Classification and Characterization of Neural Activity During Bradykinetic and Freezing of Upper Limb Movements

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Mari Hanchi

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Dr. Sommer Amundsen-Huffmaster

May 2025

Acknowledgements

I would like to first and foremost thank my advisor, Dr, Sommer Amundsen-Huffmaster, for her unwavering support and guidance throughout this project. Her input was critical to furthering my understanding of Parkinson's disease as well as my development as a researcher. I would also like to thank Dr. Colum MacKinnon for providing valuable expertise and insight into the neurophysiological aspects of my project, and Dr. Matthew Johnson for his teachings in signal processing and data science which were the foundations of the analysis performed in this study. Many thanks to Dr. Amunden-Huffmaster, Dr. MacKinnon, and Dr. Johnson for serving as members of my committee.

Additionally, I would like to thank the Project 1 team from the Udall Center of Excellence: Dr. Jing Wang, Dr. Stephanie Alberico, Dr. Joshua Aman, Kelly Brown, Dr. Scott Cooper, Dr. David Darrow, Meghan Hill, Dr. Luke Johnson, Dr. Robert McGovern, Dr. Kevin O'Neill, Dr. Paul Tuite, Dr. Michael Park, and Dr. Jerrold Vitek. This project would not have been possible without their work to perform data collections and secure funding. I would like to especially acknowledge Dr. Wang, Dr. Alberico, and Dr. O'Neill for answering my many emails and meeting requests to help me understand the data and the data collection methodologies. Finally, I would also like to acknowledge Sofia Guedes and Anjali Singh, who aided in the initial classification of movement epochs.

Funding for this study was provided by the Udall Center of Excellence in Parkinson's Research (NIH grants: P50 NS09857 and P50 NS123109-03) and the Engdahl Family Foundation.

Abstract

Bradykinesia and freezing of the upper limb (FOUL) movements are debilitating symptoms of Parkinson's disease (PD) that occur during rapid alternating movements (RAMs). When attempting to perform high-rate RAMs, movements are typically characterized by bradykinesia (maintaining amplitude by slowing velocity) or episodes of FOUL (hesitations, freezing, or festination - hastening with small amplitudes). Currently, little is known about the neural signatures of basal ganglia oscillations associated with bradykinesia and FOUL behaviors.

In this study, ten people with PD with externalized deep brain stimulation (DBS) leads (directional 1-3-3-1 configuration) implanted in the globus pallidus internus (GPi) performed a RAMs task while OFF medication and OFF stimulation. Participants performed epochs (8-12 s) of forearm pronation-supination movements (±15 deg.) that were cued by an acoustic stimulus alternating between slow (1Hz) and fast (2Hz) rates. For each of the 226 fast epochs, power spectra of bipolar local field potentials (LFPs) were calculated for all vertically adjacent DBS contact pairs using bipolar ring pairs (containing dorsal-ventral information) or directional pairs (containing circumferential information). Each epoch was visually classified from the forearm angular position traces as normal (met criteria for movement amplitude and rate), bradykinetic (low velocity at target amplitude), or FOUL (hesitations, freezing, festination). Five machine learning models were trained to distinguish between the three movement classes, and SHapley Additive exPlanations (SHAP) analysis was used to identify the frequency bands that contributed most towards classification, for each of the LFP pair sets (ring versus directional pairs).

There was no significant difference found in model performance between features using bipolar ring pairs or directional pairs. Across both pair types, the non-linear Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) models significantly outperformed the linear Gaussian Naïve Bayes (GNB) and Linear Discriminant Analysis (LDA) models. For the RF, SVM and MLP models, frequencies within β-band (13-33Hz) and θ-band (4-8Hz) were consistently ranked among the most important features, however when comparing the power in individual frequency bands between movement

classes, there were no statistical differences shown between any of the top features. This suggests that there are non-linear relationships between frequency bands that distinguish between the movement classes, and that bradykinesia and FOUL may be distinct neural phenomena.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Background

Parkinson's disease (PD) is a neurodegenerative movement disorder affecting 8-18 out of 100,000 people in industrialized countries, which is caused by the loss of dopaminergic neurons in the substantia nigra (Lee and Gilbert, 2016). Dopamine is an important neurotransmitter for motor control due to its role in the basal ganglia pathways, where it facilitates movement through the direct pathway and inhibits movement via the indirect pathway (Macpherson et al., 2014). Without this dopaminergic input, the direct pathway activity is reduced while the indirect pathway is overactivated, leading to the manifestation of motor impairments (Ramesh and Arachchige, 2023). Deep brain stimulation (DBS) is an intervention for PD where an electrode is implanted in the subthalamic nucleus (STN) or the globus pallidus internus (GPi) and provides electrical stimulation to disrupt abnormal activity in the basal ganglia pathways (Lozano et al., 2019). Traditional DBS leads consist of four layers of circumferential ring electrodes however, the relatively recent development of directional DBS leads, where the two middle levels are segmented into three electrodes, has allowed for targeted stimulation while avoiding unwanted side effects associated with omnidirectional stimulation (Schnitzler et al., 2022). Recordings of signals within the basal ganglia can be performed using externalized DBS electrodes or the implanted Medtronic Percept DBS system. These signal recordings allow researchers to study and characterize neural activity associated with impairments in motor and non-motor symptoms in PD.

Motivation

Bradykinesia and freezing of the upper limb (FOUL) are debilitating symptoms of PD that occur during repetitive movements (Espay et al., 2009). When attempting to perform repetitive movements at high rates of >2Hz, movements are typically characterized by either bradykinesia (maintaining amplitude by slowing velocity) or episodes of FOUL (hesitations, freezing, or festination – hastening with small amplitudes) (Heremans et al., 2019; Stegemöller et al., 2009). However, these phenotypes of PD motor symptoms are commonly assessed together, such as in the Movement Disorders Society-Unified Parkinson's Disease

Rating Scale (MDS-UPDRS) which evaluates for speed, amplitude, hesitations, and halts concurrently (Goetz et al., 2008).

Much work has identified elevated β-band (approximately 10-30 Hz) activity in the basal ganglia as a pathophysiological marker for PD (Cassidy et al., 2002; Weinberger et al., 2006 Yu et at., 2021), with more recent work indicating that elevated activity within high frequency oscillations (HFOs, 100-300Hz) may also be a feature of PD (Johnson et al., 2021). During repetitive movement tasks such as finger tapping, β-band oscillations in the STN and cortex have been shown to be movement-modulated (i.e. correlated with movement), which decreases with bradykinesia and freezing (Androulidakis et al., 2008; Scholten et al., 2020), However, while DBS treatment has been shown to affect performance differently for people with PD who exhibited bradykinetic and FOUL behaviors (Stegemöller et al., 2013), little is known about the neural signatures of basal ganglia oscillations associated with these two movement phenotypes.

For this study, machine learning was used to determine whether a wide range of frequency bands could be used to distinguish between normal, bradykinetic, and FOUL movements during a rapid alternating movements (RAMs) task. Compared to traditional statistics, machine learning models excel when there are many more variables (features) than there are samples (Bzodok et al., 2018). With the emergence of techniques such as SHapley Additive exPlanations (SHAP), it has become possible to gain insight into how machine learning models, traditionally thought of as black boxes, make their predictions (Lundberg and Lee, 2017). As a result, it may be possible to identify new frequency bands associated with PD motor symptoms, while also serving as a dimensionality reduction technique that identifies the most important frequencies to be used in simpler statistical models.

Hypothesis

The main objective of this study was to use machine learning to characterize neural oscillatory activity in the GPi associated with motor performance during a RAMs task in people with PD. We hypothesized that these models could be used to identify GPi LFP power spectrum features that significantly distinguished

between normal, bradykinetic, and FOUL movements triggered at high RAMs rates. We additionally hypothesized that adding in the directional component of the DBS lead would lead to better performance of the models.

METHODS

DBS Implantation and Externalization

The surgical procedure to implant and externalize DBS leads was performed as detailed in Aman et al. (2020). Prior to the surgery, 3T and 7T MRI scans were obtained to aid in the planning of the trajectory for the lead implantation. An intraoperative microelectrode mapping technique (Vitek et al., 1998) was used to locate the posterolateral (motor) region of the GPi. Directional DBS leads with a "1-3-3-1" electrode configuration from Abbott (Infinity, 0.5mm vertical spacing), Boston Scientific (Vercise Cartesia, 0.5mm vertical spacing), or Medtronic (SenSight, 0.5mm vertical spacing) were then implanted in the motor region of the GPi. An extension wire was then connected to the DBS lead and tunneled to a subcutaneous pocket in the infraclavicular pocket, which was then connected to a second extension wire that was tunneled to the abdomen and subsequently externalized (Aman et al., 2020). The participants were allowed to rest at home for 4-8 days to recover from the surgery, after which they returned to perform the externalization recordings.

Study Participants

All participants were diagnosed with idiopathic PD and had consented to the externalization study prior to undergoing DBS surgery. All 12 participants had directional DBS leads (1-3-3-1 configuration, 0.5mm vertical spacing) implanted unilaterally or bilaterally in the GPi. They each performed the RAMs task while OFF meds and OFF stim. LFP data was recorded from the externalized DBS lead(s) via a trialing cable connected to an ATLAS Neurophysiological System (NeuraLynx, Inc.) at a sampling frequency of 24 kHz. Two participants were excluded from analysis since they exhibited severe tremor. The 10 remaining participants had received either Abbott (n=5), Boston Scientific (n=3), or Medtronic (n=2) DBS leads. There was only one subject who received bilateral lead implants, however LFP recordings from their left

side were excluded from analysis due to significant noise. All other participants received unilateral implants, yielding usable recordings from 5 right-GPi and 5 left-GPi.

Task

During the RAMs task, participants were asked to hold the handle of a robotic manipulandum (Entact Robotics, Inc.) with the arm contralateral to their DBS side and actively pronate and supinate their forearm by ±15° (Figure 1A). They were cued by acoustic stimuli at slow (1 Hz) and fast (2 Hz) rates. Real-time visual feedback (Figure 1B) of the manipulandum angular position (as a horizontal moving curser) was shown on a screen directly in front of the participants, allowing them to compare their movements to the target amplitude. A single trial consisted of four alternating epochs of slow and fast cueing rates lasting between 8-12 seconds, where the starting cueing rate and epoch duration were randomized to prevent participants from predicting the rate transitions. The angular position of the handle and the cueing tones were recorded at a sampling frequency of 24 kHz during the task, synchronized with the LFP data.

Figure 1: (A) Robot manipulandum used during the RAMs task and **(B)** the accompanying visual feedback (the dot moved horizontally to indicate the participant's angular position, and the vertical lines show the target ±15° movement amplitude).

Data analysis

Feature Generation

All analysis for feature generation was performed using customized scripts in MATLAB version R2024a. The raw DBS recordings were filtered between 0.5–500 Hz using a 2nd order Butterworth filter,

then downsampled to 3 kHz. The Butterworth filter was chosen due to its flat passband, ensuring that the spectral content within the frequency range of interest was preserved with minimal distortion. LFP activity was extracted by creating bipolar pairs via the subtraction of signals from vertically adjacent DBS contacts. The directional DBS leads used in this study all had four vertical levels of contacts in a "1-3-3-1" contact configuration, meaning that levels 1 and 4 had one ring contact while levels 2 and 3 with were segmented with three contacts. To compare the use of vertical versus directional pairs on the machine learning models, two sets of bipolar pairs were created. Ring pairs refers to bipolar pairs between vertically adjacent level contacts yielding three level pairs: 01-02*, 02*-03*, 03*-04 (* denotes contacts that were averaged across the three directional contacts into a single "ring"). Directional pairs refers to bipolar pairs between vertically aligned contacts within the segmented rings only, yielding three directional pairs: 2a-3a, 2b-3b, 2c-3c.

