

# Preoperative Electroencephalography-Based Machine Learning Predicts Cognitive Deterioration After Subthalamic Deep Brain Stimulation

Victor J. Geraedts, MD, PhD,<sup>1,2\*</sup> Milan Koch, PhD,<sup>3</sup> Roy Kuiper, MSc,<sup>1,4</sup> Marios Kefalas, MSc,<sup>3</sup> Thomas H.W. Bäck, PhD,<sup>3</sup> Jacobus J. van Hilten, MD, PhD,<sup>1</sup> Hao Wang, PhD,<sup>3</sup> Huub A.M. Middelkoop, PhD,<sup>1,5</sup> Niels A. van der Gaag, MD, PhD,<sup>6,7</sup> Maria Fiorella Contarino, MD, PhD,<sup>1,4</sup> and Martijn R. Tannemaat, MD, PhD<sup>1</sup>

<sup>1</sup>Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands

<sup>2</sup>Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

<sup>3</sup>Leiden Institute of Advanced Computer Science, Leiden, The Netherlands

<sup>4</sup>Department of Neurology, Haga Teaching Hospital, Den Haag, The Netherlands

<sup>5</sup>Neuropsychology Unit, Leiden University Institute of Psychology, Leiden, The Netherlands

<sup>6</sup>Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands

<sup>7</sup>Department of Neurosurgery, Haga Teaching Hospital, Den Haag, The Netherlands

**ABSTRACT: Background:** Subthalamic deep brain stimulation (STN DBS) may relieve refractory motor complications in Parkinson's disease (PD) patients. Despite careful screening, it remains difficult to determine severity of alpha-synucleinopathy involvement which influences the risk of postoperative complications including cognitive deterioration. Quantitative electroencephalography (qEEG) reflects cognitive dysfunction in PD and may provide biomarkers of postoperative cognitive decline.

**Objective:** To develop an automated machine learning model based on preoperative EEG data to predict cognitive deterioration 1 year after STN DBS.

**Methods:** Sixty DBS candidates were included; 42 patients had available preoperative EEGs to compute a fully automated machine learning model. Movement Disorder Society criteria classified patients as cognitively stable or deteriorated at 1-year follow-up. A total of 16,674 EEG-features were extracted per patient; a Boruta algorithm selected EEG-features to reflect representative neurophysiological signatures for each class. A random forest classifier with 10-fold cross-validation with Bayesian optimization provided class-differentiation.

**Results:** Twenty-five patients were classified as cognitively stable and 17 patients demonstrated cognitive decline. The model differentiated classes with a mean (SD) accuracy of 0.88 (0.05), with a positive predictive value of 91.4% (95% CI 82.9, 95.9) and negative predictive value of 85.0% (95% CI 81.9, 91.4). Predicted probabilities between classes were highly differential (hazard ratio 11.14 [95% CI 7.25, 17.12]); the risk of cognitive decline in patients with high probabilities of being prognosticated as cognitively stable (>0.5) was very limited.

**Conclusions:** Preoperative EEGs can predict cognitive deterioration after STN DBS with high accuracy. Cortical neurophysiological alterations may indicate future cognitive decline and can be used as biomarkers during the DBS screening. © 2021 The Authors. *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society

**Key Words:** machine learning; quantitative EEG; Parkinson's disease; deep brain stimulation; cognition; prediction; subthalamic nucleus

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

\*Correspondence to: Dr. Victor J. Geraedts, Department of Neurology, Leiden University Medical Center, Albinusdreef 2 room K5-97, 2333ZA Leiden, The Netherlands; E-mail: v.j.geraedts@lumc.nl

Relevant conflicts of interest/financial disclosures: None.

**Funding agencies:** This work was supported by a grant from the "Stichting ParkinsonFonds" and the "Stichting Alkemade-Keuls".

**Received:** 9 March 2021; **Revised:** 12 April 2021; **Accepted:** 5 May 2021

Published online 3 June 2021 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.28661

Patients with Parkinson's disease (PD) may be eligible for subthalamic deep brain stimulation (STN DBS) in cases of debilitating motor complications refractory to oral therapy. STN DBS has similar, or even better, effects on motor function compared to oral therapy, reduces motor complications, and improves quality of life (QoL).<sup>1-4</sup> However, deterioration in one or more cognitive domains may be observed after STN DBS: 34 of 59 studies on STN DBS reported deterioration in  $\geq 1$  cognitive domain, and 12 of 20 studies reported cognitive deterioration in  $\geq 1$  cognitive domain in controlled settings.<sup>5</sup> Careful and accurate assessment of cognitive functioning is a crucial element during the DBS eligibility screening,<sup>6,7</sup> but mild cognitive deterioration can occur even after rigorous patient selection.<sup>8</sup> It remains difficult to determine the severity of the impact of alpha-synucleinopathy on the central nervous system, which may in turn influence the long-term impact of DBS in terms of postoperative complications, including worsening of specific cognitive functions. Biomarkers based on quantitative electroencephalography (qEEG) have been previously described to reflect cognitive functioning in both the general PD population<sup>9</sup> and DBS candidates specifically.<sup>10-14</sup> The utility of qEEG in predicting future cognitive decline has been previously described in the general PD population in studies with 3-5 years' follow-up.<sup>9</sup> Given that cognitive deterioration may already be manifest within 1 year post-surgery,<sup>5,8</sup> we hypothesize that qEEG may have potential utility during the DBS screening to prognosticate cognitive decline after DBS.

In previous studies<sup>14,15</sup> we demonstrated the utility of a fully automated machine learning algorithm to classify patients according to their cognitive status through EEGs. These models apply a series of automatic processes to sequentially extract a large number of EEG-features, select EEG-features to create a comprehensive EEG-signature of either class, and subsequently build and optimize a classification model without requiring arbitrary decisions by the researcher.<sup>14</sup> This process is considered highly efficient and reduces the need for prior knowledge of existing EEG-features, thereby increasing the likelihood of identifying previously unknown biomarkers of cognitive decline.

The aim of this study was to develop an EEG-based fully-automated machine learning model to evaluate the utility of EEG during the DBS screening to prognosticate cognitive deterioration according to Movement Disorder Society (MDS) criteria 1 year after STN DBS.<sup>16</sup>

## Methods

Consecutive patients who underwent STN DBS at the DBS Center of the Leiden University Medical Center (LUMC) and Haga Teaching Hospital between May

2017 and July 2019 were included, as part of the OPTIMIST trial (OPTIMizing patient selection for deep brain STimulation of the subthalamic nucleus) (Netherlands Trial Register NL6079). Sample size calculation was based on a different research question. All patients fulfilled MDS PD criteria.<sup>17</sup> Written informed consent was obtained from all participants. The study was approved by the local medical ethics committee of the LUMC. Anonymized data may be shared upon request. The Standards for Reporting Diagnostic Accuracy (STARD) guidelines were followed during the writing process.

