Systematic Review Article
http://mjiri.iums.ac.ir
Medical Journal of the Islamic Republic of Iran (MJIRI)

Med J Islam Repub Iran. 2025 (26 Dec);39.164. https://doi.org/10.47176/mjiri.39.164

The Application of Machine Learning to Predict Clinical Outcomes of Deep Brain Stimulation in Parkinson's Disease: A Systematic Review

Parisa Javadnia[1], Mohammad Rohani[1,2], Elahe Amini[1], Mohammad Yousefi[1], Atieh Jafarabadi Ashtiani[1], Sadra Rohani[3], Arezoo Farzi[1]*

Received: 16 Aug 2025 Published: 26 Dec 2025

Abstract

Background: Parkinson's disease (PD) is a degenerative condition of the nervous system that is primarily characterized by a gradual decline of motor function. For patients with suboptimal response to medical treatment, deep brain stimulation (DBS) is a well-recognized surgical approach. This systematic review evaluates the performance of machine learning (ML) models in classifying patients or symptoms or to predict postoperative outcomes following DBS in PD.

Methods: PubMed, Scopus, Cochrane, Embase, and Web of Science were searched in accordance with PRISMA through December 31, 2024. We included original human studies of DBS-treated PD in which ML used clinical (non-imaging) features to classify patients or symptoms, or to predict postoperative outcomes. Cohort, cross-sectional, and case-series designs were eligible. Imaging-based prediction studies were excluded.

Results: From 961 records, eight studies (n=555 patients) met the inclusion criteria. Three studies performed preoperative-to-postoperative outcome prediction, and five focused on symptom or patient classification. Targets included motor severity, speech outcomes, and gait-related measures. The Support Vector Machine (SVM) was the most frequently applied ML model, followed by the k-nearest neighbor, which was used in three studies. Commonly used assessment tools included the Mini-Mental State Examination (MMSE), the Hoehn and Yahr Scale, and the Unified Parkinson's Disease Rating Scale (UPDRS).

Conclusion: This review highlights early but exploratory application of ML for patients' or symptoms classification and predicting clinical outcomes and adverse events following DBS using preoperative clinical data. However, the current evidence is sparse, single-center, and methodologically heterogeneous, with limited external validation. Therefore, clinical translation remains premature.

Keywords: Deep Brain Stimulation, Machine Learning, Parkinson's Disease, Artificial Intelligence, Motor symptoms

Conflicts of Interest: *None declared*
Funding: *None*

Cite this article as: Javadnia P, Rohani M, Amini E, Yousefi M, Jafarabadi Ashtiani A, Rohani S, Farzi A. The Application of Machine Learning to Predict Clinical Outcomes of Deep Brain Stimulation in Parkinson's Disease: A Systematic Review. *Med J Islam Repub Iran.* 2025 (26 Dec);39:164. https://doi.org/10.47176/mjiri.39.164

Introduction

PD is a neurodegenerative disorder impacting nearly 1% of people worldwide who are aged 60 years or older (1). PD manifests with motor and non-motor symptoms, including tremor, bradykinesia, rigidity, depression, and cognitive decline (2, 3). Being affected by these symptoms significantly lowers the individual's quality of life over time and disease progression (4). First-line treatment is medical therapy in almost all patients. However, poor medication response can lead patients to surgical approaches to control their symptoms and increase their independence (5). DBS is a widely adopted surgical technique used in PD patients to help control their symptoms (6). The operation is performed by implanting electrodes into specific brain areas responsible for movement con-

Corresponding author: *Dr Arezoo Farzi, arezoofarzi77@gmail.com*

1. Skull Base Research Center, The Five Senses Health Institute, Rasoul Akram Hospital, Iran University of Medical Sciences, Tehran, Iran
2. Department of Neurology, The Five Senses Health Institute, Rasoul Akram Hospital, Iran University of Medical Sciences, Tehran, Iran
3. Neurosurgery Department, Rasool Akram Hospital, Iran University of Medical Sciences, Tehran, Iran

↑What is "already known" in this topic:
DBS is effective for motor symptom control in PD, but predicting individual clinical outcomes remains challenging.

→What this article adds:
This systematic review summarizes machine learning-based classification and early predictive modeling studies in DBS for PD, while highlighting current limitations and the need for external validation.

trol; these areas include the subthalamic nucleus (STN), globus pallidus internus (GPI), and ventral intermediate nucleus (VIN) (7, 8). Despite the benefits DBS offers, it can also lead to unwanted outcomes, especially due to unintended stimulation of nearby brain regions. Therefore some patients do not receive much benefit from this approach (8).

Although a considerable number of studies have attempted to identify factors that predict treatment success, clinicians still struggle to forecast which patients will benefit most (7, 8). One reason is that DBS effects depend on very small differences in electrode position, while electrode localization and image registration can introduce meaningful uncertainty, which makes reliable prediction harder (9, 10). In addition, patients do not respond in the same way, outcomes can vary widely across individuals, and factors such as clinical and genetic subtypes may contribute to this heterogeneity (11, 12). Results also differ across centers because practices vary in patient selection, target choice, surgical approach, and postoperative management, which limits how well prediction models generalize (13). Furthermore, studies often use different endpoints and responder definitions, and even common clinical tests like the levodopa challenge do not always predict DBS benefits well when evaluated in broader datasets (14-16). Therefore, the development of improved methods for patient selection is critically needed.

Although DBS is an effective way for controlling motor symptoms, predicting individualized outcomes is still a significant challenge.

Conventional statistical techniques, such as logistic regression, are commonly used for DBS outcome modeling, but they can fall short when relationships between variables are complex and non-linear across patients.

Recent work has explored a range of ML-based modeling strategies using preoperative clinical and neuropsychological features. As an example, Habets et al. (17) developed a multivariable logistic regression model to identify weak responders one year after STN-DBS. While promising, the evidence remains limited and requires external validation. This study suggests that ML-based approaches may offer added value for DBS outcomes prediction compared with conventional models, although evidence is still limited and model performance is not consistently validated across independent cohorts.

Beyond routine clinical assessments, non-imaging data sources such as intraoperative microelectrode recordings and wearable or sensor-derived features have been explored in ML studies in PD and DBS research. Park et al. (18), for instance, used deep learning on intraoperative microelectrode recordings to model DBS-related clinical outcomes. Similar non-imaging directions have also been explored using intraoperative recordings and wearable or sensor-derived features, although endpoints and evaluation vary across studies (19-21).

In this context, various recent studies have investigated the application of ML techniques to predict clinical outcomes of DBS in patients with PD (7, 17, 19). ML, a subfield of artificial intelligence (AI), identifies patterns in complex datasets to generate predictive insights (22). In DBS interventions, ML can be employed to model associations between preoperative clinical variables and treatment efficacy (7, 22). The approaches in ML are mainly grouped into supervised, unsupervised, semi-supervised, and reinforcement learning. Within the framework of supervised learning, commonly applied predictive techniques include Logistic Regression, Naive Bayes, Random Forest (23), SVM, Neural Network, Deep Neural Network, and Decision Tree. Supervised learning predicts the outcome using labeled datasets, whereas unsupervised learning extracts patterns from unlabeled input variables. Comparing the statistical parameters of different ML models can help identify the efficiency of each model in clinical practice applications (22).

