminergic neurons in the substantia nigra, leading to the death of these cells. Current Parkinson's disease medications treat symptoms but do not stop or slow the degeneration of dopaminergic neurons [1]. Since other neurological disorders have similar signs, doctors use a combination of medical history and medical science to make a judgment, but this is inefficient and about 25% of diagnoses are considered incorrect. Therefore, it is important to recognize the disease at an early stage and treat it promptly [2].

Although the neurologist's opinions and reviews are crucial in determining the final diagnosis, any tool that assists in comparing diagnoses is highly valued. Thus, there is an increasing need for automated procedures that can enhance the accuracy of diagnosing Parkinson's disease. Some auxiliary tests such

as imaging and body fluid markers are helpful in the diagnosis and differential diagnosis of Parkinson's disease.

Based on the pathological change of dopaminergic neuron injury in Parkinson's disease, the detection of dopamine transporters through positron emission tomography (PET) is a valuable imaging auxiliary diagnosis method [3]. However, this project is expensive, and it should be noted that it is not suitable for routine use. Besides, DAT SPECT does not distinguish Parkinson's disease from atypical Parkinson's syndrome [4].

Magnetic resonance imaging (MRI) can help identify Parkinson's disease and other diseases that cause Parkinson's syndrome. In structural MRI, neuromelanin in the substantia nigra and the high signal of the dorsal substantia nigra are reduced, and iron deposition and diffusion coefficient of the substantia nigra are increased, and quantitative susceptibility mapping is of certain value for the diagnosis of Parkinson's disease [5]. However, price, convenience and other factors limit the clinical application, and the external validity of these methods is still to be confirmed.

To monitor the electrical activity of the brain, a non-invasive method is utilized called EEG. It records the electrical signals generated by the firing of neurons in the brain using electrodes placed on the scalp. EEG is frequently employed in clinical and academic contexts to investigate cognitive processes, track sleep patterns, and diagnose neurological problems. EEG has the advantage of being affordable and portable, enabling the capture of cerebral data in real-life situations. Moreover, EEG surpasses other technologies in terms of both speed and duration in recording brain activity [6]. Swann, a professor in the Department of Human Physiology, suggested that by utilizing a secure and cost-effective method to assess and measure brain activity, it can identify differences between individuals with PD who were both on and off medication, as well as in comparison to healthy people [7]. Since Parkinson's disease is associated with brain abnormalities, an early preliminary clinical diagnosis can be made using EEG. The data collected from an EEG can provide valuable insights into brain activity and help in understanding various brain states and conditions. In this study, MLP neural networks were used to identify and classify EEG data, which has the advantage of simplicity and can deal with nonlinear problems, and finally a classification model with high accuracy was obtained.

2. Method

This section outlines the recommended approaches for analyzing EEG signals, including steps such as preprocessing, extracting useful features, and applying classification techniques by MLP neural network.

2.1. Data Acquisition

The dataset used in this study is publicly available [8] and consists of two samples from the University of New Mexico and the University of Iowa. In this paper, we used the dataset from Iowa, which consists of 14 control subjects and 14 patients with Parkinson's disease. The researchers recorded the resting state EEG signals of the subjects using 64-channel electrodes made of sintered Ag/AgCl. The sampling rate was between 0.1 to 100 Hz, and each recording lasted for about 2 minutes per trial. In the experiment, the EEG data of three groups of ordinary people were not successfully extracted in the feature extraction step, so actually the dataset of 11 control subjects and 14 PD patients was used.

2.2. EEG preprocessing

The EEGLab2023 toolbox was employed to preprocess data. Firstly, an initial filtering process was performed on the EEG signals using a band-pass filter within the range of 0.5-70 Hz and a notch filter at 60 Hz to eliminate ambient noise and power frequency noise. Then, Independent Component Analysis (ICA) is applied to the filtered signal. Independent component analysis is to do a reverse operation on the data of the recording point and decompose the data of each recording point into components. After that, any components related to eye drift and blinking were manually identified and rejected.

2.3. Feature Extraction

Feature extraction was conducted using MATLAB 2022b. Given that the dataset was obtained during the resting state, the EEG signal was divided into batches in a balanced manner. Subsequently, features are extracted from these segmented intervals. This research has identified three features from time domain and frequency domain, including beta wave band power, skewness and kurtosis.

