Gender variability in machine learning based subcortical neuroimaging for Parkinson’s disease diagnosis

Nair Ul Islam
Department of Computer Science and Engineering,
Centre of Excellence for Artificial Intelligence in Medicine, Imaging and Forensics,
Sharda University, Greater Noida, India, and
Ruqaiya Khanam
Department of Electrical Electronics and Communication Engineering,
Centre of Excellence for Artificial Intelligence in Medicine, Imaging and Forensics,
Sharda University, Greater Noida, India

Abstract
Purpose – This study evaluates machine learning (ML) classifiers for diagnosing Parkinson’s disease (PD) using subcortical brain region data from 3D T1 magnetic resonance imaging (MRI) Parkinson’s Progression Markers Initiative (PPMI database). We aim to identify top-performing algorithms and assess gender-related differences in accuracy.

Design/methodology/approach – Multiple ML algorithms will be compared for their ability to classify PD vs healthy controls using MRI scans of the brain structures like the putamen, thalamus, brainstem, accumbens, amygdala, caudate, hippocampus and pallidum. Analysis will include gender-specific performance comparisons.

Findings – The study reveals that ML classifier performance in diagnosing PD varies across subcortical brain regions and shows gender differences. The Extra Trees classifier performed best in men (86.36% accuracy in the putamen), while Naive Bayes performed best in women (69.23%, amygdala). Regions like the accumbens, hippocampus and caudate showed moderate accuracy (65–70%) in men and poor performance in women. The results point out a significant gender-based performance gap, highlighting the need for gender-specific models to improve diagnostic precision across complex brain structures.

Originality/value – This study highlights the significant impact of gender on machine learning diagnosis of PD using data from subcortical brain regions. Our novel focus on these regions uncovers their diagnostic
ACI

potential, improves model accuracy and emphasizes the need for gender-specific approaches in medical AI. This work could ultimately lead to earlier PD detection and more personalized treatment.

Keywords Parkinson’s disease, Machine learning, Sub-cortical, Medical AI, Neuroimaging

Paper type Research paper

1. Introduction

Parkinson’s disease (PD) is the most debilitating movement disorder that usually occurs in the elderly [1]. It is the second most common movement disorder in humans, preceding Essential Tremor [1]. PD is a slow, incurable and progressive neurodegenerative disease [2–4]. The characteristic feature of the PD is the early death of the dopamine-producing neurons in the Substantia Nigra region of the brain. Dopamine is a neurotransmitter that regulates and controls motion. Loss of dopamine-producing neurons in the brain leads to movement disorders that progressively start to get worse over time. What exactly causes PD remains unknown; however, it is believed that it may be caused by genetic or environmental factors such as exposure to toxins or by a combination of both [2, 5]. PD is also characterized by non-motor symptoms, which may precede the appearance of motor symptoms like cognitive impairment, sleep disturbances, depression, anxiety, fatigue, pain, irregular bowel movements and disturbances in sense of smell [2, 6, 7]. Motor symptoms of PD begin to manifest only when 60–70% of the dopamine-producing neurons have been lost [8, 9]. Most of the non-motor symptoms can be reported at least a decade before the motor symptoms begin to manifest.

There is no biomarker that can be used for the diagnosis of PD, and diagnosis can be definitively established only after a person’s death by observing the pathological changes that occur in the brain [4]. The only exception to this is genetic testing, and that is not possible in all the cases [4].

Currently, the diagnosis of PD is made based upon clinical criteria, which include the presence of resting tremor, bradykinesia, rigidity and loss of postural reflexes [10]. PD diagnosis is challenging, and the correctness of the diagnosis based on clinical criteria is not guaranteed in every case. In one study that aimed to determine the accuracy of the clinical PD diagnosis by validating the diagnosis against post-mortem neuropathology – the gold standard for diagnosis – in early cases, the accuracy of diagnosis varied between 26 and 53%, and the diagnosis was 85% accurate in cases where the patient was treated for PD for a longer time [11]. In another similar study, it was observed that only 65% of the patients were diagnosed correctly in the early stages, while in the later stages, the correctness of the PD diagnosis improved up to 76% [12]. These results highlight the importance of interpreting the diagnosis of PD with caution in early cases and challenges in the diagnosis of PD.