Since PD motor symptoms were primarily triggered during epochs with fast cued movement rates ("fast epochs"), only fast epochs were used for further data analysis. Power spectra of the LFP data during the first 8 seconds of each fast epoch were estimated using the multi-taper method in the Chronux toolbox, a method that reduces the high variance seen in conventional estimation methods by averaging the power spectra estimates computed using L orthogonal tapers (windowing functions) (Bokil et al., 2010; Prerau et al., 2017). For a signal of length N (in seconds) and a desired frequency resolution Δf (in Hz), the number of tapers L can be calculated from the time half-bandwidth product TW:

$$L = \lfloor 2TW - 1 \rfloor$$

where

$$TW = N\frac{\Delta f}{2}$$

Therefore, for 8 seconds of LFP data and a desired frequency resolution of 1 Hz, 7 tapers were used for power spectra estimation. For each spectrum, the average power contained in 51 frequency bands (11 within the 0.5-53Hz range, 40 within the 100-300Hz range) was then calculated and normalized to total power. The relative power in each frequency band at a specific bipolar pair (e.g., β_1 band power at pair 01-02*)

were used as the features for the machine learning models. Accounting for all frequency bands and bipolar pairs, there were 153 features each for the ring pair and directional pair data sets.

Movement Epoch Classification

Each fast epoch of movement was classified as normal (met target movement amplitude and velocity), bradykinetic (reduced movement velocity to maintain target amplitude), or FOUL (freezing, hesitations, or festination – hastening with small amplitudes) based on visual inspection of the angular position traces (Figure 2). The duration of PD motor symptoms was not considered, for example: an epoch where FOUL was exhibited during the first two seconds then recovered to normal movements was still

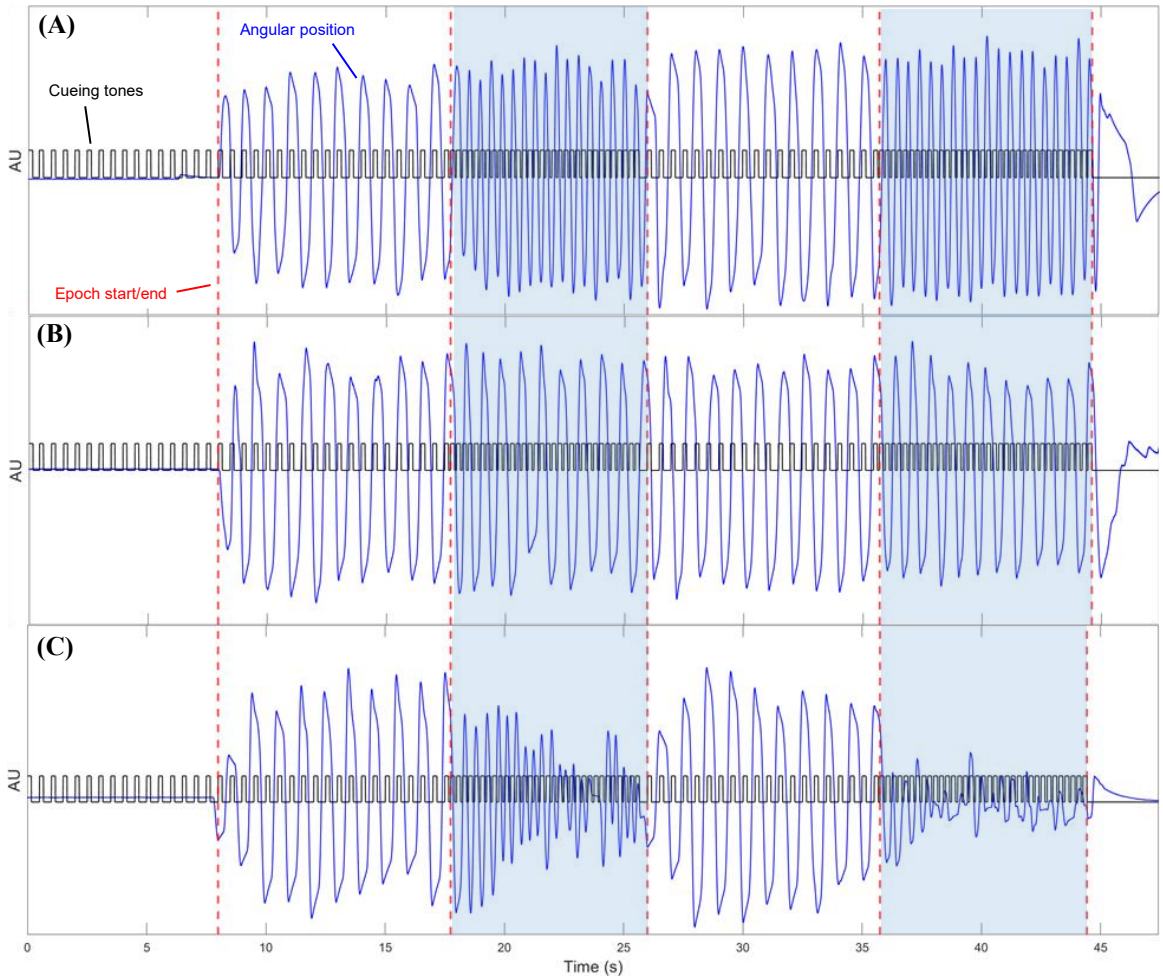

Figure 2: Example manipulandum angular position traces of RAMs trials where the fast epochs (shaded) exhibited **(A)** normal, **(B)** bradykinetic, or **(C)** FOUL movements. Example trials were chosen to have the same starting movement rate and epoch lengths but these did vary per trial.

classified as FOUL. Of the 226 total fast epochs, there were 85 epochs with normal movements (37.6%), 38 epochs with bradykinetic movements (16.8%), and 103 epochs with FOUL movements (45.6%).

Overview of Machine Learning Models

Five types of machine learning models were trained and tested in Python to classify the fast epochs based on power spectrum features: Random Forest (RF), Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and Multi-Layer Perceptron (MLP).

RF classification models consist of many decision trees (estimators), each of which are independently trained on a random subset of data and features (Breiman, 2001). After each tree predicts a class based on a series of decision nodes, the final class prediction is decided by a majority vote. RF models are a popular choice for classification, as they are often robust against high-dimensionality data and noise, and tend to have high accuracies without overfitting (Matsuki et al., 2016)).

GNB classifiers make predictions based on the assumption that the features are naïve (conditionally independent from each other) and have a Gaussian (normal) distribution (Rish, 2001). Using Bayes' Theorem, the probability of an input with *n* features being of a particular class can be calculated given the distributions of classes and features within the training dataset:

$$P(class|feature_1, \dots, feature_n) = \frac{P(class)\prod_{i=1}^{n} P(feature_n|Class)}{P(feature_1, \dots, feature_n)}$$

After calculating the probabilities for each class, the final predicted output will be the class that had the highest probability of occurring. While GNB models are relatively simple, they are sensitive to the assumptions of independence and normality.

SVM models are able to classify data by finding hyperplanes (decision boundaries) that best separate the classes linearly in the feature space (Cristianini, 2000). When linear separation is not possible in the feature space, the data can be mapped to a higher dimensionality space using a kernel function. The radial basis function (RBF) kernel, used in this study, is defined as

$$K(\mathbf{x_i}, \mathbf{x_j}) = \exp\left(-\gamma^2 \left\|\mathbf{x_i} - \mathbf{x_j}\right\|^2\right)$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are vectors with the features describing two input samples, γ is a parameter, and $\left\|\mathbf{x_i} - \mathbf{x_j}\right\|$ is the Euclidian distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ (Vert et al., 2004). Therefore, the RBF kernel is a measure of the similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$. This pairwise comparison between all training samples results in a $i \times j$ matrix describing the dataset, which can then be linearly separated.

LDA is a dimensionality reduction technique that can be used to classify data using Bayes' Theorem (Tharwat et al., 2017). Dimensionality reduction is first performed by projecting the data onto the hyperplanes that maximize the between-class scatter (i.e., the distance between the means of each class) and minimize the within-class scatter (i.e., the variance within each class). For m classes, the between-class scatter matrix is

$$\mathbf{S}_\mathrm{B} = \sum_{i=1}^{m} n_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$$

and the within-class scatter matrix is

$$\mathbf{S}_W = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\mathbf{x}_{i,j} - \boldsymbol{\mu}_i)(\mathbf{x}_{i,j} - \boldsymbol{\mu}_i)^T$$

where n_i is the number of samples in the i^{th} class, $\boldsymbol{\mu}_i$ is the mean of the i^{th} class, $\boldsymbol{\mu}$ is the overall mean, and $\mathbf{x}_{i,j}$ is the j^{th} sample in the i^{th} class. Then, the transformation matrix $\mathbf{W}$ that maximizes the ratio of the between-class scatter to within-class scatter, i.e., to solve

$$\underset{\mathbf{W}}{\operatorname{argmax}} \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}$$

is found. After the samples are projected into the lower dimensional space with $\mathbf{W}$, the distributions of the transformed data can be used to predict the class of a new sample using Bayes' Theorem, as described

earlier for the GNB models. Similarly, LDA relies on assumptions including classes with normal distributions and equal covariance matrices, and linearly separable classes.

The final classification model used was the MLP, a type of artificial neural network consisting of fully connected (dense) input, hidden, and output layers (Bishop, 1994). Each neuron in the input layer receives the data for each feature in the sample, and the output layer has as many neurons as there are classes (for multiclass classifiers). The hidden layers are found between the input and output layers, where there may be any number of hidden layers with any number of neurons. As the name suggests, artificial neural networks are modeled after biological neural networks, where neurons receive inputs and give outputs other neurons. In an MLP, each neuron receives a weighted (w_i) sum of inputs (x_i) and a bias term (w_0) from all neurons in the preceding layer, which is then transformed with an activation function ($g()$) before being passed along to as an input (z) to all neurons in the next layer:

$$z = g(a) = g\left(\sum_{i=1}^{d} w_i x_i + w_0\right)$$

The weights are analogous to the synaptic strength between biological neurons, while the bias term is analogous to the firing threshold. Activation functions determine how the neuron will respond to its outputs, and are often non-linear, allowing for artificial neural networks to learn more complex patterns. In this study, all hidden layer neurons had a Rectified Linear Unit (ReLu) activation function, while the neurons in the output layer had SoftMax activation functions to give the classification probabilities for each class. During the training process, the weights and bias terms are initialized and updated based on the error between the actual and predicted classes.

Model Parameters

The RF, GNB, SVM, and LDA classifiers were implemented using the scikit-learn library version 1.4.2, and the MLP was implemented using the Keras library version 3.8.0 (Pedregosa et al., 2011; Chollet, 2015). For the RF, GNB, SVM, and LDA models, their respective default hyperparameters were used (select

parameters are shown in Table 1). For the MLP model, hyperparameter tuning was performed using the grid search method to determine the combination of hidden layers (1, 2, or 3), neurons per hidden layer (25, 50, 75, or 100), and learning rate (0.001, 0.01, or 0.1) that yielded the highest average accuracy during cross-validation.

Table 1: Select hyperparameters for all models.

