Development and evaluation of a Parkinson's disease diagnosis system using Deep Learning of resting state EEG and computer vision analysis of brain fMRI scans in a multi-modal system.

Introduction:

Parkinson's disease (PD) is a neurodegenerative disorder that impacts the production of dopamine in the substantia nigra, leading to various effects on the brain, including motor control, memory, and mental health [1]. It affects over 8.5 million people globally and has resulted in approximately 330,000 deaths since 2000 [2]. In the UK, it affects about 1 in 37 people and is expected to increase to around 172,000 in the next six years [3]. Diagnosis involves evaluation by a general practitioner followed by specialist consultation. These assessments, as detailed on the NHS website, rely on qualitative measures. Given developments in ML, we could use some form of machine learning algorithm/combination of algorithms to process and understand quantitative information about the brain (i.e., Magnetic Resonance Imaging and/or Electroencephalography) using the six-stage crisp-dm framework and aid in the diagnosis of PD.

People have been putting time into researching the application of AI in the case of PD diagnosis [4, 5] using several pattern recognition techniques with significant classification accuracy ranging up to 95% for off-medication PD patients using RF-CNN (Random forest-convolutional neural Network). Similarly, CNNs have been used to aid in the diagnosis of PD by processing MRI (Magnetic Resonance Imagery.) [6] This has shown less consistent results than EEG analysis but given a multi-modal approach, it would still be invaluable information. We can aim to create a deep learning model which implements several different techniques which should allow us to create a comprehensive diagnostic model, to identify PD in a patient based on quantitative measurements with high certainty.

Methodology*:*

Data Gathering:

The publicly available data on PD relating directly to brain research is sparse as most data is proprietary. There were three different data sets I was able to gather on the internet for the fMRI data [7,8,9]

EEG data is even more sparse to a point where in its current state I would not be confident in rating any model I could conceive of a high degree of accuracy. This would lead me to weigh the voting system heavily in favour of the MRI dataset. Caveats aside, the two datasets

I was confident in were the “UC San Diego Resting State EEG” dataset [10]. This dataset was collected from a group of PD sufferers and healthy controls who were then told to relax and lightly focus on a screen in front of them.

I then shifted focus to a larger but shorter dataset from the University of New Mexico [11], specifically, their d002 dataset, a collection of 2-minute resting state EEG recordings (split equally between open-eyed and closed-eyed.)

MRI Data Pre-Processing:

Data-Preparation/Cleaning:

A problem I ran into early on while tackling the fMRI data was a lack of publicly available datasets, let alone a consistent file type, having to mix Neuroimaging Informatics Technology Initiative (NIfTI) and Digital Imaging and Communications in Medicine (DICOM) for MRI files meaning the data set for this project would have poor integrity. Luckily, some tools have been developed to convert between the two. In this case, I used dcm2niix [12]. This is a lossless conversion tool meaning it preserves metadata and has little, if any, impact on the data quality which is essential as that removes most risk of artifacting which could otherwise impact the analysis of the images when run through the ML algorithm later on.

We must discuss further data cleaning. Working with fMRI imaging we need to undergo a process called skull-stripping where we remove the non-brain tissues (primarily the skull but also some soft tissue e.g. eyes.) We do this to remove a lot of the unnecessary data as this can add noise to the image which may decrease the accuracy of the model in this case we will use SynthStrip[13] as it can be automated and is highly accurate.

Normalization:

The next stage is going to be data normalisationion. First, we will need to normalize the image intensity. The issue we will run into once again is a corruption of data through the introduction of noise or the removal of actual content we will need for an accurate diagnosis which in this case may cause more issues than benefits given that [14] recognised that intensity normalization can cause significant bias which would cuase the processed image to be in conflict with reality. We would also need to spatially normalize the images but our conversion tools and model viewers auto-normalize the images with low-impactimpact normalization algorithms which means that the fMRI data is ready to be run through our ML system.

EEG Data Pre-Processing:

Data-Preparation/Filtering:

With EEG data, the common practice is to run a noise filter on the complex sinusoidal waves recorded. There are multiple methods of filtering available (e.g., Artifact subspace reconstruction or a band-pass filter.) In this case, I propose we take influence from this MATLAB regression ML model [15] which would allow us to integrate it with our ML model as the first stage of two.