As mentioned earlier, PD is incurable, but its symptoms can be managed by use of medication, though there is no available mechanism that can slow the disease progression, and the disease may remain prodromal up to the onset of the clinical symptoms [4, 13, 14]. Given the nature of the disease progression and the diagnostic accuracy at the early stages, it is imperative that new techniques be exploited that can result in improved accuracy of the diagnosis or at least help physicians make informed decisions.

Artificial intelligence (AI) and machine learning (ML) have the capability to learn and recognize patterns that may not be apparent to the human eye. Because of this property, AI and ML models are being extensively reported in the scientific literature, which can help with better diagnosis of diseases and offer solutions to difficult medical problems. We propose an ML solution for PD diagnosis based on subcortical neurodegeneration, a well-established feature of PD [15, 16].

Recent studies have demonstrated the potential of machine learning and deep learning techniques for PD classification using neuroimaging modalities. Convolutional neural networks (CNNs) have shown promising performance in classifying PD using magnetic resonance imaging (MRI) data, even with limited and heterogeneous datasets from multiple centres in one case
Ensemble models combining multiple machine learning algorithms have also achieved high accuracy in PD classification using MRI-derived features [22]. Additionally, retinal fundus imaging has emerged as a potential diagnostic screening tool for PD, with machine learning models demonstrating the ability to differentiate PD individuals from healthy controls [23]. Multimodal approaches, such as combining PET and MRI radiomics with clinical characteristics, have shown promise in differentiating PD from other neurodegenerative disorders like multiple system atrophy [24]. Furthermore, advancements in deep learning architectures, such as enhanced ResNeXt models, have been applied to improve the efficiency and accuracy of image categorisation tasks relevant to PD diagnosis [25].

Subcortical neuro-degeneration or brain atrophy in PD patients has been widely reported in the scientific literature [15, 16, 26–28]. Thalamus, which is the largest subcortical structure and acts as an information relay centre for all the signals to the cerebral cortex except for the sense of smell, is affected in the PD. A shape difference between the right and left Thalami has been reported in patients with the PD [29]. Caudate and putamen atrophy have also been reported in the scientific literature [30, 31]. Globus pallidus processes both motor and non-motor information by forming connections with basal ganglia and other brain regions. Impaired dopamine transfer and insufficient dopamine levels in globus pallidus have been reported as contributing factors towards the PD [32, 33]. Changes in the structure and volume of other subcortical brain regions like hippocampus, amygdala, nucleus accumbens and brainstem have also been reported in the scientific literature [34–37].

Building upon these findings, this work aims to segment all these subcortical structures and make an attempt to answer the research questions listed below. Also, the aims and contributions of this work have been listed as well.

1.1 Research questions

(1) How accurately can machine learning models diagnose PD using neuroimaging data from subcortical brain regions?

(2) What is the comparative diagnostic performance of different machine learning algorithms based on these brain regions for PD identification?

(3) To evaluate gender-based differences in ML model performances for brain scans in PD diagnosis.

1.2 Aims and contribution

(1) Develop precise ML models for diagnosing PD using neuroimaging data from relevant subcortical brain regions.

(2) Compare the diagnostic performance of different ML algorithms based on these brain regions to identify the most effective approaches.

(3) Advance medical AI by enabling early and accurate detection of PD, potentially enhancing clinical decision-making and patient outcomes.

(4) Provide valuable insights into the neural substrates of PD pathology through the analysis of neuroimaging data from subcortical brain regions.

2. Materials and methods

2.1 Data acquisition and preprocessing

3D T1 MRI images for PD and healthy controls (HCs) were obtained from the Parkinson’s Progression Markers Initiative (PPMI) upon request after approval by the PPMI Data Access
Committee, with all data collected at baseline. Due to a significant imbalance in the count of the scans between the two classes (HC & PD), an equal number of scans were selected from each class. This is done to avoid the biases that could arise in training due to a noticeable data imbalance between the two classes. More than one scan was available for many subjects, but only one scan per subject was used to ensure that every data point in the training was unique. For training, gender segregated scans were used with the number of Male\_Scans = 220 (PD\_Male = 110, HC\_Male = 110) and Female\_Scans = 130 (PD\_Female = 65, HC\_Female = 65). The dataset contains more T1 MRI scans for men than women; this is why we use a larger sample size when training ML classifiers on male scans. For the samples, the age range was 37–82 years for women and 31–83 years for men.

