

A Task-Aligned Multimodal Machine Learning Framework for Studying Working Memory Dysfunction in Parkinson's Disease

Mercedes Terry

`mercedes.terry@und.edu`

University of North Dakota

Samuel Birkholz

Sanford Health

Jeffrey Johnson

North Dakota State University

Jau-Shin Lou

Sanford Health

Asenath Huether

Minnesota State Community and Technical College

Jessica Keller

Sanford Health

Enrique Alvarez

University of North Dakota

Research Article

Keywords: Parkinson's disease, working memory, electroencephalography, feature engineering, principal component analysis, machine learning

Posted Date: May 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9631659/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Cognitive Neurodynamics on August 13th, 2026. See the published version at <https://doi.org/10.1007/s11571-026-10526-z>.

A Task-Aligned Multimodal Machine Learning Framework for Studying Working Memory Dysfunction in Parkinson's Disease

Mercedes Terry, PhD^{1*}, Samuel A. Birkholz, PhD^{2,3}, Jeffery S. Johnson, PhD², Jau-Shin Lou, MD, PhD, MBA^{3,4}, Asenath X. Huether, PhD⁵, Jessica Keller, BS³, and Enrique Alvarez Vazquez, PhD¹

¹Biomedical Engineering Department, University of North Dakota, Grand Forks, ND 58202, USA

²Department of Psychology, North Dakota State University, Fargo, ND 58102, USA

³Department of Neurology, Sanford Health, Fargo, ND 58103, USA

⁴Department of Neurology, University of North Dakota School of Medicine and Health Science, Grand Forks, ND 58202, USA

⁵Department of Psychology, Minnesota State Community and Technical College, Moorhead, MN 56560, USA

* Correspondence:

Corresponding Author

mercedes.terry@und.edu

Abstract—Cognitive impairment is one of the most functionally debilitating non-motor symptoms in Parkinson's disease (PD). Yet, current diagnostic and clinical practices rely heavily on subjective assessments and burdensome behavioral testing, which lack the temporal resolution to detect subtle deficits. While EEG and pupillometry offer promising non-invasive insights into cognitive processing, more scalable and objective methods are needed to detect subtle and often elusive dysfunction in early PD. Machine learning (ML) leverages patterns in neurophysiological signals recorded during cognitive tasks to identify subtle impairments that may not be evident in standard evaluations. We recorded EEG and pupillometry data from 68 participants (35 PD, 33 healthy controls (HC)) during a visual Change Detection working memory task. Using our custom, standardized feature-extraction toolbox, we extracted 108 features and incorporated the resulting feature matrices into an ML pipeline. We applied SMOTE to balance the classes, then used principal component analysis (PCA) to reduce dimensionality. An automated elbow method identified optimal cutoffs for principal components (PCs) and original features, which guided subsequent recursive elimination (RE) for computational efficiency. We trained a support vector machine with a radial basis function kernel (RBF-SVM) to classify PD and evaluated the final model's performance on a hold-out dataset. The automated elbow method identified optimal cutoffs at 20 PCs and 22 original features.

The RE revealed that a model using 14 PCs and the top 7 PCA-weighted features achieved the highest performance, with 71% accuracy and an F1 score of 0.701. The classifier's performance declined slightly when tested on held-out data, with an accuracy of 63% and an F1 score of 0.626. This study demonstrates the potential of a task-aligned ML pipeline for uncovering early cognitive markers of PD. The framework supports traceable tools that link classification back to specific physiological features and task stages, supporting mechanistic investigation of cognitive dysfunction in PD.

Index Terms— Parkinson's disease, working memory, electroencephalography, feature engineering, principal component analysis, machine learning

I. INTRODUCTION

PARKINSON'S disease (PD) is a progressive neurodegenerative disorder traditionally defined by motor symptoms such as tremor, rigidity, and bradykinesia [1]. Yet cognitive impairments also emerge early and substantially affect daily functioning and quality of life. Despite their predictive value for decline and progression, these non-motor symptoms are often overlooked and undertreated [1], [2]. Working memory (WM) is one of the most consistently affected domains, supporting reasoning, decision-making, and

This work was supported in part by the Peltier Family Foundation. The Sanford Parkinson's Disease Research Laboratory assisted with recruitment of Parkinson's disease participants, while healthy control recruitment and all study protocols for both groups were conducted at the University of North Dakota. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. (Corresponding author: Mercedes A. Terry.)

Mercedes A. Terry and Enrique Alvarez Vazquez are with the Biomedical Engineering Department, University of North Dakota, Grand Forks, ND 58202 USA (e-mail: mercedes.terry@und.edu; enrique.alvarez@und.edu).

Samuel A. Birkholz is with the Department of Psychology, North Dakota State University, Fargo, ND 58102 USA, and also with the Department of Neurology, Sanford Health, Fargo, ND 58103 USA (e-mail: samuel.birkholz@sanfordhealth.org).

Jeffrey S. Johnson is with the Department of Psychology, North Dakota State University, Fargo, ND 58102 USA (e-mail: jeffrey.johnson@ndsu.edu). Jau-Shin Lou is with the Department of Neurology, Sanford Health, Fargo, ND 58103 USA, and also with the Department of Neurology, University of North Dakota School of Medicine and Health Sciences, Grand Forks, ND 58202 USA (e-mail: jau-shin.lou@sanfordhealth.org). Asenath X. A. Huether is with the Department of Neurology, Sanford Health, Fargo, ND 58103 USA, and also with the Department of Psychology, Minnesota State Community and Technical College, Moorhead, MN 56560 USA (e-mail: asenath.huether@minnesota.edu). Jessica N. Keller is with the Department of Neurology, Sanford Health, Fargo, ND 58103 USA (e-mail: jessica.keller@sanfordhealth.org).

goal-directed behavior as well as attention and executive control [3], [4]. In PD, WM deficits in updating, attentional filtering, and storage capacity are linked to frontostriatal dysfunction. They are especially evident in visual WM tasks that manipulate cognitive load, where patients perform worse than healthy controls, even without global cognitive decline [5], [6], [7]. These impairments directly affect quality of life: Lawson et al. [8] found that mild cognitive impairment (MCI), often characterized by WM dysfunction, predicted poorer outcomes, and Fan et al. [9] reported that PD-MCI patients had markedly worse quality of life than cognitively intact individuals.

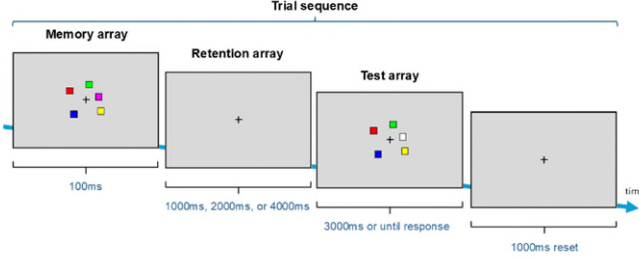


Fig. 1. Graphical image of the sequence of tasks from the WM-CDT. Participants first viewed colored squares, followed by different delay periods. This was followed by a test display in which participants responded whether the test display matched the memory display in terms of color [24].

Despite growing recognition of cognitive impairment in PD, accurate diagnosis and monitoring remain difficult [9]. Traditional neuropsychological tests are time-consuming, burdensome, and lack the temporal resolution to detect subtle neural changes [10]. Early detection is further hindered by symptom overlap with normal aging, reducing diagnostic sensitivity [11]. Cognitive decline in PD is also heterogeneous, with wide variation in onset, severity, and progression, limiting the utility of standardized assessments for diagnosis and tracking [12], [13]. Thus, there is an urgent need for reliable biomarkers to predict and measure cognitive decline, thereby supporting the development of disease-modifying therapies [12]. Physiological measures such as EEG and pupillometry have been used to address limitations in the diagnosis of cognitive impairment in PD [14]. EEG provides millisecond-level resolution to capture rapid neural dynamics during WM and can reveal subtle processing deficits [15], while pupillometry indexes autonomic and cognitive processes such as effort and attention, which are sensitive to WM load and decline [14], [16], [17]. Most studies, however, rely on unimodal data, isolated features, or conventional statistics that may miss complex relationships [10], [16], [18], [19]. Machine learning (ML) can capture nonlinear patterns in EEG and pupillometry, and early work shows promise for classifying cognitive states and detecting PD-related abnormalities during WM tasks [20], [21]. Still, existing pipelines often lack interpretability, robustness, and clinical generalizability, and few provide standardized multimodal integration, wide-range feature extraction, or models whose classification decisions can be traced back to specific physiological features and the cognitive task stages in which they occur, enabling mechanistic interpretation of group differences [21], [22].