To date, few reviews have specifically focused on ML models that use clinical and non-imaging predictors of DBS outcomes in PD, and prior work has typically been narrative in scope with limited formal quality assessment. Therefore, this study aims to synthesize studies that applied machine learning to either classify symptoms or patient subgroups in DBS-treated PD cohorts, predict postoperative clinical outcomes using preoperative clinical and non-imaging features.

Methods

The design and reporting of this research were conducted in full compliance with the PRISMA 2020 guidelines for systematic reviews and meta-analysis to uphold high standards of transparency and methodological soundness.

Literature Search Plan and Data Sources

A structured and comprehensive search of the existing scientific literature was carried out across five electronic databases: PubMed, Scopus, Cochrane, Embase, and Web of Science. The search strategy utilized a mix of keywords, MeSH terms, and Boolean operators (AND, OR) related to "artificial intelligence," "machine learning," "Parkinson's disease," "idiopathic Parkinson's disease," "deep brain stimulation," and "DBS." All records included from database inception through December 31, 2024, were considered without restrictions on publication date. An independent screening of the reference lists from eligible records was conducted to find any potentially relevant papers missed in the initial search.

Study selection

For study selection, all identified citations were transferred into EndNote version 20 to facilitate organized reference handling. After removing duplicate entries, two independent reviewers separately screened the titles and abstracts during the initial evaluation process. Following this, the full texts of the articles were assessed according to previously defined eligibility criteria. Disagreements were addressed by a third reviewer.

Eligibility criteria

Eligibility was limited to English-language primary studies involving human subjects diagnosed with idiopathic PD who had undergone DBS. ML algorithms were applied using clinical, neurophysiological, wearable, or sensor-derived features to either classify patients or symp-

[DOI: 10.47176/mjiri.39.164]

toms in DBS-treated PD cohorts or model postoperative outcomes. We required that participants had undergone DBS, but did not restrict analysis to strictly preoperative measurements when the ML task involved clinically relevant classification in DBS-treated populations. Eligible study designs included cohort studies, cross-sectional studies, and case series. We also included studies that used ML to predict postoperative adverse outcomes.

The excluded studies were non-English publications, studies on animal models, studies with fewer than four subjects, case reports, review articles, letters to the editor, and commentaries. Additionally, studies that included patients with PD without a confirmed diagnosis of idiopathic PD or those who did not receive DBS were excluded. Furthermore, studies in which ML models used imaging data instead of clinical symptoms as predictive variables were also excluded.

Data extraction

Two independent reviewers extracted and aligned data from the included studies, organizing the information into structured Excel documents. The dataset consisted of the author's name, study population size, demographic features (age and sex), country of origin, preoperative clinical symptoms used for prediction, assessment tools, ML subtypes, predicted postoperative symptoms, and outcome.

Quality Assessment

The studies included were evaluated for their methodological quality with the help of the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. The QUADAS-2 instrument evaluates potential risk of bias across four areas: participant selection, assessment of the index test, appraisal of the reference standard, and the management of study flow and timing. It also considers concerns regarding applicability (24). Two reviewers independently assessed each study, and disagreements were resolved by discussion with a third reviewer. Given that QUADAS-2 was specifically created for diagnostic test accuracy studies, we utilized it here as a structured framework to summarize key sources of bias. Findings were interpreted cautiously in light of the specific challenges of ML-based modeling studies. We acknowledge that tools such as PROBAST have been specifically developed for assessing prediction model studies. However, due to incomplete reporting and heterogeneous study designs, a full PROBAST assessment was not feasible. QUADAS-2 was therefore used as a pragmatic framework to summarize major sources of bias, and all interpretations were made cautiously in light of ML-specific limitations.

Results

Selection of Studies

The process of selecting studies, following the PRISMA 2020 guidelines, is depicted in Figure 1. A total of 961 citations were retrieved from electronic database searches. Following the elimination of 307 duplicate entries, the remaining studies were screened using their titles and abstracts. Following the screening process, eight studies met all the inclusion criteria and were selected for the final analysis.

Characteristics of the Studies

The combined study population consisted of 555 patients diagnosed with PD. Table 1 presents an overview of the main features of the selected studies. The UPDRS was the most frequently used assessment tool, used in 5 studies. The predicted outcomes varied across the studies and included motor symptoms (three studies) and speech and gait impairment (two separate studies). Among the ML methods utilized, SVM was the most common, applied in four studies out of the eight, followed by the k-nearest neighbor algorithm used in 3 studies.

Risk of Bias and Relevance Evaluation

The results of the QUADAS-2 evaluation are shown in Figures 2 and 3. The most frequent risk of bias was in the index test domain, followed by patient selection, where methodological details were often inadequately described or justified. One study showed high applicability concerns in the patient selection domain, while the remaining studies demonstrated no major issues. These findings guided our interpretation of the evidence and highlighted methodological weaknesses that challenge the generalizability of the results.

The bar chart displays how the included studies were rated for risk of bias and applicability (high/red, unclear/yellow, and low/green) across the four domains of the QUADAS-2 framework: patient selection, index test, reference standard, and study flow and timing

Each domain is evaluated for risk of bias and applicability in the eight included studies. Color coding reflects risk status: green for minimal concern, yellow for intermediate or uncertain concern, and red for elevated concern.

For clarity, we first distinguish between cross-sectional classification tasks and longitudinal prediction of postoperative outcomes. Within each category, results are organized by clinical domain (motor, speech, gait, and cognition). Studies by Angeles et al (20), Yohanandan et al. (21), Suppa et al. (25), Watt et al. (26), and Sabo et al. (27) are symptom and patient classification studies. Furthermore, studies by Alhourani et al. (28), Habets et al. (17), and Chang et al. (29) are postoperative outcome prediction studies.

Synthesized Findings

Because of substantial methodological variability across the included studies, including differences in ML algorithms, input features such as clinical and wearable data, outcome measures, and performance metrics such as accuracy, sensitivity, and specificity, F1 score, and k, a quantitative meta-analysis was not feasible. Instead, we adopted a structured approach. Findings were organized by outcome domains, namely motor symptoms, speech and voice, verbal fluency, gait, and cognitive outcomes. Within each domain, we systematically described the type of input variables, ML models applied, and the main predictive performance reported in the studies. This resulted in a clear and organized summary of the evidence, while ac-

[DOI: 10.47176/mjiri.39.164]

