



OPEN

# Enhancing early Parkinson's disease detection through multimodal deep learning and explainable AI: insights from the PPMI database

Vincenzo Dentamaro<sup>1,3</sup>✉, Donato Impedovo<sup>1,3</sup>, Luca Musti<sup>1,3</sup>, Giuseppe Pirlo<sup>1,3</sup> & Paolo Taurisano<sup>2,3</sup>

Parkinson's is the second most common neurodegenerative disease, affecting nearly 8.5M people and steadily increasing. In this research, Multimodal Deep Learning is investigated for the Prodromal stage detection of Parkinson's Disease (PD), combining different 3D architectures with the novel Excitation Network (EN) and supported by Explainable Artificial Intelligence (XAI) techniques. Utilizing data from the Parkinson's Progression Markers Initiative, this study introduces a joint co-learning approach for multimodal fusion, enabling end-to-end training of deep neural networks and facilitating the learning of complementary information from both imaging and clinical modalities. DenseNet with EN outperformed other models, showing a substantial increase in accuracy when supplemented with clinical data. XAI methods, such as Integrated Gradients for ResNet and DenseNet, and Attention Heatmaps for Vision Transformer (ViT), revealed that DenseNet focused on brain regions believed to be critical to prodromal pathophysiology, including the right temporal and left pre-frontal areas. Similarly, ViT highlighted the lateral ventricles associated with cognitive decline, indicating their potential in the Prodromal stage. These findings underscore the potential of these regions as early-stage PD biomarkers and showcase the proposed framework's efficacy in predicting subtypes of PD and aiding in early diagnosis, paving the way for innovative diagnostic tools and precision medicine.

Parkinson's disease (PD) is a progressively evolving disease of undefined etiology (idiopathic) related to neuronal degeneration affecting the substantia nigra<sup>1</sup>. The disease usually manifests between 45 and 75 years of age, with the peak age of onset in the sixth decade, affecting a slightly higher incidence in men<sup>2</sup>. The causes of PD are still unknown, but it is believed to originate from genetic and environmental factors. Numerous factors have been proposed throughout the years as predisposing to the disease, including trauma<sup>3</sup>, emotional upset<sup>4</sup>, overwork<sup>5</sup>, exposure to cold<sup>6</sup>, and rigid personality<sup>7</sup>. Idiopathic PD is observed in all countries, all ethnic groups, and all socioeconomic classes, although the incidence in African Americans is only one-quarter than in whites<sup>8</sup>.

The cardinal motor manifestations such as bradykinesia-akinesia, rigidity, resting tremor, and postural instability constitute the fundamental clinical indicators of PD<sup>9</sup>. These symptoms manifest as a gradual slowing of movement, diminished or absent motor function, muscle stiffness, facial expression masking, tremors in the hands, stooped posture, and shuffling gait. Their onset typically occurs during the advanced stages of the disease and may initially be disregarded by family members as part of normal aging. Even worse is the same patient may be unaware of his symptoms for a long period. Early indicators may include complaints of discomfort in the low back, neck, shoulders, knees, or hips, along with generalized weakness<sup>10</sup>.

Additionally, reduced frequency of blinking emerges as an early sign<sup>11,12</sup>, with Parkinsonian patients exhibiting a significant decrease in the normal blink rate (24/min) to 4 to 12/min<sup>11</sup>, accompanied by a subtle widening of the eyelids, resulting in a fixed stare.

<sup>1</sup>Dipartimento di Informatica, University of Bari Aldo Moro, 70125 Bari, Italy. <sup>2</sup>Dipartimento di Biomedicina Traslazionale e Neuroscienze (DiBraIN), University of Bari Aldo Moro, Bari, Italy. <sup>3</sup>These authors contributed equally: Vincenzo Dentamaro, Donato Impedovo, Luca Musti, Giuseppe Pirlo and Paolo Taurisano. ✉email: vincenzo.dentamaro@uniba.it

The worsening of the motor symptoms greatly hinders daily activities. Handwriting undergoes a tremulous, diminutive, and cramped alteration known as micrographia<sup>13</sup>. Speech gradually becomes softer, monotone, and indistinct, eventually diminishing to a whisper, characterized as cluttered speech<sup>14</sup>. The speech disorder was initially identified as hypokinetic dysarthria by Caekebeke et al.<sup>15</sup>, who attribute it to a combination of respiratory, phonatory, and articulatory issues. Mobility is also severely impaired almost reduced to a shuffling gait; balance is frequently compromised, with forward or backward movement characterized by a sensation of "chasing" the body's center of gravity through a succession of increasingly rapid short steps to prevent falls, a phenomenon referred to as festination<sup>16</sup>.

However, the symptoms that are most troublesome, ignored, and worrisome are those that are not related to movement<sup>17</sup>. These include loss of smell, constipation, depression, anxiety, REM sleep behaviour disorder, and visual dysfunctions can occur years before the first diagnosis<sup>18,19</sup> in the so called prodromal phase. Identifying these symptoms early on can provide a great opportunity to find potential biomarkers. The overall course of the disease is quite variable. In most patients, the mean period from inception of the disease to a chairbound state is 7.5 years, but with a wide range<sup>20</sup>.

The wide range of symptoms makes the diagnosis of PD very challenging<sup>21</sup>: it is difficult to distinguish typical PD from the many Parkinsonian syndromes caused by other degenerative conditions or by medications and toxins and to distinguish the Parkinson tremor from other types, especially essential tremor<sup>22</sup>. Clinical diagnostic accuracy has greatly improved over the last decade, going from an accuracy rate of 83.9% (follow-up diagnosis)<sup>23</sup> to 97.2% made by experts<sup>24</sup>, although it remains suboptimal. The detection of PD in the early stages, the so-called prodromal stage, is even more challenging yet extremely important because early treatment can slow down the neurodegenerative progression and can even prevent the emergence of clinical PD symptoms. In this matter, studies explored new recruitment strategies for disease-modification trials in various prodromal PD cohorts to boost the diagnosis in the early stages<sup>25</sup>. For this reason, computer-aided systems have been developed to assist experts in their diagnosis. The constant and rapid evolution in the field of Artificial Intelligence (AI) and machine learning (ML) has been proven essential in biomedical data analysis, yielding promising results in a wide variety of scenarios, such as the early detection of Alzheimer's disease using MRI<sup>26</sup>.

In the past decade, the field of medical diagnostics has witnessed a significant improvement with the integration of machine learning and deep learning algorithms for the automated detection and classification of PD. These advancements have been particularly focused in the utilization of neuroimaging techniques<sup>27</sup>, which serve as the primary data source for the early identification of PD. The recent advancements of sophisticated 3D reconstruction technologies have further enhanced the quality of imaging, thereby contributing to the accuracy and reliability of PD diagnostics<sup>28</sup>. Also advances in face recognition with improved Inception-ResNet-V1 networks can offer potential for non-invasive PD diagnostics<sup>29</sup>. Additionally, developing emotionally intelligent robots to interact with PD patients can enhance personalized care and support, promising significant improvements in patient outcomes<sup>30</sup>.

In 2022, Saravanan et al.<sup>31</sup> provided a comprehensive analysis of the influence of ML and deep learning approaches exploring new research areas ranging from various modalities, including brain signals (MRI<sup>32,33</sup>, EEG<sup>34–37</sup>, SPECT<sup>38,39</sup>, PET<sup>40,41</sup>) and physiological signals (speech<sup>42,43</sup>, gait<sup>44,45</sup>, handwriting and sensor data<sup>46</sup>). Additionally, H.W. Loh et al.<sup>47</sup> reviewed 63 studies from which they showed that deep learning models can achieve high prediction accuracy for PD, especially convolutional models for neuroimaging classification and handwriting analysis. However, most research predominantly focus on binary issues, primarily directed towards the screening and categorization of individuals as either healthy controls (HC) or PD patients. Few studies delve into the more intricate matters surrounding the staging of PD or the more complex and challenging pursuit desired by clinicians and scientists, which involves pinpointing the characteristics and early biomarkers indicative of PD. Equally important, there is also a general lack of data availability and model interpretability in existing research. Furthermore, most research focuses on a single modality of data. Given the wide range of symptoms of PD, it is beneficial to train a multimodal model combining data from multiple sources to resemble the multimodal nature of clinical-expert decision-making. Adopting multimodal learning within the medical domain facilitates the development of models aimed at enhancing the accuracy, predictability, and interpretability of medical diagnostics<sup>48</sup>. Studies reported that the analysis of multimodal data (including imaging and clinical data) through machine learning was helpful in detecting brain abnormalities in PD.

Prashanth et al.<sup>49</sup> used a combination of non-motor features of sleep behaviour disorder (RDB) and olfactory loss, cerebrospinal fluid (CSF) measurements, and Single Photon Emission Computed Tomography (SPECT) imaging markers obtained from the Parkinson's Progression Markers Initiative (PPMI) database<sup>50</sup> (401 PD, 183 HC) to develop diagnostic models to classify subjects into early PD and healthy normal, using Naive Bayes, Support Vector Machines (SVM), Boosted trees and Random forests (RF) classifiers trained with tenfold cross-validation. They achieved the best performances with SVM reaching 96.40% accuracy, 97.03% sensitivity, 95.01% specificity, and 98.88% area under ROC.

Long et al.<sup>51</sup> collected from 19 PD and 27 HC resting-state functional magnetic resonance imaging (rsfMRI) and structural images, from which they extracted six features: amplitude of low-frequency fluctuations (ALFF), regional homogeneity (ReHo), regional functional connectivity strength (RCFS), volume of grey matter, white matter, and CSF. They trained an SVM with leave-one-out cross-validation, achieving the best results in multi-modal feature combination (86.96% accuracy, 78.95% sensitivity, 92.59% specificity).

Zhu et al.<sup>52</sup> proposed a hybrid model trained (80% training, 20% testing, and validation) with the combination of symptoms data and MRI data from the PPMI database to accurately diagnose PD severity. From their findings, the hybrid model (3D CNN + Logistic Regression) achieved higher accuracy (94%) compared to single models' accuracies of 77% and 68%, respectively. More recently, the multimodal learning approach was the suggested strategy to improve early clinical diagnosis of PD from a systematic evaluation<sup>53</sup> of all the deep learning methods used for this purpose but based solely on motor signs.

Differently, Huang et al.<sup>54</sup> introduced a framework based on graph structure learning and node clustering for more interpretable multimodal structural brain networks to perform early PD detection, achieving highest accuracy of 95.5% on the PPMI dataset.