Model	Hyperparameters
RF	Default parameters: RandomForestClassifier(n_estimators=100, criterion='gini', bootstrap=True)
GNB	Default parameters: GaussianNB(priors=None, var_smoothing=1e-09)
SVM	Default parameters: SVC(kernel='rbf')
LDA	Default parameters: LinearDiscriminantAnalysis(solver='svd', n_components=2)

MLP	Model structure:
	`def build_model(n_layers, n_neurons, learning_rate, input_shape):`
	`  model = Sequential()`
	`  model.add(Input(shape=input_shape))`
	`  for layer in range(n_layers):`
	`  model.add(Dense(n_neurons, activation="relu"))`
	`  model.add(Dense(3, activation="softmax"))`
	`  optimizer = Adam(learning_rate=learning_rate)`
	`  model.compile(optimizer=optimizer, loss='categorical_crossentropy', metrics=['accuracy'])`
	`  return model`
	
	Tuned hyperparameters
	• Ring pair model: n_layers = 1, n_neurons = 25, learning_rate = 0.01
	• Directional pair model: n_layers = 1, n_neurons = 100, learning_rate = 0.01

Model Training and Evaluation

Stratified repeated k-fold cross-validation (5 folds, 2 repeats) was performed on 80% of the dataset (n_{normal}=66 (35.2%), $n_{bradykinetic}$=30 (16.0%), n_{FOUL}=91 (48.7%)). Stratified splits were utilized due to the imbalance of classes, ensuring that for each randomly selected fold, the proportion of normal, bradykinetic, and FOUL epochs was preserved. Then, final testing was performed on the remaining 20% of the dataset, simulating new and unseen data. All random processes (e.g. test/train/validation splitting, weight and bias

initialization) was seeded (random seed=42) for reproducibility. A one-sample t-test was performed between the validation and testing accuracies, as a significantly lower testing accuracy could indicate that a model was overfitting on the training dataset. The performances of each model were also compared against each other using a one-way ANOVA with the validation accuracies.

Feature Importance and Ranking

SHapley Additive exPlanations (SHAP) analysis was used to determine which features were important towards classification (Lundberg & Lee, 2017). SHAP analysis is based on Shapley values, a concept in cooperative game theory that determined the individual contribution (φ) of a player *j* within a total of *F* players based on payouts (*V*) of all coalitions (Ponce-Bobadilla et al., 2024):

$$\varphi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \big(V(S \cup \{j\}) - V(S)\big)$$

Essentially, the Shapley value represents the weighted average of a player's marginal contributions across all possible combinations of players. In the context of machine learning, the SHAP values describe the change in predictions made by two models across all subsets of features, one with the feature of interest and the other without.

For the best-performing models, SHAP values were calculated for each movement type and sorted by importance: the greater the mean magnitude of the SHAP values for all samples, the greater the importance. To determine the overall feature rankings for a model, the features were scored based on their class ranking (the top feature for a class scored 153 points, while the least important feature scored 1 point). Scores were summed across all three movement classes, from which the overall model rankings were determined.

The models were then trained (with cross-validation) and tested, with the top features sequentially added to the dataset to determine how many features were needed until model performance ceased to improve. This subset of features were compiled across each bipolar pair type, then scored and ranked as

described previously to find the overall top features for each pair type. Finally, a Kruskal-Wallis test was performed to compare the distributions of each feature across the movement classes.

RESULTS

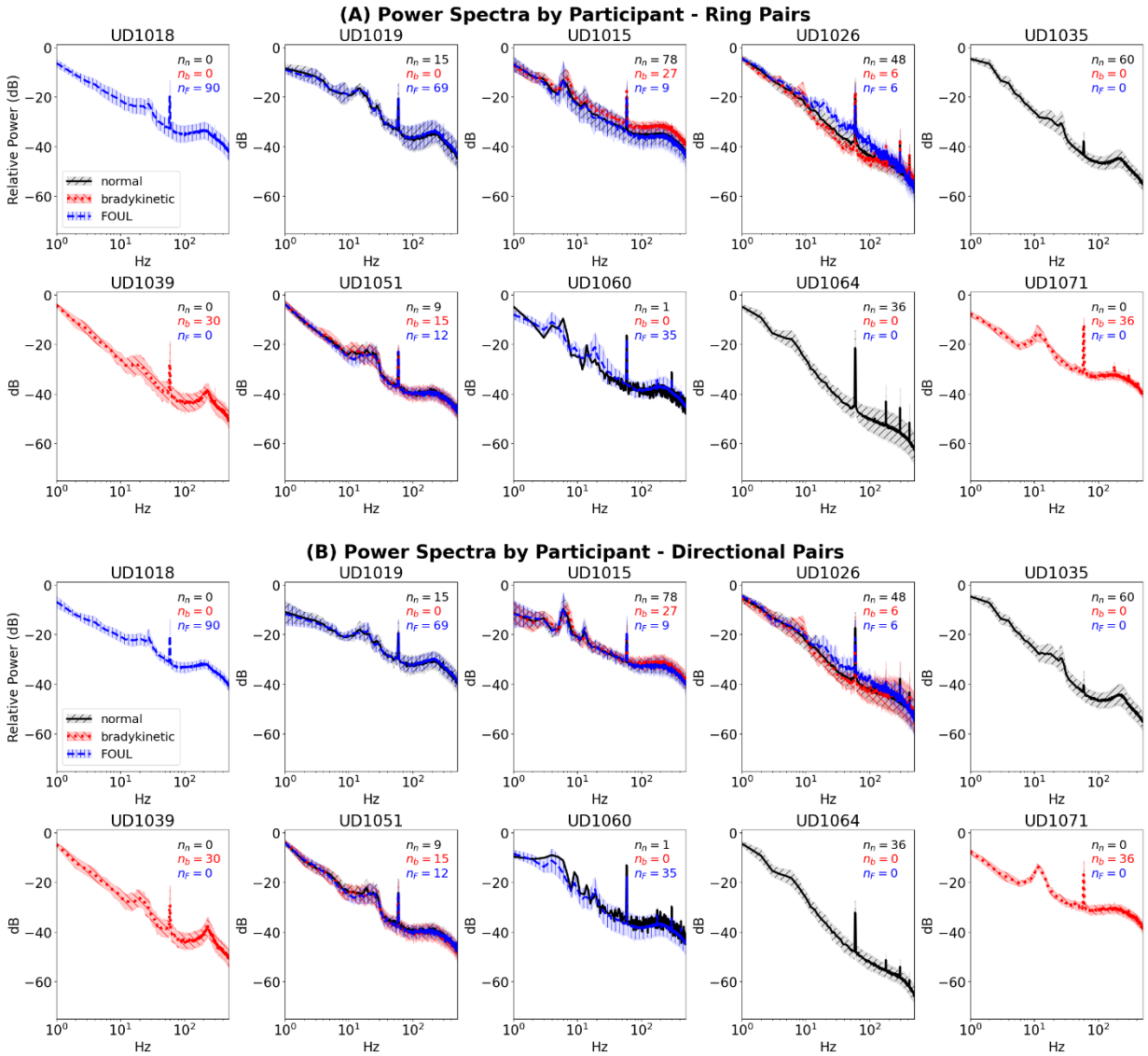

Figure 3: Mean normalized power spectra, with shaded standard deviation, for each movement type exhibited by each subject for **(A)** ring bipolar pairs and **(B)** directional bipolar pairs. The annotations n_n, n_b, and n_F indicate the total number of pair recordings during normal, bradykinetic, and FOUL epochs, respectively (there were 3 pair recordings per epoch).

As shown by the normalized power spectra for all subjects (Figure 3), most subjects exhibited elevated power within the β-band and HFO frequencies, supporting previous findings that peaks in power within these bands were pathophysiological features of PD (Yu et al., 2021; Johnson et al., 2021).

Model Evaluation

A one-way ANOVA on the accuracies during cross-validation ("validation accuracies") showed that across ring pairs and directional pairs, the RF, SVM, and MLP model performance did not significantly differ from each other, and they all significantly out-performed the GNB and LDA models ($p<0.05$) (Tables 2 and 3). To assess how well the models could generalize, classification accuracies during cross-validation versus testing were compared using a one-sample t-test. These t-tests for the RF, SVM, and MLP models showed no significant differences ($p>0.05$) between validation and testing accuracies when using either ring or directional pairs (Tables 2 and 3), indicating good model generalizability. On the other hand, the testing accuracies for GNB and LDA models were significantly lower than their validation accuracies for both pair types, indicating that the models overfit on the training dataset and had poor generalizability. The

Table 2: The validation (mean ± SD) and testing accuracies, and p-values for all models using ring pairs. * denotes $p < 0.05$.

Model	Validation Accuracy	Testing Accuracy	p-value
RF	0.83 ± 0.08	0.85	0.481
GNB	0.61 ± 0.07	0.51	0.002*
SVM	0.85 ± 0.08	0.79	0.064
LDA	0.65 ± 0.11	0.41	<0.001*
MLP	0.85 ± 0.09	0.82	0.295

Table 3: The validation (mean ± SD) and testing accuracies, and p-values for all models using directional pairs. * denotes $p < 0.05$.

Model	Validation Accuracy	Testing Accuracy	p-value
RF	0.82 ± 0.09	0.82	0.884
GNB	0.60 ± 0.07	0.54	0.027*
SVM	0.83 ± 0.08	0.85	0.615
LDA	0.69 ± 0.11	0.38	<0.001*
MLP	0.81 ± 0.09	0.79	0.693

validation accuracies between using ring pairs and directional pairs were also compared using two-sample t-tests for each model, however, none of the models showed significant differences ($p<0.05$) in performance.

Top Features

SHAP value analysis was performed on the best-performing models (RF, SVM, and MLP) to determine which features contributed most towards classification. In the SHAP summary plots (Figures 4 and 5), each dot represents power content data from a single epoch (the darker the dot, the higher the value), and the x-axis indicates the contribution of a given feature towards classification (positive SHAP values indicate positive contributions, i.e., towards a class; negative SHAP values indicate negative contributions, i.e., away from a class). For example, from Figure 4A, high β_3 power at pair 01-02* tended to contribute negatively from predicting an epoch as normal, positively from predicting as bradykinesia, and negatively from predicting as FOUL. The ranking of features in Figures 4 and 5 was determined based on the mean magnitude (absolute value) of all SHAP values, as both higher positive SHAP values and lower negative SHAP values indicate greater importance.

SHAP value analysis was performed on the top-performing RF, SVM, and MLP models to assess feature contributions and determine the top important features. For each model, the features were scored based on their ranking for each movement class (the lowest and highest possible scores were 1 point and 153 points for being ranked last and first out of 153 features, respectively). Then the scores were summed across all three classes and ranked to determine overall feature rankings. For example, from Figure 4A, the feature β_3 power at pair 01-02* scored 149 points for being ranked the 5th (out of 153) most important feature for normal classifications, 153 points each for being ranked 1st for bradykinesia and FOUL, giving an overall score of 455.