In general, individuals with Parkinson's disease tend to exhibit a reduction in the speed of their brain's overall electrical activity, specifically a decrease in the power of beta and gamma waves [7]. Besides, the symmetry and steepness are also important characteristics, in which the PD patients show differences compared to normal people. As Swann said that the EEG signals of healthy people fluctuate similar to sine waves, but with increased asymmetry, and the slope or steepness of these signals holds significance for patients with PD. This characteristic was readily observable in patients who were not taking any medication [9]. Band power reflects the energy in the corresponding band. The skewness measures the degree of deviation from the symmetry of a normal distribution, and the kurtosis represents the peakedness or flatness of a distribution. Therefore, these three features were finally selected since they carry well the above features related to PD.

2.4. Multilayer Perceptron (MLP)

The classification model is based on the MLP model with Keras in Python. The Multilayer Perceptron (MLP) is a straightforward and fundamental neural network. Based on the biological neuron model, we can derive the MLP's fundamental structure. The most common MLP consists of three layers: input layer, hidden layer, and output layer. These layers are all completely interconnected. In this experiment, we trained a 6-layer MLP and the architecture is shown in Fig 1.
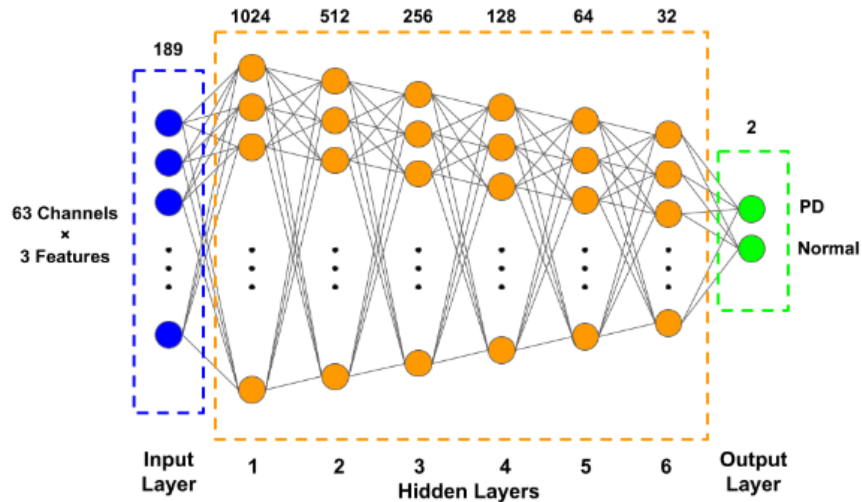

Figure 1. The structure schematic of the MLP neural network

For hidden layers, the degree of generalization and accuracy obtained by first expanding the dataset before compressing it, is higher than that obtained by continuous compression [10]. The ReLu function, which is more suitable for nonlinear activation, is used. Meanwhile, we only have two final outputs, the use of softmax activation function gives better binarization of the results and hence more accurate results. We use a simple train-test split to divide all the dataset into training sets (80%) and testing sets (20%) and optimized for gradient descent using a Nadam function with a convergence rate of 0.0001. Use batch size equals to 500 and 50 epochs to get the training results.

3. Results

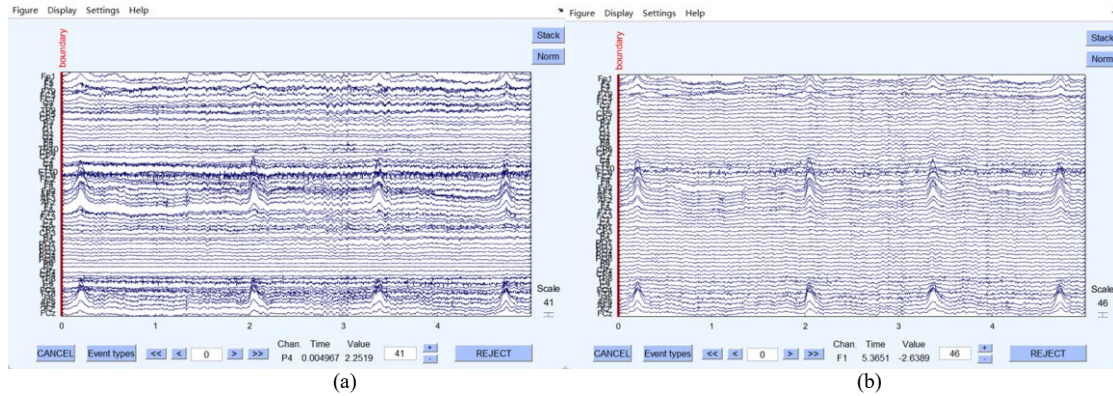(a)   (b)

Figure 2. (a) EEG signal before the filtering
(b) EEG signal after the filtering

The ambient and power frequency noise of the EEG signal were well eliminated after filtering. Fig 2 shows that the filtering worked well since the filtered EEG signal became flatter and less burred. Next, the artifacts were eliminated after ICA, and the dataset was successfully preprocessed, and all the files were output in set format.