From the literature review above, it is clear that there is a lack of studies evaluating extensively the application of multimodal deep learning to Parkinson's Prodromal stage recognition.

Based on the limitations and current breakthroughs in machine learning medical diagnosis, this paper aims to detect PD in the prodromal stage via multimodal learning. Imaging and clinical data were obtained from the PPMI database. Imaging data was used to train deep models (ResNet 3D, DenseNet 3D, Vision Transformer 3D), and tabular clinical data was used to train traditional classification algorithms (RF, SVM, XGBoost) and a novel squeeze-and-excitation<sup>55</sup> inspired network called Excitation Network (EN). Then, three multimodal models were built by fusing each of the deep models with the EN. The training of every model has been carried out using stratified tenfold cross-validation. Finally, explainable AI techniques (Integrated Gradients, Attention heatmaps, Feature Importance) have been used to understand what the models actually learned.

The innovations of the proposed framework are the following:

- Integration of the novel Excitation Network (EN) for processing clinical data, outperforming traditional machine learning models such as Random Forest and SVM.
- Utilization of joint co-learning approach for multimodal fusion, enabling end-to-end training of deep neural networks and allowing the models to learn complementary information from both modalities.
- Employing state-of-the-art deep learning architectures, including ResNet, DenseNet, and Vision Transformer (ViT), to analyze 3D MRI scans.
- Application of Explainable AI (XAI) techniques, such as Integrated Gradients for ResNet and DenseNet, and Attention Heatmaps for ViT, to interpret the decision-making processes of the deep learning models.
- Identification of brain regions critical to prodromal PD pathophysiology, such as the right temporal area and left pre-frontal lobe, through XAI analysis of the best-performing model, DenseNet.
- Recognition of the potential role of lateral ventricle enlargement as an indicator of prodromal PD, as highlighted by the Vision Transformer model.

From a practical point of view, the multimodal models proposed here would allow family doctors to promptly call for an expert by using an innovative health record software that integrates structural MRI brain scans with cardinal symptoms of PD (Boolean presence of tremor, rigidity, bradykinesia, and postural instability) as well as hereditary symptoms, highlighting all at-risk patients in one shot and without performing any psychological or motor tests.

## Results

### Data acquisition

A total of 34 (37.8%) female, and 56 (62.2%) male patients were obtained from the PPMI database (“[Methods](#)”). Among the female patients, 18 (53%) are HC and 16 (47%) are Prodromal, while 32 (57%) male patients are HC and 24 (43%) are Prodromal. The mean age of all the patients is  $62.78 \pm 8.92$  (53–72) years. From each patient, it has been retrieved their 3D MRI scan along with their clinical data. Once all the data was pre-processed (in “[Methods](#)” section), several deep and machine learning models were trained. The deep models receive as input the MRI scans and, by means of 3D convolutional layers integrated within ResNet and DenseNet architectures or Vision Transformer (ViT), can extract the high-level feature representations which are then vectorized for the final binary classification (Prodromal or HC). ResNet and DenseNet were employed as they are the state-of-the-art convolutional neural networks for computer vision tasks, while ViT was used because in the last few years emerged as a valid alternative to convolutional networks. While the traditional machine learning models (RF, SVM, XGBoost) and EN receive as input a selection of the clinical data features and output the prediction class. In total twelve features were selected from all the 79 features available from the clinical dataset.

The major reason for this choice was to train the simplest possible model with the fewest possible features that are immediately recognizable, don't need additional medical testing and are available in most situations. The twelve selected features concern the cardinal symptoms of PD (tremor, rigidity, bradykinesia, and postural instability) and hereditary symptoms from the family and grandparents. A detailed description of the dataset is included in the “[Methods](#)” section.

### Comparison between multimodal models and unimodal models

The combination of heterogeneous features can improve the diagnostic accuracy of machine learning medical models, just like a multi-disciplinary team examines data coming from multiple different sources to have a global overview of the patient's condition and recommend the best treatment. The integration of heterogeneous features may encompass supplementary data contributed by the individual modalities, alongside shared information inherent to all modalities.

There are several ways to perform multimodal learning<sup>56–58</sup>. The most prominent are:

- Early feature fusion, where features from different modalities are combined before feeding them into the model.
- Late decision fusion, where separate models, one for each modality are trained and their final predictions combined at later stage.

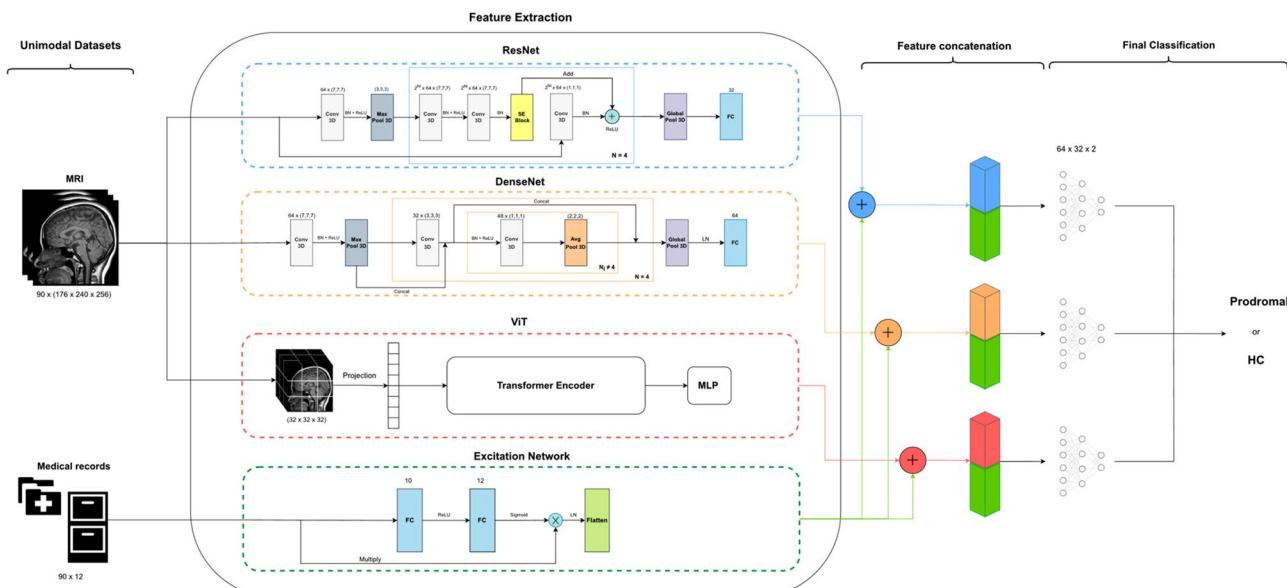
- Intermediate fusion or hybrid fusion, which combines early and late fusion by integrating the various modalities into the model at an intermediate level.
- Joint co-learning, where the model optimization is performed on all modalities simultaneously, influencing the learned representation of each modality.

In this work, the selected multimodal fusion is the joint co-learning, where the optimization problem is tackled as a whole, and as a consequence, the deep neural network is trained end-to-end with backpropagation. This allows the model to learn complementary information of both modalities; it finds intermediate representations that are robust across modalities also mitigating the noise within data, while also learning a unified representation of the problem<sup>56</sup>. On the other hand, Early fusion technique can improve the simplicity of the model thanks to the direct integration of the combined features in a singular architecture. However the process of creating a comprehensive representation from the heterogeneous features is not a trivial task and could also lead to the amplification of irrelevant features or noise, thereby impacting negatively the model performances. Instead, Late fusion allows to separately optimize individual modality by training each model in separate processing paths but it is not inherently capable of learning correlation between modalities, potentially reducing overall performances and synergies.

In this specific case, embeddings coming from Deep Learning models for MRI classification are concatenated with the intermediate representations of the health records (tabular data) learned using the EN as shown in (Fig. 1). The main reason of adopting intermediate fusion over early and late fusion, is its better suitability for deep learning approaches and the ability to learn cross-modality interactions from the marginal representations<sup>57</sup> leading to an improvement in the overall classification task.

### Models evaluation

The models have been tenfold cross-validated (using stratified cross validation) and evaluated by monitoring the average values and standard deviation of traditional performance metrics such as Accuracy, AUC, Precision, Recall, and F1-macro (“Methods”). The results in Table 1 show overall better performances for the multimodal



**Figure 1.** Architecture overview. Multimodal network architectures (BN = batch normalization, LN = layer normalization, FC = fully connected layer, SE = squeeze-excitation).

Model	Accuracy (%)	AUC (%)	Precision (%)	Recall (%)	F1-macro (%)
ResNet	95.5 ( $\pm 5.4$ )	98 ( $\pm 4.3$ )	96.3 ( $\pm 4.7$ )	94.8 ( $\pm 6.7$ )	94.9 ( $\pm 6.2$ )
DenseNet	93.3 ( $\pm 7$ )	97.9 ( $\pm 4.8$ )	94.2 ( $\pm 6.6$ )	93.6 ( $\pm 6.8$ )	92.9 ( $\pm 7.7$ )
ViT	76.1 ( $\pm 17.5$ )	80.9 ( $\pm 24.8$ )	66 ( $\pm 31$ )	73 ( $\pm 20.8$ )	66.7 ( $\pm 26.5$ )
ResNet + EN	95.5 ( $\pm 5.4$ )	98.3 ( $\pm 3.5$ )	96.2 ( $\pm 4.8$ )	95.2 ( $\pm 6.1$ )	95.1 ( $\pm 5.9$ )
DenseNet + EN	96.6 ( $\pm 5$ )	98.9 ( $\pm 3.3$ )	96.9 ( $\pm 4.8$ )	96.9 ( $\pm 4.8$ )	96.5 ( $\pm 5.2$ )
ViT + EN	93.4 ( $\pm 7$ )	96.4 ( $\pm 6$ )	94.2 ( $\pm 6$ )	93.8 ( $\pm 7.6$ )	93 ( $\pm 8$ )

**Table 1.** Average performances in stratified 10-Fold CV for MRI models and multimodal models with respective standard deviation.

models with respect to their unimodal counterpart with a slight exception for ResNet. ResNet proved to be the best architecture for training on only the MRI scans, reaching an average F1-Score of 94.9%. However, when paired also with the clinical data, there is a very small improvement of just 0.02% in F1-Score. Instead, DenseNet performed slightly worse than ResNet with just the MRI data (~ 2% F1) in unimodal learning but reached the highest performances in all metrics (96.5% F1-Score) among all models with the help of clinical data, with a + 3.6% increase in F1-Score. Continuing, it was expected ViT to perform worse than ResNet (~ 28.2% F1-Score) and DenseNet (~ 26.2% F1-Score), primarily based on its dependence on being trained on huge datasets to reach state-of-the-results, which is rarely possible with medical datasets where data augmentation is not suggested. ViT results also have the highest standard deviation (minimum  $\pm$  26.5% F1-Score, maximum  $\pm$  31% precision) in all metrics across all models, meaning that the model's stability is very low across data. Additionally, the concatenation with the clinical features boosted its performances (+ 26.3% F1-Score) reaching 93% F1-Score, almost comparable with the other deep architectures.