2.2 Segmentation and isolation of subcortical structures
Subcortical structures were segmented using FMRIB’s Integrated Registration and Segmentation Tool (FIRST), which is present in FSL (version 6.0.5) [38]. FIRST employs a Bayesian model incorporating shape and intensity information to segment subcortical brain structures effectively. The MR images were registered to the MNI152 template using affine registration, aligning all the scans to a standard brain space to facilitate accurate comparison. The model iteratively searches over a predefined number of significant modes of variation in the statistical shape model, ensuring the most accurate fit in order to generate binary segmentation masks of the subcortical regions. More information on this can be found in the work of Patenaude et al. [38]. These masks were then applied to the whole brain MR images to zero out non-structure voxels. The remaining segmented structure was then cropped to its bounding box, extracting it into a separate image file for each structure.

2.3 Machine learning model development and evaluation
The isolated subcortical structures were split into gender-specific training and testing sets of ratio 80:20. Each ML classifier was trained and tested separately on gender-specific scans. Prior to model training, feature scaling was performed on all the scans to ensure that all the features were on the same scale. This step was performed using StandardScalar, which is available in the scikit-learn library. Parameters were defined and optimized when training on male brains to maximize accuracy and minimize over-fitting. These optimized parameters were also used when training on the brain scans of the women. ML classifiers that were used are Random Forest, Extra Trees, AdaBoost, Gradient Boosting, XGBoost, Logistic Regression, SVMs (linear and polynomial), K-Nearest Neighbours (KNN) and Gaussian Naive Bayes (NB).

Figure 1 represents the graphical outline of how FSL–FIRST generates masks of the subcortical structures and the isolation of these structures, Figure 2 illustrates all the subcortical structures that were isolated and Figure 3 illustrates the general outline of the entire process of the isolation of the subcortical structures and the subsequent ML training and model development.

3. Results and discussion
This work focused on evaluating the proficiency of ML classifiers in identification of the PD by making use of MRI scans of the distinct brain structures, with a focus on gender-specific variations. Analyses were performed separately on male and female brain scans, utilizing ML classifiers such as Random Forest, Extra Trees, AdaBoost, Gradient Boosting, XGBoost, Logistic Regression, SVMs (linear and polynomial), KNN and Naive Bayes.

Brain structures that were used are putamen, thalamus, brainstem, accumbens, amygdala, caudate, hippocampus and pallidum. Performance of ML classifiers was measured based on the following metrics: accuracy, specificity and sensitivity. Complete and gender-specific results
Figure 1. Workflow detailing the FSL-FIRST segmentation of subcortical brain regions, followed by post-segmentation processing for extraction of the actual brain structures from binary masks, which FIRST generates.

Source(s): Authors’ own creation

Figure 2. Images representing the actual extracted brain structures.

Note(s): The caption of each of these images is the name of the subcortical structure that it represents.

Source(s): Authors’ own work
obtained on different classifiers using the subcortical brain regions have been reported in Supplementary Materials in Table S1 and Table S2. Table 1 outlines the best performing ML classifier for each brain structure and the obtained accuracy for each gender.

Variation in classifier performance between male and female brain scans was observed. This difference can be attributed to several factors, such as biological differences in brain structure and functioning, variation in disease progression, inherent biases in the training dataset, or a smaller number of female brain scans in the training data. These results align with recent studies that have emphasized the importance of considering gender differences in PD diagnosis and treatment [39, 40].

In male brains, the Extra Trees classifier demonstrated superior accuracy across several brain structures, such as the putamen (86.36%) and thalamus (77.27%) that could possibly

<table>
<thead>
<tr>
<th>Brain structure</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best algorithm</td>
<td>Accuracy (%)</td>
</tr>
<tr>
<td>Putamen</td>
<td>Extra Trees</td>
<td>86.36</td>
</tr>
<tr>
<td>Brainstem</td>
<td>Extra Trees</td>
<td>84.09</td>
</tr>
<tr>
<td>Thalamus</td>
<td>Extra Trees</td>
<td>77.27</td>
</tr>
<tr>
<td>Palladium</td>
<td>Extra Trees</td>
<td>78.00</td>
</tr>
<tr>
<td>Amygdala</td>
<td>Extra Trees</td>
<td>72.73</td>
</tr>
<tr>
<td>Caudate</td>
<td>Extra Trees</td>
<td>70.45</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>Random Forest</td>
<td>70.45</td>
</tr>
<tr>
<td>Accumbens</td>
<td>Extra Trees</td>
<td>65.91</td>
</tr>
</tbody>
</table>