Then, through MATLAB, the three features of the preprocessed data have been extracted, and files were converted to csv format. After merging all the files into one, the final dataset used for training was obtained.

After the import of data, many trials have been done on the parameters to get the most ideal model, which especially focused on the number of the layers. By using Nadam function with a batch size of 500 and an epoch of 50, we experimented 10 times with each number of the layers, and Figure 3 shows the corresponding 10 sets of test loss and accuracy and their average values when the MLP neural network has 4 to 9 layers respectively, and the model with 6 layers is the most stable and accurate one with an average test loss of 12.75% and test accuracy of 96.8%, and the training results are shown in the figure 4.

4 Layer		5 Layer		6 Layer		7 Layer		8 Layer		9 Layer	
Test Loss	Test Accuracy	Test Loss	Test Accuracy	Test Loss	Test Accuracy	Test Loss	Test Accuracy	Test Loss	Test Accuracy	Test Loss	Test Accuracy
16.4	96.5	8.3	98.2	15.6	96.5	6.1	98.1	42.8	93.4	8.75	98.08
17.9	96.4	14.3	96.4	11.7	97.5	11.3	98.5	12.1	98	19.9	94
14.5	96	19.4	94.9	16.6	96.5	16.1	95	9.9	97.3	42.9	86.1
218.9	64.5	9.6	98.4	14.9	95.4	15.5	96.5	17.6	96.1	10.1	97
9.8	97.3	17.2	94.1	16.3	94.4	23.4	94.2	9.4	97.7	19.3	93.8
6	98.6	44.9	88.6	7	98.4	9.7	98.1	11.1	97	9.5	98.1
10.7	97.2	16.7	95.9	13.1	97	11.5	97.3	18.3	97	17.9	95.1
9.8	98.3	10.1	97.6	12.6	97	18.9	94	20.4	95	38.4	88.5
23.1	95.3	45.7	79.6	8.1	98	10.2	97.3	16.7	95.2	17.1	95
10.6	97.8	403.6	52.1	11.6	97	12.4	97.5	9.6	97	44.7	89.8
Average		Average		Average		Average		Average		Average	
33.77	93.79	58.98	89.58	12.75	96.8	13.51	96.65	13.9	96.7	24.42	93.04

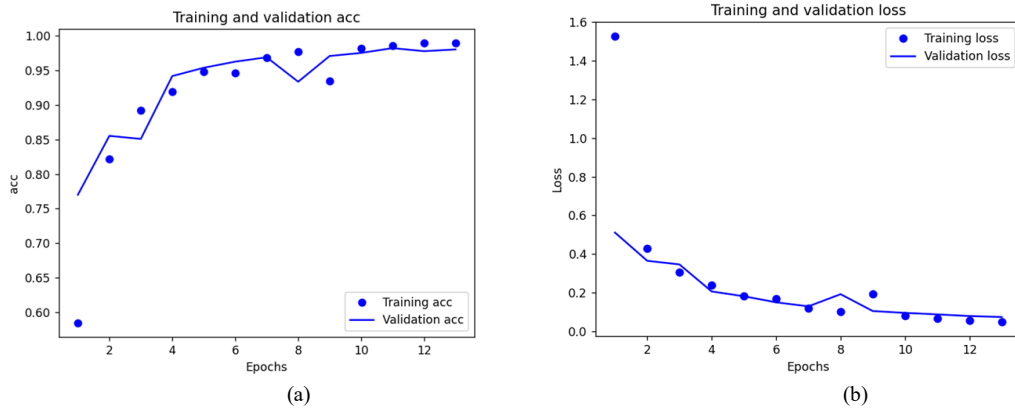**Figure 3.** The table showing the test loss and accuracy of the model with 4 to 9 layers

Figure 4. (a) Training and validation loss and (b) Training and validation accuracy

Since there is no new data to use for the experiment, there is no test on whether the model is overfitted or under-fitted, but despite the small number of subjects (11 normal people and 14 PD patients), the model is still in a good fitting condition.