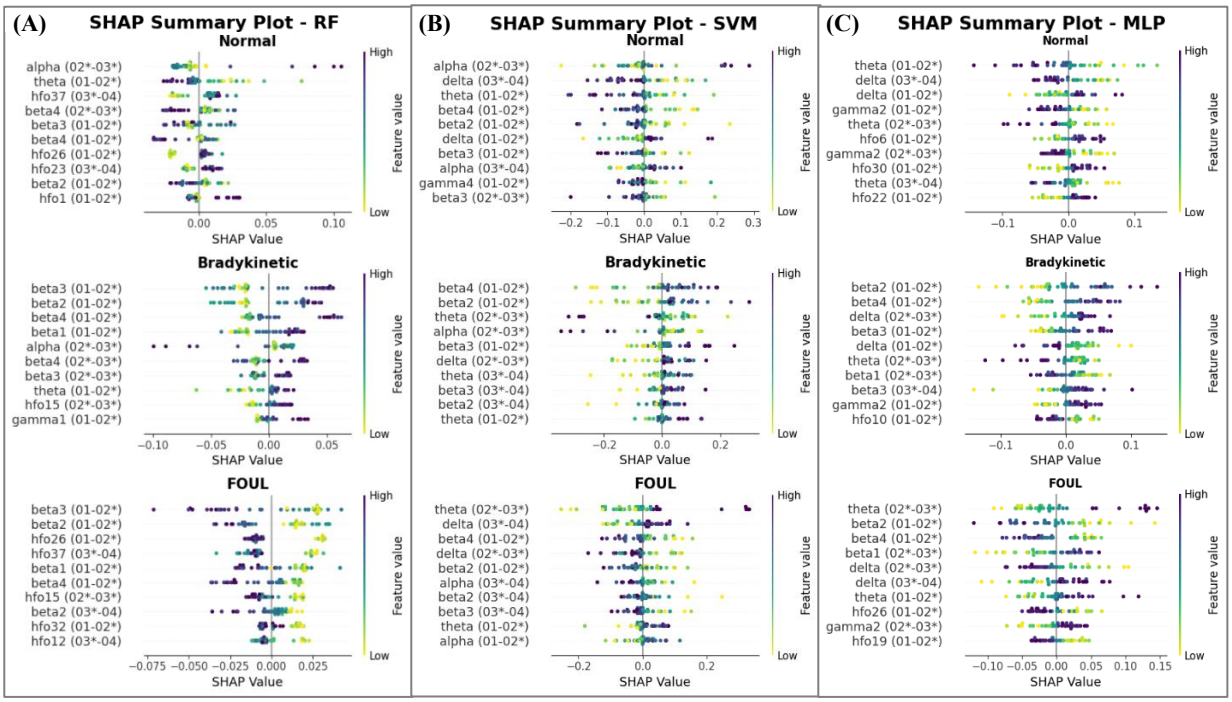

Figure 4: SHAP summary plots of the top 10 features (most important features are ordered highest) for each movement class for the **(A)** RF, **(B)** SVM, and **(C)** MLP models using ring pairs.

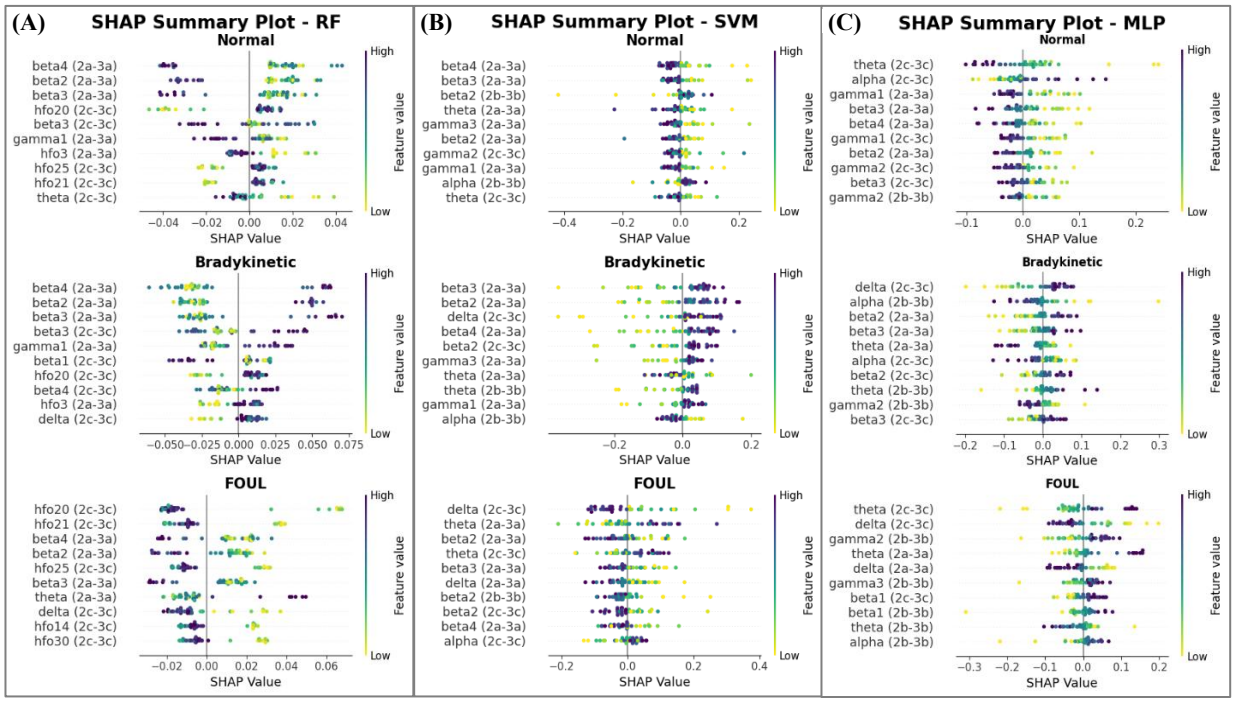

Figure 5: SHAP summary plots of the top 10 features for each movement class for the **(A)** RF, **(B)** SVM, and **(C)** MLP models using directional pairs.

When the RF, SVM, and MLP models were trained with only their respective top overall features, model performance plateaued once the top approximately 10 features were included (Figures 6 and 7), indicating that the remaining features contributed little or redundant information. When the top 10 features from each model were combined from all ring pair models, there were 18 unique features, of which 10 appeared for multiple models (Table 4) For the directional pairs, there were 20 unique features, of which 8 appeared for multiple models (Table 5).

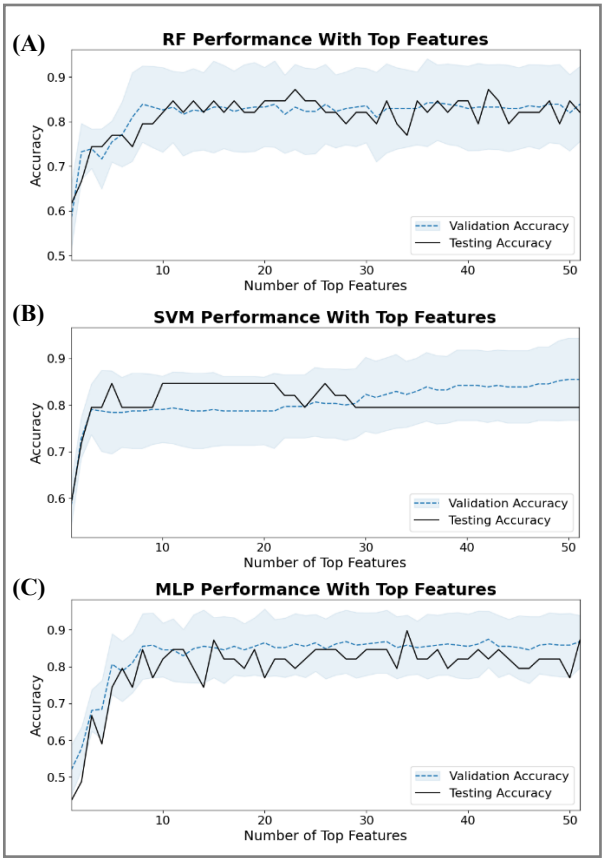

Figure 6: (A) RF, **(B)** SVM, and **(C)** MLP model performance with ring pairs when using only the top-ranked features.

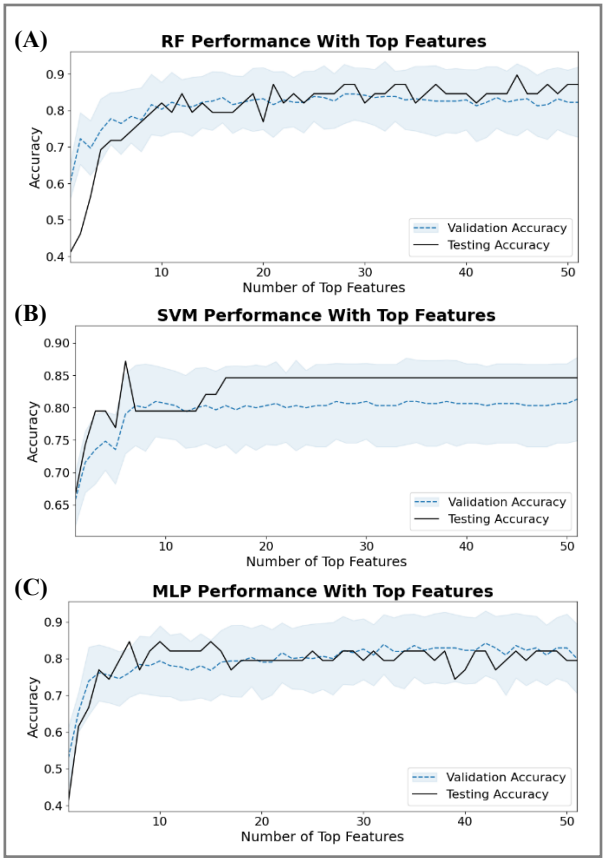

Figure 7: (A) RF, **(B)** SVM, and **(C)** MLP model performance with directional pairs when using only the top-ranked features.

Table 4: Overall feature rankings for ring pair data across the RF, SVM, and MLP models (features that only ranked within the top 10 most important features of one model are not included).

Feature	Model Rankings	Overall Rank
β_2 (01-02*)	2 (RF), 2 (SVM), 7 (MLP)	1
β_3 (01-02*)	1 (RF), 7 (SVM), 5 (MLP)	2
β_4 (01-02*)	3 (RF), 1 (SVM)	3
θ (01-02*)	5 (SVM), 2 (MLP)	4
θ (02*-03*)	6 (SVM), 1 (MLP)	4
δ (03*-04)	4 (SVM), 4 (MLP)	6
α (02*-03*)	10 (RF), 3 (SVM)	7
HFO_{26} (01-02*)	6 (RF), 8 (MLP)	8
δ (01-02*)	8 (SVM), 10 (MLP)	13
δ (02*-03*)	9 (SVM), 9 (MLP)	13

Table 5: Overall feature rankings for directional pair data across the RF, SVM, and MLP models (features that only ranked within the top 10 most important features of one model are not included).

Feature	Model Rankings	Overall Rank
β_2 (2a-3a)	2 (RF), 2 (SVM), 4 (MLP)	1
β_3 (2a-3a)	3 (RF), 1 (SVM)	2
β_4 (2a-3a)	1 (RF), 4 (SVM)	3
θ (2a-3a)	9 (RF), 3 (SVM), 6 (MLP)	4
θ (2c-3c)	7 (SVM), 1 (MLP)	5
β_2 (2c-3c)	6 (SVM), 9 (MLP)	8
γ_1 (2c-3c)	7 (RF), 10 (SVM)	13
β_1 (2c-3c)	10 (RF), 10 (MLP)	19

A Kruskal-Wallis test was performed on each top feature to determine whether there were differences in relative power between the three movement classes. Since the Kruskal-Wallis test assumed independent samples, the mean power content for each subject were used as samples. However, when considered individually, none of the top features for neither ring or directional pairs showed statistically significant differences between classes (Figure 8).