II. METHODS

A. Data Collection and Behavior Data Analysis

EEG and pupillometry were recorded during a visuospatial WM Change Detection Task (CDT) [23]. On each trial, participants viewed 2–5 colored squares, followed by a variable delay (1000, 2000, or 4000 ms), and then a test display requiring a button press to indicate whether the array matched the original or contained a color change (probability of change = 0.5; see Figure 1). EEG was collected from 64 scalp electrodes using a BioSemi ActiveTwo system (512 Hz, 0.01–100 Hz bandpass), with additional mastoid and EOG electrodes for artifact monitoring. Data from 68 participants were analyzed (PD: $n = 33$; HC: $n = 35$). EEG was segmented into 1195 ms epochs (195 ms pre- to 1000 ms post-memory array onset), baseline-corrected, and preprocessed with independent component analysis (ICA). Pupillometry was embedded in the LAREA channel (an auxiliary ERPLAB channel for time-locking external signals), enabling synchronized feature extraction alongside EEG. ICA was used to identify and remove artifact-related components, including eye blinks, saccades, and muscle activity, based on a combination of component topography, time course, and power spectra, following established EEGLAB guidelines. Components reflecting ocular artifacts were identified by correlating them with EOG channels and by their characteristic frontal scalp distributions and were removed before further analysis. For pupillometry, blink periods were detected and interpolated using standard pupil preprocessing procedures, and trials with excessive signal loss was excluded to ensure a reliable estimation of pupil-based features. WM capacity was estimated with a modified Pashler’s k formula [24], and group differences were tested with repeated-measures ANOVA [23].

To capture stage-specific processing, EEG and pupillometry features were grouped into task-condition cells defined by memory stage (encoding vs. retrieval), set size (2–5 items), delay (1000, 2000, or 4000 ms), and response accuracy (correct vs. incorrect). This yielded 48 task-relevant task-condition cells, allowing fine-grained alignment of features with specific cognitive demands of the WM-CDT. This task-condition cell structure is central to the pipeline’s interpretability: by retaining the correspondence between each feature and its originating task stage, the framework allows classification decisions to be traced back to specific physiological signals occurring at specific moments during the working memory task, enabling post-hoc investigation of the neural and autonomic dynamics underlying group differences.

B. Machine Learning Pipeline

We used our custom MATLAB toolbox to extract 108 EEG and pupillometry features across ten categories [25]. The toolbox preserves temporal alignment with task events and focuses on physiologically meaningful markers identified in prior literature [5], [10], [14], [17], [18], [19], [25], [26]. For example, frontal alpha asymmetry (FAA) was included as it

reflects altered frontal alpha power linked to executive control deficits in PD [5], [10], [27], while pupil diameter variability was included as a marker of cognitive effort during WM tasks [14], [16]. Table I lists all extracted features by category.

The toolbox extracted features within task-relevant task-condition cells for all participants, preserving the task-condition cell-by-feature structure and temporal dynamics of each trial. Channel-specific measures were computed at electrodes designated by prior literature, while global features were derived from the full signal within each task-condition cell. The same task-condition cell loop was applied across participants, ensuring comparable feature matrices that could then be input into ML models to identify not only which features were most predictive, but also when they occurred. By aligning features

20% of participants ($n = 14$; 7 PD, 7 HC) prior to any model training, feature selection, or preprocessing. This ensured evaluation of participants who were entirely unseen and provided an unbiased estimate of generalization performance. All feature matrices for a withheld participant were excluded, ensuring that no participant contributed data to both the training and evaluation sets and eliminating the risk of data leakage. The remaining 54 participants (28 PD, 26 HC) formed the training pool. See Table II for a full breakdown of the dataset's composition.

Although participant counts were nearly balanced overall (35 PD, 33 HC), the 80/20 participant-level split produced slightly unequal class sizes in the training set (28 PD vs. 26 HC feature

Table I

ALL 108 EEG AND PUPILLOMETRY FEATURES ACROSS TEN CATEGORIES USED IN THE FEATURE EXTRACTION TOOLBOX

CATEGORY (N)	FEATURES
TIME DOMAIN (21)	MEAN AMPLITUDE, STANDARD DEVIATION OF AMPLITUDE, PEAK-TO-PEAK AMPLITUDE, ROOT MEAN SQUARE AMPLITUDE, VARIANCE OF AMPLITUDE, ZERO CROSSINGS, SKEWNESS, KURTOSIS, MAXIMUM PEAK AMPLITUDE, MINIMUM PEAK AMPLITUDE, MINIMUM SIGNAL VALUE, MAXIMUM SIGNAL VALUE, SIGNAL LATENCY, AREA UNDER THE CURVE, LINE LENGTH OF WAVEFORM, SIGNAL ENERGY, SIGNAL POWER, NOISE POWER, SIGNAL-TO-NOISE RATIO, AUTOCORRELATION AT LAG 1, INTERQUARTILE RANGE OF AMPLITUDE
HJORTH PARAMETERS (4)	HJORTH ACTIVITY, HJORTH MOBILITY, HJORTH COMPLEXITY, HJORTH COMPLEXITY/MOBILITY RATIO
FREQUENCY DOMAIN (19)	DELTA BAND POWER, THETA BAND POWER, ALPHA BAND POWER, BETA BAND POWER, GAMMA BAND POWER, RELATIVE BAND POWER, DELTA PEAK FREQUENCY, THETA PEAK FREQUENCY, ALPHA PEAK FREQUENCY, BETA PEAK FREQUENCY, GAMMA PEAK FREQUENCY, ALPHA/BETA POWER RATIO, THETA/BETA POWER RATIO, ALPHA/THETA POWER RATIO, GAMMA/BETA POWER RATIO, MEAN FREQUENCY, MEDIAN FREQUENCY, PETROSIAN FRACTAL DIMENSION, BURST PERCENTAGE
PUPILLOMETRY (15)	MEAN PUPIL DIAMETER, STANDARD DEVIATION OF PUPIL DIAMETER, MINIMUM PUPIL DIAMETER, MAXIMUM PUPIL DIAMETER, RANGE OF PUPIL DIAMETER, VARIANCE OF PUPIL DIAMETER, MEAN PUPIL VELOCITY, MEAN PUPIL ACCELERATION, BLINK RATE, BLINK DURATIONS, MEAN BLINK DURATION, TASK-EVOKED PUPILLARY RESPONSE LATENCY, TASK-EVOKED PUPILLARY RESPONSE PEAK AMPLITUDE, HALF RECOVERY TIME OF PUPIL RESPONSE, AREA UNDER THE CURVE OF TASK-EVOKED PUPILLARY RESPONSE
ERP COMPONENTS (12)	N100 AMPLITUDE, N100 LATENCY, N200 AMPLITUDE, N200 LATENCY, P200 AMPLITUDE, P200 LATENCY, P300 AMPLITUDE, P300 LATENCY, AVERAGE N200 AMPLITUDE, AVERAGE P300 AMPLITUDE, COMPONENT AMPLITUDE RATIO, P300 VARIABILITY
PHASE AND SYNCHRONY METRICS (10)	THETA SYNCHRONY, ALPHA DESYNCHRONIZATION, PHASE DIFFERENCE, PHASE SYNCHRONY, PHASE RESETTING INDEX, ALPHA BAND COHERENCE, BETA BAND COHERENCE, PHASE LOCKING VALUE, PHASE LOCKING VALUE MEAN, CROSS-CHANNEL COUPLING
ENTROPY AND COMPLEXITY (7)	ERP ENTROPY, TSALLIS ENTROPY, HURST EXPONENT, TEAGER ENERGY OPERATOR SIGNAL, MEAN TEAGER ENERGY OPERATOR, GLOBAL FIELD POWER, ROOT MEAN SQUARE OF SUCCESSIVE DIFFERENCES
SPIKE AND WAVEFORM FEATURES (5)	WAVEFORM SHARPNESS, SPIKE FREQUENCY, SPIKE AMPLITUDE, SPIKE/SLOW WAVE AMPLITUDE RATIO, SLOW WAVE ACTIVITY
ASYMMETRY AND INDICES (3)	FRONTAL ALPHA ASYMMETRY, PREFRONTAL ASYMMETRY INDEX, PHASE- AMPLITUDE COUPLING
SPECTRAL FEATURES (12)	SPECTRAL CENTROID, SPECTRAL ENTROPY, SPECTRAL FLATNESS, SPECTRAL ROLL-OFF FREQUENCY, SPECTRAL KURTOSIS, SPECTRAL VARIABILITY, SPECTRAL EDGE FREQUENCY (95%), EVENT-RELATED SPECTRAL PERTURBATION, POWER SPECTRAL DENSITY (PSD) SLOPE, PSD SKEWNESS, PSD KURTOSIS, CUMULATIVE SPECTRAL POWER

with cognitive task stages, the toolbox facilitated the development of interpretable models that can help link EEG and pupillary signals to underlying processes.