The integration of clinical features is posited to have facilitated the reclassification of erroneous predictions into accurate ones, thereby engendering a notable enhancement in the consistency of the outcomes. This assertion is corroborated by the observed reduction in the standard deviation across the evaluated metrics. ViT concatenated with EN strongly confirms the initial hypothesis of the rise in performances obtained with multimodal learning compared with unimodal learning.

On the other hand, the results reported in Table 2 using just the clinical data proved to be worse than just the training with the MRI scans, reaching the highest F1-Score with XGBoost (78.4% F1-Score) outperforming all the other tabular models (SVM, Random Forest and EN). However, apart from XGBoost, EN is the second-best choice outperforming Random Forest and SVM with 72.8% F1-Score. Additionally, EN is also the only candidate for the multimodal co-learning, thus allowing to train the multimodal model end-to-end within the same framework. Concluding, DenseNet combined with EN proved to be the best model with 96.5% of F1-Score and just 5.2% of standard deviation.

### Explainable artificial intelligence (XAI) results

Reliability and trust are essential when building medical diagnostic deep models. Knowing why and how a black-box model came to its conclusion is extremely important in healthcare, where decisions can be a matter of life and death. XAI techniques can help understand how the network came to its decision. Integrated Gradients (IG) for ResNet and DenseNet and Attention Heatmap (in “Methods”) for ViT have been used for this purpose, that is, to identify the regions where the networks focus the most and examine if, in current literature, there is a known correlation between such region and PD. 23 images (10 HC, 13 Prodromal) have been tested for explainability to discover any underlying pattern from an axial and sagittal point of view (“Methods”). In total 84 maps have been generated for the Prodromal class (Fig. 2a) and 66 maps for the Healthy class (Fig. 2b). Additionally, twelve more maps have been generated from the ReLU difference of the Prodromal and Healthy class (and vice versa) (Fig. 2c).

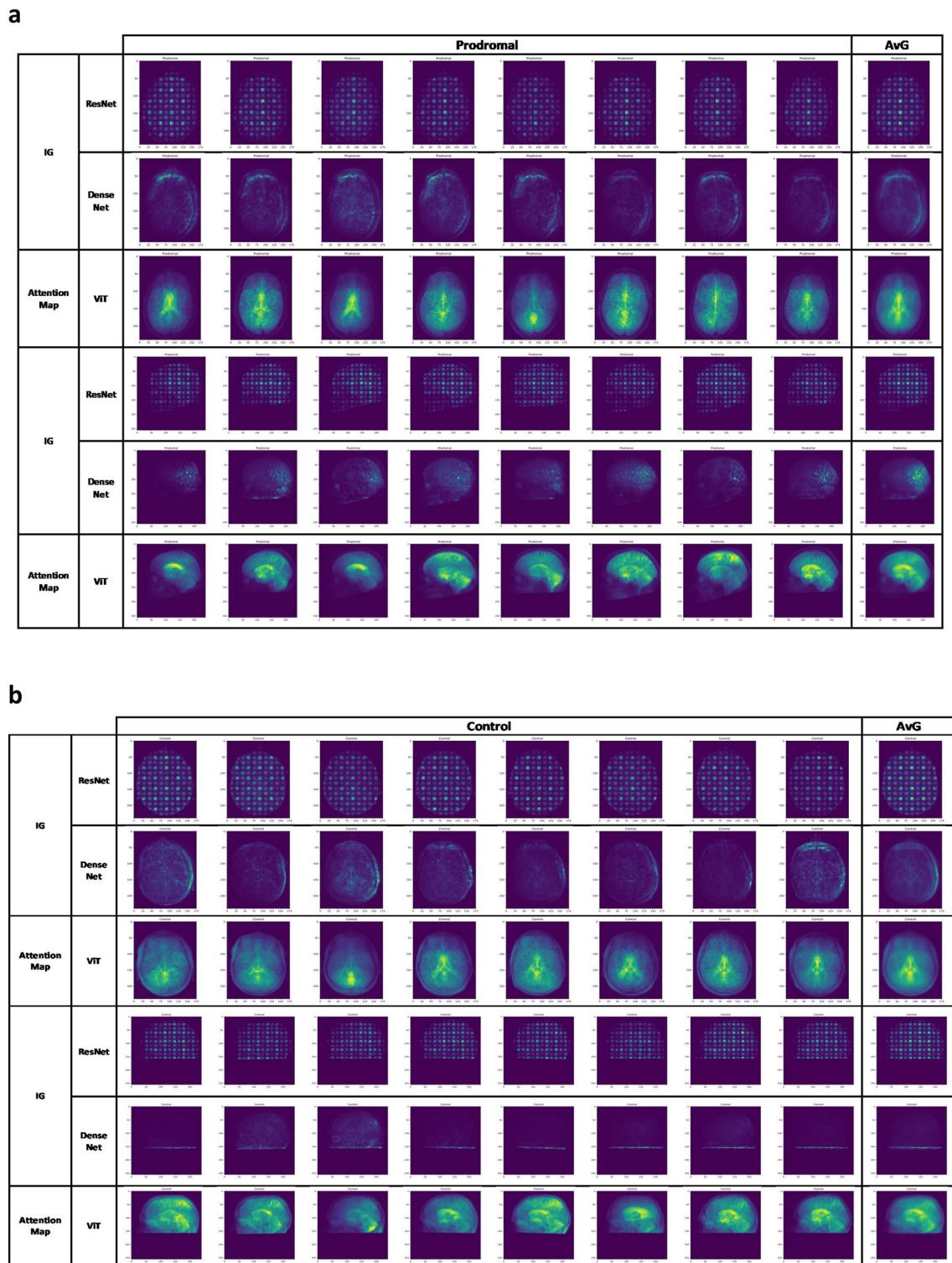
Despite ResNet being the best unimodal model, it provided the worst explainable maps in general from a human interpretability perspective. Its representations are not easily interpretable and there is no clear distinction between Control and Prodromal maps. Much more promising are the DenseNet and ViT maps. The DenseNet maps, seen both from an axial and sagittal point of view, show some common patterns. The singular prodromal maps seem to focus on the right temporal area, while others also focus on the left pre-frontal lobe. The average map shows both regions, which are even more marked in Prodromal-HC difference, meaning that those regions helped to discriminate prodromal patients over healthy patients.

According to the literature, PD patients can have cognitive impairment that is associated with prefrontal dopaminergic dysfunction<sup>58</sup>. Moreover, there is also a correlation between the freezing of gait and the pre-frontal cortex, whose activation becomes greater when trying to compensate for the reduction of automatic control of movement<sup>59</sup>. Left pre-frontal area stimulation via electrodes area has also been connected to a worsening of depressive symptoms<sup>60</sup>. Additionally, a study<sup>61</sup> showed that patients who converted to the prodromal stage with mild cognitive impairment showed significant right temporal atrophy at baseline and extensive atrophy at follow-up. On the other hand, the DenseNet Control maps seem to focus on the part of the skull that covers the right temporal lobe. It is hypothesized that, since a healthy brain is being examined, the focus on regions external to the brain is considered to be a positive result, meaning the network found no abnormalities in the brain.

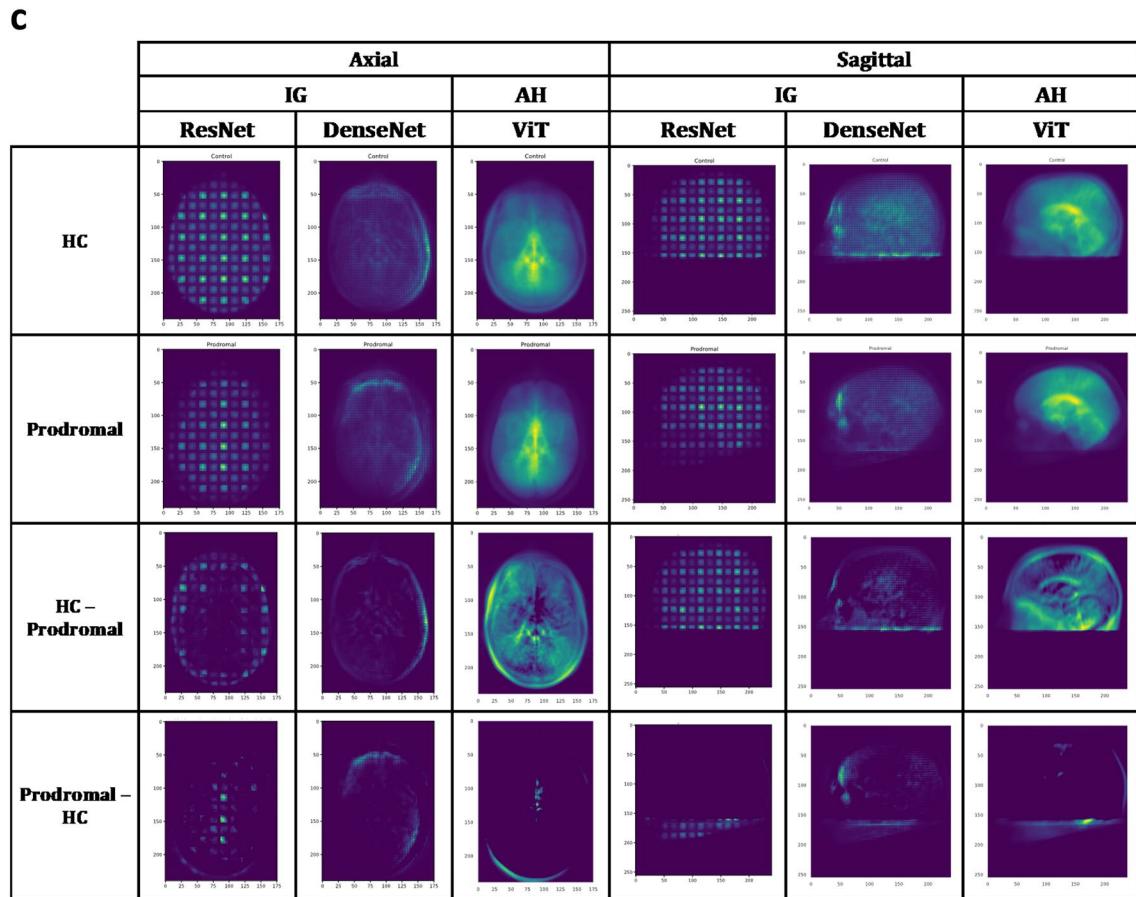
The ViT maps also show some common patterns but there is less distinction between prodromal and control maps compared to DenseNet. The singular prodromal ViT maps focus on the lateral ventricles. In literature, enlargement of the lateral ventricles has been associated with the developing of prodromal PD in the future<sup>62,63</sup>. From the left and right view of the 3D brain reconstruction of ViT (Fig. 3a) it is also possible to slightly see the third ventricle whose width has been correlated to cognitive performance<sup>64</sup>. The 3D brain reconstructions have been generated by converting the 3D attribution maps into the Nifti format and visualized through the Slicer