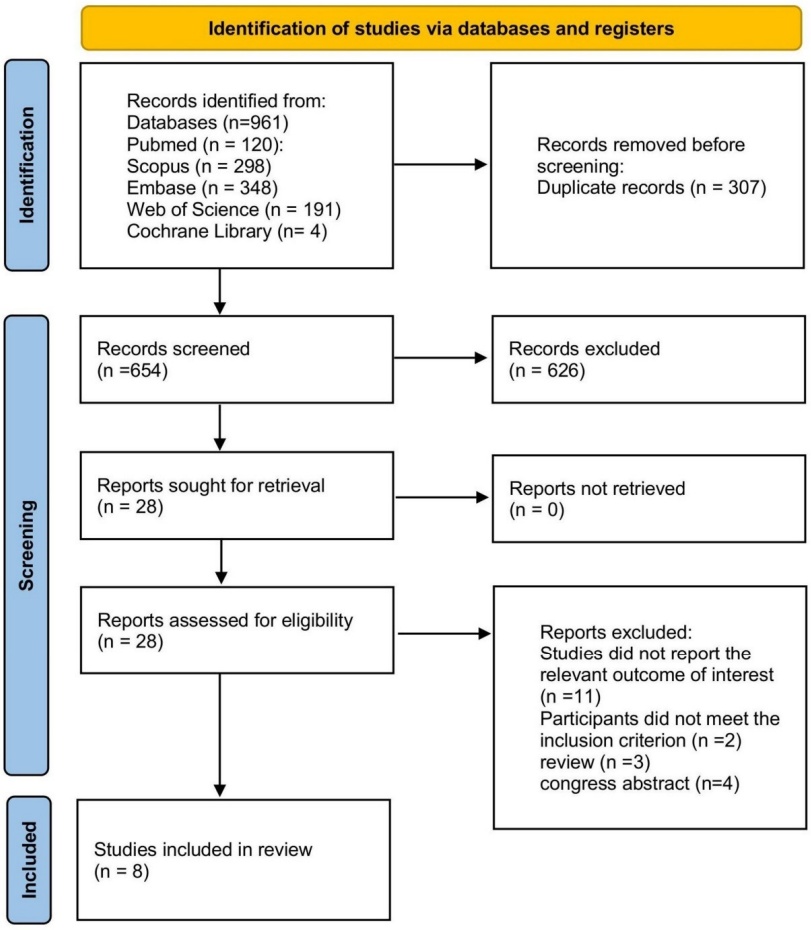

Figure 1. Diagrammatic representation of the literature selection workflow guided by the PRISMA 2020 guidelines.

knowledging the heterogeneity and limitations of the available data.

Motor symptoms

Two studies explored the use of wearable device sensor data to predict motor symptom severity using standardized clinical scales such as the UPDRS and the Bain-Findley Tremor Rating Scale (BTRS). Angeles et al. (20) examined 13 PD patients who underwent DBS and three healthy controls using wearable devices placed on their more affected arms. The sensor data were processed to calculate the UPDRS scores. After that, the correlation was assessed using different types of ML (Simple trees, linear SVM, and fine k-nearest neighbors (KNN)) with clinician feedback. The fine KNN model achieved very high accuracies across rigidity and bradykinesia subscores (up to 100% for elbow rigidity), while linear SVM performed best for postural tremor (82.9%).

In a related study, Yohanandan et al. (21) used the BTRS rather than the UPDRS for tremor evaluation. They found that random forest ML classifiers achieved the highest agreement with clinical scores (κw = 0.81). These findings suggest that ML models may, in principle, approximate clinician-rated motor scores from wearable sensors, but the evidence remains strictly exploratory, with no external validation and considerable risk of overfitting. In another study, a logistic regression model was developed to identify weak responders to STN-DBS using preoperative variables in addition to neuropsychological variables. The model achieved a diagnostic accuracy of 78%, and high UPDRS scores in the on-medication state emerged as the strongest predictor of post-DBS outcomes (17).

Speech and verbal fluency

Alhourani et al. (28) applied ML techniques to identify neuropsychological predictors of postoperative verbal fluency decline in a cohort of 90 PD patients who underwent DBS. Among various linear and non-linear algorithms, support vector regression (SVR) and the least absolute shrinkage and selection operator were the most effective. Additionally, it was shown that greater deficits in

[DOI: 10.47176/mjiri.39.164]

Table 1. A Summary of the Studies Included Evaluating Machine Learning Models to Predict Clinical Outcomes After Deep Brain Stimulation in Parkinson's Disease

Author (year)	Predicted symptoms	ML type and measurement method	Results
Alhourani et al. (2022)	Post-op Verbal fluency by preoperative cognition	ML type: SVR, LASSO, extra-trees, KNN, ordinary least squares; Measurement method: DKEF, Digit span, QUIP	-Greater scores in pre-surgical fluency, digit span, education, and MMSE are predictors of higher post-op verbal fluency score -higher frontal system deficit scores, older age, elevated impulsive-compulsive disorder questionnaire scores, disease duration, and behavioral inhibitory are predictors of lower post-op verbal fluency scores
Angeles et al. (2017)	Bradykinesia, rigidity, tremor	ML type: Supervision machine learning (simple decision tree, linear SVM, fine KNN; Measurement method: UPDRS, sensory system attached to arm	-ML, by using data recorded by a sensor system, with 90.9% accuracy, could predict the clinician's severity score -the highest accuracy was for the fine KNN model for elbow rigidity (100%), wrist rigidity (95%), bradykinesia (92.5%), kinetic tremor (87.3%), rest tremor (87.8%) -for postural tremor, the highest accuracy was for linear SVM (82.9%)
Chang et al. (2022)	Cognition	ML type: Nomogram; Measurement method: MocA, MMSE, HAMA, HAMD	ROC: 0.98, AUC:0.987 C-index:0.98; The nomogram effectively predicted the chance of substantial cognitive enhancement one year after STN-DBS in PD patients
Habets et al. (2020)	UPDRS part I-IV, H&Y scores, LEDD, and neuropsychological measures evaluating executive function (in particular verbal fluency (semantic and lexical) and response inhibition).	ML type: multivariate logistic regression; Measurement method: UPDRS I-IV scores, H&Y scale, LEDD, category fluency test, verbal fluency test, interference score of the Stroop Color Word test.	Accuracy: 0.78, Sensitivity: 0.80, Specificity: 0.76, AUC=0.79 (SD=0.08), PPV=0.63, NPV=0.88; These results support the proof-of-concept that machine learning can predict individual motor outcomes after STN DBS for PD using preoperative clinical variables.
Sabo et al. (2023)	Gait improvement	ML type: Spatial-temporal graph CNN; Measurement method: MDS-UPDRS-gait scores, Video recording	Although the vision-derived model, developed using Parkinsonian gait data, failed to accurately predict MDS-UPDRS-gait scores in a different cohort of PD patients, it nevertheless captured weak but significant proportional fluctuations associated with medication and DBS interventions.
Suppa et al. (2023)	Voice impairment	ML type: SVM, ANN; Measurement method: Voice recordings, UPDRS-III sub-item voice	From a clinical perspective, individuals with STN-DBS exhibited more severe vocal disturbances than those managed with oral pharmacotherapy. Using machine learning-based analysis, it was possible to distinguish the vocal patterns of the DBS group from those receiving medication with high objectivity and precision
Watt et al. (2024)	Freezing of gait	ML type: KNN, Naïve Bayes, Random Forest, SVM; Measurement method: MDS-UPDRS III, Wearing device, Video recording	Machine learning algorithms show high effectiveness in distinguishing individuals with advanced PD as freezers or non-freezers using Stand-and-walk trials performed in the absence of both pharmacological treatment and active stimulation.
Yohanandan et al. (2016)	Tremor	ML type: RF, Multilayer perceptron, SVM, DT, Bayesian network, radial basis network, Naïve Bayes; Measurement method: BTRS, Wearing device, Video recording	This study shows that the RF was the most accurate model (κw = 0.81) at transforming tremor information into BTRS ratings.