## Procedures and Inclusion

All patients received standard questionnaires, a neuropsychological assessment, and routine EEG as part of the DBS screening procedure.<sup>6</sup> After acceptance for STN DBS, patients were invited for participation in the study (1 patient declined during the inclusion period). Inclusion criteria included age  $>18$  years, diagnosis of idiopathic PD according to established criteria, clinical indication for STN DBS, written informed consent, ability to comply with the study assessments, and ability to read or understand Dutch. Exclusion criteria included Hoehn & Yahr stage 5, Mattis Dementia Rating Scale (MDRS) scores  $<120$ , psychiatric contraindications for STN DBS, and general contraindications for stereotactic surgery. Surgery took place approximately 1-2 months after the DBS eligibility screening. Surgical procedures have been published elsewhere.<sup>18</sup> Lead implantation was performed during an awake procedure, withdrawn from sedatives and dopaminergic medication. On average, 2-3 cannulas and microelectrodes were inserted simultaneously in a Ben-Gun array. Individual adjustments were made to avoid ventricles, blood vessels, and sulci. All procedures were performed bilaterally and during the same session. Permanent leads were usually positioned with the middle two contacts located at the site of best therapeutic effect. Follow-up visits were carried out 1 year after surgery ( $\pm 6$  weeks) and included similar assessment of cognition and questionnaires. EEG was only performed preoperatively.

## EEG Acquisition

EEG acquisition and preprocessing procedure has been detailed previously.<sup>10</sup> EEGs were recorded with patients in a supine position, eyes closed, and in a state of relaxed wakefulness. Twenty-one Ag/AgCl EEG electrodes were positioned according to a standard 10-20 setup. Medication was continued according to individual schedules (ie, "ON"); dyskinesias were not observed. All data were re-referenced towards a source-derivation approaching the surface Laplacian derivation to amplify spatial resolution.<sup>19,20</sup> Five consecutive artifact-free non-overlapping epochs (4096 points, 8.192 seconds, sampling rate 500 Hz) were manually

selected for offline analysis; patients with fewer artifact-free epochs were excluded. Brainwave software was used to calculate global peak frequencies and spectral slowing ratios (relative  $[\delta + \theta]/[\alpha + \beta]$  band-power) (BrainWave version 0.9.152.12.26, C. J. Stam; available at <https://home.kpn.nl/stam7883/brainwave.html>).

### Used Scales and Determination of Outcome

From the neuropsychological assessment, six tests were selected to reflect cognitive functioning: Rey Auditory Verbal Learning Test (RAVLT),<sup>21</sup> Verbal Fluency,<sup>22</sup> Trail Making Test (TMT) B corrected for A,<sup>23</sup> Stroop Color and Word Test (Stroop) section 3 - interference,<sup>24</sup> and Digit Cancellation Test (DCT) sections “correct” and “wrong + corrected”.<sup>25</sup> All six scores provided a Test-score (T-score) with normally distributed population means of 50 and standard deviations (SD) of 10. A global cognitive composite score (Cog-score) was calculated by averaging scores from all six tests; a maximum of two missing subscores was allowed. Cognitive deterioration was based on MDS level I criteria, that is, presence of significant cognitive decline in  $\geq 2$  tests (ie,  $\geq 10$  points deterioration).<sup>16</sup>

Secondary outcomes included motor function (Movement Disorder Society-Unified Parkinson’s Disease Rating Scale [MDS-UPDRS] III),<sup>26</sup> motor fluctuations (MDS-UPDRS IV),<sup>26</sup> Levodopa Equivalent Dose (LED),<sup>27</sup> DBS impairment scale (DBS IS),<sup>28</sup> (SENS-PD),<sup>29</sup> Parkinson’s Disease Quality of Life scale 39 (PDQ39),<sup>30</sup> MDRS,<sup>31</sup> Montreal Cognitive Assessment (MoCA),<sup>32</sup> psychiatric symptoms (depression: Becks Depression Inventory [BDI]),<sup>33</sup> Parkinson Anxiety Scale (PAS),<sup>34</sup> Apathy Evaluation Scale (AES),<sup>35</sup> impulse-control disorder (Quip-RS and Quip-RS-ICD),<sup>36</sup> autonomic symptoms (SCOPA-AUT),<sup>37</sup> nighttime sleeping problems (SCOPA-SLEEP),<sup>37</sup> and excessive daytime sleepiness (EDS).<sup>38</sup>

Motor outcomes after STN DBS were assessed using a Stimulation Challenge Test (SCT) as published previously,<sup>39</sup> in the conditions “Med-OFF-Stim-ON”, “Med-OFF-Stim-OFF”, and “Med-ON-Stim-ON”. Baseline levodopa-response was determined through a Levodopa Challenge Test (LCT) using a 120% suprathreshold dosage of the early morning LED.<sup>40</sup>

### Machine Learning Classification Algorithm

The machine learning algorithm has been previously reported in relation to EEG research.<sup>14,15</sup> The machine learning pipeline consists of four fully automated phases: feature-extraction, feature-selection, classifier-training, and hyperparameter-optimization. EEG-features are derived from the “Time Series Feature Extraction on basis of Scalable Hypothesis tests” (tsfresh) library,<sup>41,42</sup> resulting in 16,674 features per patient (794 features per EEG-channel). Features from

all five epochs were averaged per patient. A Boruta selection algorithm was used for feature-selection, which tests the variable importance (VIMP) against fictive features created by random shuffling of the extracted features. Multiple independent trials test whether the VIMP of the real features are consistently higher than the VIMP of the fictive features during a random forest classifier (RFC).<sup>43</sup> After saturation of the feature-selection (ie, only “real features” are retained), a RFC was trained using the resulting feature set; a majority vote from all individual decision trees was used for binary classification.<sup>44</sup> Hyperparameter-optimization was performed using a Bayesian optimization method (Mixed Integer Parallel Efficient Global Optimization).<sup>45,46</sup> The robustness of the resulting model was validated using 10-fold cross-validation; the resulting T-scores were averaged to obtain cross-validated metrics. Split-sample validation was considered, but demonstrated to be inferior to our cross-validation approach.<sup>14</sup> Each model was developed five times sequentially to evaluate its robustness. All models were built, tested, and validated without manual interference in relation to feature-extraction or feature-selection.

### Statistical Analyses

Differences between baseline and follow-up were compared using paired *t* tests or Wilcoxon signed-rank tests and Pearson  $\chi^2$  tests, if appropriate. *P* values were reported using Monte Carlo estimation in the case of Wilcoxon signed-rank tests, and Fisher’s exact test in cases of dichotomous  $\chi^2$  tests. Continuous demographic and clinical data, including change indices (ie, follow-up – baseline values), was compared among the cognitive groups in a similar fashion.