Model	Accuracy (%)	AUC (%)	Precision (%)	Recall (%)	F1-macro (%)
EN	74.7 ( $\pm$ 10.9)	72.6 ( $\pm$ 16.3)	78.2 ( $\pm$ 11.7)	74 ( $\pm$ 10.1)	72.8 ( $\pm$ 10.7)
RF	74.7 ( $\pm$ 9.8)	72.1 ( $\pm$ 9.1)	78.7 ( $\pm$ 11.7)	72.4 ( $\pm$ 9)	71.9 ( $\pm$ 9.5)
SVM	74.7 ( $\pm$ 9.8)	70.8 ( $\pm$ 13.2)	69.5 ( $\pm$ 20)	70.9 ( $\pm$ 13.4)	68.6 ( $\pm$ 16.7)
XGBoost	80.2 ( $\pm$ 13.8)	78.3 ( $\pm$ 13.8)	82.9 ( $\pm$ 14)	78.4 ( $\pm$ 13.7)	78.4 ( $\pm$ 14.4)

**Table 2.** Average performances in stratified 10-Fold CV for tabular models.



**Figure 2.** XAI images. (a) XAI of the Axial and Sagittal images of 8 Prodromal patients. Each image per column belongs to the same patient, while the AvG map corresponds to mean of all the images on the same row. Only 8 images out of 13 are shown for the sake of space, but the AvG is still calculated based on all the patients. (b) XAI of the Axial and Sagittal images of 8 HC patients. Same format as (a). (c) Overview of the AvG XAIs and difference between the Prodromal and Control class. The last two rows show the ReLU difference between the HC class and the Prodromal class (and vice versa).



**Figure 2.** (continued)

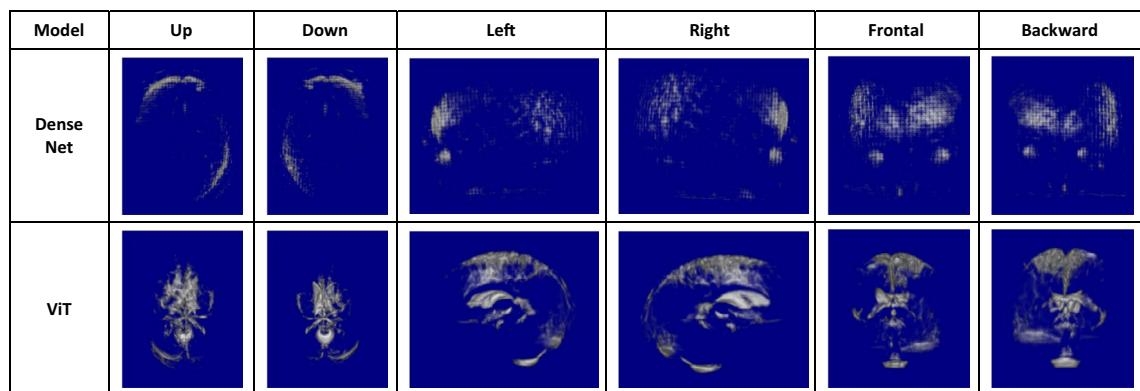
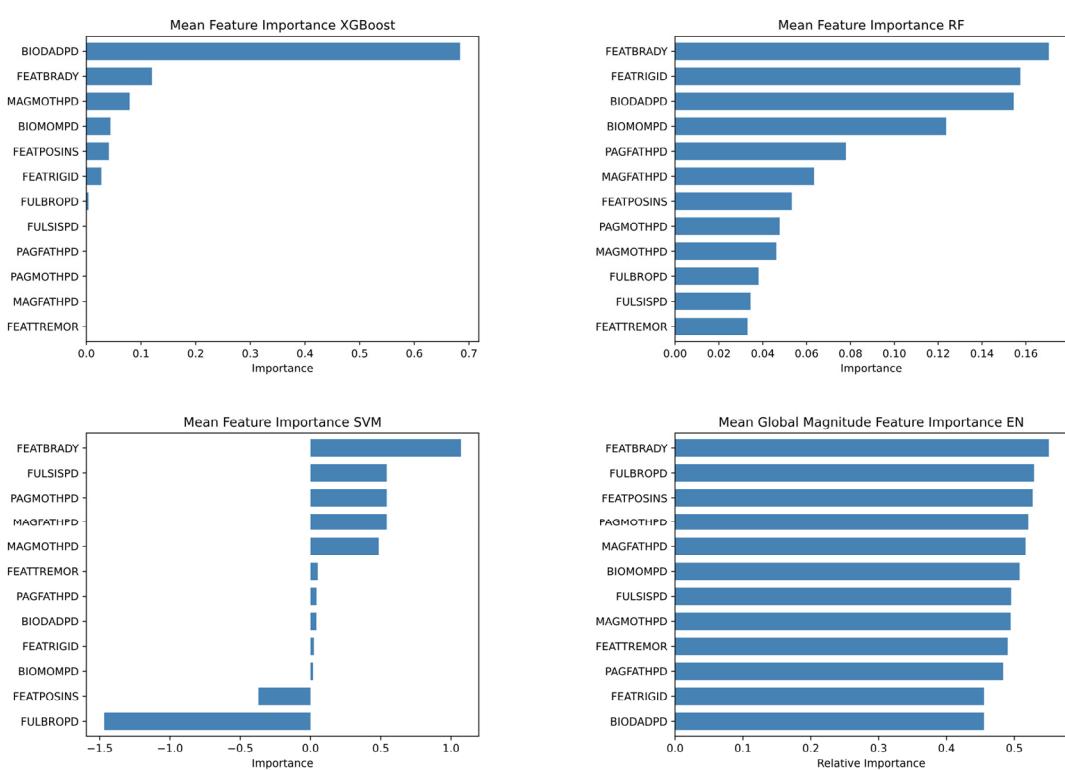
5.4.1 software for Windows in order to capture precisely all the regions of interest and their location in a volumetric view. The 3D brain reconstructions have been generated solely for DenseNet and ViT as ResNet does not provide any meaningful insight.

The differences observed in the regions highlighted by the various models and attention mechanisms can be attributed to various aspects. The interpretations achieved by the models may be reliant on the quantity and quality of the training data. Few data can lead to overfitting and therefore the interpretations may look very consistent between different examples of the same class. This limitation can be partially mitigated by employing cross-validation techniques, like in this work, the stratified cross-validation was used to allow the models to better generalize on unseen data. Equally important is the correlation between the architectural unit (convolution or attention mechanism) that generates the feature maps and the corresponding highlighted region. Convolutional Neural Networks (CNNs) like ResNet and DenseNet tend to focus on local and hierarchical features, with DenseNet's dense connections leading to more fine-grained region highlighting compared to ResNet. In contrast, the Vision Transformer (ViT) utilizes a global attention mechanism, allowing it to capture broader contextual information, thus highlighting larger areas that may not be as granular as those identified by CNNs. These architectural differences influence the granularity and specificity of the highlighted regions, which in turn impacts the interpretability and reliability of our findings. For instance, the CNN models' focus on specific, localized areas like the right temporal and left pre-frontal regions aligns well with existing literature on early PD biomarkers. This specificity enhances the interpretability of these models in a clinical context. On the other hand, ViT's broader focus on areas like the lateral ventricles, while less specific, corroborates with studies linking ventricular enlargement to cognitive decline in PD, offering a complementary perspective.

Lastly, from the explainability of the tabular models (Fig. 3b), XGBoost highly discriminated one feature (BIODADPD) over the others. However, only 10–15% of PD cases are inherited<sup>65</sup>, but people who get early-onset PD are more likely to have inherited it<sup>66</sup>. The most important feature that is shared across all models is clearly FEATBRADY which indicates the presence of bradykinesia symptoms. According to UKPDSBB guidelines<sup>67</sup>, bradykinesia is considered the main motor symptom for a diagnostic criterion, followed by muscular rigidity.

## Discussion

In this paper, multimodal learning has been applied to recognize PD in its prodromal stage. MRI data and clinical data have been merged using intermediate fusion in multimodal co-learning fashion to improve performances, resulting in a significant boost over the unimodal clinical counterpart.

**a****b**

**Figure 3.** 3D brain reconstruction images and Tabular XAI. **(a)** 3D brain reconstruction from the AvG difference of Prodromal and Control of DenseNet and ViT seen from all angles. **(b)** Explainability of the tabular models, ordered by the most important features.

The co-learning fusion strategy facilitated the learning of robust representations across modalities, mitigating noise and enhancing overall model performance. In accordance with the published medical literature, these methods demonstrated that the models focus on brain regions clinically significant to PD. There have been several scientific studies that have reported on the complexity of brain structure and function in patients with PD<sup>68</sup>.

These studies have found evidence of atrophy in the bilateral medial temporal lobe and the left dorsolateral prefrontal cortex<sup>69</sup>. More recently, there is growing evidence that changes in both structural and functional connectivity are responsible for cognitive deficits in Parkinson's patients<sup>70</sup>. From the explainability analysis, it is possible to infer that the connectivity between subcortical structures and the left dorsal prefrontal cortex and parietal lobe appears impaired, contributing to cognitive deficits in PD. As shown in (Fig. 2c), the focus of the DenseNet model on the left pre-frontal lobe and right temporal region during prodromal disease aligns with known cognitive deficits and motor control difficulties associated with PD.