Each participant contributed one feature matrix of dimensions 108 features $\times$ 48 task-condition cells, yielding 68 feature matrices in total: 35 from PD participants and 33 from HC participants. A hold-out set was constructed by withholding

matrices). In the context of a small training set ($n = 54$ total matrices), this 7% imbalance is sufficient to bias a decision boundary toward the majority class, as even moderate class imbalance has been shown to increase majority-class prediction bias in small, high-dimensional datasets [28]. Accordingly, the Synthetic Minority Over-Sampling Technique (SMOTE) was applied to generate 2 synthetic minority-class samples and fully balance the training distribution prior to model fitting, yielding a balanced training set of 56 matrices (28 PD, 28 HC). SMOTE generates synthetic minority-class samples by interpolating between existing samples and their k nearest neighbors in feature space ($k = 5$ in this study), producing novel synthetic points rather than duplicating existing observations. SMOTE was applied solely to balance the class distribution and improve

Table II

DATASET COMPOSITION AND TRAIN/HOLD OUT PARTITION SUMMARY

CATEGORY	PD	HC	TOTAL
ALL PARTICIPANTS	35	33	68
TRAINING PARTICIPANTS (80%)	28	26	54
HOLD OUT PARTICIPANTS (20%)	7	7	14
FEATURE MATRICES PER PARTICIPANT	1(108x48)	1(108x48)	68

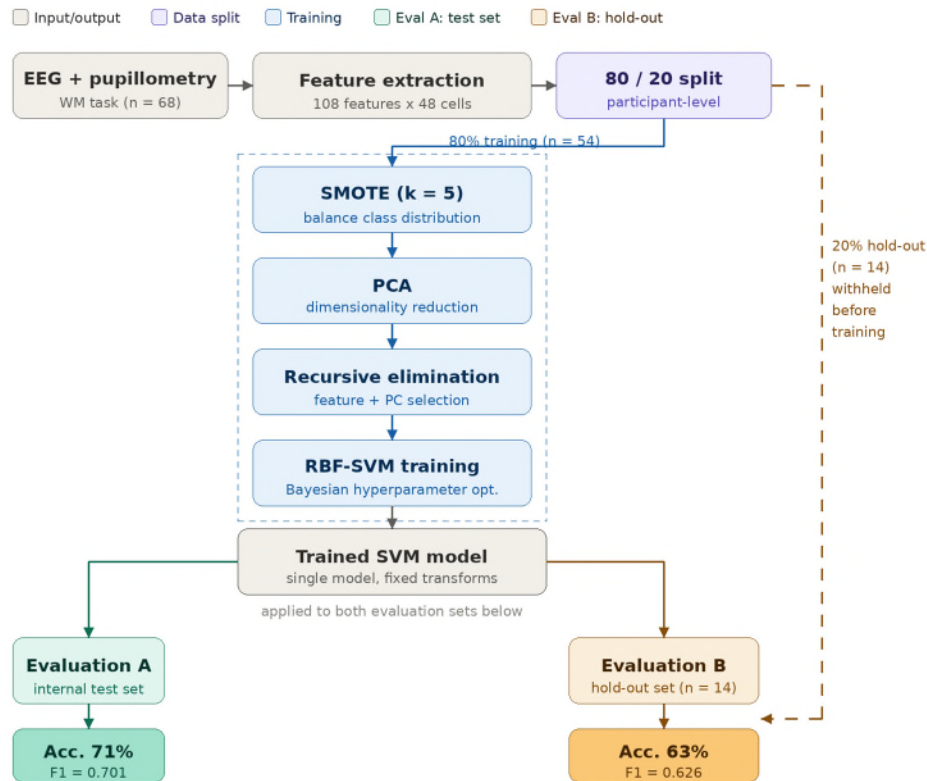


Fig. 2. Overview of the ML pipeline developed for classifying PD using EEG and pupillometry data. The process begins with feature extraction using a standardized, temporally aligned toolbox that captures a broad range of neural and physiological features during a WM-CDT. Dimensionality reduction is performed using PCA and automated elbow point detection, followed by RE to identify the most informative components. The optimized feature set is then used to train an SVM-RBF classifier, which is validated using test and holdout datasets.

recall on the minority class (PD); it was not applied to the hold-out set, which was evaluated on its original unmodified class distribution. Alternative rebalancing strategies, including ADASYN and class-weight adjustment, were considered; SMOTE was selected for its well-characterized interpolation behavior and its established use in small-sample neurophysiological classification tasks. Subsequent PCA further compressed the feature space prior to SVM training, consistent with recommendations that dimensionality reduction precede or accompany oversampling in high-dimensional, low-sample-size settings [28].

Principal Component Analysis (PCA) and recursive elimination (RE) were applied exclusively within the training set to compress correlated features and limit the number of free parameters exposed to the classifier. Generalization was then evaluated using a strict participant-level hold-out set, providing a conservative assessment of performance in a setting where the risk of overfitting is inherently high.

PCA was applied to reduce dataset complexity by transforming the 108 original features into orthogonal components ranked by explained variance. PCA was fit using only training data, and the resulting loading matrix was then applied to the project validation and hold-out feature matrices to map them into the same reduced-dimensional space, ensuring

that no information from unseen participants influenced component estimation. PCA was preferred over direct feature ranking because the extracted features were highly correlated across task-condition cells and modalities. Dimensionality reduction via orthogonal components mitigated multicollinearity and overfitting while preserving distributed variance that would be lost by selecting a small subset of individual features.

An automated elbow method with a geometric knee-point algorithm was applied to both the explained variance curve (PCs) and the ranked feature contributions (summed absolute loadings) to identify objective cutoffs for when additional inputs added minimal value. This procedure does not rely on a manually defined threshold. Instead, the knee-point algorithm automatically detects the point of maximum curvature in the variance and feature-importance curves, marking where additional components or features contribute minimal

incremental information. RE was then applied to iteratively prune the least informative PCs and features, focusing only on the reduced search space and maximizing classification accuracy while improving computational efficiency. RE was validated by systematically running models with varying PC subsets to confirm the optimal selection.

A support vector machine (SVM) with a radial basis function

(RBF) kernel was then trained to classify PD and HC groups. The RBF kernel enabled the model to capture complex,

Full univariate statistics are provided in Supplementary Table III. Effect sizes and p-values for two features could not be

Table III
TOP 7 FEATURES FOR THE WM-CDT RANKED BY THEIR TOTAL WEIGHT CONTRIBUTION TO THE TOP 20 PRINCIPAL COMPONENTS

RANK	FEATURE NAME	WEIGHT CONTRIBUTION	TASK-CONDITION CELL	MEMORY PHASE	SET SIZE	ACCURACY	DELAY PERIOD (MS)
1	FAA	2.7537	36	RETRIEVAL	3	INCORRECT	4000
2	PAI	2.5115	34	RETRIEVAL	3	INCORRECT	2000
3	ALPHA THETA RATIO	2.3127	6	ENCODING	2	INCORRECT	4000
4	ERP LATENCY	2.2865	28	RETRIEVAL	2	INCORRECT	2000
5	ALPHA BETA RATIO	2.2710	28	RETRIEVAL	2	INCORRECT	2000
6	HALF RECOVERY TIME	2.2710	36	RETRIEVAL	3	INCORRECT	4000
7	DELTA PEAK FREQUENCY	2.2065	48	RETRIEVAL	5	INCORRECT	4000

nonlinear boundaries that linear methods would miss, improving separability in high-dimensional data. For the RBF-SVM classifier, the BoxConstraint (C) and KernelScale (γ) parameters were optimized within MATLAB's built-in Bayesian optimization framework over default logarithmic search ranges, using five-fold cross-validation on the training set only. Cross-validation was used exclusively for hyperparameter tuning and model selection, and not to estimate final classification performance. All dimensionality-reduction and feature-selection steps were learned on the training data and saved as part of the trained model. Test and hold-out data were processed using the same fixed transformations and evaluated without refitting, preventing dimensionality mismatch or information leakage.

Model performance was evaluated using accuracy, F1-scores, and confusion matrices. Final validation was conducted on a hold-out set using the same metrics. Fig. 2 illustrates the full pipeline, and the top features identified by RE were mapped back to their temporal task-condition cells via the toolbox's alignment, clarifying when during the task the most discriminative features occurred.

III. RESULTS

Behavioral analysis revealed that overall WM capacity was lower in PD participants compared to HCs ($F(1,67) = 4.59$, $p = 0.036$, $\eta^2 = 0.064$). Both groups, however, showed a paradoxical improvement in performance with increasing retention intervals ($F(2,134) = 7.001$, $p = 0.001$, $\eta^2 = 0.095$), such that capacity estimates were higher at 2000 ms and 4000 ms than at the shortest delay. The group $\times$ delay interaction was not significant ($F(2,134) = 0.286$, $p = 0.752$, $\eta^2 = 0.004$). Simple effects analysis revealed that the only statistically significant group difference emerged at the 1000 ms delay ($p = 0.015$), where PD participants performed worse than HCs; however, no differences were found at the longer delays (2000 ms: $p = 0.102$; 4000 ms: $p = 0.087$).