4. Discussion

There are many ways to diagnose PD medically, but there are few clinical cases of using neural networks to analyze EEG and combine brain function with deep learning. In this research we used the MLP model, which has the advantage of being simple and easy to study, and good performance can be obtained with few subjects.

In the future, when faced with more complex images, we will try to use the GCN model which is unified in a structurally simple and effective way and guarantees diagnostic accuracy and robustness to individual differences [11].

5. Conclusion

In this experiment, we first acquired the EEG dataset from the University of Iowa and preprocessed it by EEGLAB. For the feature extraction part, we selected three features that are highly correlated with Parkinson's disease after consulting the literature. In this way, the computational efficiency of the subsequent model can be improved, the risk of overfitting can be reduced, and the accuracy of the model for Parkinson's disease can be increased.

Then we trained a 6-layer MLP neural network to diagnose PD by using the preprocessed EEG data. Through multiple adjustments of the model parameters such as the number layers, the batch size and the epoch to avoid problems such as underfitting, overfitting and gradient disappearance, we obtained the model with an average test accuracy of 96.8% and an average test loss of 12.7%. Based on the good performance obtained, the model can be used as a reliable and long-term tool to help early clinical diagnosis of Parkinson's disease.

In the future, our group intends to obtain more diverse data from people of different ages, races, and genders to enhance the diversity and accuracy of our model. This may extend to the diagnosis of many different brain diseases.

References

[1]   William Dauer, Serge Przedborski, Parkinson's Disease: Mechanisms and Models, Neuron, Volume 39, Issue 6, 2003, Pages 889-909, ISSN 0896-6273, https://doi.org/10.1016/S0896-6273(03)00568-3.(https://www.sciencedirect.com/science/article/pii/S0896627303005683)

[2]   Vyas, T. *et al.* (2021) 'Deep learning-based scheme to diagnose parkinson's disease', *Expert Systems*, 39(3). doi:10.1111/exsy.12739.

[3]   SuwijnSR, van BoheemenCJ, de HaanRJ, et al. The diagnostic accuracy of dopamine transporter SPECT imaging to detect nigrostriatal cell loss in patients with Parkinson′s disease or clinically uncertain parkinsonism: a systematic review[J]. EJNMMI Res, 2015, 5: 12. DOI: 10.1186/s13550-015-0087-1.

[4]   BajajN, HauserRA, GrachevID. Clinical utility of dopamine transporter single photon emission CT (DaT-SPECT) with (123I) ioflupane in diagnosis of parkinsonian syndromes[J]. J Neurol Neurosurg Psychiatry, 2013, 84(11): 1288-1295. DOI: 10.1136/jnnp-2012-304436.

[5]   PrangeS, MetereauE, ThoboisS. Structural imaging in Parkinson′s disease: new developments[J]. Curr Neurol Neurosci Rep, 2019, 19(8): 50. DOI: 10.1007/s11910-019-0964-5.

[6]   Gupta, G., Pequito, S. & Bogdan, P. Re-thinking EEG-based non-invasive brain interfaces: Modeling and analysis. In 2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS), Porto, Portugal 275–286 (2018).

[7]   Barlow, J. (2023) *UO-led study finds angles in raw EEG data point to parkinson's*, *Around the O*. Available at: https://around.uoregon.edu/content/uo-led-study-finds-angles-raw-eeg-data-point-parkinsons (Accessed: 02 August 2023).

[8]   M. F. Anjum, S. Dasgupta, R. Mudumbai, A. Singh, J. F. Cavanagh and N. S. Narayanan, "Linear predictive coding distinguishes spectral EEG features of Parkinson's disease", Parkinsonism & Related Disorders, vol. 79, pp. 79-85, Oct. 2020.

[9]   Wang, Q. *et al.* (2020) 'Characterization of EEG data revealing relationships with cognitive and motor symptoms in parkinson's disease: A systematic review', *Frontiers in Aging Neuroscience*, 12. doi:10.3389/fnagi.2020.587396.

[10] M. Noshad, Y. Zeng and A. O. Hero, "Scalable Mutual Information Estimation Using Dependence Graphs," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 2962-2966, doi: 10.1109/ICASSP.2019.8683351.

[11] T. Lyu and H. Guo, "BGCN: An EEG-based Graphical Classification Method for Parkinson's Disease Diagnosis with Heuristic Functional Connectivity Speculation," 2023 11th International IEEE/EMBS Conference on Neural Engineering (NER), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/NER52421.2023.10123796.