Recent studies<sup>71</sup> have indicated that involvement of the left prefrontal cortex in surface-based regional homogeneity may be more prevalent in individuals with early-onset PD. On the other hand, bilateral effects have been observed in late-onset subjects. These findings suggest that these data could provide significant potential for

predicting different subtypes of PD. Consistent with existing literature that proposes lateral ventricle enlargement as an indicator of prodromal PD, ViT's emphasis on this region is also noteworthy. A key point of discussion is that the integration of deep learning models in clinical applications, such as early PD detection, necessitates a careful balance between model accuracy and the clinical relevance of the regions highlighted by the models. This trade-off is crucial to ensure that the model's outputs are not only statistically robust but also meaningful and actionable in a clinical setting. The reliability of the interpretations provided by the models may be dependent on the quality and quantity of the training data used. It is known that limited datasets can potentially lead to overfitting, and the highlighted regions may appear highly consistent within the same class despite not generalizing well to unseen data. This limitation was mitigated by employing stratified cross-validation techniques to enhance the models' ability to generalize. Additionally, given the extensive literature regarding early biomarkers of PD as support of these findings and the collaboration made with clinical experts, it has been ensured that the model's focus aligns with the clinical expectations and known disease markers.

Overall, the results reported in this work are encouraging and consistent with the expectation of using a multimodal learning approach for Parkinson's prodromal recognition. In future works, integrating other modalities, such as genetic data, will be explored by applying specific deep learning models tailored for genetic data classification. These data-driven findings imply that the proposed framework could significantly aid in predicting different subtypes of PD.

## Conclusions and future research directions

This study has demonstrated the potential of multimodal deep learning approaches in detecting Parkinson's Disease (PD) at its prodromal stage. By combining advanced 3D neural network architectures with a novel Excitation Network (EN) for processing clinical data, achieving a significant improvements in early PD detection accuracy compared to unimodal approaches. Findings highlight the effectiveness of the DenseNet architecture combined with the proposed Excitation Network, which outperformed other models in terms of accuracy and F1-score. This multimodal approach not only improved classification performance but also provided valuable insights into the brain regions potentially associated with early-stage PD. The application of Explainable AI (XAI) techniques, including Integrated Gradients and Attention Heatmaps, revealed that our models focused on clinically relevant areas such as the right temporal and left pre-frontal regions, as well as the lateral ventricles. These findings align with current literature on PD pathophysiology and suggest promising directions for identifying early biomarkers of the disease.

However, it is crucial to acknowledge the limitations and challenges inherent in applying deep learning models to medical diagnostics. The reliability of model interpretations is intrinsically linked to the quality and quantity of training data. In this work, the stratified cross-validation has been used to mitigate potential overfitting, i.e. the risk of model interpretations not generalizing well to unseen data remains a concern, particularly given the limited size of the dataset. Furthermore, the differences in highlighted regions across various models and attention mechanisms underscore the complexity of interpreting deep learning outputs in a clinical context. These differences, while providing complementary perspectives on potential PD biomarkers, also underscore the importance of maintaining clinical expertise in the interpretative process. The integration of domain knowledge from medical professionals is crucial for transposing AI models into clinically meaningful and actionable insights.

Future research directions include expanding the dataset to include more diverse patient populations from multiple centers for improving model generalizability. Longitudinal studies tracking the progression of highlighted brain regions in individuals transitioning from prodromal to clinical PD could provide valuable insights into the predictive power of these biomarkers. The application of more sophisticated XAI techniques tailored for multimodal medical imaging analysis will enhance understanding of model decision-making processes. Additionally, exploring the integration of additional data modalities, such as genetic information or detailed clinical assessments, may enhance the accuracy and interpretability of early PD detection.

## Methods

### Clinical dataset

The clinical data in the PPMI collection includes a wide range of information about individuals' health, symptoms, and medical history. The clinical evaluations are designed to provide a thorough understanding of PD and its progression. The following are some major clinical data components:

1. *Demographic information*: basic demographic details such as Age, Gender, Ethnicity help characterize the study population.
2. *Medical history*: contains details about the diagnosis assessments of PD participants and their time of diagnosis. Additionally, information related to family history of PD or other neurodegenerative disorders of the patients is present to provide more insights regarding the diagnosis.
3. *Motor assessments*: the Unified Parkinson's Disease Rating Scale (UPDRS<sup>72</sup>) is a widely recognized standardized scale used to assess the severity of motor symptoms associated with PD. It includes sections for mentation, behaviour, mood, activities of daily living and motor function. Moreover, motor function tests were conducted for the assessments of motor skills such as the cardinal symptoms of PD (bradykinesia, tremor, rigidity and postural instability)
4. *Non-motor assessments*: Contains the evaluation of non-motor symptoms such as sleep disturbance, autonomic dysfunction and mood disorders through questionnaires, and also the assessments of cognitive functions as its decline can be associated with PD.
5. *Follow-up data*: PPMI systematically archives patient data at various intervals to meticulously monitor the progression and trajectory of symptoms over time. This longitudinal approach enables comprehensive

Feature	Description	Values
FEATBRADY	Bradykinesia	Yes, no
FEATRIGID	Rigidity	Yes, no
FEATPOSINS	Postural disturbance	Yes, no
FEATTREMOR	Rest tremor	Yes, no,
BIODADPD	Biological father with PD or Parkinsonism	Yes, no
BIOMOMPD	Biological mother with PD or Parkinsonism	Yes, no
FULBROPD	Full brothers with PD or Parkinsonism	Number
FULSISPD	Full sisters with PD or Parkinsonism	Number
MAGFATHPD	Maternal grandfather with PD or Parkinsonism	Yes, no
MAGMOTHPD	Maternal grandmother with PD or Parkinsonism	Yes, no
PAGFATHPD	Paternal grandfather with PD or Parkinsonism	Yes, no
PAGMOTHPD	Paternal grandmother with PD or Parkinsonism	Yes, no

**Table 3.** Selected features for training.

observation of changes and progression in symptomatology, facilitating a deeper understanding of disease dynamics and informing potential treatment strategies.

Out of all the 79 features extracted from the clinical dataset (Appendix) only twelve were selected for training (Table 3). The selected features refer to the cardinal motor symptoms such as FEATBRADY, FEATRIGID, FEATTREMOR, FEATPOSINS and family information limited to parents, siblings and grandparents. These features were chosen as they are characteristic of the Parkinson's Disease and easily obtainable without deep medical examinations, as the scope of the paper is detection of PD in its prodromal stage by family doctors.

### Data extraction

The PPMI database constitutes a rich repository containing longitudinal data encompassing clinical, imaging, and biomarker information from individuals across various stages of PD, including prodromal PD, and healthy controls. Therefore, a rigorous procedure was adopted to choose only the candidates that met the research question.

The procedure involved the focus on specific parameters concerning the classification Group, the Visit, the Image Description and Modality. Initially, only the patients classified within the Prodromal and Healthy groups were exclusively chosen, reflecting the study's focus on early-stage Parkinson's detection. This prior classification was established based on the clinical and diagnostics assessments provided within the PPMI database. Next, the research query was further refined by selecting only the patients whose visit time corresponded to the Baseline stage. By focusing on Baseline visits, it has been ensured that only initial assessments and evaluations were included, thereby facilitating the observation of early disease manifestations. Lastly, only the MRI images described as T2 in T1-anatomical space were included in the selection process, to ensure homogeneity in imaging modality and spatial representation. This selection process resulted in the identification of 90 patients with their clinical data alongside their MRI scans.

### Data pre-processing

All the data downloaded from the PPMI database underwent pre-processing steps to ensure it is suitable for machine learning approaches. For the clinical data, several steps have been performed to address missing values, standardize the dataset and detect potential outliers. Initially, all missing values were replaced with -1. This specific value was chosen to have a clear distinction between real data and imputed data, while also retaining the integrity and interpretability of the data preprocessing step. Replacing missing values with -1 ensures that the model can still learn from the data while being aware of the missing entries, which might hold relevant information about the dataset. Moreover, categorical variables were converted into a format suitable for ML algorithms through one-hot encoding. This process creates binary columns for each category, allowing algorithms to process categorical data effectively. Finally all clinical data was standardized to have mean equal to 0 and standard deviation of 1 to boost the convergence and overall better performances for many ML algorithms that may be sensitive to the scale of the input data.

Instead the MRI data, stored in the NifTI format, was only normalized by dividing each pixel value by the highest pixel value calculated among all the images. This normalization step scales the pixel values to a range between 0 and 1, which is particularly beneficial for deep learning models. Such models, including convolutional neural networks, perform more efficiently when input data is within a standardized range. This normalization also helps in mitigating issues related to varying image intensities across different scans. Finally, the original dimensions of the MRI scans (176 height, 240 width, 256 depth) have been preserved. Maintaining these dimensions is critical to ensure that no spatial information is lost, which is vital for accurate analysis and diagnosis. Altering the dimensions could lead to the loss of important anatomical details, adversely affecting the model's performance. By performing these steps, it has been ensured that all the models operated in the most optimal initial conditions.

### Excitation network

The Excitation Network exploits the advantages of the squeeze-and-excitation<sup>55</sup> (SE) building block for convolutional neural networks but is adopted in a different way to tabular data. As the tabular data is already one-channelled it is not necessary to first perform the squeeze operation. Next, the excitation operation is designed to highlight the most important features by passing them through two dense layers. A nonlinear compressed embedding space of the original feature space is created by compressing the features  $f$  by a squeeze ratio  $r \in (0, 1)$  (in this case  $r = 0.9$ ) in the first dense layer. The compressed features are then passed through the second Dense layer via a Rectified Linear Unit (ReLU). The second layer restores the compressed features back to their input space and applies a sigmoid activation. Formally, the output of the excitation block is represented in Eq. (1):

$$Exc(f) = \text{sigmoid}(W_2 \text{ReLU}(W_1 f)) \quad (1)$$

where  $f$  are the input features,  $W_1 \in \mathbb{R}^{f \times r \times f}$  and  $W_2 \in \mathbb{R}^{f \times (f \times r)}$ . The output of the excitation block is then multiplied to the original input to account for residual connections. The merged representations are then normalized through a Layer Normalization to help EN to focus on features without being affected by their scale. The normalized feature representations are then passed to a flattening layer to obtain a one-dimensional vector which is fed through three final Dense layers to get the classification. Thus, the EN architecture is a simple architecture that focuses on a simple but effective attention mechanism<sup>73–76</sup>.

### Experimental setup and model implementation

The deep learning models were implemented using Tensorflow and Python 3.10.14. Specifically, the standard architecture of ResNet-18 with the SE block and a compact DenseNet-18 were employed for training on volumetric data by adapting the existing 2D convolution with their 3D counterparts. ResNet leverages residual connections to mitigate the vanishing gradient problem, enabling effective training of deep neural networks while DenseNet employs dense connectivity to enhance feature reuse and gradient propagation, which is particularly beneficial for deep volumetric data processing.