MDS-UPDRS, movement disorder society-Unified Parkinson's Disease Rating Scale; SVM, support vector machine; STN-DBS, subthalamic nuclei-deep brain stimulation; LEDD, levodopa equivalent daily dosage; BTRS, Bain-findley tremor rating scale; LASSO, ; KNN, K-Nearest Neighbors; PD, parkinsonian disease; DT, decision tree; RF, Random forest; MMSE, Mini-Mental State Examination ;Y, year; H & Y, Hoehn and Yahr Scale; ANN, artificial neural network; CNN, convolutional neural network; HAMA, Hamilton anxiety; HAMD, Hamilton depression; MocA, Montreal Cognitive Assessment; QUIP, Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease; DKEF, Delis-Kaplan Executive Function; LASSO, Least Absolute Shrinkage and Selection Operator

the frontal system, older age, higher scores on the impulsive-compulsive disorder questionnaire, longer disease duration, and increased behavioral difficulties were related to a more serious risk of verbal fluency problem following DBS.

In another study, Suppa et al. (25) used ML models to compare voice impairment severity between 50 DBS-treated PD patients and 51 patients treated with oral medication. They used UPDRS-III scores and ML-based voice analysis for their assessment. The study reported that patients with STN-DBS exhibited greater voice impairment, and the SVM model could successfully distinguish between the voice profiles of the DBS and medication groups with high accuracy.

Gait impairment

Watts et al. (26) examined 21 individuals diagnosed with idiopathic PD who received bilateral STN-DBS therapy and had gait impairment defined by a score of 2 or 3 on the MDS-UPDRS gait component. The study employed wearable sensors and applied various ML models (KNN, random forest, logistic regression, Naive Bayes, and SVM) to distinguish between patients who exhibited freezing of gait, known as freezers, and those who did not have freezing of gait, known as non-freezers. All models

[DOI: 10.47176/mjiri.39.164]

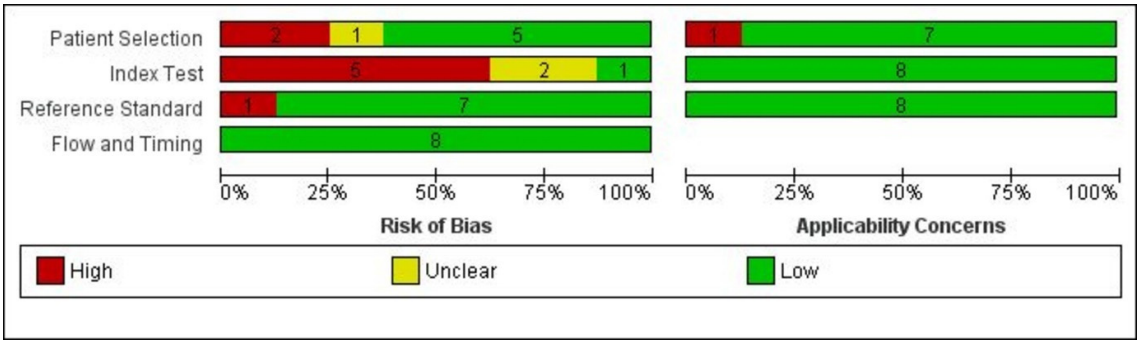

Figure 2. Quality assessment of included studies (QUADAS-2)

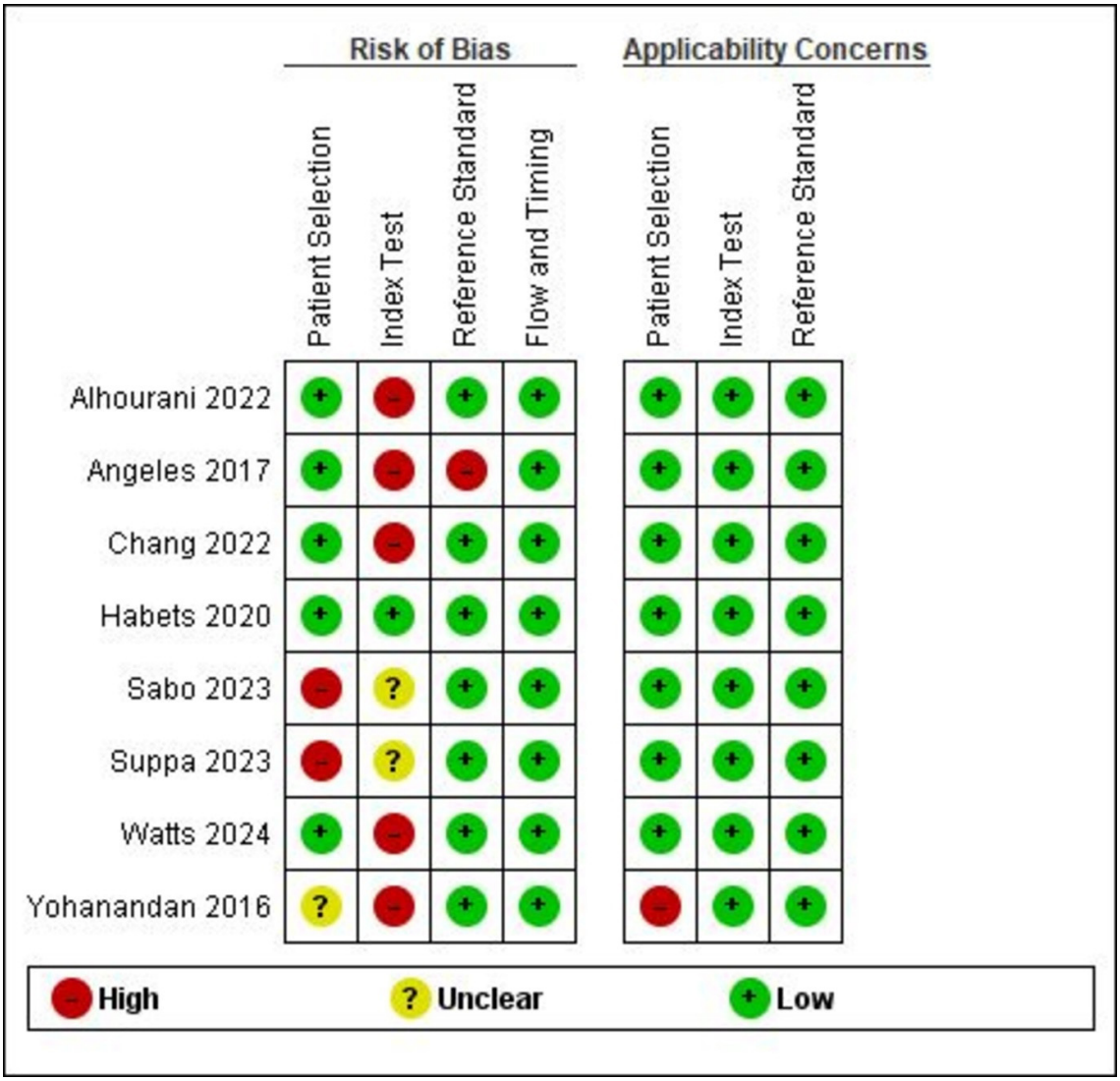

Figure 3. Detailed QUADAS-2 risk of bias and applicability assessment for each included study.

revealed similar performance in classification.