The machine learning classification algorithm was evaluated using sensitivity, specificity, accuracy, and F1 scores; receiver–operator characteristic (ROC) curves were used for visualization. Kaplan–Meier curves were used to demonstrate the differentiation in predicted probabilities between the groups, in which the predicted probability of being classified as cognitively stable was used as “time” and the label as “event”. The y-axis is the inverse likelihood of being classified (100% - % classified), separate lines show at what levels of the predicted probability of being classified as cognitively stable patients were indeed classified (ie, green: stable cognition; red: cognitive deterioration). The assumption of proportional hazards was confirmed through visual inspection of the log minus log plots. Curves were compared using Cox regression analyses.

All analyses were performed using IBM Statistical Package for the Social Sciences (SPSS) 25 Software and Python software.

## Results

Seventy-three patients were originally included in this study. Two patients had missing baseline Cog-scores, and 11 patients were lost-to-follow-up (reasons for lost-to-follow-up were not systematically documented): 60 patients were included for analyses, of which 42 had usable preoperative EEG-recordings; 18 patients had EEGs with insufficient artifact-free epochs (see Fig. S1 for the STARD flowchart). Mean age at baseline was 60.3 (8.3) years, with 10.0 (5.7)

years disease duration, 33% were female ( $n = 20$ ) (see Table 1 for a full overview). Mean duration between screening and follow-up was 14.5 (2.0) months. Cognitive performance was significantly reduced after STN DBS, mean Cog-score reduction  $-3.6$  (4.3) points (FU-baseline) ( $P < 0.001$ ). Patients without available EEG-recordings did not differ significantly from patients with EEG-recordings for any demographic or clinical characteristic.

Based on a cut-off of significant deterioration of  $\geq 2$  tests, 25/60 patients were classified as having clinically established

**TABLE 1** Cohort description

Clinical characteristic	Baseline	Follow-up	P value
Age at baseline, y <sup>a</sup>	60.0 (55.0, 67.8)		
% female sex (n)	33 (20)		
Disease duration, y <sup>a</sup>	8.8 (6.6, 11.7)		
% opt for DBS again (n)		81 (46)	
MDS-UPDRS III (Med-OFF [Stim-OFF])	43.3 (11.1)	50.6 (13.6)	0.001
MDS-UPDRS III (Med-ON [Stim-ON])	19.0 (8.4)	17.2 (9.9)	0.140
Therapy-responsiveness <sup>b c</sup>	56.4 (15.1)	66.0 (15.2)	<0.001
MDS-UPDRS IV <sup>a</sup>	9.0 (6.3, 11.0)	1.0 (0.0, 5.0)	<0.001
LEDD <sup>b</sup>	1129 (482)	435 (273)	<0.001
LEDD-DA <sup>a</sup>	80 (0, 248)	30 (0, 120)	0.036
Cog-score	49.2 (5.4)	45.5 (6.2)	<0.001
MDRS <sup>a</sup>	140.0 (137.0, 142.0)	140.0 (138.0, 142.8)	0.658
MoCA <sup>b</sup>	26.4 (2.1)	25.7 (2.5)	0.025
BDI <sup>b</sup>	12.2 (7.0)	10.1 (6.3)	0.005
PAS <sup>a</sup>	10.0 (6.0, 18.0)	8.0 (3.0, 14.0)	0.017
Apathy Scale <sup>b</sup>	10.5 (3.0)	12.0 (3.8)	0.014
Quip-RS <sup>a</sup>	8.0 (2.0, 18.8)	4.0 (0.0, 13.0)	0.007
SENS-PD <sup>a</sup>	11.0 (7.3, 14.0)	11.0 (6.8, 15.0)	0.660
DBS IS <sup>b</sup>	17.3 (9.9)	18.6 (12.8)	0.568
Scopa-AUT <sup>b</sup>	16.2 (7.1)	13.9 (7.2)	0.004
Scopa-Sleep <sup>b</sup>	6.0 (3.0)	3.7 (3.4)	<0.001
Scopa-EDS <sup>b</sup>	4.2 (3.3)	3.4 (2.7)	0.021
PDQ39 <sup>b</sup>	47.4 (21.8)	36.1 (23.4)	<0.001
EQ5D <sup>a</sup>	11.0 (9.0, 12.0)	8.0 (7.0, 10.0)	<0.001

<sup>a</sup>Median (interquartile range), Wilcoxon signed-rank tests.

<sup>b</sup>Mean (standard deviation), paired  $t$  tests.

<sup>c</sup>Therapy-responsiveness: percentage (%) response to dopaminergic therapy (baseline) or oral dopaminergic therapy + stimulation (follow-up) ([OFF-ON]/[OFF]).

Abbreviations: DBS, deep brain stimulation; MDS-UPDRS, Movement Disorder Society Unified Parkinson's Disease Rating Scale; LEDD, Levodopa Equivalent Dose (DA, dopamine agonists); Cog-score, global composite cognitive scores; MDRS, Mattis Dementia Rating Scale; MoCA, Montreal Cognitive Assessment; BDI, Becks Depression Inventory; PAS, Parkinson's Disease Anxiety Scale; SENS-PD, Severity of Predominantly Non-Dopaminergic Symptoms in Parkinson's Disease; DBS IS, Deep Brain Stimulation Impairment Scale; Scopa, Scales for Outcomes of Parkinson's Disease; AUT, autonomic symptoms; EDS, Excessive Daytime Sleepiness; PDQ39, Parkinson's Disease Questionnaire 39; EQ5D, EuroQoL 5 Dimensions.