Furthermore, a modified version of the standard architecture of ViT, as described in the paper<sup>77</sup>, was employed to accommodate 3D patching. Table 4 provides a detailed overview of the models' parameters.

It is immediate that DenseNet emerges as the most lightweight deep neural network while also delivering the best performances (Table 1). Additionally, given the inherently lightweight nature of EN, the increase in parameter count when combining the networks is negligible compared to the improvement in performances. Finally, Table 5 provides a description of the used hyperparameters per each model.

DenseNet and ResNet were instantiated with the default hyperparameters of their respective standard architectures. In contrast, ViT underwent iterative tuning via a trial-and-error methodology. Multiple training sessions of ViT were conducted, varying hyperparameter combinations, and through empirical analysis, the aforementioned hyperparameters were selected as optimal, balancing performance, memory demands, and training

Model	# of parameters
ResNet	14,428,578
DenseNet	203,046
ViT	69,359,106
EN	3264
ResNet + EN	14,437,884
DenseNet + EN	212,320
ViT + EN	69,382,300

**Table 4.** Model parameters.

Model	Parameter	Description	Value
DenseNet	Growth rate	Factor to determine the number of filters for each dense block	32
	Compression factor	Factor to reduce the number of output feature maps. Range [0,1]	0.5
ResNet	SE ratio	Squeeze-excitation ratio	16
ViT	Patch size	Dimension of each patch. Patches are cubic	32
	Projection dim	Dimension for linear projection from the flattened patches	256
	Heads	Number of attention heads for each multi-head attention layer	12
	Transformer layers	Number of transformer encoders	7
EN	Ratio	Squeeze ratio	0.9

**Table 5.** Model hyperparameters.

durations. All networks were trained utilizing two NVIDIA RTX A6000 GPUs, each equipped with 48GB of dedicated RAM.

### Model training

The training was conducted with stratified tenfold cross-validation. The datasets get split into ten class-balanced subsets and the training procedure is repeated ten times so that at each iteration, one new subset is used as test set and the rest as training set. The model's overall performance is calculated by averaging all the per-iteration results. This method was purposefully adopted to get more reliable results from the small dataset size and reduce bias and variance as every data is used for fitting and validating. All the deep models and EN were trained for 200 epochs with categorical cross-entropy as loss and Adam as optimizer with learning rate of 0.001. Random Forest and XGBoost were trained with 100 sub-trees and SVM with a linear kernel.

### Performance metrics

All the selected models have been evaluated using a variety of performance metrics, each providing a meaningful insight regarding the different behaviours and effectiveness of the models. The metrics considered are:

- *Accuracy*: the accuracy is defined as the ratio of all correctly predicted instances to the total instances. It is a straightforward measure of the overall model performance but can be misleading for unbalanced datasets,
- *Area under the receiver operating characteristics curve (AUC)*: the AUC evaluates the model's ability to discriminate between positive and negative classes. A value equal to 0.5 suggests that there is no discrimination (equivalent to random guessing) while an AUC of 1.0 means perfect discrimination,
- *Precision*: the precision measures the accuracy of the positive predictions. This metric is very important when the cost of false positive is high like in medical diagnosis. A high precision indicates that most of the positive cases are indeed positive, reducing the risk of unnecessary treatments for false positive cases,
- *Recall*: the recall measures the ability of the model of capturing all relevant examples. Similarly to the precision, the recall is crucial for ensuring that most of the actual positive cases are detected, minimizing the chances of missing true positives instances during disease screening,
- *F1-macro score*: the F1 is the harmonic mean of precision and recall. The F1-macro score averages the F1 scores of the individual classes to treat them equally. For this reason this metric is valuable in evaluating performance on imbalanced datasets.

Formally, in formulas:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$AUC = \int_0^1 TPRd(FPR) \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

$$F1\text{-macro} = \frac{1}{N} \sum_{i=0}^N F1_i \quad (7)$$

where:

- TP (True Positive): the number of correctly predicted positive examples
- TN (True Negative): the number of correctly predicted negative examples
- FP (False Positive): the number of incorrectly predicted positive examples
- FN (False Negative): the number of incorrectly predicted negative examples
- TPR (True Positive Ratio):
- FPR (False Positive Ratio):
- N: number of classes

### Integrated gradients

IG is an attribution technique that works by comparing the prediction output to a baseline to quantify how much each pixel contributed to the final classification. Generally, a black image is used as baseline because it acts as a neutral prediction. The computation of IG is defined as in Eq. (8):

$$\mathbf{IG}_i(x) = (x_i - x_{i'}) \times \sum_{k=1}^m \frac{\partial F\left(x' + \frac{k}{m} \times (x - x')\right)}{\partial x_i} \times \frac{1}{m} \quad (8)$$

where  $x$  is the input image,  $x'$  is the baseline image,  $F$  is the neural network and  $m$  is the number of steps for the Riemann approximation. According to the authors<sup>78</sup>, 20 to 300 steps are enough for an accurate approximation based on the network complexity. In this work, 50 steps were used as increasing the number of steps did not improve the resulting heatmap.

### Attention heatmap

The explainability of Vision Transformers is performed by examining the attention scores obtained by the multi-head attention layers from all the transformer blocks. The attention scores get averaged along all heads, added to an identity matrix to account for residual connections, as in Eq. (9), and re-normalized, as in Eq. (10)<sup>79</sup>. Formally:

$$AH(i) = \left[ \left( \frac{1}{n} \sum_n MHA(i) \right) + I \right] \quad (9)$$

$$Norm_{AH(i)} = \frac{AH(i)}{\sum_{i,j} AH(i)_{i,j}} \quad (10)$$

where  $i$  is the input image of shape (176,240,256),  $MHA(i)$  are the attention scores for the input image,  $n$  is the number of attention heads for each multi-head attention ( $n = 12$ ) and  $I$  is the identity matrix (9). The resulting matrix is then normalized (10). The intermediate result  $Norm_{AH(i)}$  is then recursively multiplied from all the transformer blocks to obtain a final attention matrix. Additionally, a mask creates an attention map highlighting specific regions in the input image that the model considers important for making predictions. This mask is generated by extracting the output token from the attention matrix and reshaped into a 3D tensor with dimensions equal to the number of patches along each axis. The extracted attention weights are normalized by dividing each tensor element by the tensor's maximum value. This normalization ensures that the attention weights are scaled between 0 and 1, making it suitable for interpreting them as a probability-like distribution. Finally, the mask is resized to the original input image space and element-wise multiplied to the original input image to obtain the final attention heatmap.

### Feature importance extraction

Random Forest, Support Vector Machines and XGBoost are intrinsically explainable models. No further processing is required to obtain the feature importance as they are easily retrievable from the model outputs. For Random Forest and XGBoost, the importance of a feature is calculated by how much it reduces the impurity in the tree. In support vector machines, the predicted class is represented by the direction of the coefficients, better, coordinates of the orthogonal vector to the hyperplane, which indicate the feature relevance. A feature is considered more significant the higher its absolute value.

EN is a semi-interpretable model compared to traditional deep neural networks. The interpretability of EN comes from the feature attentions generated by the excited block in Eq. (1). The overall feature importance generated by a EN is calculated through the global magnitude feature attention metric in Eq. (11). This metric computes the overall attention of each feature across all instances as the mean of the absolute values of the attention weights of the excited block.

$$\tau_i = \frac{1}{n} \sum_{i=1}^n |Exc(J_i)| \quad (11)$$

where  $J_i$  is the  $i$ -th input instance,  $n$  is the total number of instances and  $Exc(J_i)$  is the output of the excitation block. The absolute value is taken because attention weights are calculated globally rather than with respect to a specific class, so a strong negative value for a specific class on specific features implies that these features are very sensitive for the global task. As EN was trained with stratified tenfold cross validation, the global magnitude feature attention has been calculated for each iteration and averaged across all iterations as in Eq. (12).

$$\tau = \frac{\sum_{i=1}^n \tau_i}{n} \quad (12)$$

where  $n$  is the number of iterations.

### XAI image generation

The application of Integrated gradients and attention heatmap to the MRI scan generates a 3D attribution map whose brightest pixels correspond to the most important area where the network is focused in the classification. The size of the attribution map is of the same size as the MRI, and by fixing one axis, it is possible to view the

attribution map per each slice from all points of view (axial, sagittal, and coronal). Afterward, the 3D attribution maps of each tested patient were normalized to the standard template MNI152 so that it would be possible to accurately locate which anatomical structure of the brain corresponds to the given coordinates.

From the 23 tested images (10 HC, 13 Prodromal), 69 3D attribution maps were generated for the deep models. Each 3D attribution map per patient has been further compressed in a two-dimensional space by overlapping all the slice on a given axis (axial or sagittal) to have a clear view of the most highlighted areas, as shown in Eq. (13). Additionally, each compressed map has been averaged across all the other compressed maps (grouped per model, technique, and class) to obtain a single image that is representative of the respective class. This was performed using Eq. (14). Finally, the Prodromal average maps have been subtracted to the Control average maps (and vice versa) to understand what are the regions that helped to discriminate that specific class and not the other class as in Eqs. (15) and (16).

$$C_{sag}^{(k)} = \sum_x M_{x,y,z}^{(k)} \text{ or } C_{ax}^{(k)} = \sum_z M_{x,y,z}^{(k)} \quad (13)$$

$$A_{sag}^{(k)} = \frac{\sum_n C_{sag}^{(k)}}{n} \quad A_{ax}^{(k)} = \frac{\sum_n C_{ax}^{(k)}}{n} \quad (14)$$

$$Diff_{sag}^{(P-C)} = A_{sag}^P - A_{sag}^C \quad Diff_{sag}^{(C-P)} = A_{sag}^C - A_{sag}^P \quad (15)$$

$$Diff_{ax}^{(P-C)} = A_{ax}^P - A_{ax}^C \quad Diff_{ax}^{(C-P)} = A_{ax}^C - A_{ax}^P \quad (16)$$

where:

- $M_{x,y,z}^{(k)}$  is the 3D attribution map defined by its three axes ( $x, y, z$ ),
- $(k)$  is the classification class, which can be  $P$  for Prodromal and  $C$  for Control,
- $C_{axis}^{(k)}$  is the compressed map, both by a sagittal and axial point of view,
- $axis$  is the reference axis for the point of view, which can be sagittal or axial,
- $A_{axis}^{(k)}$  is the average map, grouped per class, divided by the number  $n$  of maps
- $Diff_{axis}^{(k)}$  is the difference map between the Prodromal and Control (and vice versa) average maps.