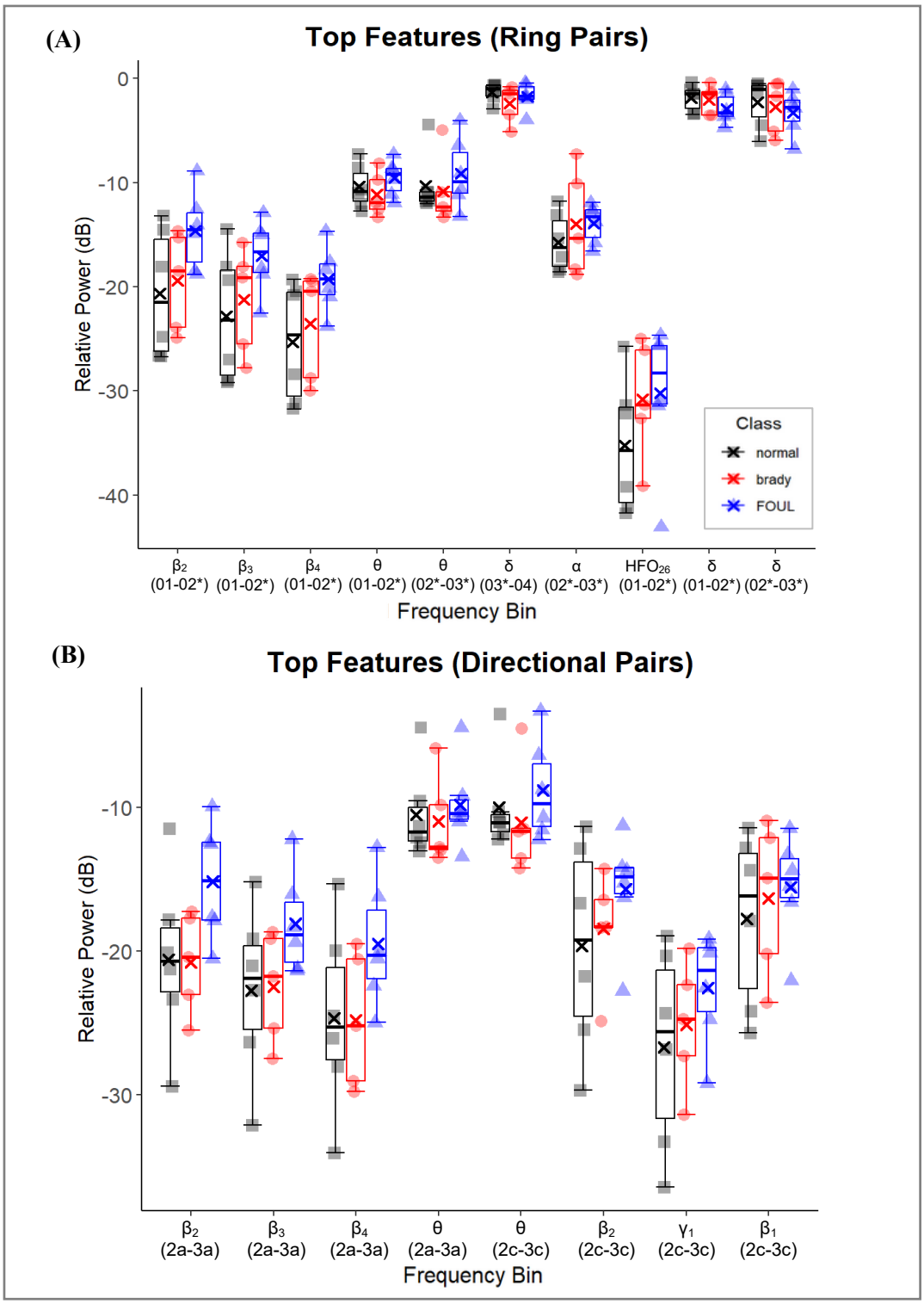

Figure 8: Distribution of power content (each point represents the mean value for a subject) across the top overall features (ranked highest to lowest, from left to right) for **(A)** ring pairs and **(B)** directional pairs. The boxplots show the median and quartiles of the data, and the cross shows the mean. See Table A1 for specific definitions of each frequency band.

DISCUSSION

In this study, SHAP value analysis revealed the spectral features of GPi LFP data that were most important for machine learning models that accurately predicted between normal, bradykinetic, and FOUL movements. Out of five machine learning models, the RF, SVM, and MLP models significantly outperformed the linear GNB and LDA classifiers. However, the use of ring or directional bipolar pairs in the features provided to the models had no significant effect on prediction accuracy. When the most important features across all models and pair types were compared, β-band and θ-band oscillations dominated, with additional top features within the δ, α, γ, and HFO frequency bands.

Model Evaluation

Regardless of bipolar pair type, the RF, SVM, and MLP models had similarly high accuracies that significantly outperformed the GNB and LDA models. This was likely because GNB and LDA models relied on the assumptions that the features were normally distributed, independent, and linearly separable in the feature space (Rish, 2001; Tharwat et al., 2017). However, the distribution of features was not normally distributed due to the variation between subjects (Figure 8), and many of the features were not independent (e.g., power in the β_1 frequencies would be correlated across pairs). While SVM also used linear hyperplanes to classify the data, the additional transformation with the non-linear RBF kernel allowed more complex relationships to be captured (Cristianini, 2000). The additional high performances by RF and MLP models serve as a strong indication that the characterization of normal, bradykinetic, and FOUL movements requires non-linear methods, as the linear classifications made by the GNB and LDA models were insufficient for accurately distinguishing between the movement classes.

For all models, bradykinesia consistently had the lowest F1 scores (see Tables B1 and B2), which is the harmonic mean of the precision (a ratio describing how many samples were correctly classified out of all samples predicted to a class) and recall (a ratio describing how many samples of a class could be correctly identified) (Hicks et al., 2022). This could be due to the class imbalance, where bradykinetic epochs only accounted for 16.0% (30 samples) of the training dataset. As such, the models may not have had enough

data to learn the neural signatures of bradykinesia, leading to confusion with normal or FOUL epochs. However, another possibility is that bradykinetic neural activity is more similar to normal or FOUL(for example, note how β-band power in Figure 8 are more similar between normal and bradykinetic movements), or that the participants were moving slowly in the unlikely event of a lack of motivation, rather than inability, to hit the target movement frequency.

For the RF, SVM, and MLP models, it is worth noting that there was little confusion between bradykinetic and FOUL epochs (i.e., very few bradykinetic epochs were misclassified as FOUL, and vice versa). When only considering the bradykinetic and FOUL classifications, the RF, SVM, and MLP models were able to distinguish between the two movement classes with >90% accuracy (see Appendix B). This distinction between both kinematic and neuronal features provides early evidence that bradykinesia and FOUL may need to be evaluated as separate phenomena during clinical assessments such as MDS-UPDRS. If these two phenotypes have distinct neural mechanisms, these findings could help explain previous studies that showed DBS stimulation had distinct effects on patients whose motor symptoms during a RAMs task manifested as slowing versus hastening (Stegemöller et al., 2013).

Bipolar Pairs

By comparing the model accuracies when using ring versus directional pair features, it was shown that there was no statistical difference in model performance between the two pair types. We hypothesized that the directional pairs would yield higher accuracies since previous studies have recorded greatest power and modulation in directional contact that faced posterolateral (sensorimotor) region of the GPi, while ring contacts could diminish spatially selective neural signatures (Aman et al., 2020). This lack of difference in prediction accuracy may have been due to the sheer quantity of features that were provided to the models. As shown in Figures 6 and 7, model performance stopped increasing once approximately 10 of the top features were included, meaning that the remaining features contained little or redundant information. Although the ring pairs features may have not individually contained as much spectral information as a directional pair feature, the patterns in power within each frequency band may have been more important

for the machine learning models rather than the specific dorsal-ventral or circumferential direction. There were also variations in DBS implantation depth and orientation, as well as variations in the size and shape of the sensorimotor GPi region for each participant (Aman et al., 2020; Patriat et al., 2018), which could have minimized the effect of contact pair as a feature. Further comparison with imaging data would be needed to confirm the consistency of the DBS lead implantation.

Important Frequencies

Across both ring and directional pairs, features from the β-band frequencies (spanning 13-33Hz) consistently ranked highest, consistent with findings showing changes in β-band modulation during PD motor symptoms (Scholten et al., 2020). However, it is still unclear to what degree any observed differences in β-band power could be attributed specifically to neural mechanisms associated with each movement type, or whether they were movement-related (e.g., it could be expected that the higher β-band power exhibited during FOUL versus bradykinesia could result from a near/complete arrest versus a slowness of movement, respectively) (Androulidakis et al., 2008; Neuville et al., 2021).

Features from the θ-band (4-8Hz) were also ranked highly for both pair types, although these oscillations are thought to be associated with tremor (Brown, 2003). There were additional features within the δ (0.5-4Hz), α (8-13Hz), and HFO_{26} (225-230Hz) frequency bands for the models using ring pairs, and within the γ_1 (33-38Hz) for directional pairs. A peak in power within the HFO (100-300Hz) in the GPi has been shown to be a feature of PD, and increased HFO modulation has been associated with faster movement during a reach task (Johnson et al., 2021). While this peak was observed in the power spectra for all participants (Figure 3), power in the HFO_{26} band was higher (although not statistically significant) during epochs of motor impairment than during normal epochs. If HFO modulation were to persist during repetitive movements, it could have been expected that the power, when computed across an entire epoch, would have been higher for normal epochs. However, the roles of δ, α, γ oscillations in the GPi are still unclear.

When the distributions of the top features were analyzed, there were no significant differences between classes for any features, likely due to subject-by-subject variability. Although it would have been ideal to compare the power content in each feature on a subject-by-subject basis, this was not possible since five of the participants only exhibited one movement type, and only three out of the ten participants exhibited all three movements. While the features did not individually show statistically significant differences, it was clear that there must be more complex, non-linear relationships between the top features that the RF, SVM, and MLP models leveraged to predict with such high accuracies. This could indicate that current adaptive DBS systems, which primarily monitor β-band activity (Feldmann et al., 2022), will require more complex algorithms than are currently available commercially to detect and tailor stimulation parameters for specific complex motor symptoms.

Limitations and Future Directions

One of the primary limitations of this study was sample size. Since there were only 226 samples (epochs) across 10 participants, the top features identified by the machine learning models may have been specific to the cohort in the study. Additionally, with more participants, it would have been possible to determine if the discrepancy between model accuracy and the lack of statistical significance within the top features was truly due to non-linear relationships or simply due to a large variation within a small sample size. It would also allow for the comparison of features across different participants who exhibited different phenotypes of PD motor symptoms (e.g., those who only had bradykinesia, or FOUL, or both).

Another limitation was that the method to initially assign classes was qualitative and did not take into account the duration and severity of PD motor symptoms. For example, epochs where slight hesitations were observed and had severe freezing were both classified as FOUL. If the behavioral and neural activity were mostly "normal" during the epoch with slight hesitations, training on its data could lead to confusion between normal and FOUL classes. To address this issue in the future, measures such as the density of jerk singularities (Fimbel et al., 2003, 2009) or a freezing index (Mancini et al., 2012) should be used to quantify

the RAMs movement, from which machine learning regressors (instead of classifiers) could be used to predict the degree of motor impairment.

Continuation of this study could compare the neural activity during different medication and DBS stimulation statuses. If PD motor symptoms persist even when ON meds and/or ON stim, this methodology could reveal which features or frequency bands are affected (i.e., which ones become more/less important). On the other hand, if the symptoms improve, changes in the distributions of the top features identified in this study could be analyzed to determine whether they are pathophysiological features of PD.

By analyzing the power spectra, one of the implicit assumptions in this study was that neural activity was stationary throughout an epoch. However, as indicated by numerous studies cited throughout that showed modulation of neural activity be modulated by movement, this assumption was likely insufficient. Further work could be done to analyze the changes in spectral features with time, such as through the use of spectrograms or coherence measures.

CONCLUSION

This study introduced a machine learning methodology to identify GPi neural oscillatory features that could characterize and distinguish between normal, bradykinetic, and FOUL movements during an upper limb RAMs task. SHAP analysis was performed on models that accurately predicted the three movement types, describing and ranking the contributions of each feature towards classification. Although it was hypothesized that the use of directional contact pairs would improve model performance, there was not significant difference between the accuracy of models trained on ring or directional pair features. For both pair types, the non-linear RF, SVM, and MLP models performed significantly better than the linear GNB and LDA models, indicating that there may be more complex, non-linear relationships between neural features that characterize each movement type. Of the top features, frequencies within β-band and θ-band were predominantly ranked highest, although there were no statistical differences in any of the top features

between movement classes. These findings could be used to improve closed-loop DBS systems, which currently only monitor β-band activity, to more accurately detect and treat different PD motor symptoms.