**TABLE 2** Clinical differences between the cognitive classes

Clinical characteristic		Stable cognition	Cognitive deterioration	P value
Age at baseline, y <sup>a</sup>		58.5 (53.5, 68.0)	62.0 (55.5, 67.0)	0.470
% female sex (n)		41 (14)	20 (5)	0.100
Disease duration, y <sup>a</sup>		8.1 (6.3, 11.4)	9.6 (6.6, 12.3)	0.342
% opt for DBS again (n)		100 (30)	67 (16)	0.030
MDS-UPDRS III OFF <sup>b</sup>	Baseline	41.5 (11.9)	45.7 (10.0)	0.155
	Follow-up	47.6 (13.1)	54.3 (13.7)	0.068
	Change <sup>c</sup>	5.8 (14.9)	8.3 (14.8)	0.535
MDS-UPDRS III ON <sup>b</sup>	Baseline	18.4 (8.7)	19.7 (8.2)	0.565
	Follow-up	15.6 (9.5)	19.0 (10.4)	0.215
	Change <sup>c</sup>	-3.2 (9.0)	-0.9 (12.7)	0.456
Therapy-responsiveness <sup>b d e</sup>	Baseline	56.1 (16.0)	57.0 (14.3)	0.816
	Follow-up	66.9 (14.8)	64.9 (16.3)	0.645
	Change <sup>c</sup>	0.1 (0.2)	0.1 (0.2)	0.611
MDS-UPDRS IV <sup>a</sup>	Baseline	9.0 (5.0, 12.0)	9.0 (7.0, 11.0)	0.787
	Follow-up	2.0 (0.0, 5.0)	0.0 (0.0, 6.0)	0.853
	Change <sup>c</sup>	-5.1 (6.6)	-6.4 (4.6)	0.503
LEDD <sup>d</sup>	Baseline	1009 (411)	1276 (535)	0.039
	Follow-up	391 (266)	500 (276)	0.140
	Change <sup>c</sup>	-596 (437)	-798 (535)	0.136
LEDD-DA <sup>a</sup>	Baseline	80 (0, 240)	60 (0, 240)	0.789
	Follow-up	58 (0, 120)	0 (0, 160)	0.327
	Change <sup>c</sup>	5.0 (305)	-66.0 (151.5)	0.347
Cog-score <sup>d</sup>	Baseline	49.3 (5.7)	49.3 (5.2)	0.994
	Follow-up	47.3 (6.0)	43.0 (5.7)	0.008
	Change <sup>c</sup>	-2.0 (3.9)	-6.3 (4.6)	<0.001
RAVLT <sup>d</sup>	Baseline	41.9 (10.5)	44.4 (10.6)	0.381
	Follow-up	41.8 (10.7)	38.0 (11.5)	0.189
	Change <sup>c</sup>	-0.1 (8.9)	-6.4 (10.0)	0.013
Verbal Fluency <sup>d</sup>	Baseline	59.3 (13.5)	60.6 (13.1)	0.703
	Follow-up	49.2 (11.5)	41.6 (10.1)	0.011
	Change <sup>c</sup>	-10.1 (10.5)	-19.5 (14.1)	0.005
TMT B corrected for A <sup>d</sup>	Baseline	46.0 (12.6)	41.2 (12.6)	0.169
	Follow-up	45.1 (11.6)	43.9 (9.1)	0.708
	Change <sup>c</sup>	-2.5 (14.8)	1.3 (11.9)	0.349
Stroop 3 – interference <sup>d</sup>	Baseline	52.1 (7.6)	53.0 (8.3)	0.676
	Follow-up	53.2 (8.0)	48.4 (9.5)	0.040
	Change <sup>c</sup>	1.1 (5.6)	-4.9 (10.7)	0.016
DCT – total correct <sup>d</sup>	Baseline	45.7 (9.4)	43.3 (9.1)	0.342

(Continues)

TABLE 2 Continued

Clinical characteristic		Stable cognition	Cognitive deterioration	P value
	<i>Follow-up</i>	44.1 (10.8)	37.7 (9.1)	0.021
	<i>Change<sup>c</sup></i>	−0.9 (6.8)	−4.4 (6.2)	0.058
DCT – total wrong and corrected <sup>d</sup>	<i>Baseline</i>	50.9 (9.2)	52.6 (8.4)	0.484
	<i>Follow-up</i>	50.6 (10.4)	49.4 (9.1)	0.648
	<i>Change<sup>c</sup></i>	−0.3 (8.6)	−3.2 (8.1)	0.218
MDRS <sup>a</sup>	<i>Baseline</i>	140.0 (138.5, 142.5)	139.0 (13.6, 141.8)	0.294
	<i>Follow-up</i>	140.4 (138.0, 143.0)	140.0 (136, 142.0)	0.548
	<i>Change<sup>c</sup></i>	−0.2 (3.0)	0.7 (3.7)	0.323
MoCA <sup>d</sup>	<i>Baseline</i>	26.6 (2.3)	26.2 (1.7)	0.442
	<i>Follow-up</i>	25.9 (2.7)	25.6 (2.3)	0.655
	<i>Change<sup>c</sup></i>	−0.7 (2.3)	−0.6 (2.1)	0.762
BDI <sup>d</sup>	<i>Baseline</i>	12.8 (7.2)	11.4 (6.9)	0.452
	<i>Follow-up</i>	9.7 (5.5)	10.2 (7.0)	0.730
	<i>Change<sup>c</sup></i>	−3.1 (6.3)	−1.1 (4.2)	0.181
PAS <sup>a</sup>	<i>Baseline</i>	11.0 (6.0, 18.0)	10.0 (8.0, 17.5)	0.730
	<i>Follow-up</i>	5.5 (3.0, 15.5)	10.0 (4.5, 14.0)	0.549
	<i>Change<sup>c</sup></i>	−2.6 (7.7)	−1.3 (6.1)	0.510
Apathy Scale <sup>d</sup>	<i>Baseline</i>	10.4 (3.4)	10.4 (2.4)	0.948
	<i>Follow-up</i>	12.2 (3.4)	11.9 (4.4)	0.806
	<i>Change<sup>c</sup></i>	1.1 (3.7)	1.5 (3.5)	0.710
Quip-RS <sup>a</sup>	<i>Baseline</i>	9.0 (1.5, 18.8)	7.0 (2.0, 23.0)	0.849
	<i>Follow-up</i>	4.5 (0.0, 12.8)	2.5 (0.0, 13.3)	0.614
	<i>Change<sup>c</sup></i>	−3.8 (8.0)	−4.2 (17.6)	0.922
SENS-PD <sup>a</sup>	<i>Baseline</i>	9.0 (6.8, 13.0)	12.0 (8.5, 15.0)	0.162
	<i>Follow-up</i>	11.0 (6.0, 15.0)	11.0 (7.3, 15.0)	0.666
	<i>Change<sup>c</sup></i>	0.1 (4.2)	−0.4 (3.7)	0.665
DBS IS <sup>d</sup>	<i>Baseline</i>	18.3 (10.2)	15.8 (9.5)	0.350
	<i>Follow-up</i>	18.3 (11.9)	18.8 (14.3)	0.875
	<i>Change<sup>c</sup></i>	−0.8 (10.6)	3.0 (13.0)	0.224
Scopa-AUT <sup>d</sup>	<i>Baseline</i>	14.6 (6.6)	17.9 (7.2)	0.077
	<i>Follow-up</i>	13.3 (7.0)	14.6 (7.7)	0.505
	<i>Change<sup>c</sup></i>	−1.3 (5.7)	−3.2 (5.9)	0.199
Scopa-Sleep <sup>d</sup>	<i>Baseline</i>	6.0 (2.9)	5.9 (2.9)	0.906
	<i>Follow-up</i>	3.7 (2.9)	3.3 (3.2)	0.598
	<i>Change<sup>c</sup></i>	−2.3 (3.2)	−2.6 (4.2)	0.729
Scopa-EDS <sup>d</sup>	<i>Baseline</i>	4.3 (3.4)	3.9 (3.0)	0.639
	<i>Follow-up</i>	3.7 (2.7)	2.8 (2.2)	0.223
	<i>Change<sup>c</sup></i>	−0.7 (2.8)	−1.1 (2.4)	0.563

(Continues)

TABLE 2 Continued

Clinical characteristic		Stable cognition	Cognitive deterioration	P value
PDQ39 <sup>d</sup>	Baseline	47.4 (24.1)	46.9 (18.9)	0.931
	Follow-up	33.1 (20.4)	38.1 (25.4)	0.400
	Change <sup>c</sup>	-14.3 (22.2)	-8.8 (21.4)	0.339
EQ5D <sup>a</sup>	Baseline	10.5 (9.0, 12.0)	11.0 (9.0, 12.0)	0.852
	Follow-up	8.0 (7.0, 10.0)	8.0 (7.0, 10.5)	0.913
	Change <sup>c</sup>	-1.7 (2.6)	-1.5 (2.2)	0.739

<sup>a</sup>Median (interquartile range), Mann–Whitney *U* tests.