After feature extraction, univariate screening of all 108 features revealed small-to-moderate group differences across multiple EEG and pupillometric measures. Effect sizes (Cohen's d) were generally modest, and few features survived FDR correction, consistent with a diffuse cognitive phenotype rather than a small number of highly discriminative biomarkers.

computed and are reported as not available in the supplementary table. Specifically, global field power exhibited near-zero between-participant variance after task-aligned segmentation and normalization, resulting in effectively constant values across subjects. Phase difference was undefined because stable phase estimation requires consistent oscillatory power and phase locking across trials, which were not uniformly present across participants and task-condition cells in

this dataset. Importantly, such features were effectively downweighted during dimensionality reduction, ensuring that features with minimal variance or unreliable estimates did not meaningfully influence the multivariate classification model.

The elbow point for the PCs occurred at 20 components, indicating that additional components beyond this point contributed minimally to the overall variance structure. Similarly, when evaluating feature importance based on the summed absolute PCA loadings, the elbow point was identified at 22 features. These results suggest that the top 20 principal components and the top 22 original features capture the most critical information for distinguishing between PD and HC groups in WM-CDT. RE found that the model achieved its best performance with 14 principal components and the top 7 PCA-weighted features, yielding an accuracy of 71%, precision of 0.679, recall of 0.669, and an F1 score of 0.701 on the test set.

To validate the model's robustness, we systematically reduced the number of PCs from 13 to 1 while keeping the top 7 features fixed. Performance declined as fewer PCs were used, confirming that 14 components and 7 features provided the most effective and generalizable representation for distinguishing PD from HC in the WM-CDT. The final SVM model, tested on the hold-out set, achieved 63% accuracy with a precision of 0.632, recall of 0.619, and F-score of 0.626. To assess whether the learned feature representation captured group-level cognitive signal beyond chance, we performed a participant-level permutation test on balanced accuracy using

the fully held-out subject predictions. The observed balanced accuracy was 0.701. This value exceeded the permutation distribution (10,000 permutations; permutation $p = 0.0013$), indicating that the learned multimodal feature representation captured statistically significant group information despite

Table VI
Present pipeline in the context of published EEG-based PD classification studies

STUDY	MODALITY/FEATURES	CLASSIFIER	N	ACCURACY	OTHER EVALUATION METRIC
PRESENT STUDY	EEG + PUPILLOMETRY	RBF-SVM (TEST SET)	54	71%	0.701
PRESENT STUDY	EEG + PUPILLOMETRY	RBF + SVM (HOLD OUT)	14	63%	0.626
BERA ET AL. [29]	RESTING EEG,	SVM (TEST SET)	88 (SUBJ. DEPEND.)	82-94%	-
BERA ET AL., [29]	RESTING EEG	SVM (HOLD OUT)	88 (SUBJ. INDEP)	68%	-
ANJUM ET AL. [26]	RESTING EEG, LPC FEATURES	PC + ML ENSEMBLE	106	85.7%	AUC = 0.85
OH ET AL. [20]	RESTING EEG	13-LAYER CNN	40	88.3%	SENS. 84.7%
ALTHAM ET AL. [21]	EEG, MULTI-MODAL REVIEW	RF (BEST REPORTED)	40	91%	AUC = 0.98
SRINIVASAN ET AL. [30]	VOICE SIGNALS	CNN/FMM + SMOTE	195	98%	F1 = 0.97
ALIJALAL ET AL. [31]	RESTING EEG, CSP + ENTROPY	SVN/KNN	50	75-88%	-

PRESENT STUDY. TEST-SET = INTERNAL VALIDATION; HOLD-OUT = STRICTLY INDEPENDENT PARTICIPANT-LEVEL SET, NEVER SEEN DURING TRAINING.

modest absolute classification performance.

The seven most discriminative features were frontal alpha asymmetry (FAA), prefrontal alpha index (PAI), alpha/theta ratio, ERP latency, alpha/beta ratio, pupillary half recovery time, and delta peak frequency. These features emerged in task-condition cells with the greatest group-level variance, most often during retrieval under high load (set sizes 3–5), on incorrect trials, and with longer delays (2000–4000 ms) (Table III). Notably, while behavioral performance converged between groups at the longer delays, the most discriminative physiological features were observed in these intervals, suggesting that late-stage, high-effort memory processes provide the clearest markers of group separation.

IV. DISCUSSION

The present pipeline successfully demonstrates that task-aligned, multimodal feature extraction combined with structured dimensionality reduction enables identification of which physiological features contribute most strongly to PD vs. HC classification and precisely when during the working memory task those differences emerge. This is a meaningful advance over prior approaches, and one that directly addresses a gap in the field: cognitive impairment in PD is heterogeneous, temporally dynamic, and distributed across neural and autonomic systems, yet most existing EEG and pupillometry studies rely on isolated features or temporally averaged signals, thereby collapsing the task structure critical for mechanistic interpretation. The novelty of the present framework lies in its standardized, task-aligned, multimodal feature-extraction strategy, which preserves the temporal structure of working memory processing while enabling integration of EEG and pupillometry data. By organizing features within predefined task-condition cells, the pipeline provides a principled means of asking not only whether groups differ, but also when during the task those differences emerge and which physiological systems drive them.

Within this context, interpretability is defined in terms of feature origin, task alignment, and post-hoc traceability rather than at the level of individual model parameters. The key

interpretable property of this pipeline is its ability to trace classification decisions back to specific physiological features

occurring at specific moments during the working memory task: by retaining the correspondence between each feature and its originating task-condition cell, the framework enables post-hoc identification of which signal domains, task stages, and cognitive contexts contributed most to group separation. This allows mechanistically grounded hypotheses about the underlying neural dynamics to be generated directly from the model output — a capability that purely predictive approaches, such as deep networks or ensemble methods, do not provide, as their decision processes cannot be meaningfully related to neurophysiology or cognitive theory. Although PCA mixes features and the RBF-SVM does not yield directly interpretable weights, interpretability is preserved by mapping the retained principal components back to their original contributing features using summed absolute loadings and by maintaining temporal alignment throughout the pipeline. This approach is consistent with established frameworks distinguishing post-hoc traceability from model-parameter transparency as a legitimate and clinically relevant form of interpretability [32], [33]. In the next section, we demonstrate this aspect of the pipeline by walking through the top-ranked features and offering provisional interpretations to illustrate the model's potential utility.

The pipeline's hold-out accuracy of 63% and F1 of 0.626 are best understood by comparison to the most methodologically parallel published work. Bera et al., [29] evaluated SVM classification on resting-state EEG reporting accuracy of 82–94% under subject-dependent evaluation and 68.09% under strict subject-independent evaluation — a result directly comparable to the present hold-out, which was obtained under equivalent participant-level separation. The fact that the present pipeline achieves a similar floor to that of the subject-independent resting-state SVM, despite targeting the substantially harder problem of task-evoked cognitive classification in a small, heterogeneous cohort, suggests the pipeline is performing at the level supported by this validation approach and signal type.

Broader comparisons across Table VI should be interpreted holistically rather than as direct benchmarks, for three reasons. First, studies reporting 88–98% accuracy [20, 21, 31] predominantly use resting-state EEG or voice signals, which capture well-established motor signatures of PD that are neurologically more robust and less variable than the task-evoked cognitive signals targeted here. Second, most published benchmarks rely on subject-dependent evaluation, which is known to overestimate performance on truly unseen participants. Third, sample sizes, preprocessing strategies, and validation designs vary substantially across studies, making numerical comparison uninformative without accounting for these factors [22]. Taken together, the present results are consistent with the range expected for strict participant-level validation on task-based cognitive EEG in a small clinical cohort, and the permutation-validated balanced accuracy of 0.701 confirms that the pipeline captures genuine group-level signal rather than noise.

A. Feature Level Insights

The RE analysis identified several features that differentiated PD from HC. As shown in Table II, task-condition cells are defined by task phase, set size, accuracy, and delay length, thereby linking features to specific cognitive contexts. The full behavioral results are reported in detail in the original study on which this ML model was trained and tested [23]. Still, several general findings are relevant for interpreting feature-level patterns. We emphasize that these findings demonstrate the pipeline’s potential rather than definitive biomarkers. Accordingly, feature-level interpretations are presented to illustrate how the pipeline can generate mechanistically grounded hypotheses when linked to task structure and prior behavioral findings, rather than to assert causal or diagnostic conclusions. Furthermore, while PCA components are linear combinations of features, interpretation in this study was not based on individual components themselves. Instead, we examined which original, physiologically defined features consistently contributed to retained components and survived RE, grounding interpretation in the original feature space rather than in abstract PCA axes. This strategy preserves interpretability at the feature level by maintaining direct correspondence to known neurophysiological measures, while still benefiting from dimensionality reduction to mitigate redundancy and noise. This interpretation strategy is supported by univariate statistics, which revealed small-to-moderate effect sizes across multiple features rather than a single dominant discriminator, reinforcing the need for a multivariate framework.