In contrast, Sabo et al. (27) used a spatial-temporal graph convolutional network (ST-GCN) model trained on video recordings to predict gait scores in another group of PD patients. The model could not reliably predict MDS-UPDRS gait scores, but it successfully detected measurable changes in response following both medication and DBS. The observed discrepancies can be explained by heterogeneity in data acquisition methods; Watt et al. (26) used wearable sensors, while Sabo et al. (27) relied on

[DOI: 10.47176/mjiri.39.164]

video-based analysis.

Predicting DBS outcomes

Only a subset of the included studies directly modeled postoperative DBS-related outcomes using preoperative patient-level variables. Habets et al. (17) used a multivariable logistic regression approach based on preoperative clinical and neuropsychological variables and reported an accuracy of 0.78 with an AUC of 0.79 (PPV 0.63; NPV 0.88) for predicting postoperative outcomes. Chang et al. (29) developed a nomogram model to predict postoperative cognitive outcome and reported high discrimination on ROC-based metrics (ROC 0.98; AUC 0.987; C-index 0.98). Alhourani et al. (28) investigated postoperative verbal fluency outcomes using multiple regression ML approaches (SVR, LASSO, extra-trees, KNN), where preoperative cognitive and demographic measures were identified as relevant predictors. However, performance metrics were not consistently reported. Overall, while these findings suggest potential for preoperative prediction, the evidence remains preliminary due to limited reporting consistency across outcomes and model evaluation, and the lack of external validation in the included predictive studies.

Discussion

Several prior systematic reviews have explored the application of AI in managing patients with PD. They explore multiple key aspects, from identifying clinical symptoms and tracking disease progression to optimizing treatment effectiveness by adjusting stimulation parameters.

Huang et al. (30) executed a systematic review examining the use of wearable sensors in combination with ML. The goal of this work was to detect freezing of gait and predict fall risk in individuals diagnosed with PD. This study revealed that these technologies effectively identify freezing of gait episodes and predict fall risk with notable efficacy. In Addition, Sun et al. (31) investigated the use of ML in predicting cognitive impairment in PD following DBS in a systematic review and meta-analysis. The study considered various clinical parameters, including the MMSE, Montreal Cognitive Assessment (MOCA), age, sex, disease duration, and imaging data. The study demonstrated that among the ML models, SVM achieved the highest sensitivity (83%), while artificial neural networks (ANN) showed the highest specificity (93%).

A review by Oliveria et al. (32) suggested that ML can help create a personalized closed-loop DBS system by analyzing different electrophysiological biomarkers, thereby addressing the symptoms of individual PD patients. In this systematic review, our aim was to evaluate how effective ML algorithms are in classifying patients and their symptoms, and predicting how patients will respond to DBS and forecasting the likelihood of adverse events following DBS based on their preoperative clinical profiles. We excluded studies that used imaging parameters for assessments and focused instead on research where ML models relied on clinical symptoms. This approach helps make the study's results more clinically feasible.

Previous narrative reviews, such as Watts et al. (33), have provided a broad overview of ML applications across multiple aspects of DBS in PD, including candidate selection, programming optimization, surgical targeting, and mechanistic insights. While these reviews highlight the interdisciplinary potential of ML in DBS, they did not perform a structured quality assessment and were limited to studies published up to 2020. In contrast, our systematic review focuses specifically on patient or symptoms classification and clinically relevant predictors of postoperative outcomes, applies formal risk of bias evaluation, and incorporates the most recent literature up to 2024, thereby providing a more targeted and up-to-date synthesis.

Recent studies have suggested that ML may support DBS-related decision-making; however, it is important to distinguish between models using preoperative clinical or wearable features and imaging-driven approaches that rely on neuroimaging data and, in some cases, postoperative electrode localization. Because these approaches address different tasks and use different input modalities, their reported performance is not directly comparable. Within the studies reviewed, performance ranged from symptom-severity classification using sensor-derived features (with accuracies reported up to 90.9%) (20) to preoperative clinical prediction of postoperative motor response (78% accuracy for identifying weak responders after STN-DBS) (17). By contrast, imaging-based approaches that fall outside our eligibility criteria have reported accuracies such as 62.5% utilizing patient-specific 3D point clouds generated from preoperative MRI and postoperative CT (34) and 88% using fMRI (35). These findings are promising, but should be interpreted cautiously given differences in outcomes, inputs, and evaluation protocols. Therefore, we present them here only as context and not as part of the systematic synthesis (34, 35). Accordingly, to avoid conflating fundamentally different ML tasks and data modalities, the remainder of this Discussion focuses on evidence that aligns with our eligibility criteria and separates symptom classification from postoperative outcome prediction.

Motor response

Strong responsiveness to DBS in PD patients, as indicated by symptom reduction, is associated with better levodopa responsiveness, lower baseline tremor severity, and a younger age (36). However, the strength and consistency of these associations remain controversial (36-39). ML effectively identifies patterns critical for surgical decision-making and is used to optimize DBS programming and electrode placement (40, 41).

Speech

Patients with PD exhibit a spectrum of voice disorders, typically including hypophonia, monoloudness, and monopitch, as well as hypophonic and hypotonic articulation, which together are referred to as hypokinetic dysarthria (42, 43). Despite being a well-established therapy for advanced PD, the influence of STN-DBS on axial symptoms, such as vocal impairments, remains poorly defined

[DOI: 10.47176/mjiri.39.164]

(44-46). Following STN-DBS, the incidence of speech disturbances has been documented to range from 1% at six months to as high as 70% by three years of follow-up (36, 45, 47).

Several mechanisms have been proposed to explain the development of dysarthria in individuals with PD following STN-DBS. First, a decrease in the levodopa equivalent daily dosage (LEDD) may negatively impact speech production. Second, antidromic activation of the hyper-direct pathway may lead to aberrant activation of cortical regions, thereby contributing to phenomena such as stuttering and spastic speech patterns (48, 49). Third, stimulation may extend to adjacent structures, such as the corticobulbar and corticospinal tracts, further disrupting motor control of speech (50-52). ML methodologies offer valuable tools for identifying vocal abnormalities associated with neurological conditions, including PD (25, 43, 50, 53).

Suppa et al. (25) report that SVM ML can accurately differentiate the vocal characteristics of patients treated with STN-DBS from those managed with oral medication, using post-stimulation voice deterioration as the key discriminative factor.

Verbal fluency

Verbal fluency (VF) relies on executive functions beyond verbal skills (52), including working memory (53), cognitive flexibility, and response inhibition (54). These functions can be impaired initially in PD (55) and may also be affected by DBS. Thus, understanding pre-existing executive function deficits in patients is crucial for predicting DBS-related changes in VF (28).