<sup>b</sup>Follow-up: MDS-UPDRS III Med-OFF-Stim-OFF or Med-ON-Stim-ON.

<sup>c</sup>Follow-up – baseline (mean [standard deviation], Student *t* tests).

<sup>d</sup>Mean (standard deviation), Student *t* tests.

<sup>e</sup>Therapy-responsiveness: percentage (%) response to dopaminergic therapy (baseline) or oral dopaminergic therapy + stimulation (follow-up) ([OFF-ON]/[OFF]).

Abbreviations: DBS, deep brain stimulation; MDS-UPDRS, Movement Disorder Society Unified Parkinson's Disease Rating Scale; LEDD, Levodopa Equivalent Dose (DA, dopamine agonists); Cog-score, global composite cognitive scores; RAVLT, Rey Auditory Verbal Learning Test; TMT, Trail Making Test; Stroop 3, Stroop Color and Word Test section 3; DCT, Digit Cancellation Test; MDRS, Mattis Dementia Rating Scale; MoCA, Montreal Cognitive Assessment; BDI, Beck's Depression Inventory; PAS, Parkinson's Disease Anxiety Scale; SENS-PD, Severity of Predominantly Non-Dopaminergic Symptoms in Parkinson's Disease; DBS IS, Deep Brain Stimulation Impairment Scale; Scopa, Scales for Outcomes of Parkinson's Disease; AUT, autonomic symptoms; EDS, Excessive Daytime Sleepiness; PDQ39, Parkinson's Disease Questionnaire 39; EQ5D, EuroQoL 5 Dimensions.

cognitive decline (of which 17 had EEG-recordings), and 34 patients were classified as cognitively stable (of which 25 had EEG-recordings) (1 patient was missing due to missing data on two follow-up cognitive tests, such that the criteria for cognitive decline could not be assessed).

### Clinical Differences Between Cognitive Classes

Both cognitive classes were not significantly different in any baseline characteristic, including baseline individual cognitive T-scores (see Table 2). Of note, mean baseline Cog-score was 49.3 points in both groups, relatively similar to the population average of 50.0.

At follow-up, groups differed in terms of Cog-score (47.3 (6.0) points versus 43.0 (5.7) points,  $P = 0.008$ ); patients classified as having cognitive decline at 1 year follow-up had similarly significantly lower T-scores for the Verbal Fluency, Stroop 3, and DCT – total correct tests. Patients with stable cognition were unanimous in choosing DBS again if given the opportunity, however only two-thirds of patients with cognitive deterioration would opt for DBS again ( $P = 0.030$ ), despite no significant difference in QoL. Apart from cognitive performance, the groups did not differ at 1 year follow-up or in change index in any other domain. Both groups did not differ in spectral EEG band-powers (see Table S1).

### Machine Learning Performance

Eighteen EEG-features were selected by the machine learning pipeline during each of the five cross-validation runs. None of the features were retained in each of the individual runs; one feature was selected four times, two features were selected three times, and six features were selected twice (see Fig. S2 for an overview of each of the relative feature-importances; the cerebral location

TABLE 3 Model performance

Dataset	Accuracy	Sensitivity	Specificity	F1
Training-set <sup>a</sup>	0.97 (0.02)	0.99 (0.01)	0.93 (0.05)	0.97 (0.02)
Validation results <sup>a</sup>	0.88 (0.05)	0.95 (0.04)	0.76 (0.07)	0.84 (0.06)

<sup>a</sup>Mean (standard deviation); all metrics were averaged over five separate computation runs.

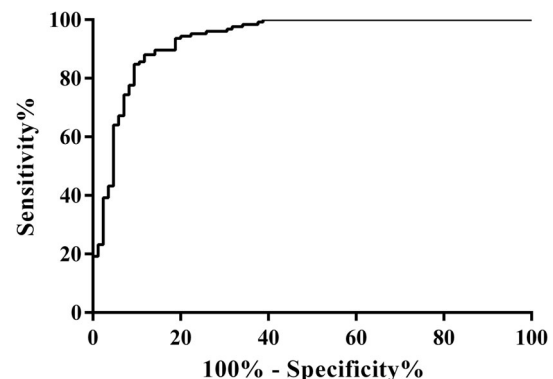
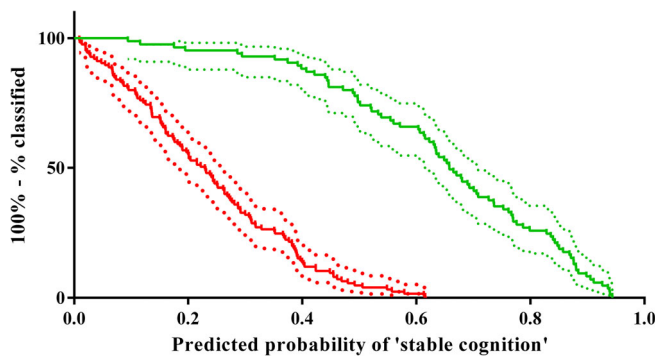


FIG. 1. Receiver–operator characteristic (ROC) curve of all individual cross-validation runs combined. Individual ROC curves of the cross-validation runs separately are shown in Figure S2.

of each feature is shown in Figure S3; see Table S2 for a brief explanation of the mathematical background).