Table 1. Accuracy of algorithms by brain structure and gender

Source(s): Authors’ own work
suggest a more distinct manifestation of PD-related features in these regions among males. Both of these structures play a part in movement control [41, 42]. This finding is consistent with recent studies that have reported gender differences in the presentation and progression of motor symptoms in PD [43–45]. Tree-based classifiers consistently seem to outperform other classifiers among men.

On the other hand, in women, best-performing classifiers seem to exhibit more variation depending on the specific brain region, even though tree-based classifiers seem to be more consistent in general. Naive Bayes classifier demonstrated improved accuracy in the amygdala (69.23%) for female brains. Amygdala in conjunction with other brain regions plays a part in the processing of emotions and has also been linked with the non-motor symptoms of PD [46, 47]. AdaBoost classifier demonstrated a reasonable accuracy of (61.54%) in putamen.

The thalamus and hippocampus, along with other interconnected brain regions, are critical for various cognitive functions such as attention, sensory processing, learning and memory. They demonstrated moderate performance in male brains and poor performance in female brains. These results may be suggestive of a more pronounced neurodegeneration in males in these brain regions, which is in line with recent findings that indicate a higher prevalence of cognitive impairment in men with PD [48, 49]. This inconsistency in algorithmic performance across brain regions and genders calls for the use of diverse and gender-balanced datasets in the development of diagnostic tools, and it further indicates the necessity of considering gender and brain region-specific variations in the expression of PD.

Tree-based classifiers, such as Extra Trees and Random Forest, outperformed other classifiers in male brains, indicating their effectiveness in capturing complex patterns in neuroimaging data associated with PD. However, their relatively lower performance in female brains suggests the need for further algorithmic refinements or more data to accurately interpret PD-associated neuroimaging data in women. Linear models, such as LR and SVM, may not be able to identify most of the nonlinear relationships compared to tree-based classifiers, leading to a lower performance.

To the best of our knowledge, we could find only one preexisting ML study [50] that trained gender-specific models using MRI data for the classification of the PD, and at the time of writing, only the preprint of this work was available. Both morphological and radiomic features were extracted in this study with a number of; Female scans = 70 (PD Female = 46, HC Female = 24). MLP was the best-performing algorithm, achieving an accuracy of 94.88% on male scans. Similarly for female scans, SVM was the top performer, with an accuracy of 96.31%. Our work employed a relatively larger and balanced sample size (Male scans = 220; PD Male = 110, HC Male = 110; Female scans = 130, PD Female = 65, HC Female = 65). This could provide our work with more generalization, but we would also point out that our study only used the subcortical structures, while the previous study used a much larger subset of morphological and radiomic features, allowing it to register and work with important features from many brain regions. While by the nature of our study, it was limited only to subcortical structures.

4. Conclusion
This study provides valuable insights into the potential of ML classifiers for the identification of PD using MRI scans of subcortical brain structures, with a specific focus on gender-specific variations. Our findings signify the importance of considering gender differences in the development of ML models for PD diagnosis. Variation in algorithmic performance between male and female brain scans emphasizes the need for gender-specific approaches in medical AI.
Our findings also reflect upon the complex nature of PD and the variable performance of ML classifiers across different brain regions. The superior performance of tree-based classifiers, particularly in male brains, suggests their ability to capture complex patterns in neuroimaging data associated with PD. However, the lower performance in female brains indicates the need for further algorithmic refinements and the inclusion of more diverse data to accurately interpret PD-associated neuroimaging data in women. In some brain regions, the pathological changes might be more pronounced, thus making it easier for the classifiers to detect, while the subtler changes in some brain regions make it harder to detect, which could be the reason for varying performance of the classifiers across different brain regions and sexes.

References


**Supplementary material**

The supplementary material for this article can be found online at: https://github.com/NairUlIslam/Supp_Material/blob/main/ACI-02-2024-0080_suppl1.pdf

**Corresponding author**

Ruqaiya Khanam can be contacted at: dr.kruqaiya@gmail.com