Alhourani et al. (28) investigated three types of VF, including letter, semantic, and action fluency, employing an ML approach to assess neuropsychological variables that predict VF deterioration following DBS. LASSO and SVR were the most effective predictive methods. However, the simple regression model also provided comparable variance, offering a more straightforward option for clinical predictions. Across all three predictive models, greater baseline levels of fluency, digit span performance, education, and Mini-Mental State Examination scores were linked to superior fluency outcomes following surgery.

Freezing of gait

Gait disturbance can manifest in the early stages of PD, with research indicating that subtle alterations may be detectable during the prodromal phase (26, 56, 57). In patients with PD, common gait abnormalities include diminished stride length, reduced walking velocity, absence of arm swing, and difficulties with multi-step turning (58, 59).

Freezing of gait (FOG) episodes can either be triggered by specific actions or occur paroxysmally, significantly elevating the likelihood of falling and negatively impacting patients' quality of life (60, 61). The majority of research on FOG utilizes ML approaches to detect and forecast freezing episodes based on signals captured by wearable sensors (62).

Watt et al. (26) studied PD patients undergoing STN-DBS who exhibited freezing of gait when medication and DBS were both discontinued. A random forest model identifies ten key predictive features, encompassing spatial parameters such as foot strike angle, trunk and lumbar range of motion, stride length, and toe-off angle, along with temporal parameters such as gait speed and lateral step variability. Various ML models accurately classified patients based on instrumented stand and walk trials, including KNN, naïve Bayes, random forest, logistic regression, and SVM.

Cognitive decline

Along with the characteristic motor symptoms, PD often includes non-motor features such as cognitive impairment and cognitive decline (63, 64). Mild cognitive impairment affects approximately 25% of PD patients, with dementia impacting 20-70% (65, 66). Although DBS primarily targets motor symptoms, it may adversely affect cognitive domains, including memory, visual function, and executive performance (67, 68).

Chang et al. (29) developed a nomogram model to assess postoperative cognitive improvement in 103 PD patients after one year of STN-DBS, using both univariate and multivariate logistic regression. The multivariate analysis revealed four key predictors of cognitive improvement: years of education, MoCA scores, MMSE scores, and UPDRS part III. The resulting model demonstrated strong predictive power, with a concordance index (C-index) of 0.985 and a sensitivity of 98% on the receiver operating characteristic curve.

The effectiveness of ML models in predicting clinical outcomes following DBS shows promising results. Chang et al. (29) achieved an impressive AUC (area under the curve) of 0.987 and a C-index of 0.98, demonstrating strong predictive power. On the other hand, Habets et al. (17) reported a more modest AUC of 0.79, with a PPV (positive predictive value) of 0.63 and an NPV (negative predictive value) of 0.88, indicating that while the models are still valuable, there is room for improvement in prediction accuracy.

Limitations

This work offers the first detailed assessment of ML applications for predicting clinical outcomes following DBS in PD. However, several limitations should be acknowledged. The number of eligible studies was limited, and most were single-center and retrospective in design. It should also be noted that many existing ML studies in DBS focus on imaging-based predictors, which were outside the scope of our research question. By restricting inclusion to clinical and symptom-based predictors, only eight studies met the eligibility criteria, further limiting the generalizability of our conclusions. External validation was rarely performed, limiting the clinical applicability of these results.

Studies varied considerably in their choice of specific symptoms as input parameters, and outcome measures were inconsistently reported.

Furthermore, methodologies for evaluating preoperative symptoms varied across studies, making it difficult for reviewers to compare them. Further research should prior-

[DOI: 10.47176/mjiri.39.164]

itize large-scale, multicenter prospective studies to identify the most reliable symptom-based predictors for specific outcomes and to determine the most effective ML models for these predictions. Additionally, there is a lack of research comparing the efficacy of ML models in predicting symptom changes following DBS versus traditional oral medications. Another limitation is that we did not apply ML-specific risk-of-bias tools such as PROBAST or QUADAS-AI, and instead we adapted QUADAS-2 as a structured framework, which may not capture all sources of bias unique to prediction modeling studies. This knowledge gap underscores the need for further investigations in this field.

Conclusion

The role of ML techniques in predicting clinical outcomes for PD patients who have undergone DBS has been increasingly investigated in recent studies. These models might help clinicians select patients, facilitate consultations, and design individualized treatment plans. These results should be considered carefully, as the available studies are limited by small sample sizes, predominantly retrospective designs, and substantial methodological heterogeneity. Future validation through large, multicenter prospective investigations is required to confirm the reliability of ML models and to support their safe and practical integration into routine clinical practice.

Authors' Contributions

PJ: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review and editing.

MR: Methodology, Project administration, Supervision, Writing - review and editing.

EA: Conceptualization, Methodology, Formal analysis, Supervision, Writing - review and editing.

MY: Methodology, Formal analysis, Writing - original draft, Writing - review and editing.

AJ: Investigation, Supervision, Writing – original draft.

SR: Investigation, Supervision, Writing – review and editing.

AF: Investigation, Supervision, Writing – original draft, Writing – review and editing.

All authors have read and approved the final manuscript.

Ethical Considerations

The present study is a systematic review based only on existing published evidence. It did not involve the collection or analysis of new data from human participants. Therefore, ethics committee approval was not necessary in line with institutional and international guidelines.

Acknowledgment

The authors wish to convey their deep appreciation to all individuals who contributed to the preparation of this review.

Conflict of Interests

The authors declare that they have no competing interests.