The machine learning algorithm differentiated patients with stable cognition at 1 year follow-up from patients with cognitive deterioration with a mean (SD) accuracy of 0.88 (0.05) in the test-set (see Table 3), with a positive predictive value (PPV) (ie, likelihood of stable cognition 1 year after surgery) of





**FIG. 2.** Kaplan–Meier curve using predicted probability as “time” and labeled class as “event”, combined over all cross-validation runs. The dispersion between the curves demonstrates the differentiation in predicted probabilities between the classes (hazard ratio [95% CI] 11.14 [95% CI 7.25, 17.12]). For high predicted probabilities of being classified as cognitively stable (ie, >0.5), the likelihood of cognitive deterioration appears slim. Kaplan–Meier curves for all cross-validation runs separately are shown in Figure S3. Green: cognitively stable patients; red: patients with cognitive deterioration; dotted line: 95% CI. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

91.4% (95% CI 82.9, 95.9) and a negative predictive value (NPV) (ie, likelihood of cognitive deterioration) of 85.0% (95% CI 81.9%, 91.4%). Performance metrics of all five individual model-runs, as well as overall, are shown in Figure S4. A ROC curve based on the predicted probabilities of all the individual model-runs combined is shown in Fig. 1; ROC curves for the individual runs are shown in Figure S5; the robustness of the cross-validation runs indicates no large effect of overfitting. A Kaplan–Meier curve (Fig. 2) using the pertaining class as “event” and the predicted probability of being classified as cognitively stable as “time” shows the differentiation between the predicted probabilities of the two classes, combined over all cross-validation runs, with a hazard ratio (HR) 11.14 (95% CI 7.25, 17.12) (individual Kaplan–Meier curves for each of the individual cross-validation runs are shown in Fig. S6). For high predicted probabilities of being classified as cognitively stable (ie, >0.5), the likelihood of cognitive deterioration approached zero.

## Discussion

The aim of this study was to develop a fully automated machine learning model based on preoperative EEGs to predict cognitive deterioration 1 year after STN DBS. Our model differentiates patients with stable cognition with 88% accuracy from patients with cognitive decline according to MDS criteria, with a PPV of 91.4% and a NPV of 85.0%. The robustness of this approach is demonstrated by relatively stable performance-metrics for each of the individual cross-validation runs, and the stable difference in predicted class-probabilities between the groups. Survival

analyses demonstrate that the cognitive classes had highly differential predicted probabilities; the chance of suffering from cognitive decline 1 year after STN DBS was particularly small in cases of probabilities >0.5.

To our knowledge, no previous study has investigated the actual model-performance of EEG to predict cognitive decline after STN DBS. Although the exact retest reliability of the individual tests used to compose the Cog-score in patients both with and without neuromodulation is unknown, previous literature reported cognitive test performance as relatively stable in PD patients without DBS.<sup>47</sup> A previous paper investigated the potential of qEEG to predict cognitive decline after STN DBS in 17 patients<sup>13</sup> by evaluating the predictive potential of relative  $\theta$  power in predicting cognitive decline. Although results indicated the utility of qEEG to predict postoperative decline, only one spectral metric was studied and correlated with only 3/13 cognitive tests, whereas its predictive potential was not discussed. Another paper described Grand Total EEG (GTE) scores to predict cognitive decline after STN DBS in 30 patients,<sup>48</sup> and reported an odds ratio of 2.78 (95% CI 1.25–6.18) in predicting cognitive decline, although the minority class (cognitive decline based on Mini-Mental State Examination [MMSE]/DemTect scores) consisted of only six patients and model performance was not reported. Moreover, GTE scores are semiquantitative in nature and still require expert knowledge of EEGs, limiting generalizability compared to automated machine learning.

Eighteen features were retained in total during cross-validation runs, none of which were selected in all five models. This suggests that there is not one representative neurophysiological signature of cognitive deterioration, but that several patterns of preoperative cognitive (dys)function exist with roughly equal potential to reflect postoperative cognitive decline. Eleven of 18 features were Fast Fourier Transformation (FFT)-based, although baseline spectral features that are more routinely studied in PD did not differ between the two groups. We hypothesize that our more in-depth spectral features are more suitable as novel biomarkers reflecting neurophysiological signaling speed than conventional band-powers. Three other metrics were related to the mass quantile (ie, calculation of a relative index in which  $q\%$  of the mass of the time series is left of the index). To our knowledge, quantile functions have not been studied previously in relation to qEEG in PD, or in cognitive research, and represent a novel mathematical approach in this setting. The retained EEG-features for modeling cognitive state in a cross-sectional model (using the same methodology and feature-library) are different from those retained for modeling future cognitive deterioration,<sup>14</sup> however both congregate by prioritizing FFT-features. We speculate that different FFT-aspects underlie either current



cognitive functioning or future cognitive dysfunction. Conventional EEG band-powers are probably too inter-related to optimally reflect this distinction. It should be noted that our model pertains to cognitive decline after STN DBS specifically and has not been validated for use in prognosticating cognition without DBS. Future research should evaluate its utility in non-surgical cohorts as well, and investigate cognitive functioning after a longer follow-up. However, including a non-surgical cohort would have introduced substantial bias as ineligibility for surgery is often based on poor cognitive functioning, which would have set these patients apart from the groups characterized here.

A limitation of many machine learning models is the “black-box-problem”, in which it is unclear how the model reaches a decision or what these decisions are based on.<sup>49</sup> As we minimized arbitrary choices in our model, the resulting EEG-features can be considered novel biomarkers in relation to EEG research in PD, and results can be traced back to pathophysiological EEG alterations (eg, EEG-slowing, mass quantile alterations, etc.). However, we realize that the exact mathematical background of the resulting features may be relatively unclear and, as a consequence, their pathophysiological substrate. In terms of etiology, we cannot readily explain why some features were selected and others were not; however, in terms of implications for clinical practice the prognostic performance of the selected features currently supersedes their content. Despite our relatively small sample size, we have demonstrated good accuracy in differentiating cognitively stable patients from those with cognitive decline. We demonstrate that overfitting did not play a major role in our models despite an approximate drop in accuracy of 9% relative to the training-set, and differentiation of the groups based on predicted probability was stable.

Our results further demonstrate that a standard 21-channel EEG probably suffices to predict cognitive decline after STN DBS. Further insight into pathophysiological alterations underlying cognitive decline may be provided through source localization to more closely pinpoint anatomical substrates,<sup>50,51</sup> especially when combined with high-density EEG setups or pre–post EEG settings.<sup>52</sup> However, these setups are less applicable in a clinical prediction setting which would have limited its current utility.

In this study, cognitive deterioration was classified based on MDS criteria for cognitive decline,<sup>16</sup> which increases the generalizability of our results. MDS level I criteria for cognitive decline (ie,  $\geq 2$  tests showing significant deterioration) are less strict than those for PD dementia (PDD),<sup>53</sup> which is the likely reason for our observed high incidence of cognitive decline. Similar to previous studies,<sup>5</sup> the observed cognitive decline is still within the range of normal cognitive functioning. Our classification therefore reflects “deterioration” rather

than actual clinical impairment. Although class-transition to mild cognitive impairment or PDD would have been clinically more relevant, this transition within 1 year is most likely rare after DBS and would require larger sample sizes than currently available, as well as introducing a significant class imbalance. Whether the observed cognitive deterioration is a result of disease progression, a differential impact of stimulation, or a postoperative hypodopaminergic state cannot be determined. However, since the groups did not differ in any baseline characteristic, our results indicate the prognostic utility of EEG regardless of the mechanism behind cognitive deterioration.