Machine Learning Techniques and Algorithms:

Machine learning algorithms relating to MRI brain scans:

Given our two distinctly different data types, image data and raw numeric data, we must take different approaches. s. As mentioned in the introduction we know that PD affects specific known brain areas. This will help us massively as we know where we need to analyze, which reduces the chances of us encountering situations where the system may notice other irrelevant variations between patients' brains that may be caused by other neurological diseases which gives us an idea of the results we expect when segmenting the imaging. For our purposes, we could adapt the system Boustani, A.E. *et al* used to detect brain tumours [16]. In this paper, they used rand to segment the image and selected elected for the regions consistent with what you would expect of a tumour.

Given the subject matter of disease (which is very personal and would have a massive impact on the individual) I would work to focus on accuracy which is why, while algorithms like YOLO or SSD are generally accurate for their applications, I would recommend using the following algorithms and considering the outputs of each in something like a majority voting system. Doing some further reading I have concluded that using DenseNet as one participant and a 3D-CNN (e.g., U-Net, V-net) as the other.[17] Each of these models will be trained against all 3 of the data sets above and for evaluation, I will reserve a randomly selected group of HC and PD MRIs to use to gauge the accuracy of the system

Machine learning algorithms relating to EEG data:

EEG processing is a little more straightforward as DeepConvNet is a purpose-made CNN tool for EEG data processing [18]. I will train this on the datasets listed above while reserving a part of them for validation.

Weighted voting system:

This system, being based around three AI systems being coalesced into one, will return three different certainties. This means we need a way to consolidate our results. Given we know the accuracy of the systems from our validation and testing just prior I would apply a weighted voting system. The equation I will use to do this is simple and listed below.

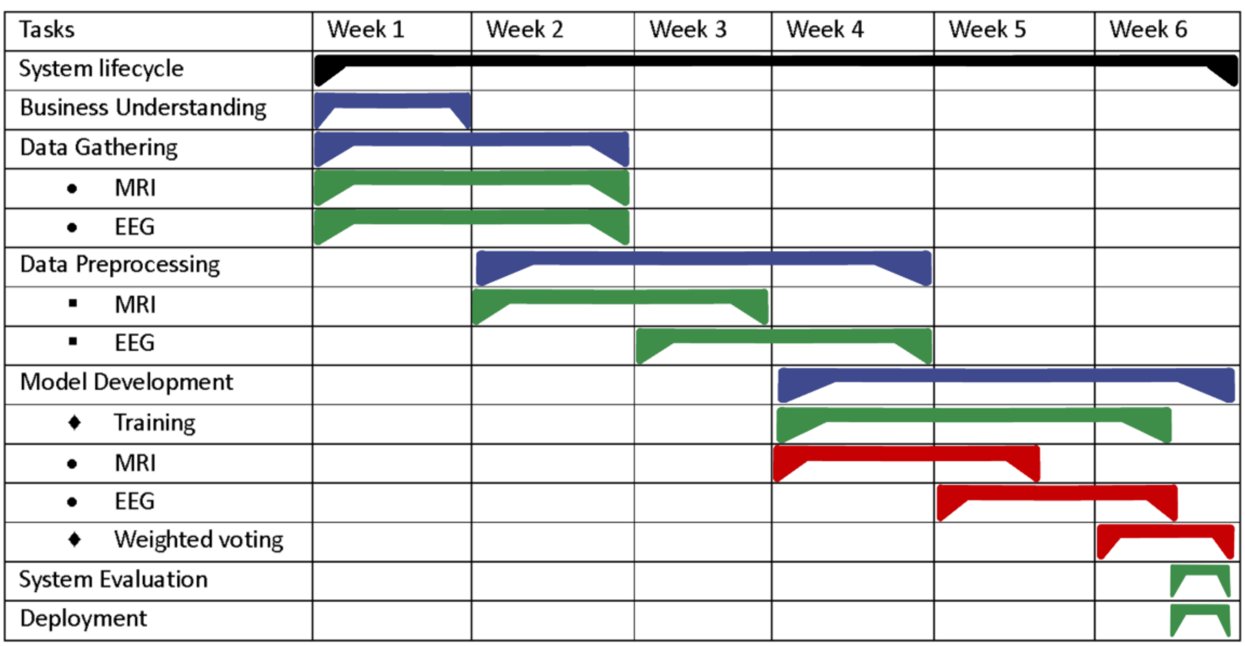
$$C_{\text{sys total}} = \frac{\sum_{i=1}^{n}(C_i \times A_i)}{\sum_{i=1}^{n}(A_i)}$$

Where $C_{\text{sys total}}$ is the total certainty of the system,
C_i is the certainty of the model currently being assessed,
A_i is the Accuracy of the system currently being tested

This equation returns the total certainty of the system based on a normalized weighting of each system. This equation works as a general weighted voting system to determine the certainty of any number of models (n). In this case, n = 3.