Among the top-ranked features, frontal asymmetry measures emerged at later task stages and under higher cognitive load. PAI emerged during retrieval at 2000 ms, while FAA emerged at 4000 ms, suggesting a shift from fine-grained prefrontal control to broader frontal recruitment under higher demand [34]. Behavioral results supported this interpretation: WM capacity was significantly lower in PD at 1000 ms but not at longer delays. Performance improved from 1000 to 2000 ms in both groups, with a plateau in HCs and a slight dip in PD. This

suggests that the shortest delay may not allow sufficient time for response preparation, whereas additional time enables compensatory processes. However, not all studies report this pattern. Fallon et al. [7] found that PD patients showed persistent deficits at longer delays, with recall precision degrading more quickly over time, regardless of current dopaminergic medication status (ON/OFF). This discrepancy likely reflects task differences; our CDT involves discrete change detection with variable set sizes, whereas Fallon’s continuous reproduction task was more sensitive to gradual memory decay. Together, these findings suggest PD-related WM impairments may arise from either early encoding/retrieval inefficiency or accelerated decay, depending on task demands and response format [35]. Importantly, the emergence of FAA and PAI at later task stages underscores the value of temporally aligned feature extraction, as these effects would be obscured in analyses that collapse across task phases or delay conditions.

Oscillatory ratios also differentiated groups. Alpha/theta and alpha/beta ratios showed moderate effect sizes (α/θ : Cohen’s $d = 0.779$, $p = 0.00279$; α/β : Cohen’s $d = 0.405$, $p = 0.1060$). Alpha/theta, emerging during encoding, and alpha/beta, observed during retrieval, were both elevated in PD, consistent with altered attentional control and inhibitory processes [35]. The improvement from 1000 to 2000 ms and subsequent stabilization parallels these oscillatory dynamics, suggesting they may support compensatory performance [36], [37].

To further validate the spectral-ratio features, we performed an additional spectral-parameterization analysis using FOOOF to separate periodic oscillatory components from the aperiodic $1/f$ background. This analysis showed group differences in alpha peak parameters, including peak height and center frequency, indicating that differences in the aperiodic slope did not solely drive the observed α/θ and α/β elevations. Additionally, FOOOF-derived features revealed large effect sizes for alpha peak height at both frontal and posterior sites (PostFz_AlphaHeight: Cohen’s $d = 0.84$, $p = 0.0021$; PostCz_AlphaHeight: Cohen’s $d = 0.84$, $p = 0.0018$; PostFz_AlphaHHeight: Cohen’s $d = 1.01$, $p = 0.0088$), demonstrating that group differences reflect genuine alterations in alpha oscillatory magnitude rather than aperiodic slope alone. Additionally, aperiodic offset differences (PrePz_offset: Cohen’s $d = 0.65$, $p = 0.0093$) indicate altered broadband excitation–inhibition balance in PD, further motivating the inclusion of spectral parameterization analyses to separate oscillatory from aperiodic contributions. Although individual alpha frequency was not explicitly normalized during feature extraction, shifts in alpha peak frequency are themselves well documented in PD. They may reflect disease-related alterations rather than methodological bias [35], [36], [37]. These findings suggest that spectral-ratio features capture meaningful oscillatory changes and highlight opportunities for future refinement through explicit IAF normalization.

Time-sensitive features added further evidence where signal latency, defined in our toolbox as the time of maximum positive deflection within a task-condition cell, was prolonged in PD

during retrieval at 2000 ms [18], [38] with a small-to-moderate effect (Cohen's $d = 0.213$, $p = 0.4122$). Although component-agnostic, this measure can be mapped to ERP epochs and electrode sites for comparison with canonical markers (e.g., P200, P300). To complement this, ERP-specific latencies (N100, P200, P300) were also extracted, directly targeting well-characterized cognitive stages.

In addition, pupillary half recovery time was reduced in PD at 4000 ms, indicating altered autonomic recovery dynamics during sustained task engagement. This feature exhibited a small effect size ($d = 0.117$, $p = 0.6552$), reflecting substantial inter-individual variability and reinforcing the idea that autonomic features act as complementary rather than primary discriminators. Prior work has shown that pupillary recovery timing reflects locus coeruleus–norepinephrine signaling and task-dependent regulation of arousal and autonomic control, rather than a single cognitive construct [14], [16]. Within this framework, reduced recovery time in PD may reflect compensatory autonomic adjustments engaged under prolonged task demands [14], [16].

Taken together, these findings suggest a tentative compensatory sequence in which PD participants show deficits under high temporal demands, followed by oscillatory, frontal, and autonomic adjustments that help sustain performance at longer delays. The pattern of Cohen's d values across features demonstrates that cognitive dysfunction in PD is characterized by multiple moderate-to-large, yet distributed, effects, rather than a single dominant marker, providing empirical justification for the multivariate, task-aligned approach adopted here. Importantly, this interpretation is intended to demonstrate how temporally aligned, multimodal feature extraction can contextualize ML-derived feature rankings within known cognitive and neurophysiological frameworks. Rather than relying on a single dominant marker, the results indicate that discriminative information in PD is distributed across modalities, task stages, and signal domains and becomes detectable only when integrated within a multivariate framework. While preliminary, these patterns illustrate the pipeline's utility for hypothesis generation and motivate further validation in larger, clinically stratified cohorts.

As with all aggregate cohort studies, clinical heterogeneity within the PD cohort, including disease duration, motor severity, cognitive status, and dopaminergic medication status, was not explicitly modeled in the present analyses. These factors are known to influence EEG and pupillometric measures and may contribute to within-group variability in the extracted features. Accordingly, the current results reflect aggregate group-level patterns. They should be interpreted as demonstrating the feasibility of a standardized analytic framework rather than isolating effects attributable to specific clinical subtypes, which will be an important focus of future studies with richer clinical phenotyping. Future work incorporating detailed clinical stratification will be critical for determining whether the observed compensatory patterns differ systematically across disease stages, cognitive profiles, or medication states.

B. Implications

This study demonstrates that task-aligned, multimodal EEG and pupillometric features successfully capture distributed cognitive signatures of PD that become detectable only when integrated across modalities, task stages, and signal domains. The pipeline's core strength lies in its ability to link classification decisions to specific physiological processes at specific moments in the working memory task — a property that enables mechanistically grounded hypothesis generation rather than treating the model as a black box. The univariate analyses confirm that no single feature dominates group separation; instead, discriminative information is distributed across multiple moderate-effect features spanning oscillatory, frontal asymmetry, and autonomic domains. The multivariate model integrates these distributed signals into statistically above-chance classification performance, with a permutation-validated balanced accuracy of 0.701 ($p = 0.0013$) on the fully held-out cohort, confirming that the learned representation captures genuine group-level cognitive signal. This is a particularly meaningful result given the small cohort size and the inherent difficulty of task-evoked cognitive classification, and it underscores the value of the multimodal, task-aligned integration strategy.

The pipeline's performance is best understood in the context of what it is designed to do. Rather than maximizing classification accuracy on a well-characterized resting-state signal, it is designed to reveal the temporal and physiological structure of cognitive dysfunction in PD during an active working memory task — a substantially more challenging and less explored problem. The results establish that this is feasible: multimodal task-aligned features, organized within predefined cognitive task stages and reduced to their most informative dimensions, provide statistically reliable group separation even in a small, clinically heterogeneous cohort under conservative participant-level hold-out evaluation. The feature-level insights generated by the pipeline — including the emergence of frontal asymmetry, changes in oscillatory ratios, and differences in autonomic recovery at specific task stages illustrate the framework's capacity to yield targeted, physiologically grounded hypotheses — a capability that positions this pipeline as a practical tool for hypothesis-driven investigation of cognitive dysfunction in PD.

Accordingly, the framework's primary contribution is methodological and mechanistic: it provides a reproducible, standardized platform for examining how multimodal physiological signals relate to working memory dysfunction in PD, with temporal resolution that conventional analyses cannot achieve. With further validation in larger, more diverse cohorts, this approach has the potential to support both biomarker discovery and the development of task-based cognitive monitoring tools to track disease progression and treatment response in PD.

Taken together, the present framework makes three contributions that distinguish it from existing EEG-based ML approaches for PD. First, it introduces a standardized, task-aligned feature-extraction strategy organized within cognitive

task-condition cells, enabling systematic comparison across encoding, maintenance, and retrieval stages that prior unimodal or temporally averaged pipelines cannot achieve. Second, it provides post hoc feature traceability, the ability to identify which physiological features contributed most to classification and precisely when during the working memory task those contributions occurred — thereby linking ML outputs directly to neurophysiologically interpretable processes rather than treating the classifier as a black box. Third, it demonstrates that distributed, temporally structured cognitive signatures of PD are detectable under conservative participant-level hold-out validation, establishing that statistically reliable group separation is achievable even in small, heterogeneous cohorts when multimodal features are integrated across task stages. These properties position the framework as a principled research platform for cognitive biomarker discovery in PD, with a clear path toward validation in larger cohorts and translation into longitudinal monitoring applications.