Strikingly, generic tests such as the MDRS and MoCA did not show the same deterioration as the Cog-score, even in patients classified as cognitively deteriorated. This is in line with previous literature, as 27/29 studies reported no significant cognitive decline in these global tests after STN DBS.<sup>5</sup> Both generic tests are subject to confounding such as educational level.<sup>54-57</sup> We hypothesize that T-score-based cognitive tests resulted in more sensitive metrics to reflect cognitive functioning. Our observed cognitive deterioration was predominantly fueled by deterioration in the domains “executive functioning” and “language”<sup>58</sup> in line with previous research.<sup>5</sup> As there are multiple possible cognitive tests to quantify cognition in PD, the ideal classification would be based on  $\geq 2$  tests per cognitive domain according to Diagnostic and Statistical Manual of Mental Disorders (DSM-5) and MDS level II criteria to provide a more complete overview of cognitive functioning.<sup>58</sup>

Our results imply that EEG may be used to prognosticate cognitive decline after STN DBS in a clinical setting. A risk of (mild) cognitive decline does not necessarily equal ineligibility for surgery, but improved prognostication may contribute to better patient education and informed consent. Moreover, a risk of postoperative decline as indicated by our machine learning algorithm could skew clinicians towards targets with possible less impact on cognition such as pallidal DBS, although our model has not been tested for prognostication in these settings. Strikingly, patients classified as cognitively stable were unanimous in choosing DBS again if given the opportunity, whereas this was significantly lower in patients with cognitive deterioration despite no differences in any other characteristic. This underscores the importance of cognitive function for patient satisfaction after DBS.<sup>39</sup> Our results indicate that EEG-based prognostication could be included in the DBS screening to improve shared-decision-making.

Strengths of this study include using a fully automated machine learning algorithm which enhances the generalizability and potential clinical utility of our results. Based on the nature of the Boruta selection algorithm which uses multiple independent trials during feature-selection, we are confident that the resulting

model provides an accurate and robust electrophysiological signature underlying the cognitive classes, despite a relatively small sample size. A limitation of our study is a possibly selective loss-to-follow-up as several patients that declined the follow-up assessments reported that cognitive assessments would be burdensome for them at this stage, which may result in an underrepresentation of cognitively impaired patients in the entire cohort, although this would not affect the prediction analyses. Although patients without EEG-recordings did not differ clinically from patients with available recordings, the utility of EEG may be limited due to unavailability of artifact-free epochs. The amount of missing data due to EEG-artifacts is similar in previous literature.<sup>10</sup> Given our lack of external validation, our results are best interpreted as proof-of-concept that machine learning algorithms can accurately predict cognitive deterioration after STN DBS.

Our results indicate that cortical neurophysiological alterations prior to STN DBS may reflect a risk of postoperative cognitive deterioration, and the use of algorithms such as those detailed here may be helpful during the DBS eligibility screening to directly improve clinical practice. Automated analyses such as these have the ability to identify new biomarkers. Further exploration of the features selected by our model may ultimately provide greater insight into pathophysiological processes and the mechanism behind the association between EEG and cognitive decline. ■

**Acknowledgments:** The authors would like to thank the members of the DBS team of LUMC/Haga Teaching Hospital (G. E. L. Hendriks, A. Mosch, R. Zutt, C. F. Hoffmann, E. Marks, E. Lauwen, G. van Holten, P. de Maa, and S. van der Gaag) for patient care and the EEG technicians of the LUMC for their help with the data acquisition.

## References

1. Deuschl G, Schade-Brittinger C, Krack P, et al. A randomized trial of deep-brain stimulation for Parkinson's disease. *N Engl J Med* 2006;355:896–908.
2. Schuepbach WM, Rau J, Knudsen K, et al. Neurostimulation for Parkinson's disease with early motor complications. *N Engl J Med* 2013;368:610–622.
3. Okun MS, Gallo BV, Mandybur G, et al. Subthalamic deep brain stimulation with a constant-current device in Parkinson's disease: an open-label randomised controlled trial. *Lancet Neurol* 2012;11:140–149.
4. Geraedts VJ, Feleus S, Marinus J, van Hilten JJ, Contarino MF. What predicts quality of life after subthalamic deep brain stimulation in Parkinson's disease? A systematic review. *Eur J Neurol* 2020; 27:419–428.
5. Mehanna R, Bajwa JA, Fernandez H, Wagle Shukla AA. Cognitive impact of deep brain stimulation on Parkinson's disease patients. *Parkinsons Disease* 2017;2017:3085140.
6. Geraedts VJ, Kuijff ML, van Hilten JJ, Marinus J, Oosterloo M, Contarino MF. Selecting candidates for deep brain stimulation in Parkinson's disease: the role of patients' expectations. *Parkinsonism Relat Disord* 2019;66:207–211.
7. Lang AE, Houeto JL, Krack P, et al. Deep brain stimulation: preoperative issues. *Mov Disord* 2006;21(Suppl 14):S171–S196.
8. Cernera S, Okun MS, Gunduz A. A review of cognitive outcomes across movement disorder patients undergoing deep brain stimulation. *Front Neurol* 2019;10:419.
9. Geraedts VJ, Boon LI, Marinus J, et al. Clinical correlates of quantitative EEG in Parkinson disease: a systematic review. *Neurology* 2018;91:871–883.
10. Geraedts VJ, Marinus J, Gouw AA, et al. Quantitative EEG reflects non-dopaminergic disease severity in Parkinson's disease. *Clin Neurophysiol* 2018;129:1748–1755.
11. Bočková M, Lamoš M, Klimeš P, et al. Suboptimal response to STN-DBS in Parkinson's disease can be identified via reaction times in a motor cognitive paradigm. *J Neural Transm (Vienna)* 2020; 127(12):1579–1588.
12. Hatz F, Meyer A, Roesch A, Taub E, Gschwandtner U, Fuhr P. Quantitative EEG and verbal fluency in DBS patients: comparison of stimulator-on and -off conditions. *Front Neurol* 2018;9:1152.
13. Yakufujiang M, Higuchi Y, Aoyagi K, et al. Predictive potential of preoperative electroencephalogram for neuropsychological change following subthalamic nucleus deep brain stimulation in Parkinson's disease. *Acta Neurochir* 2019;161:2049–2058.
14. Geraedts VJ, Koch M, Contarino MF, 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* 2021;132(5):1041–1048.
15. Koch M, Geraedts V, Wang H, Tannemaat MR, Bäck T. Automated machine learning for EEG-based classification of Parkinson's disease patients. In: *2019 IEEE International Conference on Big Data*. Los Angeles; 2019;4845–4852.
16. Litvan I, Goldman JG, Troster AL, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society task force guidelines. *Mov Disord* 2012;27:349–356.
17. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord* 2015;30:1591–1601.
18. Geraedts VJ, van Ham RAP, Marinus J, et al. Intraoperative test stimulation of the subthalamic nucleus aids postoperative programming of chronic stimulation settings in Parkinson's disease. *Parkinsonism Relat Disord* 2019;65:62–66.
19. Hjorth B. Source derivation simplifies topographical EEG interpretation. *Am J EEG Technol* 1980;20:121–132.
20. Burle B, Spieser L, Roger C, Casini L, Hasbroucq T, Vidal F. Spatial and temporal resolutions of EEG: is it really black and white? A scalp current density view. *Int J Psychophysiol* 2015;97:210–220.
21. Vakil E, Blachstein H. Rey auditory-verbal learning test: structure analysis. *J Clin Psychol* 1993;49:883–890.
22. Huppert FA, Brayne C, Gill C, Paykel ES, Beardsall L. CAMCOG—a concise neuropsychological test to assist dementia diagnosis: socio-demographic determinants in an elderly population sample. *Br J Clin Psychol* 1995;34(Pt 4):529–541.
23. Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol* 2004;19:203–214.
24. Scarpina F, Tagini S. The Stroop color and word test. *Front Psychol* 2017;8:557–557.
25. Dekker R, Mulder JL, Dekker PH. De ontwikkeling van vijf nieuwe Nederlandstalige tests. Leiden: PITS; 2007.
26. Goetz CG, Tilley BC, Shaftman SR, 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* 2008;23:2129–2170.
27. Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov Disord* 2010;25:2649–2653.
28. Maier F, Lewis CJ, Eggers C, et al. Development and validation of the deep brain stimulation impairment scale (DBS-IS). *Parkinsonism Relat Disord* 2017;36:69–75.
29. van der Heeden JF, Marinus J, Martinez-Martin P, van Hilten JJ. Evaluation of severity of predominantly non-dopaminergic symptoms in Parkinson's disease: the SENS-PD scale. *Parkinsonism Relat Disord* 2016;25:39–44.