All features	Description
Age	Age of the patient
Sex	Sex of the patient
<b>FEATBRADY*</b>	<b>Bradykinesia</b>
<b>FEATRIGID*</b>	<b>Rigidity</b>
<b>FEATPOSINS*</b>	<b>Postural disturbance</b>
<b>FEATTREMOR*</b>	<b>Rest tremor</b>
PSGLVL	Abnormalities consistent with a PD syndrome
BIODAD	Biological father
<b>BIODADPD*</b>	<b>Biological father with PD or Parkinsonism</b>
BIOMOM	Biological mother
<b>BIOMOPPD*</b>	<b>Biological mother with PD or Parkinsonism</b>
DISFAMPD	Do you have a more distant relative not listed above who has/had Parkinson's disease or Parkinsonism?
FULBRO	Full brothers
<b>FULBROPD*</b>	<b>Full brothers with PD or Parkinsonism</b>
FULSIB	Full siblings
FULSIBPD	Full siblings with PD
FULSIS	Full sisters
<b>FULSISPD*</b>	<b>Full sisters with PD or Parkinsonism</b>
HAFSIB	Half siblings
HAFSIBPD	Half siblings with PD
KIDSNUM	Children
KIDSPD	Children with PD or Parkinsonism
<b>MAGFATHPD*</b>	<b>Maternal grandfather with PD or Parkinsonism</b>
<b>MAGMOTHPD*</b>	<b>Maternal grandmother with PD or Parkinsonism</b>
MAGPAR	Maternal grandparents
MAGPARPD	Maternal grandparents with PD
MAHAFSIB	Maternal half siblings
MAHAFSIBPD	Maternal half siblings with PD or Parkinsonism
MATAU	Maternal aunts and uncles
MATAUPD	Maternal aunts and uncles with PD or Parkinsonism
MATCOUS	Maternal cousins
MATCOUSPD	Maternal Cousins with PD or Parkinsonism
<b>PAGFATHPD*</b>	<b>Paternal Grandfather with PD or Parkinsonism</b>
<b>PAGMOTHPD*</b>	<b>Paternal grandmother with PD or Parkinsonism</b>
PAGPAR	Paternal grandparents
PAGPARPD	Paternal grandparents with PD
PAHAFSIB	Paternal half siblings
PAHAFSIBPD	Paternal half siblings with PD or Parkinsonism
PATAU	Paternal aunts and uncles
PATAUPD	Paternal aunts and uncles with PD or Parkinsonism
PATCOUS	Paternal cousins
PATCOUSPD	Paternal cousins with PD or Parkinsonism
FEATACTTRM	Prominent action tremor
FEATALCREP	Definite response to alcohol
FEATALNLMB	Alien limb phenomenon
FEATANTCOL	Disproportionate anterocollis
FEATANXITY	Anxiety
FEATAPATHY	Apathy
FEATBWLDYS	Bowel dysfunction
FEATCLRLEV	Clear and dramatic response to levodopa
FEATCOGFLC	Cognitive fluctuations
FEATCRTSNS	Unequivocal cortical sensory loss (i.e., graphesthesia, stereognosis with intact primary sensory modalities)
FEATDCRARM	Decreased arm swing
FEATDELHAL	Systematized delusions or visual hallucinations unrelated to medications
FEATDEPRES	Depression
FEATDIMOLF	Diminished olfaction
Continued	

All features	Description
FEATDYSART	Dysarthria
FEATDYSKIN	Dyskinesia
FEATDYSPHG	Dysphagia
FEATDYSTNA	Dystonia
FEATGZEPAL	Supranuclear gaze palsy
FEATINSPST	Inspiratory stridor
FEATLMBAPX	Limb ideomotor apraxia
FEATMCRGRA	Micrographia
FEATMTRFLC	Motor fluctuations
FEATMYCLNS	Myoclonus
FEATNEURSS	Neuroleptic super-sensitivity
FEATNOLEVO	Little or no response to levodopa
FEATPOSHYP	Postural hypotension
FEATPST3YR	Postural instability or gait freezing in the first 3 years
FEATPYRTCT	Otherwise unexplained pyramidal tract signs
FEATSBDERM	Seborrheic dermatitis
FEATSEXDYS	Sexual dysfunction
FEATSHGAIT	Shuffling gait
FEATSTPPOS	Stooped posture
FEATSUGRBD	Dream enactment suggestive of REM sleep behavior disorder
FEATURNDYS	Urinary dysfunction
FEATWDGAIT	Wide based gait/cerebellar features

**Table 6.** Features list. \*Selected features for training.

## Data availability

The datasets analysed during the current study are available in the Parkinson's Progression Markers Initiative PPMI repository at the link <https://www.ppmi-info.org/access-data-specimens/download-data>. Online application for accessing the dataset is required.

## Appendix

See Table 6.

Received: 21 March 2024; Accepted: 13 August 2024  
Published online: 09 September 2024

## References

- Prajwal, P. *et al.* Parkinson's disease updates: Addressing the pathophysiology, risk factors, genetics, diagnosis, along with the medical and surgical treatment. *Ann. Med. Surg.* **85**, 4887–4902 (2023).
- Cerri, S., Mus, L. & Blandini, F. Parkinson's disease in women and men: What's the difference?. *J. Parkinsons Dis.* **9**, 501–515 (2019).
- Chan, Y. L. E. *et al.* Post-traumatic stress disorder and risk of parkinson disease: A nationwide longitudinal study. *Am. J. Geriatr. Psychiatry* **25**, 917–923 (2017).
- Sotgiu, I. & Rusconi, M. L. Investigating emotions in Parkinson's disease: What we know and what we still don't know. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2013.00336> (2013).
- Sieurin, J. *et al.* Occupational stress and risk for Parkinson's disease: A nationwide cohort study. *Mov. Disord.* **33**, 1456–1464 (2018).
- Tsai, Y. J. *et al.* Cold Exposure after exercise impedes the neuroprotective effects of exercise on thermoregulation and UCP4 expression in an MPTP-induced Parkinsonian mouse model. *Front. Neurosci.* <https://doi.org/10.3389/fnins.2020.573509> (2020).
- Luca, A., Nicoletti, A., Mostile, G. & Zappia, M. The Parkinsonian personality: More than just a "trait". *Front. Neurol.* **9**, 1191 (2018).
- Bailey, M., Anderson, S. & Hall, D. A. Parkinson's disease in African Americans: A review of the current literature. *J. Parkinsons Dis.* **10**, 831–841 (2020).
- Kouli, A., Torsney, K. M. & Kuan, W.-L. Parkinson's disease: Etiology, neuropathology, and pathogenesis. In *Parkinson's Disease: Pathogenesis and Clinical Aspects* 3–26 (2018) <https://doi.org/10.15586/CODONPUBLICATIONS.PARKINSONSDISEASE.2018.CH1>.
- Tueth, L. E. & Duncan, R. P. Musculoskeletal pain in Parkinson's disease: A narrative review. *Neurodegener. Dis. Manag.* **11**, 373–385 (2021).
- Iwaki, H. *et al.* Using spontaneous eye-blink rates to predict the motor status of patients with Parkinson's disease. *Intern. Med.* **58**, 1417 (2019).
- Goubault, E. *et al.* Cardinal motor features of Parkinson's disease coexist with peak-dose choreic-type drug-induced dyskinesia. *J. Parkinsons Dis.* **8**, 323 (2018).
- Thomas, M., Lenka, A. & Kumar Pal, P. Handwriting analysis in Parkinson's disease: Current status and future directions. *Mov. Disord. Clin. Pract.* **4**, 806–818 (2017).
- Knowles, T., Adams, S. G. & Jog, M. Effects of speech rate modifications on phonatory acoustic outcomes in Parkinson's disease. *Front. Hum. Neurosci.* **18**, 1331816 (2024).