References

1. Zhu J, Cui Y, Zhang J, Yan R, Su D, Zhao D, et al. Temporal trends in the prevalence of Parkinson's disease from 1980 to 2023: a systematic review and meta-analysis. Lancet Healthy Longev. 2024;5(7):e464-e79.
2. Kumar R, Tripathy M, Kumar N, Anand RS. Management of Parkinson's disease dysarthria: Can artificial intelligence provide the solution? Ann Indian Acad Neurol. 2022;25(5):810-6.
3. Kubu CS. The role of a neuropsychologist on a movement disorders deep brain stimulation team. Arch Clin Neuropsychol. 2018;33(3):365-74.
4. Khobragade N, Graupe D, Tuninetti D. Towards fully automated closed-loop Deep Brain Stimulation in Parkinson's disease patients: A LAMSTAR-based tremor predictor. Conf proc IEEE Eng Med Biol Soc. 2015;2015:2616-9.
5. Fox SH, Katzenschlager R, Lim SY, Barton B, De Bie RM, Seppi K, et al. International Parkinson and movement disorder society evidence-based medicine review: update on treatments for the motor symptoms of Parkinson's disease. Mov Disord. 2018;33(8):1248-66.
6. Farrokhi F, Buchlak QD, Sikora M, Esmaili N, Marsans M, McLeod P, et al. Investigating Risk Factors and Predicting Complications in Deep Brain Stimulation Surgery with Machine Learning Algorithms. World Neurosurg. 2020;134:e325-e38.
7. Krause KJ, Phibbs F, Davis T, Fabbri D. Predicting Motor Responsiveness to Deep Brain Stimulation with Machine Learning. AMIA Annu Symp Proc. 2021:651-9.
8. Peralta M, Haegelen C, Jannin P, Baxter JSH. PassFlow: a multimodal workflow for predicting deep brain stimulation outcomes. Int J Comput Assist Radiol Surg. 2021;16(8):1361-70.
9. Yearley AG, Chua M, Horn A, Cosgrove GR, Rolston JD. Deep Brain Stimulation Lead Localization Variability Comparing Intraoperative MRI Versus Postoperative Computed Tomography. Oper Neurosurg. 2023;25(5):441-8.
10. Abbass M, Taha A, Gilmore G, Santyr B, Chalil A, Jog M, et al. The impact of localization and registration accuracy on estimates of deep brain stimulation electrode position in stereotactic space. Imaging Neurosci (Camb). 2025;3.
11. He S, West TO, Plazas FR, Wehmeyer L, Pogosyan A, Deli A, et al. Cortico-thalamic tremor circuits and their associations with deep brain stimulation effects in essential tremor. Brain. 2025;148(6):2093-107.
12. Rački V, Hero M, Papić E, Rožmarić G, Čizmarević NS, Chudy D, et al. Applicability of clinical genetic testing for deep brain stimulation treatment in monogenic Parkinson's disease and monogenic dystonia: a multidisciplinary team perspective. Front Neurosci. 2023;17:1282267.
13. Mahajan A, Butala A, Okun MS, Mari Z, Mills KA. Global Variability in Deep Brain Stimulation Practices for Parkinson's Disease. Front Hum Neurosci. 2021;15:667035.
14. Ferrea E, Negahbani F, Cebi I, Weiss D, Gharabaghi A. Machine learning explains response variability of deep brain stimulation on Parkinson's disease quality of life. npj Digital Medicine. 2024;7(1):269.
15. Koivu M, Sihvonen AJ, Eerola-Rautio J, Pauls KAM, Resendiz-Nieves J, Vartiainen N, et al. Clinical and Brain Morphometry Predictors of Deep Brain Stimulation Outcome in Parkinson's Disease. Brain Topography. 2024;37(6):1186-94.
16. Wolke R, Becktepe JS, Paschen S, Helmers AK, Kübler-Weller D, Youn J, et al. The Role of Levodopa Challenge in Predicting the Outcome of Subthalamic Deep Brain Stimulation. Mov Disord Clin Pract. 2023;10(8):1181-91.
17. Habets JGV, Janssen MLF, Duits AA, Sijben LCJ, Mulders AEP, De Greef B, et al. Machine learning prediction of motor response after deep brain stimulation in Parkinson's disease-proof of principle in a retrospective cohort. PeerJ. 2020;8:e10317.
18. Park KH, Sun S, Lim YH, Park HR, Lee JM, Park K, et al. Clinical outcome prediction from analysis of microelectrode recordings using deep learning in subthalamic deep brain

[DOI: 10.47176/mjiri.39.164]