VI. LIMITATIONS AND FUTURE WORK

While the present pipeline addresses several gaps in existing EEG-based approaches, its generalizability is constrained by the relatively small, single-task cohort and the high dimensionality of the feature space. Although model selection and hyperparameter tuning were confined to the training set and final evaluation was performed on a strictly independent hold-out set, the modest performance on unseen participants indicates that overfitting remains a concern and highlights the need for larger and more diverse datasets. Importantly, the WM-CDT was chosen to probe specific WM processes rather than to maximize diagnostic sensitivity; thus, the extracted features should be interpreted as candidate markers of cognitive dysfunction rather than definitive diagnostic biomarkers.

Interpretation of spectral features is also limited by the frequent absence of clear oscillatory peaks outside the alpha band, which complicates the use of ratio-based metrics and peak counting, particularly for delta and theta activity. To address this, we incorporated spectral parameterization analyses to separate periodic oscillatory activity from the aperiodic 1/f component, improving interpretability and reducing the likelihood that group differences reflect broadband spectral slope rather than true oscillatory changes. Nonetheless, future work should more fully integrate parameterized spectral features and individual alpha frequency normalization to validate oscillatory findings further.

Additional limitations include clinical heterogeneity within the PD group, including variability in disease duration, cognitive status, and medication effects, which were not modeled at the individual level and may contribute to variability in features. Future studies should explicitly incorporate clinical covariates and examine relationships between model outputs and behavioral measures to link neural features to cognitive performance better. More broadly, future research should prioritize replication across larger, more heterogeneous cohorts, evaluate generalizability across cognitive tasks and recording paradigms, and assess longitudinal sensitivity to disease

progression. Collectively, these steps are necessary to determine which features are stable, task-general, and clinically meaningful, and to support the eventual translation of this framework from a mechanistic research tool to a clinically relevant platform for monitoring cognitive dysfunction in PD.

VII. CONCLUSION

This study presents a multimodal ML pipeline that integrates task-aligned EEG and pupillometry features using a standardized MATLAB-based extraction toolbox to distinguish individuals with PD from HC during a WM-CDT. By explicitly aligning neural and autonomic features to defined task stages and applying structured dimensionality reduction, the pipeline enables identification of temporally specific and physiologically grounded patterns that reflect distributed, compensatory cognitive processes in PD.

Although classification performance was modest and constrained by sample size, task specificity, and cohort heterogeneity, the results demonstrate that cognitive dysfunction in PD is characterized by temporally structured effects that only become detectable when integrated across modalities and task phases. Accordingly, the primary contribution of this work is not a deployable diagnostic classifier, but a principled framework for systematically examining how multimodal physiological signals relate to cognitive function in PD.

Together, these findings establish a foundation for future studies aimed at validating feature stability, expanding to larger and more diverse cohorts, and extending the approach across cognitive paradigms and longitudinal designs. With further development and validation, this framework may ultimately support research efforts to track cognitive changes in PD and to inform the design of clinically meaningful biomarkers.

ETHICAL APPROVAL STATEMENT

All procedures were approved by the North Dakota State University Institutional Review Board (Protocol IRB0004611, approved February 9th, 2023). Written informed consent was obtained from all participants before participation, and their privacy rights were protected throughout the study. All research procedures were conducted in accordance with the principles embodied in the Declaration of Helsinki and in compliance with applicable institutional and national guidelines governing research involving human participants.

ACKNOWLEDGEMENTS

We thank the Sanford Parkinson's Disease Research Laboratory for their assistance in recruiting participants with Parkinson's disease. Recruitment of healthy control participants and all study protocols for both groups were conducted in the Cognitive Neuroscience Laboratory at North Dakota State University.

FUNDING DECLARATION

Part of this work was supported by the Peltier Endowment

(Grant number not applicable), which funded the Sanford Parkinson's Disease Research Laboratory, where PD participant recruitment was supported. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DATA AVAILABILITY

De-identified feature-level data, behavioral metrics, and analysis scripts are available on GitHub [<https://github.com/XX/EEG-and-Pupillary-Feature-Extraction-Toolbox>] and from the corresponding author upon reasonable request. Access to raw EEG and pupillometry recordings requires additional approval and may be obtained by contacting co-authors Dr. Jeffrey Johnson or Dr. Samuel Birkholz. All data sharing will comply with institutional and ethical regulations for human subject research.

REFERENCES

- [1] M. J. Armstrong and M. S. Okun, "Diagnosis and Treatment of Parkinson Disease: A Review," *JAMA*, vol. 323, no. 6, pp. 548–560, Feb. 2020, doi: 10.1001/jama.2019.22360.
- [2] K. Gerner *et al.*, "Visual processing capacity and cognitive decline in Parkinson's disease," *BMJ Neurol Open*, vol. 7, no. 1, p. e000953, Mar. 2025, doi: 10.1136/bmjno-2024-000953.
- [3] A. Baddeley, "Working memory," *Science*, vol. 255, no. 5044, pp. 556–559, Jan. 1992, doi: 10.1126/science.1736359.
- [4] J. R. Fang, R. Pahwa, K. E. Lyons, T. Zantotto, and J. J. Sosnoff, "Examining the validity of smart glasses in measuring spatiotemporal parameters of gait among people with Parkinson's disease," *Gait & Posture*, vol. 113, pp. 139–144, Sep. 2024, doi: 10.1016/j.gaitpost.2024.06.001.
- [5] E.-Y. Lee, N. Cowan, E. K. Vogel, T. Rolan, F. Valle-Inclán, and S. A. Hackley, "Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information," *Brain*, vol. 133, no. 9, pp. 2677–2689, Sep. 2010, doi: 10.1093/brain/awq197.
- [6] J. Salmi, L. Ritakallio, D. Fellman, U. Ellfolk, J. O. Rinne, and M. Laine, "Disentangling the Role of Working Memory in Parkinson's Disease," *Front. Aging Neurosci.*, vol. 12, Sep. 2020, doi: 10.3389/fnagi.2020.572037.
- [7] S. J. Fallon, R. M. Mattiesing, K. Muhammed, S. Manohar, and M. Husain, "Fractionating the Neurocognitive Mechanisms Underlying Working Memory: Independent Effects of Dopamine and Parkinson's Disease," *Cereb Cortex*, vol. 27, no. 12, pp. 5727–5738, Dec. 2017, doi: 10.1093/cercor/bhx242.
- [8] R. A. Lawson *et al.*, "Cognitive decline and quality of life in incident Parkinson's disease: The role of attention," *Parkinsonism Relat Disord*, vol. 27, pp. 47–53, Jun. 2016, doi: 10.1016/j.parkreldis.2016.04.009.
- [9] I. Litvan *et al.*, "Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines," *Mov Disord*, vol. 27, no. 3, pp. 349–356, Mar. 2012, doi: 10.1002/mds.24893.
- [10] S. Grissmann, J. Faller, C. Scharinger, M. Spüler, and P. Gerjets, "Electroencephalography Based Analysis of Working Memory Load and Affective Valence in an N-back Task with Emotional Stimuli," *Front Hum Neurosci*, vol. 11, p. 616, Dec. 2017, doi: 10.3389/fnhum.2017.00616.
- [11] P. A. Beach, J. T. Huck, D. C. D. Zhu, and A. C. Bozoki, "Altered behavioral and autonomic pain responses in Alzheimer's disease are associated with dysfunctional affective, self-reflective and salience network resting-state connectivity," *Frontiers in Aging Neuroscience*, vol. 9, no. SEP, 2017, doi: 10.3389/fnagi.2017.00297.
- [12] D. Aarsland *et al.*, "Parkinson disease-associated cognitive impairment," *Nat Rev Dis Primers*, vol. 7, no. 1, p. 47, Jul. 2021, doi: 10.1038/s41572-021-00280-3.
- [13] M. Carceles-Cordon, D. Weintraub, and A. S. Chen-Plotkin, "Cognitive Heterogeneity in Parkinson's Disease: A Mechanistic View," *Neuron*, vol. 111, no. 10, pp. 1531–1546, May 2023, doi: 10.1016/j.neuron.2023.03.021.
- [14] M. Kahya, S. Moon, K. E. Lyons, R. Pahwa, A. E. Akinwuntan, and H. Devos, "Pupillary Response to Cognitive Demand in Parkinson's Disease: A Pilot Study," *Front Aging Neurosci*, vol. 10, p. 90, 2018, doi: 10.3389/fnagi.2018.00090.
- [15] G. Marsicano, C. Bertini, and L. Ronconi, "Decoding cognition in neurodevelopmental, psychiatric and neurological conditions with multivariate pattern analysis of EEG data," *Neuroscience & Biobehavioral Reviews*, vol. 164, p. 105795, Sep. 2024, doi: 10.1016/j.neubiorev.2024.105795.
- [16] N. Unsworth and M. K. Robison, "A locus coeruleus-norepinephrine account of individual differences in working memory capacity and attention control," *Psychon Bull Rev*, vol. 24, no. 4, pp. 1282–1311, Aug. 2017, doi: 10.3758/s13423-016-1220-5.
- [17] M. Borys *et al.*, "An analysis of eye-tracking and electroencephalography data for cognitive load measurement during arithmetic tasks," in *2017 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, Mar. 2017, pp. 287–292, doi: 10.1109/ATEE.2017.7905130.
- [18] T. Annanmaki, K. Palmu, K. Murros, and J. Partanen, "Altered N100-potential associates with working memory impairment in Parkinson's disease," *J Neural Transm (Vienna)*, vol. 124, no. 10, pp. 1197–1203, Oct. 2017, doi: 10.1007/s00702-017-1758-z.
- [19] J. A. Carmona Arroyave *et al.*, "Resting Functional Connectivity and Mild Cognitive Impairment in Parkinson's Disease. An Electroencephalogram Study," *Future Neurology*, vol. 14, no. 2, p. FNL18, May 2019, doi: 10.2217/fnl-2018-0048.
- [20] S. L. Oh *et al.*, "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Computing and Applications*, vol. 32, pp. 10927–10933, Aug. 2018, doi: 10.1007/s00521-018-3689-5.
- [21] C. Altham, H. Zhang, and E. Pereira, "Machine learning for the detection and diagnosis of cognitive impairment in Parkinson's Disease: A systematic review," *PLOS ONE*, vol. 19, no. 5, p. e0303644, May 2024, doi: 10.1371/journal.pone.0303644.
- [22] S. Shokrpour, A. MoghadamFarid, S. Bazzaz Abkenar, M. Haghi Kashani, M. Akbari, and M. Sarvzadeh, "Machine learning for Parkinson's disease: a comprehensive review of datasets, algorithms, and challenges," *npj Parkinsons Dis.*, vol. 11, no. 1, p. 187, Jul. 2025, doi: 10.1038/s41531-025-01025-9.
- [23] S. Birkholz, "Attention and Memory Function in Healthy Adults and Adults With Parkinson's Disease - ProQuest," North Dakota State University, 2024. Accessed: Jul. 11, 2025. [Online]. Available: <https://www.proquest.com/openview/3c6fa3478827442006c780ce2064f7d2/1?pq-origsite=gscholar&cbl=18750&diss=y>
- [24] H. Pashler, "Familiarity and visual change detection," *Perception & Psychophysics*, vol. 44, no. 4, pp. 369–378, Jul. 1988, doi: 10.3758/BF03210419.
- [25] M. Terry, S. Birkholz, J.-S. Lou, A. Huether, and J. Keller, "A Toolbox for Task-Based EEG and Pupillometry Feature Extraction for Cognitive Biomarker Discovery in Parkinson's Disease," *J. Neurosci. Methods*.
- [26] M. F. Anjum *et al.*, "Resting-state EEG measures cognitive impairment in Parkinson's disease," *npj Parkinsons Dis.*, vol. 10, no. 1, pp. 1–13, Jan. 2024, doi: 10.1038/s41531-023-00602-0.
- [27] H. Bokura, S. Yamaguchi, and S. Kobayashi, "Event-related potentials for response inhibition in Parkinson's disease," *Neuropsychologia*, vol. 43, no. 6, pp. 967–975, Jan. 2005, doi: 10.1016/j.neuropsychologia.2004.08.010.
- [28] R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," *BMC Bioinformatics*, vol. 14, p. 106, Mar. 2013, doi: 10.1186/1471-2105-14-106.
- [29] S. Bera, Z. W. Geem, Y.-I. Cho, and P. K. Singh, "A Comparative Study of Machine Learning and Deep Learning Models for Automatic Parkinson's Disease Detection from Electroencephalogram Signals," *Diagnostics (Basel)*, vol. 15, no. 6, p. 773, Mar. 2025, doi: 10.3390/diagnostics15060773.
- [30] S. Srinivasan, P. Ramadass, S. K. Mathivanan, K. Panneer Selvam, B. D. Shivahare, and M. A. Shah, "Detection of Parkinson disease using multiclass machine learning approach," *Sci Rep*, vol. 14, p. 13813, Jun. 2024, doi: 10.1038/s41598-024-64004-9.
- [31] M. Aljalal, S. A. Aldosari, K. AlSharabi, A. M. Abdurraqueeb, and F. A. Alturki, "Parkinson's Disease Detection from Resting-State EEG