References

Aman, J. E., Johnson, L. A., Sanabria, D. E., Wang, J., Patriat, R., Hill, M., Marshall, E., MacKinnon, C. D., Cooper, S. E., Schrock, L. E., Park, M. C., Harel, N., & Vitek, J. L. (2020). Directional deep brain stimulation leads reveal spatially distinct oscillatory activity in the globus pallidus internus of Parkinson's disease patients. Neurobiology of disease, 139, 104819. https://doi.org/10.1016/j.nbd.2020.104819

Androulidakis, A. G., Brücke, C., Kempf, F., Kupsch, A., Aziz, T., Ashkan, K., Kühn, A. A., & Brown, P. (2008). Amplitude modulation of oscillatory activity in the subthalamic nucleus during movement. The European journal of neuroscience, 27(5), 1277–1284. https://doi.org/10.1111/j.1460-9568.2008.06085.x

Bishop, C. M. (1994). Neural networks and their applications. Review of Scientific Instruments, 65(6), 1803–1832. https://doi.org/10.1063/1.1144830

Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S., & Mitra, P. P. (2010). Chronux: A platform for analyzing neural signals. Journal of Neuroscience Methods, 192(1), 146–151. https://doi.org/10.1016/j.jneumeth.2010.06.020

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brown P. (2003). Oscillatory nature of human basal ganglia activity: relationship to the pathophysiology of Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society, 18(4), 357–363. https://doi.org/10.1002/mds.10358

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. Nature methods, 15(4), 233–234. https://doi.org/10.1038/nmeth.4642

Cassidy, M., Mazzone, P., Oliviero, A., Insola, A., Tonali, P., Di Lazzaro, V., & Brown, P. (2002). Movement-related changes in synchronization in the human basal ganglia. Brain : a journal of neurology, 125(Pt 6), 1235–1246. https://doi.org/10.1093/brain/awf135

Chollet, F. (2015). Keras. Retrieved from https://keras.io

Cristianini, N., & Shawe-Taylor, J. (2000). Support Vector Machines. In An Introduction to Support Vector Machines and Other Kernel-based Learning Methods (pp. 93–124). chapter, Cambridge: Cambridge University Press.

Espay, A. J., Beaton, D. E., Morgante, F., Gunraj, C. A., Lang, A. E., & Chen, R. (2009). Impairments of speed and amplitude of movement in Parkinson's disease: a pilot study. Movement disorders : official journal of the Movement Disorder Society, 24(7), 1001–1008. https://doi.org/10.1002/mds.22480

Feldmann, L. K., Lofredi, R., Neumann, W. J., Al-Fatly, B., Roediger, J., Bahners, B. H., Nikolov, P., Denison, T., Saryyeva, A., Krauss, J. K., Faust, K., Florin, E., Schnitzler, A., Schneider, G. H., & Kühn, A. A. (2022). Toward therapeutic electrophysiology: beta-band suppression as a biomarker in chronic local field potential recordings. NPJ Parkinson's disease, 8(1), 44. https://doi.org/10.1038/s41531-022-00301-2

Fimbel, E. J., Domingo, P. P., Lamoureux, D., & Beuter, A. (2005). Automatic detection of movement disorders using recordings of rapid alternating movements. Journal of neuroscience methods, 146(2), 183–190. https://doi.org/10.1016/j.jneumeth.2005.02.007

Fimbel, E., Dubarry, A. S., Philibert, M., & Beuter, A. (2003). Event identification in movement recordings by means of qualitative patterns. Neuroinformatics, 1(3), 239–257. https://doi.org/10.1385/NI:1:3:239

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang, A. E., Lees, A., Leurgans, S., LeWitt, P. A., Nyenhuis, D., Olanow, C. W., … Movement Disorder Society UPDRS Revision Task Force (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Movement disorders : official journal of the Movement Disorder Society, 23(15), 2129–2170. https://doi.org/10.1002/mds.22340

Heremans, E., Broeder, S., Nieuwboer, A., Bekkers, E. M., Ginis, P., Janssens, L., & Nackaerts, E. (2019). When motor control gets out of hand: Speeding up triggers freezing in the upper limb in Parkinson's disease. Parkinsonism & related disorders, 64, 163–168. https://doi.org/10.1016/j.parkreldis.2019.04.005

Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. Scientific reports, 12(1), 5979. https://doi.org/10.1038/s41598-022-09954-8

Johnson, L. A., Aman, J. E., Yu, Y., Escobar Sanabria, D., Wang, J., Hill, M., Dharnipragada, R., Patriat, R., Fiecas, M., Li, L., Schrock, L. E., Cooper, S. E., Johnson, M. D., Park, M. C., Harel, N., & Vitek, J. L. (2021). High-Frequency Oscillations in the Pallidum: A Pathophysiological Biomarker in Parkinson's Disease?. Movement disorders : official journal of the Movement Disorder Society, 36(6), 1332–1341. https://doi.org/10.1002/mds.28566

Lee, A., & Gilbert, R. M. (2016). Epidemiology of Parkinson Disease. Neurologic clinics, 34(4), 955–965. https://doi.org/10.1016/j.ncl.2016.06.012

Lozano, A. M., Lipsman, N., Bergman, H., Brown, P., Chabardes, S., Chang, J. W., Matthews, K., McIntyre, C. C., Schlaepfer, T. E., Schulder, M., Temel, Y., Volkmann, J., & Krauss, J. K. (2019). Deep brain stimulation: current challenges and future directions. Nature reviews. Neurology, 15(3), 148–160. https://doi.org/10.1038/s41582-018-0128-2

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 30 (pp. 4765–4774). Retrieved from http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Macpherson, T., Morita, M., & Hikida, T. (2014). Striatal direct and indirect pathways control decision-making behavior. Frontiers in Psychology, 5, 1301–1301. https://doi.org/10.3389/fpsyg.2014.01301

Mancini, M., Priest, K. C., Nutt, J. G., & Horak, F. B. (2012). Quantifying freezing of gait in Parkinson's disease during the instrumented timed up and go test. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2012, 1198–1201. https://doi.org/10.1109/EMBC.2012.6346151

Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The Random Forests statistical technique: An examination of its value for the study of reading. Scientific studies of reading : the official journal of the Society for the Scientific Study of Reading, 20(1), 20–33. https://doi.org/10.1080/10888438.2015.1107073

Neuville, R. S., Petrucci, M. N., Wilkins, K. B., Anderson, R. W., Hoffman, S. L., Parker, J. E., Velisar, A., & Bronte-Stewart, H. M. (2021). Differential Effects of Pathological Beta Burst Dynamics Between Parkinson's Disease Phenotypes Across Different Movements. Frontiers in neuroscience, 15, 733203. https://doi.org/10.3389/fnins.2021.733203

Patriat, R., Cooper, S. E., Duchin, Y., Niederer, J., Lenglet, C., Aman, J., Park, M. C., Vitek, J. L., & Harel, N. (2018). Individualized tractography-based parcellation of the globus pallidus pars interna using 7T MRI in movement disorder patients prior to DBS surgery. NeuroImage, 178, 198–209. https://doi.org/10.1016/j.neuroimage.2018.05.048

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://doi.org/10.5555/1953048.2078195

Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. Clin Transl Sci. 2024 Nov;17(11):e70056. doi: 10.1111/cts.70056. PMID: 39463176; PMCID: PMC11513550.

Prerau, M. J., Brown, R. E., Bianchi, M. T., Ellenbogen, J. M., & Purdon, P. L. (2017). Sleep Neurophysiological Dynamics Through the Lens of Multitaper Spectral Analysis. Physiology (Bethesda, Md.), 32(1), 60–92. https://doi.org/10.1152/physiol.00062.2015

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. Scientific reports, 14(1), 6086. https://doi.org/10.1038/s41598-024-56706-x

Ramesh, S., & Arachchige, A. S. P. M. (2023). Depletion of dopamine in Parkinson's disease and relevant therapeutic options: A review of the literature. AIMS neuroscience, 10(3), 200–231. https://doi.org/10.3934/Neuroscience.2023017

Rish, I. (2001). An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 3, 41–46. Seattle, USA.

Schnitzler, A., Mir, P., Brodsky, M. A., Verhagen, L., Groppa, S., Alvarez, R., Evans, A., Blazquez, M., Nagel, S., Pilitsis, J. G., Pötter-Nerger, M., Tse, W., Almeida, L., Tomycz, N., Jimenez-Shahed, J., Libionka, W., Carrillo, F., Hartmann, C. J., Groiss, S. J., Glaser, M., … PROGRESS Study Investigators (2022). Directional Deep Brain Stimulation for Parkinson's Disease: Results of an International Crossover Study With Randomized, Double-Blind Primary Endpoint. Neuromodulation : journal of the International Neuromodulation Society, 25(6), 817–828. https://doi.org/10.1111/ner.13407

Scholten, M., Schoellmann, A., Ramos-Murguialday, A., López-Larraz, E., Gharabaghi, A., & Weiss, D. (2020). Transitions between repetitive tapping and upper limb freezing show impaired movement-related beta band modulation. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology, 131(10), 2499–2507. https://doi.org/10.1016/j.clinph.2020.05.037

Stegemöller, E. L., Simuni, T., & MacKinnon, C. (2009). Effect of movement frequency on repetitive finger movements in patients with Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society, 24(8), 1162–1169. https://doi.org/10.1002/mds.22535

Stegemöller, E. L., Zadikoff, C., Rosenow, J. M., & Mackinnon, C. D. (2013). Deep brain stimulation improves movement amplitude but not hastening of repetitive finger movements. Neuroscience letters, 552, 135–139. https://doi.org/10.1016/j.neulet.2013.07.056

Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. Ai Communications, 30(2), 169–190. https://doi.org/10.3233/AIC-170729

Weinberger, M., Mahant, N., Hutchison, W. D., Lozano, A. M., Moro, E., Hodaie, M., Lang, A. E., & Dostrovsky, J. O. (2006). Beta oscillatory activity in the subthalamic nucleus and its relation to dopaminergic response in Parkinson's disease. Journal of neurophysiology, 96(6), 3248–3256. https://doi.org/10.1152/jn.00697.2006

Yu, Y., Escobar Sanabria, D., Wang, J., Hendrix, C. M., Zhang, J., Nebeck, S. D., Amundson, A. M., Busby, Z. B., Bauer, D. L., Johnson, M. D., Johnson, L. A., & Vitek, J. L. (2021). Parkinsonism Alters Beta Burst Dynamics across the Basal Ganglia-Motor Cortical Network. The Journal of neuroscience : the official journal of the Society for Neuroscience, 41(10), 2274–2286. https://doi.org/10.1523/JNEUROSCI.1591-20.2021

Vert, J. P., Tsuda, K., & Schölkopf, B. (2004). A Primer on Kernel Methods. In B Schölkopf and K Tsuda and JP Vert (Ed.), Kernel Methods in Computational Biology (pp. 35–70). Cambridge, MA, USA: MIT Press.

Vitek, J. L., Bakay, R. A., Hashimoto, T., Kaneoke, Y., Mewes, K., Zhang, J. Y., Rye, D., Starr, P., Baron, M., Turner, R., & DeLong, M. R. (1998). Microelectrode-guided pallidotomy: technical approach and its application in medically intractable Parkinson's disease. Journal of neurosurgery, 88(6), 1027–1043. https://doi.org/10.3171/jns.1998.88.6.1027

Appendix A: Definition of Frequency Bands

Table A1: Description of frequency bands (bolded rows indicate frequencies listed in Tables 4 and 5, and Figure 8).