stimulation for Parkinson`s disease. PLoS One. 2021;16(1):e0244133.
19. Kostoglou K, Michmizos KP, Stathis P, Sakas D, Nikita KS, Mitsis GD. Classification and Prediction of Clinical Improvement in Deep Brain Stimulation From Intraoperative Microelectrode Recordings. IEEE Trans Biomed Eng. 2017;64(5):1123-30.
20. Angeles P, Tai Y, Pavese N, Vaidyanathan R. Assessing Parkinson's disease motor symptoms using supervised learning algorithms. Mov Disord. 2017;32(Suppl 2).
21. Yohanandan S, Peppard R, Tan J, McDermott H, Perera T. Evaluating machine learning algorithms estimating tremor severity ratings on the Bain–Findley scale. Measurement Science and Technology. 2016;27:125702.
22. Buchlak QD, Esmaili N, Leveque JC, Farrokhi F, Bennett C, Piccardi M, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. Neurosurg Rev. 2020;43(5):1235-53.
23. Alshammari T, Alseraye S, Alqasim R, Rogowska A, Alrasheed N, Alshammari M. Examining anxiety and stress regarding virtual learning in colleges of health sciences: A cross-sectional study in the era of the COVID-19 pandemic in Saudi Arabia. Saudi Pharm J. 2022;30(3):256-64.
24. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
25. Suppa A, Asci F, Costantini G, Bove F, Piano C, Pistoia F, et al. Effects of deep brain stimulation of the subthalamic nucleus on patients with Parkinson's disease: a machine-learning voice analysis. Front Neurol. 2023;14:1267360.
26. Watts J, Niethammer M, Khojandi A, Ramdhani R. Machine learning model comparison for freezing of gait prediction in advanced Parkinson's disease. Front Aging Neurosci. 2024;16:1431280.
27. Sabo A, Iaboni A, Taati B, Fasano A, Gorodetsky C. Evaluating the ability of a predictive vision-based machine learning model to measure changes in gait in response to medication and DBS within individuals with Parkinson's disease. Biomed Eng Online. 2023;22(1):120.
28. Alhourani A, Wylie SA, Summers JE, Phibbs FT, Bradley EB, Neimat JS, et al. Developing Predictor Models of Postoperative Verbal Fluency After Deep Brain Stimulation Using Preoperative Neuropsychological Assessment. Neurosurgery. 2022;91(2):256-62.
29. Chang B, Ni C, Zhang W, Mei J, Xiong C, Chen P, et al. Nomogram to Predict Cognitive State Improvement after Deep Brain Stimulation for Parkinson's Disease. Brain Sciences. 2022;12(6):759.
30. Huang T, Li M, Huang J. Recent trends in wearable device used to detect freezing of gait and falls in people with Parkinson's disease: A systematic review. Front Aging Neurosci. 2023;15:1119956.
31. Sun M, Yan T, Liu R, Zhao X, Zhou X, Ma Y, et al. Predictive value of machine learning in diagnosing cognitive impairment in patients with Parkinson's disease: a systematic review and meta-analysis. Ann Palliat Med. 2022;11(12):3775-84.
32. Oliveira AM, Coelho L, Carvalho E, Ferreira-Pinto MJ, Vaz R, Aguiar P. Machine learning for adaptive deep brain stimulation in Parkinson's disease: closing the loop. J Neurol. 2023;270(11):5313-26.
33. Watts J, Khojandi A, Shylo O, Ramdhani RA. Machine Learning's Application in Deep Brain Stimulation for Parkinson's Disease: A Review. Brain Sci. 2020;10(11):809.
34. Zhu Y, Wang R, Wang Y, Ge M, Shen B, Sun Y, et al. A PointNet++-Based Deep Learning Approach Using Patient-Specific 3D Point Clouds for Personalized DBS Efficacy Prediction in Parkinson's Disease2025. Research Square. 2025 Sep 21.
35. Boutet A, Madhavan R, Elias GJB, Joel SE, Gramer R, Ranjan M, et al. Predicting optimal deep brain stimulation parameters for Parkinson's disease using functional MRI and machine learning. Nat Commun. 2021;12(1):3043.
36. Kleiner-Fisman G, Herzog J, Fisman DN, Tamma F, Lyons KE, Pahwa R, et al. Subthalamic nucleus deep brain stimulation: summary and meta-analysis of outcomes. Mov Disord. 2006;21 Suppl 14:S290-304.
37. Frizon LA, Hogue O, Achey R, Floden DP, Nagel S, Machado AG, et al. Quality of Life Improvement Following Deep Brain Stimulation for Parkinson Disease: Development of a Prognostic Model. Neurosurgery. 2019;85(3):343-9.
38. Schuepbach WMM, Tonder L, Schnitzler A, Krack P, Rau J, Hartmann A, et al. Quality of life predicts outcome of deep brain stimulation in early Parkinson disease. Neurology. 2019;92(10):e1109-e20.
39. Zaidel A, Bergman H, Ritov Y, Israel Z. Levodopa and subthalamic deep brain stimulation responses are not congruent. Mov Disord. 2010;25(14):2379-86.
40. Park SC, Cha JH, Lee S, Jang W, Lee CS, Lee JK. Deep Learning-Based Deep Brain Stimulation Targeting and Clinical Applications. Front Neurosci. 2019;13:1128.
41. Houston B, Thompson M, Ko A, Chizeck H. A machine-learning approach to volitional control of a closed-loop deep brain stimulation system. J Neural Eng. 2019;16(1):016004.
42. Ma A, Lau KK, Thyagarajan D. Voice changes in Parkinson's disease: What are they telling us? J Clin Neurosci. 2020;72:1-7.
43. Suppa A, Costantini G, Asci F, Di Leo P, Al-Wardat MS, Di Lazzaro G, et al. Voice in Parkinson's Disease: A Machine Learning Study. Front Neurol. 2022;13:831428.
44. Baudouin R, Lechien JR, Carpentier L, Gurruchaga JM, Lisan Q, Hans S. Deep Brain Stimulation Impact on Voice and Speech Quality in Parkinson's Disease: A Systematic Review. Otolaryngol Head Neck Surg. 2023;168(3):307-18.
45. Fabbri M, Natale F, Artusi CA, Romagnolo A, Bozzali M, Giulietti G, et al. Deep brain stimulation fine-tuning in Parkinson's disease: Short pulse width effect on speech. Parkinsonism Relat Disord. 2021;87:130-4.
46. Limousin P, Foltynie T. Long-term outcomes of deep brain stimulation in Parkinson disease. Nat Rev Neurol. 2019;15(4):234-42.
47. Tripoliti E, Limousin P, Foltynie T, Candelario J, Aviles-Olmos I, Hariz MI, et al. Predictive factors of speech intelligibility following subthalamic nucleus stimulation in consecutive patients with Parkinson's disease. Mov Disord. 2014;29(4):532-8.
48. Lange F, Eldebakey H, Hilgenberg A, Weigl B, Eckert M, DeSunda A, et al. Distinct phenotypes of stimulation-induced dysarthria represent different cortical networks in STN-DBS. Parkinsonism Relat Disord. 2023;109:105347.
49. Tanaka Y, Tsuboi T, Watanabe H, Nakatsubo D, Maesawa S, Kato S, et al. Longitudinal Speech Change After Subthalamic Nucleus Deep Brain Stimulation in Parkinson's Disease Patients: A 2-Year Prospective Study. J Parkinsons Dis. 2020;10(1):131-40.
50. Saggio G, Costantini G. Worldwide Healthy Adult Voice Baseline Parameters: A Comprehensive Review. Journal of Voice. 2022;36(5):637-49.
51. Shao Z, Janse E, Visser K, Meyer AS. What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. Front Psychol. 2014;5:772.
52. Henry JD, Crawford JR. Verbal fluency deficits in Parkinson's disease: a meta-analysis. J Int Neuropsychol Soc. 2004;10(4):608-22.
53. Asci F, Costantini G, Saggio G, Suppa A. Fostering Voice Objective Analysis in Patients with Movement Disorders. Mov Disord. 2021;36(4):1041.
54. Hirshorn EA, Thompson-Schill SL. Role of the left inferior frontal gyrus in covert word retrieval: neural correlates of switching during verbal fluency. Neuropsychologia. 2006;44(12):2547-57.
55. Dubois B, Pillon B. Cognitive deficits in Parkinson's disease. J Neurol. 1997;244(1):2-8.
56. McDade EM, Boot BP, Christianson TJ, Pankratz VS, Boeve BF, Ferman TJ, et al. Subtle gait changes in patients with REM sleep behavior disorder. Mov Disord. 2013;28(13):1847-53.
57. Mirelman A, Gurevich T, Giladi N, Bar-Shira A, Orr-Urtreger A, Hausdorff JM. Gait alterations in healthy carriers of the LRRK2 G2019S mutation. Ann Neurol. 2011;69(1):193-7.

[DOI: 10.47176/mjiri.39.164]

58. Galna B, Lord S, Burn DJ, Rochester L. Progression of gait dysfunction in incident Parkinson's disease: impact of medication and phenotype. Mov Disord. 2015;30(3):359-67.
59. Hausdorff JM. Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. Chaos. 2009;19(2):026113.
60. Banks SJ, Bayram E, Shan G, LaBelle DR, Bluett B. Non-motor predictors of freezing of gait in Parkinson's disease. Gait Posture. 2019;68:311-6.
61. D'Cruz N, Vervoort G, Fieuws S, Moreau C, Vandenberghe W, Nieuwboer A. Repetitive Motor Control Deficits Most Consistent Predictors of Conversion to Freezing of Gait in Parkinson's Disease: A Prospective Cohort Study. J Parkinsons Dis. 2020;10(2):559-71.
62. Pardoel S, Kofman J, Nantel J, Lemaire ED. Wearable-Sensor-based Detection and Prediction of Freezing of Gait in Parkinson's Disease: A Review. Sensors (Basel). 2019;19(23):5141.
63. Leimbach F, Atkinson-Clement C, Socorro P, Jahanshahi M. The Effects of Subthalamic Nucleus Deep Brain Stimulation in Parkinson's Disease on Associative Learning of Verbal and Non-Verbal Information by Trial and Error or with Corrective Feedback. J Parkinsons Dis. 2022;12(3):885-96.
64. Wei X, Shen Q, Litvan I, Huang M, Lee RR, Harrington DL. Internetwork Connectivity Predicts Cognitive Decline in Parkinson's and Is Altered by Genetic Variants. Front Aging Neurosci. 2022;14:853029.
65. Aarsland D, Bronnick K, Williams-Gray C, Weintraub D, Marder K, Kulisevsky J, et al. Mild cognitive impairment in Parkinson disease: a multicenter pooled analysis. Neurology. 2010;75(12):1062-9.
66. Ding W, Ding LJ, Li FF, Han Y, Mu L. Neurodegeneration and cognition in Parkinson's disease: a review. Eur Rev Med Pharmacol Sci. 2015;19(12):2275-81.
67. David FJ, Munoz MJ, Corcos DM. The effect of STN DBS on modulating brain oscillations: consequences for motor and cognitive behavior. Exp Brain Res. 2020;238(7-8):1659-76.
68. Xie Y, Meng X, Xiao J, Zhang J, Zhang J. Cognitive Changes following Bilateral Deep Brain Stimulation of Subthalamic Nucleus in Parkinson's Disease: A Meta-Analysis. Biomed Res Int. 2016;2016:3596415.

[DOI: 10.47176/mjiri.39.164]