15. Caekebeke, J. F. V., Jennekens-Schinkel, A., Van der Linden, M. E., Buruma, O. J. S. & Roos, R. A. C. The interpretation of dysprosody in patients with Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **54**, 145 (1991).
16. Racagni, C. *et al.* Gait and postural disorders in parkinsonism: A clinical approach. *J. Neurol.* **267**, 3169–3176 (2020).
17. Todorova, A., Jenner, P. & RayC haudhuri, K. Non-motor Parkinson's: Integral to motor Parkinson's, yet often neglected. *Pract. Neurol.* **14**, 310–322 (2014).
18. Tibar, H. *et al.* Non-motor symptoms of Parkinson's disease and their impact on quality of life in a cohort of Moroccan patients. *Front. Neurol.* <https://doi.org/10.3389/fneur.2018.00170> (2018).
19. Goldman, J. G. & Postuma, R. Premotor and non-motor features of Parkinson's disease. *Curr. Opin. Neurol.* **27**, 434 (2014).
20. Váradí, C. Clinical features of Parkinson's disease: The evolution of critical symptoms. *Biology* **9**, 103 (2020).
21. Tolosa, E., Garrido, A., Scholz, S. W. & Poewe, W. Challenges in the diagnosis of Parkinson's disease. *Lancet Neurol.* **20**, 385 (2021).
22. Thenganatt, M. A. & Louis, E. D. Distinguishing essential tremor from Parkinson's disease: Bedside tests and laboratory evaluations. *Expert Rev. Neurother.* **12**, 687–696 (2012).
23. Rizzo, G. *et al.* Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis. *Neurology* **86**, 566–576 (2016).
24. Virameteekul, S., Revesz, T., Jaunmuktane, Z., Warner, T. T. & De Pablo-Fernández, E. Clinical diagnostic accuracy of Parkinson's disease: Where do we stand?. *Mov. Disord.* **38**, 558–566 (2023).
25. Mahlknecht, P., Marini, K., Werkmann, M., Poewe, W. & Seppi, K. Prodromal Parkinson's disease: Hype or hope for disease-modification trials?. *Transl. Neurodegener.* <https://doi.org/10.1186/s40035-022-00286-1> (2022).
26. Pan, D. *et al.* Early detection of Alzheimer's disease using magnetic resonance imaging: A novel approach combining convolutional neural networks and ensemble learning. *Front. Neurosci.* <https://doi.org/10.3389/fnins.2020.00259> (2020).
27. Bhat, S., Acharya, U. R., Hagiwara, Y., Dadmehr, N. & Adeli, H. Parkinson's disease: Cause factors, measurable indicators, and early diagnosis. *Comput. Biol. Med.* **102**, 234–241 (2018).
28. Shang, L. & Wang, C. *Three-Dimensional Reconstruction and Protection of Mining Heritage Based on Lidar Remote Sensing and Deep Learning.* <https://doi.org/10.1155/2022/2412394> (2022).
29. Mu, Z. *et al.* Algorithm analysis of face recognition robot based on deep learning. *Int. J. Pattern Recogn. Artif. Intell.* <https://doi.org/10.1142/S0218001423560049> (2023).
30. Cavallo, F. *et al.* Emotion modelling for social robotics applications: A review. *J. Bionic Eng.* **15**, 185–203 (2018).
31. Saravanan, S. *et al.* A systematic review of artificial intelligence (AI) based approaches for the diagnosis of Parkinson's disease. *Arch. Comput. Methods Eng.* **29**(6), 3639–3653 (2022).
32. Zhang, X., Yang, Y., Wang, H., Ning, S. & Wang, H. Deep neural networks with broad views for Parkinson's disease screening. In *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019* 1018–1022 <https://doi.org/10.1109/BIBM47256.2019.8983000> (2019).
33. Prasuhn, J., Heldmann, M., Münte, T. F. & Brüggemann, N. A machine learning-based classification approach on Parkinson's disease diffusion tensor imaging datasets. *Neurol. Res. Pract.* <https://doi.org/10.1186/s42466-020-00092-y> (2020).
34. Vanegas, M. I., Ghilardi, M. F., Kelly, S. P. & Blangero, A. Machine learning for EEG-based biomarkers in Parkinson's disease. In *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018* 2661–2665 <https://doi.org/10.1109/BIBM.2018.8621498> (2019).
35. Geraedts, V. J. *et al.* Machine learning for automated EEG-based biomarkers of cognitive impairment during deep brain stimulation screening in patients with Parkinson's disease. *Clin. Neurophysiol.* **132**, 1041–1048 (2021).
36. Shaban, M., Cahoon, S., Khan, F. & Polk, M. Exploiting the differential wavelet domain of resting-state EEG using a deep-CNN for screening Parkinson's disease. In *2021 IEEE Symposium Series on Computational Intelligence, SSCI 2021 - Proceedings* <https://doi.org/10.1109/SSCI50451.2021.9660178> (2021).
37. Yuvaraj, R., Rajendra Acharya, U. & Hagiwara, Y. A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals. *Neural Comput. Appl.* **30**, 1225–1235 (2018).
38. Chien, C. Y., Hsu, S. W., Lee, T. L., Sung, P. S. & Lin, C. C. Using artificial neural network to discriminate Parkinson's disease from other Parkinsonisms by focusing on putamen of dopamine transporter SPECT images. *Biomedicines* **9**, 1–11 (2020).
39. Hsu, S. Y. *et al.* Classification of the multiple stages of Parkinson's disease by a deep convolution neural network based on 99mTc-TRODAT-1 SPECT images. *Molecules* **25**, 4792 (2020).
40. Meyer, P. T., Frings, L., Rücker, G. & Hellwig, S. 18F-FDG PET in Parkinsonism: Differential diagnosis and evaluation of cognitive impairment. *J. Nucl. Med.* **58**, 1888–1898 (2017).
41. Piccardo, A. *et al.* The role of the deep convolutional neural network as an aid to interpreting brain [18F]DOPA PET/CT in the diagnosis of Parkinson's disease. *Eur. Radiol.* **31**, 7003–7011 (2021).
42. Goyal, J., Khandnor, P. & Aseri, T. C. A hybrid approach for Parkinson's disease diagnosis with resonance and time-frequency based features from speech signals. *Expert Syst. Appl.* **182**, 115283 (2021).
43. Khaskhoussy, R. & Ayed, Y. B. Improving Parkinson's disease recognition through voice analysis using deep learning. *Pattern Recogn. Lett.* **168**, 64–70 (2023).
44. El Maachi, I., Bilodeau, G. A. & Bouachir, W. Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Syst. Appl.* **143**, 113075 (2020).
45. Bobić, V., Durić-Jovičić, M., Dragašević-Mišković, N., Kostić, V. S. & Kvaščev, G. Comparison of two deep learning models for the recognition of parkinson's disease gait patterns. In *Proceedings - 10th International Conference on Electrical, Electronic and Computing Engineering, ICETRAN 2023* <https://doi.org/10.1109/ICETRAN59631.2023.10192156> (2023).
46. Kamran, I., Naz, S., Razzak, I. & Imran, M. Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease. *Future Gener. Comput. Syst.* **117**, 234–244 (2021).
47. Loh, H. W. *et al.* Application of deep learning models for automated identification of Parkinson's disease: A review (2011–2021). *Sensors* **21**, 7034 (2021).
48. Yan, K., Li, T., Lobo Marques, J. A., Gao, J. & Fong, S. J. A review on multimodal machine learning in medical diagnostics. *Math. Biosci. Eng.* **20**, 8708–8726 (2023).
49. Prashanth, R., Dutta Roy, S., Mandal, P. K. & Ghosh, S. High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. *Int. J. Med. Inform.* **90**, 13–21 (2016).
50. Marek, K. *et al.* The Parkinson progression marker initiative (PPMI). *Prog. Neurobiol.* **95**, 629–635 (2011).
51. Long, D. *et al.* Automatic classification of early Parkinson's disease with multi-modal MR imaging. *PLoS ONE* **7**, e47714 (2012).
52. Zhu, S. Early diagnosis of Parkinson's disease by analyzing magnetic resonance imaging brain scans and patient characteristic. In *2022 10th International Conference on Bioinformatics and Computational Biology, ICBCB 2022* 116–123 <https://doi.org/10.1109/ICBCB55259.2022.9802132> (2022).
53. Skaramagkas, V., Pentari, A., Kefalopoulou, Z. & Tsiknakis, M. Multi-modal deep learning diagnosis of Parkinson's disease—a systematic review. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 2399–2423 (2023).
54. Huang, L., Ye, X., Yang, M., Pan, L. & Zheng, S. H. MNC-Net: Multi-task graph structure learning based on node clustering for early Parkinson's disease diagnosis. *Comput. Biol. Med.* **152**, 106308 (2023).
55. Hu, J., Shen, L., Albanie, S., Sun, G. & Wu, E. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 2011–2023 (2017).

56. Baltrušaitis, T., Ahuja, C. & Morency, L. P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 423–443 (2017).
57. Stahlschmidt, S. R., Ulfenborg, B. & Synnergren, J. Multimodal deep learning for biomedical data fusion: A review. *Brief Bioinform.* <https://doi.org/10.1093/bib/bbab569> (2022).
58. Narayanan, N. S., Rodnitzky, R. L. & Uc, E. Y. Prefrontal dopamine signaling and cognitive symptoms of Parkinson's disease. *Rev. Neurosci.* <https://doi.org/10.1515/revneuro-2013-0004> (2013).
59. Stuart, S., Belluscio, V., Quinn, J. F. & Mancini, M. Pre-frontal cortical activity during walking and turning is reliable and differentiates across young, older adults and people with Parkinson's disease. *Front. Neurol.* <https://doi.org/10.3389/fneur.2019.00536> (2019).
60. Irmén, F. *et al.* Left prefrontal connectivity links subthalamic stimulation with depressive symptoms. *Ann. Neurol.* **87**, 962–975 (2020).
61. Zhou, C. *et al.* Progressive brain atrophy in Parkinson's disease patients who convert to mild cognitive impairment. *CNS Neurosci. Ther.* **26**, 117 (2020).
62. Dalaker, T. O. *et al.* Ventricular enlargement and mild cognitive impairment in early Parkinson's disease. *Mov. Disord.* **26**, 297–301 (2011).
63. Apostolova, L. *et al.* Hippocampal and ventricular changes in Parkinson's disease mild cognitive impairment. *Neurobiol. Aging* **33**, 2113 (2012).
64. Behnke, S. *et al.* Third ventricular width assessed by transcranial ultrasound correlates with cognitive performance in Parkinson's disease. *Parkinsonism Relat. Disord.* **66**, 68–73 (2019).
65. Klein, C. & Westenberger, A. Genetics of Parkinson's disease. *Cold Spring Harb. Perspect. Med.* **2**, a008888 (2012).
66. Kolicheski, A. *et al.* Early-onset Parkinson's disease: Creating the right environment for a genetic disorder. *J. Parkinsons Dis.* **12**, 2353 (2022).
67. Clarke, C. E. *et al.* UK Parkinson's Disease Society Brain Bank Diagnostic Criteria. (2016).
68. Sokolowski, A. *et al.* Longitudinal brain structure changes in Parkinson's disease: A replication study. *PLoS ONE* **19**, e0295069 (2024).
69. Hanganu, A. *et al.* Mild cognitive impairment is linked with faster rate of cortical thinning in patients with Parkinson's disease longitudinally. *Brain* **137**, 1120–1129 (2014).
70. Liu, J. *et al.* Altered connectivity in the cognitive control-related prefrontal cortex in Parkinson's disease with rapid eye movement sleep behavior disorder. *Brain Imaging Behav.* **17**, 702–714 (2023).
71. Wang, M. *et al.* Surface-based functional alterations in early-onset and late-onset Parkinson's disease: A multi-modal MRI study. *Diagnostics* **13**, 2969 (2023).
72. Goetz, C. G. *et al.* Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Mov. Disord.* **23**, 2129–2170 (2008).
73. Luong, M. T., Pham, H. & Manning, C. D. Effective approaches to attention-based neural machine translation. In *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing* 1412–1421 <https://doi.org/10.18653/v1/d15-1166> (2015).
74. Jetley, S., Lord, N. A., Lee, N. & Torr, P. H. S. Learn to pay attention. (2018).
75. Vaswani, A. *et al.* Attention is all you need. *Adv. Neural Inf. Process Syst.* **2017**, 5999–6009 (2017).
76. Hernández, A. & Amigó, J. M. Attention mechanisms and their applications to complex systems. *Entropy* **23**, 283 (2021).
77. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR 2021 - 9th International Conference on Learning Representations* (2020).
78. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *34th International Conference on Machine Learning, ICML 2017* Vol. 7, 5109–5118 (2017).
79. Abnar, S. & Zuidema, W. Quantifying attention flow in transformers. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics* 4190–4197 <https://doi.org/10.18653/v1/2020.acl-main.385> (2020).

## Acknowledgements

This research was co-funded by the Complementary National Plan PNC-I.1 "Research initiatives for innovative technologies and pathways in the health and welfare sector" D.D. 931 of 06/06/2022, DARE—DigitAl lifelong pRevEntion initiative, code PNC0000002, CUP B53C22006420001.

## Author contributions

All authors contributed equally.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to V.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.