Frequency Band Name	Range of Frequencies (Hz)
δ	**0.5-4**
θ	**4-8**
α	**8-13**
β_1	13-18
β_2	**18-23**
β_3	**23-38**
β_4	**28-33**
γ_1	**33-38**
γ_2	38-43
γ_3	43-48
γ_4	48-53
HFO_1	100-105
HFO_2	105-110
HFO_3	110-115
HFO_4	115-120
HFO_5	120-125
HFO_6	125-130
HFO_7	130-135
HFO_8	135-140
HFO_9	140-145
HFO_{10}	145-150
HFO_{11}	150-155
HFO_{12}	155-160
HFO_{13}	160-165
HFO_{14}	165-170
HFO_{15}	170-175

Frequency Band Name	Range of Frequencies (Hz)
HFO_{16}	175-180
HFO_{17}	180-185
HFO_{18}	185-190
HFO_{19}	190-195
HFO_{20}	195-200
HFO_{21}	200-205
HFO_{22}	205-210
HFO_{23}	210-215
HFO_{24}	215-220
HFO_{25}	220-225
HFO_{26}	**225-230**
HFO_{27}	230-235
HFO_{28}	235-240
HFO_{29}	240-245
HFO_{31}	245-250
HFO_3	250-255
HFO_{32}	255-260
HFO_{33}	260-265
HFO_{34}	265-270
HFO_{35}	270-275
HFO_{36}	275-280
HFO_{37}	280-285
HFO_{38}	285-290
HFO_{39}	290-295
HFO_{40}	295-300

Appendix B: Classification Reports and Confusion Matrices

The performance of classification models can be evaluated through several metrics such as overall accuracy, precision, recall, and F1-score based on the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) (Rainio et al., 2024). For multiclass models, the overall accuracy (Acc) is defined as

$$\text{Acc} = \frac{\text{number of accurate predictions}}{\text{total predictions made}}$$

and are displayed in Figures B1-B10. Precision (Pre), recall (Rec), and F1-score (shown in Tables B1 and B2) are calculated for each class individually:

$$\text{Pre}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \qquad \text{Rec}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \qquad \text{F1} = \frac{2 \cdot \text{Pre}_i \cdot \text{Rec}_i}{\text{Pre}_i + \text{Rec}_i}$$

If there are a total of N samples, distributed among m classes with n_i samples (termed "support") in each class, the precision, recall, and F1-scores can be averaged across a model using a macro average:

$$\text{Avg}_{\text{macro}} = \frac{\sum_{i=1}^{m} \text{metric}_i}{m}$$

or a weighted average:

$$\text{Avg}_{\text{weighted}} = \frac{\sum_{i=1}^{m} \text{metric}_i \cdot n_i}{N}$$

Table B1: Classification reports for the RF, GNB, SVM, LDA, and MLP models using ring pairs.

Model	Class	Precision	Recall	F1-score	Support
RF	Normal	0.94	0.79	0.86	19
	Bradykinetic	0.75	0.75	0.75	8
	FOUL	0.80	1.00	0.89	12
	Model Macro Avg	0.83	0.85	0.83	39
	Model Weighted Avg	0.86	0.85	0.84	39
GNB	Normal	0.73	0.58	0.65	19
	Bradykinetic	0.00	0.00	0.00	8
	FOUL	0.43	0.75	0.55	12
	Model Macro Avg	0.39	0.44	0.40	39
	Model Weighted Avg	0.49	0.51	0.48	39
SVM	Normal	0.93	0.68	0.79	19
	Bradykinetic	0.60	0.75	0.67	8
	FOUL	0.80	1.00	0.89	12
	Model Macro Avg	0.78	0.81	0.78	39
	Model Weighted Avg	0.82	0.79	0.79	39
LDA	Normal	0.64	0.37	0.47	19
	Bradykinetic	0.25	0.38	0.30	8
	FOUL	0.38	0.50	0.43	12
	Model Macro Avg	0.42	0.41	0.40	39
	Model Weighted Avg	0.48	0.41	0.42	39
MLP	Normal	0.93	0.74	0.82	19
	Bradykinetic	0.75	0.75	0.75	8
	FOUL	0.75	1.00	0.86	12
	Model Macro Avg	0.81	0.83	0.81	39
	Model Weighted Avg	0.84	0.82	0.82	39

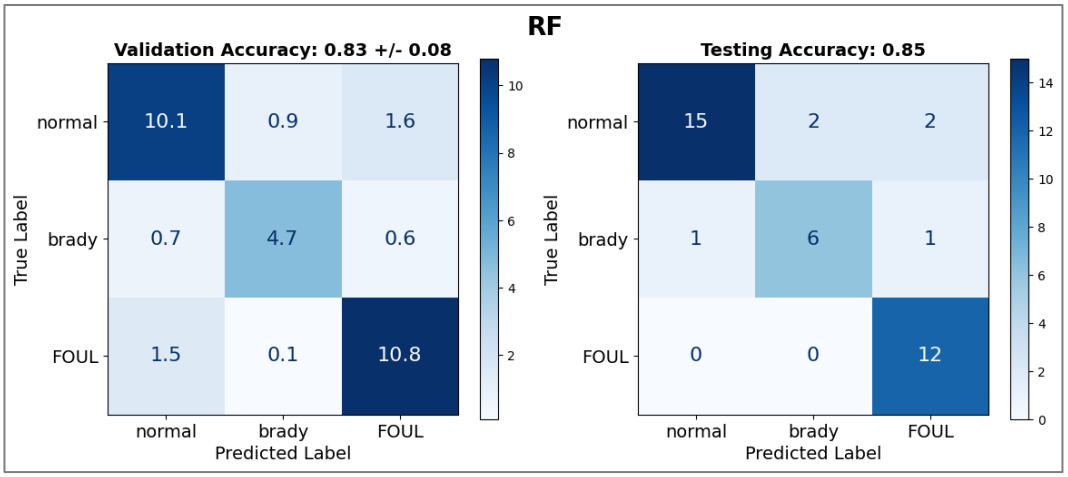

Figure B1: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing (right) for the RF model using ring pairs.

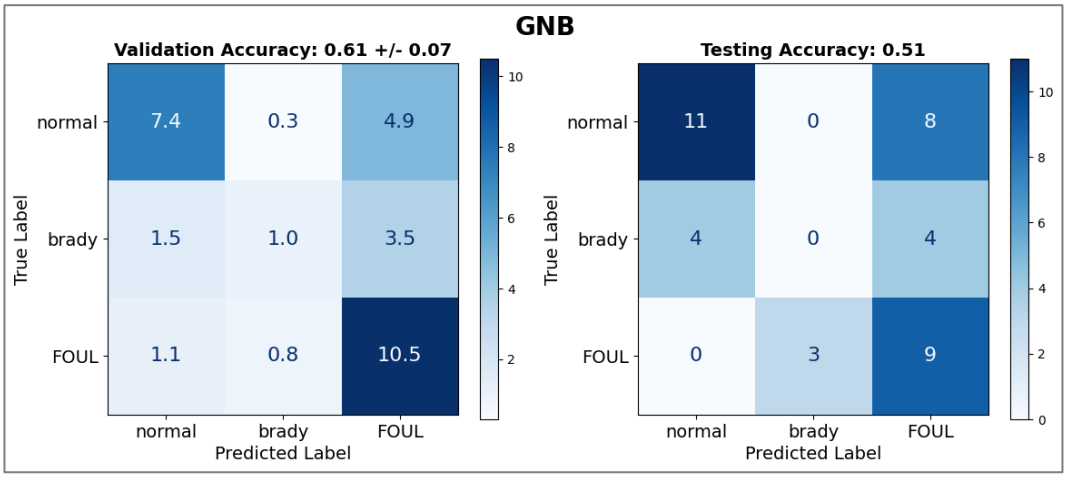

Figure B2: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing (right) for the GNB model using ring pairs.

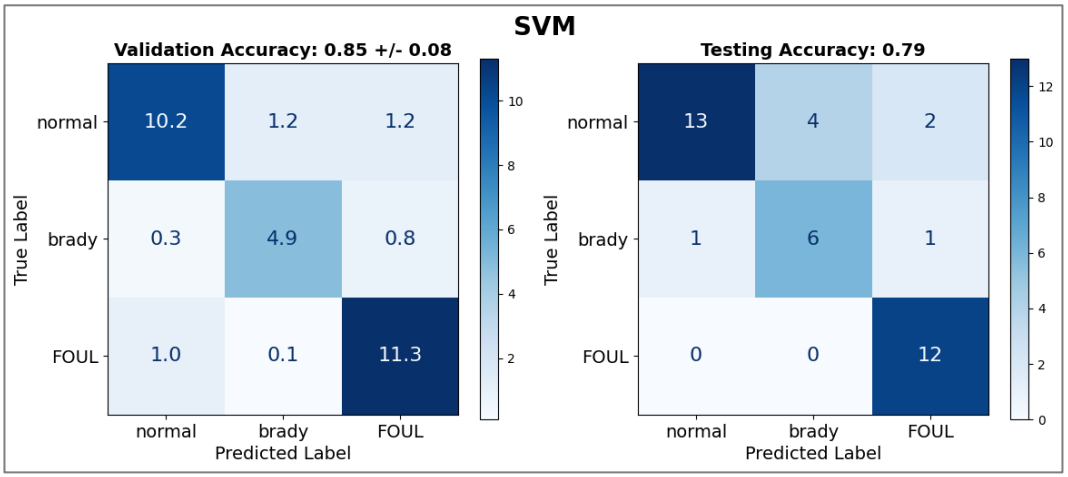

Figure B3: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing (right) for the SVM model using ring pairs.

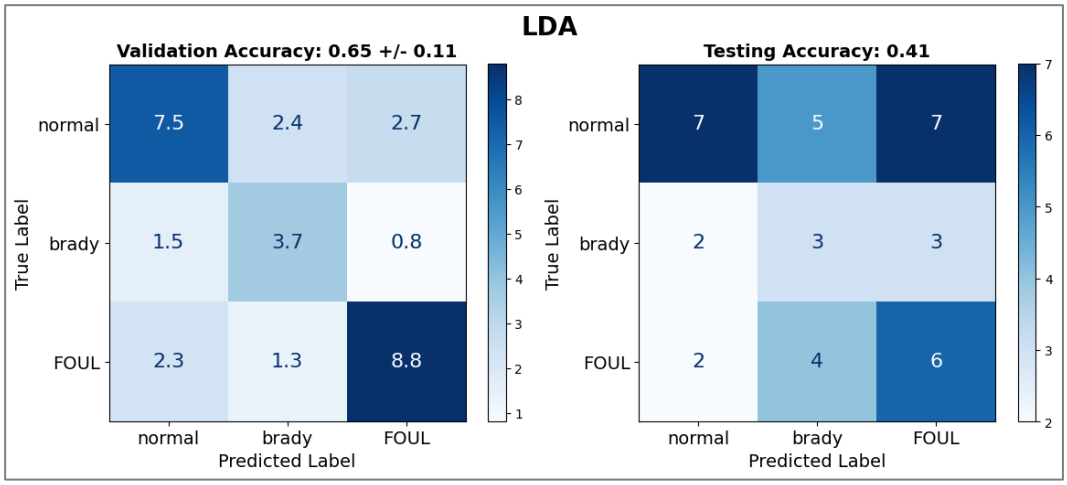

Figure B4: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing (right) for the LDA model using ring pairs.