30. Peto V, Jenkinson C, Fitzpatrick R, Greenhall R. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. *Qual Life Res* 1995;4:241–248.
31. Mattis S. *Dementia Rating Scale: Professional Manual*. Odess, FL: Psychological Assessment Resources Inc; 1988.
32. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;53:695–699.
33. Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *J Pers Assess* 1996;67:588–597.
34. Leentjens AF, Dujardin K, Pontone GM, Starkstein SE, Weintraub D, Martinez-Martin P. The Parkinson Anxiety Scale (PAS): development and validation of a new anxiety scale. *Mov Disord* 2014;29:1035–1043.
35. Marin RS, Biedrzycki RC, Firinciogullari S. Reliability and validity of the Apathy Evaluation Scale. *Psychiatry Res* 1991;38:143–162.
36. Weintraub D, Mamikonyan E, Papay K, Shea JA, Xie SX, Siderowf A. Questionnaire for impulsive-compulsive disorders in Parkinson's disease-rating scale. *Mov Disord* 2012;27:242–247.
37. Visser M, Marinus J, Stiggelbout AM, Van Hilten JJ. Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT. *Mov Disord* 2004;19:1306–1312.
38. Marinus J, Visser M, van Hilten JJ, Lammers GJ, Stiggelbout AM. Assessment of sleep and sleepiness in Parkinson disease. *Sleep* 2003;26:1049–1054.
39. Geraedts VJ, van Hilten JJ, Marinus J, et al. Stimulation challenge test after STN DBS improves satisfaction in Parkinson's disease patients. *Parkinsonism Relat Disord* 2019;69:30–33.
40. Saranza G, Lang AE. Levodopa challenge test: indications, protocol, and guide. *J Neurol* 2020. Online ahead of print.
41. Christ M, Kempa-Liehr AW, Feindt M. Distributed and parallel time series feature extraction for industrial big data applications. *arXiv e-prints* [serial online]. 2016. <https://ui.adsabs.harvard.edu/abs/2016arXiv161007717C>. Accessed October 1, 2016.
42. Christ M, Braun N, Neuffer J, Kempa-Liehr AW. Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). *Neurocomputing* 2018;307:72–77.
43. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw* 2010;36:1–13.
44. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer; 2009.
45. Wang H, Emmerich M, Bäck T. Cooling strategies for the moment-generating function in Bayesian global optimization. In: *2018 IEEE Congress on Evolutionary Computation (CEC)*. 2018;1–8.
46. Wang H, Stein By, Emmerich M, Back T. A new acquisition function for Bayesian optimization based on the moment-generating function. In: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2017;507–512.
47. Tröster AI, Woods SP, Morgan EE. Assessing cognitive change in Parkinson's disease: development of practice effect-corrected reliable change indices. *Arch Clin Neuropsychol* 2007;22:711–718.
48. Markser A, Maier F, Lewis CJ, et al. Deep brain stimulation and cognitive decline in Parkinson's disease: the predictive value of electroencephalography. *J Neurol* 2015;262:2275–2284.
49. Price WN. Big data and black-box medical algorithms. *Sci Transl Med* 2018;10:eaa05333.
50. Jatoi MA, Kamel N, Malik AS, Faye I, Begum T. A survey of methods used for source localization using EEG signals. *Biomed Signal Process Control* 2014;11:42–52.
51. Asadzadeh S, Rezaei TY, Beheshti S, Delpak A, Meshgini S. A systematic review of EEG source localization techniques and their applications on diagnosis of brain abnormalities. *J Neurosci Methods* 2020;108740.
52. Buril J, Burilova P, Pokorna A, Balaz M. Use of high-density EEG in patients with parkinson's disease treated with deep brain stimulation. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2020;164:366–370.
53. Emre M, Aarsland D, Brown R, et al. Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord* 2007;22:1689–1707; quiz 1837.
54. Foss MP, Vale Fde A, Speciali JG. Influence of education on the neuropsychological assessment of the elderly: application and analysis of the results from the Mattis Dementia Rating Scale (MDRS). *Arq Neuropsiquiatr* 2005;63:119–126.
55. Borda MG, Reyes-Ortiz C, Pérez-Zepeda MU, Patino-Hernandez D, Gómez-Arteaga C, Cano-Gutiérrez CA. Educational level and its association with the domains of the Montreal Cognitive Assessment test. *Aging Ment Health* 2019;23:1300–1306.
56. Wu Y, Wang M, Ren M, Xu W. The effects of educational background on Montreal Cognitive Assessment screening for vascular cognitive impairment, no dementia, caused by ischemic stroke. *J Clin Neurosci* 2013;20:1406–1410.
57. Yancar Demir E, Özcan T. Evaluating the relationship between education level and cognitive impairment with the Montreal Cognitive Assessment test. *Psychogeriatrics* 2015;15:186–190.
58. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.

## Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.