Signals Using Common Spatial Pattern, Entropy, and Machine Learning Techniques,” *Diagnostics (Basel)*, vol. 12, no. 5, p. 1033, Apr. 2022, doi: 10.3390/diagnostics12051033.

- [32] Z. C. Lipton, “The mythos of model interpretability,” *Commun. ACM*, vol. 61, no. 10, pp. 36–43, Sep. 2018, doi: 10.1145/3233231.
- [33] A. J. Barnett *et al.*, “Improving Clinician Performance in Classifying EEG Patterns on the Ictal-Interictal Injury Continuum Using Interpretable Machine Learning,” *NEJM AI*, vol. 1, no. 6, Jun. 2024, doi: 10.1056/aioa2300331.
- [34] B. Ghaddar and J. Naoum-Sawaya, “High dimensional data classification and feature selection using support vector machines,” *European Journal of Operational Research*, vol. 265, no. 3, pp. 993–1004, Mar. 2018, doi: 10.1016/j.ejor.2017.08.040.
- [35] Z. Ye, M. Heldmann, L. Herrmann, N. Brüggemann, and T. F. Münte, “Altered alpha and theta oscillations correlate with sequential working memory in Parkinson’s disease,” *Brain Commun*, vol. 4, no. 3, p. fcac096, Apr. 2022, doi: 10.1093/braincomms/fcac096.
- [36] M. Zhu, A. Hajihosseini, T. Baumeister, S. Garg, S. Appel-Cresswell, and M. McKeown, “Altered EEG alpha and theta oscillations characterize apathy in Parkinson’s disease during incentivized movement,” *NeuroImage: Clinical*, vol. 23, p. 101922, Jul. 2019, doi: 10.1016/j.nicl.2019.101922.
- [37] W. Klimesch, “Alpha-band oscillations, attention, and controlled access to stored information,” *Trends in Cognitive Sciences*, vol. 16, no. 12, pp. 606–617, Dec. 2012, doi: 10.1016/j.tics.2012.10.007.
- [38] P. Demirayak, İ. Kiyi, Y. Ö. İşbitiren, and G. Yener, “Cognitive load associates prolonged P300 latency during target stimulus processing in individuals with mild cognitive impairment,” *Sci Rep*, vol. 13, no. 1, p. 15956, Sep. 2023, doi: 10.1038/s41598-023-43132-8.



Mercedes A. Terry, MSc holds a Bachelor of Science in Psychology with a minor in Criminal Justice from Ball State University, Muncie, Indiana, USA (2021), and a Master of Science in Biomedical Neuroscience from the University of Florida, Gainesville, Florida, USA (2022). She is currently a

Ph.D. student in Biomedical Engineering at the University of North Dakota, focusing on wearable technology and neuroengineering applications for Parkinson’s disease management.

Mercedes has experience as a Graduate Teaching Assistant at the University of North Dakota (UND), mentoring undergraduate students in biomedical engineering. She is also the CEO and Co-founder of MindMend Biotech LLC, where she leads research and development efforts for innovative wearable devices designed to improve healthcare for individuals with neurodegenerative diseases. Her research interests span neuroengineering, wearable sensors, machine learning for healthcare, and understanding the misfolding mechanisms of alpha-synuclein in Parkinson’s disease.

Mrs. Terry is an active member the Society of Women Engineers (SWE) participating in the ALWE (Advancing Women in Engineering) program. She has participated in programs such as the National Science Foundation’s I-Corps. She has authored Introduction to Neuroengineering (2024), multiple ASEE conference papers, and a recent IEEE publication, Efficacy of

Wearable Devices in Monitoring and Treating Parkinson’s Disease Symptoms: A Systematic Review and Meta Analyses (IEEE Access, 2025). Mrs. Terry has been recognized for her entrepreneurial innovation and her dedication to advancing research in biomedical engineering.



Samuel A. Birkholz, PhD received the B.A. degree in psychology from Concordia College, Moorhead, MN, USA, in 2017, and the Ph.D. degree in psychology from North Dakota State University, Fargo, ND, USA, in 2024. His major field of study is visual and cognitive neuroscience. He is currently a

Research Scientist with the Department of Neurology, Sanford Health, Fargo, ND, USA. His work has focused on attention and memory function in healthy adults and in adults with Parkinson’s disease. His research interests include working memory, sustained attention, and the cognitive neuroscience of neurodegenerative diseases. Dr. Birkholz is a Member of the Psychonomic Society and the Cognitive Neuroscience Society.



Jeffrey S. Johnson, PhD (Member, IEEE), received the Ph.D. degree in cognitive neuroscience from the University of Minnesota, Minneapolis, MN, USA, in 2006. He is currently a Professor of Psychology with North Dakota State University, Fargo, ND, USA. His research interests include

visual working memory, attention, and neurophysiological methods, including EEG and MEG. He has authored over 50 peer-reviewed publications. Dr. Johnson is a Member of the Cognitive Neuroscience Society.