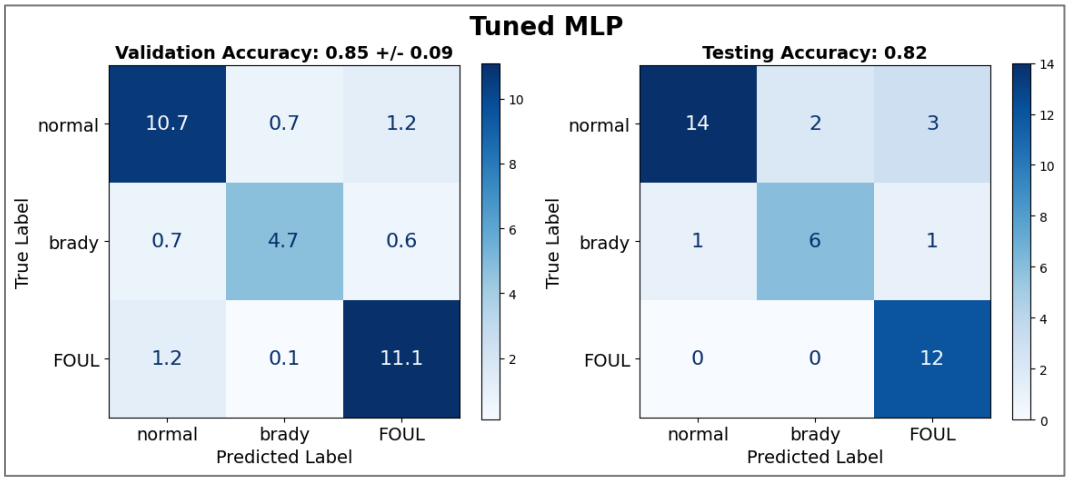

Figure B5: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing (right) for the MLP model (with tuned hyperparameters) using ring pairs.

Table B2: Classification reports for the RF, GNB, SVM, LDA, and MLP models using directional pairs

Model	Class	Precision	Recall	F1-score	Support
RF	Normal	0.84	0.84	0.84	19
	Bradykinetic	0.83	0.62	0.71	8
	FOUL	0.79	0.92	0.85	12
	Model Macro Avg	0.82	0.79	0.80	39
	Model Weighted Avg	0.82	0.82	0.82	39
GNB	Normal	0.65	0.58	0.61	19
	Bradykinetic	0.50	0.12	0.20	8
	FOUL	0.45	0.75	0.56	12
	Model Macro Avg	0.53	0.48	0.46	39
	Model Weighted Avg	0.56	0.54	0.51	39
SVM	Normal	0.85	0.89	0.87	19
	Bradykinetic	1.00	0.50	0.67	8
	FOUL	0.80	1.00	0.89	12
	Model Macro Avg	0.88	0.80	0.81	39
	Model Weighted Avg	0.87	0.85	0.83	39
LDA	Normal	0.50	0.53	0.51	19
	Bradykinetic	0.30	0.38	0.33	8
	FOUL	0.22	0.17	0.19	12
	Model Macro Avg	0.34	0.36	0.35	39
	Model Weighted Avg	0.37	0.38	0.38	39
MLP	Normal	0.79	0.79	0.79	19
	Bradykinetic	0.83	0.62	0.71	8
	FOUL	0.79	0.92	0.85	12
	Model Macro Avg	0.80	0.78	0.78	39
	Model Weighted Avg	0.80	0.79	0.79	39

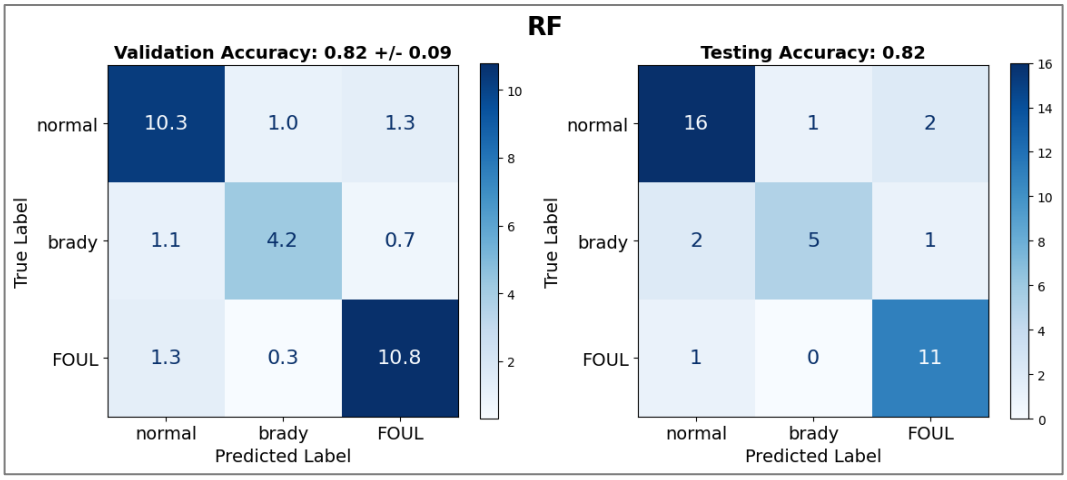

Figure B6: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing (right) for the RF model using directional pairs.

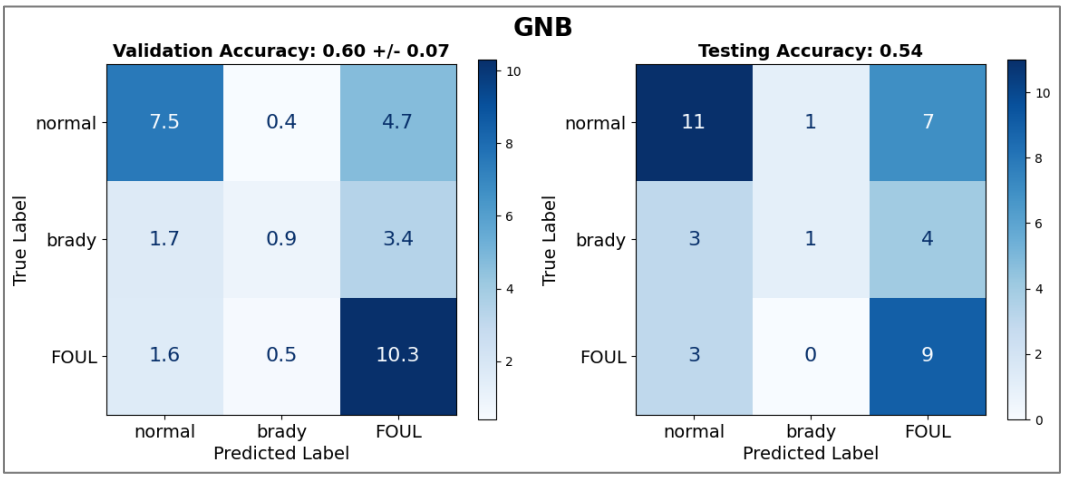

Figure B7: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing (right) for the GNB model using directional pairs.

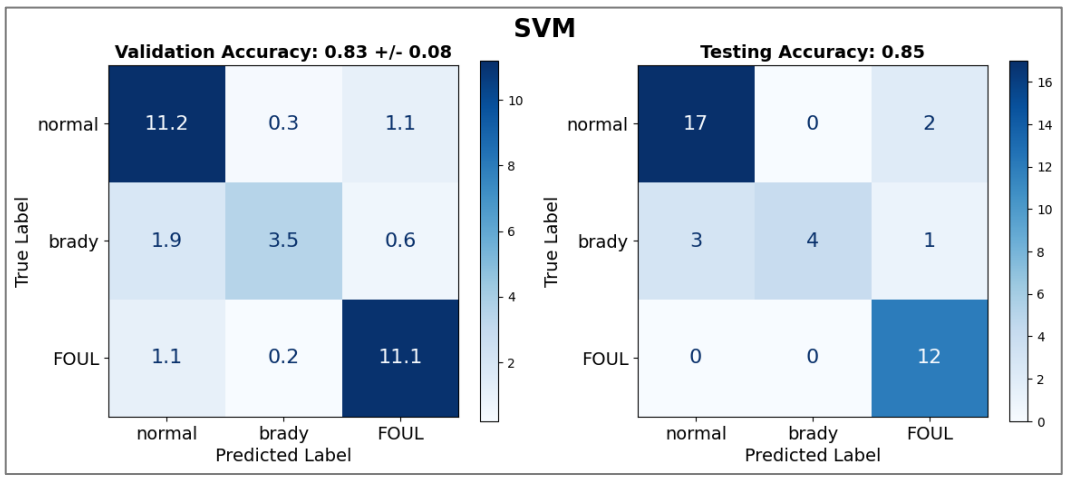

Figure B8: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing (right) for the SVM model using directional pairs.

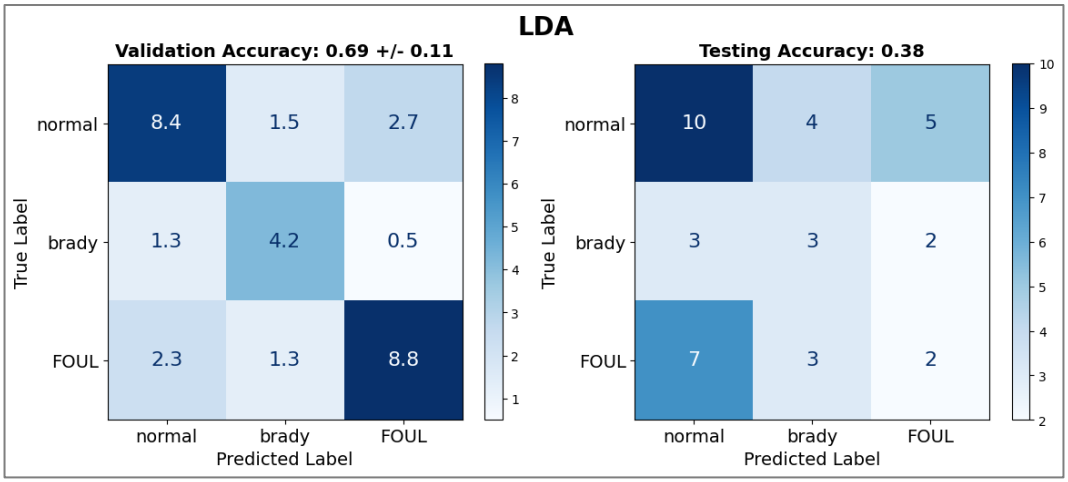

Figure B9: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing (right) for the LDA model using directional pairs.

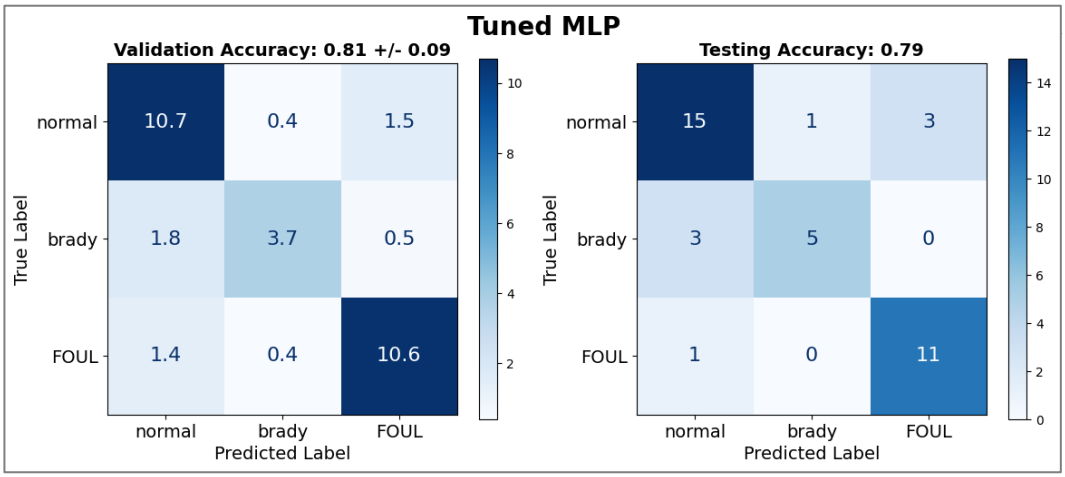

Figure B10: Average confusion matrix during cross-validation (left) and the confusion matrix during final testing for the MLP (with tuned hyperparameters) using directional pairs.