Project Plan:

This project, while diverse, is fairly simple with a lot of elements which can be done in parallel. See the diagram below for my projected workflow.



[1] National Health Service (2022) Parkinson's Disease Available at https://www.nhs.uk/conditions/parkinsons-disease/

[2] World health Organization (2023), Parkinson's Disease, Available at: https://www.who.int/news-room/fact-sheets/detail/parkinson-disease

[3] Parkinson's UK (2023), Reporting on Parkinson's: information for journalists, Available at: https://www.parkinsons.org.uk/about-us/reporting-parkinsons-information-journalists

[4] Aljalal, M. *et al.* (2022) 'Parkinson's Disease Detection from Resting-State EEG Signals Using Common Spatial Pattern, Entropy, and Machine Learning Techniques,' *Diagnostics* 12(5), p. 1033. https://doi.org/10.3390/diagnostics12051033

[5] Vanegas, M.I. *et al.* (2018) 'Machine learning for EEG-based biomarkers in Parkinson's Disease,' *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2661–2665. https://doi.org/10.1109/bibm.2018.8621498

[6] S. Zhu, "Early Diagnosis of Parkinson's Disease by Analyzing Magnetic Resonance Imaging Brain Scans and Patient Characteristic," *2022 10th International Conference on Bioinformatics and Computational Biology (ICBCB)*, Hangzhou, China, 2022, pp. 116-123, doi: 10.1109/ICBCB55259.2022.9802132

[7] Kumar, R. (2024). Parkinson's Disease Dat and MRI Scans. Kaggle. Available at: https://www.kaggle.com/datasets/rishikjha/parkinsons-disease-dat-and-mri-scans

[8] Madge, V. *et al.* (2023) 'A dataset of multi-contrast unbiased average MRI templates of a Parkinson's disease population,' *Data in Brief*, 48, p. 109141. https://doi.org/10.1016/j.dib.2023.109141

[9] Eunus, S.I. (2023). Parkinson's disease_FMRI+Image. Kaggle. Available at: https://www.kaggle.com/datasets/salmaneunus/parkinsons-disease-fmri-images

[10] Rockhill, A. et al., 2020. UC San Diego Resting State EEG Data from Patients with Parkinson's Disease. [Dataset] OpenNeuro. Available at: https://doi.org/10.18112/openneuro.ds002778.v1.0.2

[11] Rockhill, A. et al., 2020. UC San Diego Resting State EEG Data from Patients with Parkinson's Disease. [Dataset] OpenNeuro. Available at: https://doi.org/10.18112/openneuro.ds002778.v1.0.2

[12] C. Rorden (2024), dcm2niix, GitHub, Available at: http://github.com/rordenlab/dcm2niix

[13] Hoopes, A. *et al.* (2022) 'SynthStrip: skull-stripping for any brain image,' *NeuroImage*, 260, p. 119474. https://doi.org/10.1016/j.neuroimage.2022.119474

[14] López-González, F.J. *et al.* (2020) 'Intensity normalization methods in brain FDG-PET quantification,' *NeuroImage*, 222, p. 117229. https://doi.org/10.1016/j.neuroimage.2020.117229

[15] *Denoise EEG Signals Using Deep Learning Regression with GPU Acceleration - MATLAB & Simulink - MathWorks United Kingdom* (no date). https://uk.mathworks.com/help/signal/ug/ocular-artifact-removal-from-electroencephalogram-signals-using-deep-learning-regression.html

[16] Boustani, A.E. *et al.* (2020) 'MRI brain images classification using convolutional neural networks,' in *Advances in intelligent systems and computing*, pp. 308–320. https://doi.org/10.1007/978-3-030-36674-2_32

[17] Huang, G. *et al.* (2016) 'Densely connected convolutional networks,' *arXiv (Cornell University)* [Preprint]. https://doi.org/10.48550/arxiv.1608.06993

[18] Autthasan, P. *et al.* (2022) 'MIN2NET: End-to-End Multi-Task Learning for Subject-Independent Motor Imagery EEG Classification,' *IEEE Transactions on Biomedical Engineering*, 69(6), pp. 2105–2118. https://doi.org/10.1109/tbme.2021.3137184