Jau-Shin Lou, MD, PhD, MBA, received the M.D. degree from National Taiwan University, Taipei, Taiwan, the Ph.D. degree in neuroscience from Oregon Health & Science University, Portland, OR, USA, and the M.B.A. degree from the University of Oregon, Eugene, OR, USA. He is currently a

Professor of Neurology with the University of North Dakota School of Medicine and Health Sciences, Grand Forks, ND, USA, and a Neurologist with Sanford Health, Fargo, ND, USA. He has extensive experience in movement disorders and clinical research in Parkinson’s disease. His research interests include clinical neurology, neurorehabilitation, and translational neuroscience.



Asenath X. A. Huether, PhD, received the Ph.D. degree in psychology from the University of Minnesota, Minneapolis, MN, USA, in 2015. She is currently a Clinical Neuropsychologist with Sanford Health, Fargo, ND, USA, and a Faculty Member with Minnesota State Community and Technical College,

Moorhead, MN, USA. Her research interests include cognitive assessment, neuropsychology of movement disorders, and rehabilitation science.



Jessica Keller, BSc, received the B.Sc. degree in neuroscience from Minnesota State University Moorhead, Moorhead, MN, USA, in 2021. She is currently a Clinical Research Coordinator with Sanford Health, Fargo, ND, USA. Her research interests include Parkinson's disease, neurophysiology, and cognitive neuroscience.



Enrique Alvarez Vazquez, PhD (Senior Member, IEEE), received the Telecommunication Engineering degree from Universidad Politecnica de Madrid, Madrid, Spain, in 2006, and the B.Sc., M.Sc., and Ph.D. degrees in Electrical Engineering from North Dakota State University (NDSU), Fargo,

ND, USA, in 2008, 2012, and 2022, respectively. He also received Graduate Certificates in Software Engineering and College Teaching from NDSU. From 2008 to 2021, he served as a Computer Systems Engineer with the Center for Visual and Cognitive Neuroscience (a Center of Biomedical Research Excellence) in Fargo. Since 2022, he has been an Assistant Professor in the Biomedical Engineering Program at the University of North Dakota (UND).

Dr. Alvarez Vazquez's research spans digital and analog communications, embedded systems, computer programming, and systems design. His projects include computer-controlled vibrotactile and electrodermal stimulation devices for multisensory research, controlled-illumination enclosures for visual psychophysical experiments, and electromechanically actuated motion platforms for vestibular stimulation studies. His current interests lie in health informatics, digital health, and software engineering.

SUPPLEMENTARY MATERIAL

Table III

PARTICIPANT-LEVEL UNIVARIATE FEATURE DIFFERENCES BETWEEN PARKINSON'S DISEASE AND HEALTHY CONTROLS

FEATURE	COHEN'S D	P VALUE	FEATURE	COHEN'S D	P VALUE
MEAN AMPLITUDE	0.429	0.09934	TASK-EVOKED RELATED POTENTIAL (TEPR) LATENCY	-0.205	0.4268
STANDARD DEVIATION AMPLITUDE	0.473		TEPR PEAK AMPLITUDE	0.424	0.09461
PEAK TO PEAK	0.498	0.04853	HALF RECOVERY TIME	0.117	0.6552
ROOT MEAN SQUARE (RMS) AMPLITUDE	0.486	0.05961	AUC TEPR	0.429	0.09959
ZERO CROSSING	0.000	0.9996	ALPHA COHERENCE	0.188	0.4529
VARIANCE AMPLITUDE	0.370	0.1395	PHASE LOCKING VALUE (PLV)	0.275	0.2928
SKEWNESS	-0.119	0.6488	PREFRONTAL ALPHA INDEX (PAI)	-0.455	0.1098
KURTOSIS	0.173	0.4986	FRONTAL ALPHA ASYMMETRY (FAA)	0.058	0.8166
MAX PEAK	0.424	0.09461	PHASE AMPLITUDE COUPLING (PAC)	0.429	0.09934
MIN PEAK	0.543	0.03107	SHARPNESS	0.583	0.02317
HJORTH ACTIVITY	0.370	0.1395	SPIKE FREQUENCY	0.166	0.5212
HJORTH MOBILITY	0.202	0.4360	SPIKE AMPLITUDE	0.498	0.04853
HJORTH COMPLEXITY	-0.435	0.1049	SPIKE SLOW WAVE RATIO	-0.034	0.8934
HJORTH RATIO	-0.531	0.05504	SLOW WAVE ACTIVITY (SWA)	0.131	0.6200
DELTA POWER	0.020	0.9412	CROSS COUPLING	0.445	0.07061
THETA POWER	0.577	0.02217	GLOBAL FIELD POWER (GFP)	NAN	NAN
ALPHA POWER	0.734	0.00401	TSALLIS ENTROPY	0.304	0.2458
BETA POWER	0.391	0.1103	BETA COHERENCE	0.160	0.5229
GAMMA POWER	0.271	0.2632	PHASE LOCKING VALUE MEAN	0.275	0.2928
RELATIVE BAND POWER	0.230	0.3605	ERP ENTROPY	0.219	0.3618
SLOPE PSD	0.407	0.1195	HURST EXPONENT	0.003	0.9922
DELTA PEAK FREQUENCY	0.059	0.8188	PHASE DIFFERENCE	NAN	NAN
THETA PEAK FREQUENCY	1.032	0.00014	PHASE SYNCHRONY	0.275	0.2928
ALPHA PEAK FREQUENCY	-0.029	0.9119	PHASE RESETTING	0.433	0.09665
BETA PEAK FREQUENCY	-0.093	0.7162	TEAGER ENERGY OPERATOR (TEO) SIGNAL	0.296	0.2219
GAMMA PEAK FREQUENCY	-0.032	0.9021	MEAN TEO	0.296	0.2219
ALPHA BETA RATIO	0.405	0.1060	CUMULATIVE POWER	0.224	0.3925
THETA BETA RATIO	-0.299	0.2516	SPECTRAL EDGE FREQUENCY (95%),	0.157	0.5411
ALPHA THETA RATIO	0.779	0.00279	MEAN FREQUENCY	0.157	0.5450
GAMMA BETA RATIO	-0.030	0.9065	MEDIAN FREQUENCY	0.163	0.5316
THETA SYNC	0.275	0.2928	INTERQUARTILE RANGE (IQR) OF AMPLITUDE	0.466	0.06923
ALPHA DESYNCHRONY	0.275	0.2928	LINE LENGTH	0.423	0.08913
SPECTRAL CENTROID	0.157	0.5450	ENERGY	0.367	0.1504
SPECTRAL ENTROPY	0.152	0.5649	POWER SPECTRAL DENSITY (PSD) SKEWNESS	-0.050	0.8453
SPECTRAL FLATNESS	0.017	0.9468	PSD KURTOSIS	-0.098	0.7000
SPECTRAL ROLL OFF	0.147	0.5694	PETROSIAN FRACTAL DIMENSION	0.273	0.2967
SPECTRAL KURTOSIS	-0.098	0.7000	BURST PERCENTAGE	-0.040	0.8743
SPECTRAL VARIABILITY	0.213	0.4178	ROOT MEAN SQUARE OF SUCCESSIVE DIFFERENCE (RMSSD)	0.423	0.08888
EVENT-RELATED SPECTRAL PERTURBATION (ERSP)	-0.371	0.1671	SIGNAL POWER	0.367	0.1504
X MIN	0.543	0.03107	NOISE POWER	0.293	0.2261
X MAX	0.424	0.09461	SIGNAL-TO-NOISE RATIO (SNR) VALUE	-0.046	0.8597
ERP LATENCY	0.213	0.4122	AUTOCORRELATION LAG1	0.230	0.3787

AREA UNDER CURVE (AUC)	0.429	0.09959		N100 AMPLITUDE	0.394	0.1187
MEAN PUPIL DIAMETER	0.429	0.09934		N100 LATENCY	0.238	0.3664
STANDARD DEVIATION OF PUPIL DIAMETER	0.473	0.06392		N200 AMPLITUDE	0.128	0.6116
MIN PUPIL DIAMETER	0.543	0.03107		N200 LATENCY	-0.267	0.3045
MAX PUPIL DIAMETER	0.424	0.09461		P200 AMPLITUDE	0.547	0.03117
RANGE OF PUPIL DIAMETER	0.498	0.04853		P200 LATENCY	0.234	0.3814
VARIANCE OF PUPIL DIAMETER	0.370	0.1395		P300 AMPLITUDE	0.367	0.1479
PUPIL VELOCITY MEAN	0.583	0.02317		P300 LATENCY	-0.325	0.2207
PUPIL ACCELERATION MEAN	0.402	0.1059		N200 AMPLITUDE AVERAGE	0.441	0.09861
BLINK RATE	0.231	0.3785		P300 AMPLITUDE AVERAGE	0.212	0.4152
BLINK DURATIONS	0.007	0.9792		COMPONENT RATIO	0.200	0.4110
MEAN BLINK DURATION	0.007	0.9792		P300 VARIABILITY	0.332	0.1803