# Project Proposal: Machine Learning-Based Analysis of Vowel Phonation to Detect Neurological Diseases

**Evgenii Opryshko, Mahri Kadyrova, Hon Wa Ng, Jiuyang Fu**

## 1 Problem Definition

Neurological disorders like Parkinson's disease (PD) and Amyotrophic Lateral Sclerosis (ALS) often cause dysarthria (motor speech disorder) and dysphonia (voice disorder), affecting speech production. Early symptoms include subtle declines in speech intelligibility and rate, progressing to severe impairments where patients struggle to communicate clearly. Computational tools are essential for tracking speech decline and identifying critical changes to support timely therapeutic interventions. Traditional analysis methods rely on subjective evaluations by speech language pathologists and neurologists, which are inaccessible, time-consuming and variable. Thus, automated, objective methods are needed to detect, classify, and monitor these changes efficiently. This study will use vowel phonation data and clinical information to train models for disease and dysarthria severity classification.

### 1.1 Research Questions

- **Aim 1:** Evaluate various machine learning classifiers to distinguish between ALS, PD, and healthy individuals, and assess dysarthria severity using hand-crafted acoustic and clinical features from audio recordings. Given the limited number of study participants, these classifiers will serve as a baseline for disease and severity classification. We will assess both in-distribution and out-of-distribution performance across multiple datasets.
- **Aim 2:** Explore the use of neural network models for classifying disease and dysarthria severity directly from audio recordings of patients and healthy controls. Given the ability of neural networks to extract complex features, we hypothesize that pre-trained audio embeddings will enhance classification performance.
- **Aim 3:** Compare the outputs of Aim 1 and Aim 2 to identify the most effective models and the most predictive features for distinguishing healthy individuals from patients and for assessing dysarthria severity.
- **Aim 4:** Develop custom audio embeddings using vowel pronunciation data to further improve classification of disease and severity. These embeddings aim to complement and enhance the models from Aim 1 and Aim 2.

## 2 Prior Work

Automated speech analysis has emerged as a crucial tool for diagnosing and monitoring rapidly progressing neurodegenerative diseases. By leveraging machine learning and signal processing, it identifies subtle shifts in speech patterns, aiding in the detection of conditions like PD and ALS. Research has focused on vowel articulation, phonetic nuances, and a range of acoustic features, using models like SVMs, random forests, and deep learning. The creation of specialized datasets, such as VOC-ALS, has enhanced these techniques, supporting the development of remote, accessible diagnostics.

Recent studies have made significant contributions to this area. For example, Illner et al. [2023], Escobar-Grisales et al. [2024], Khatwad et al. [2024] focused on formant analysis and phonetic features for speech analysis in neurological conditions, particularly Parkinson's disease. Meanwhile, Tanglay et al. [2024] and Vishniakou and Xia [2023] employed machine learning models for detecting neurological speech deficits, emphasizing model comparison and performance. The VOC-ALS database Federico II University Hospital of Naples [2023] and telediagnosis approaches Tsanas et al. [2014] focused on dataset collection and analysis for diagnosing specific conditions such as PD and ALS.

In addition, Vachhani et al. [2017] and Appakaya and Sankar [2021] highlighted the effectiveness of deep autoencoders for feature extraction in speech data, enhancing the classification and recognition of speech patterns related to neurological conditions like PD and dysarthria.

## 3 Proposed Approach and Objectives

The objectives of this project include gathering data on ALS and PD from online sources and the University of Toronto's Speech Production Lab (SPL). After acquiring the data, we will pre-process and normalize the audio recordings to address differences in acquisition methods across the various datasets. We will utilize common signal processing methods to extract audio features, such as formant frequencies, formant slope, fundamental frequencies, jitter, shimmer, harmonic-to-noise ratio, and energy. We will experiment with existing audio embeddings, such as those from Whisper, and may also develop custom embeddings using techniques like autoencoders.

For **Aim 1**, machine learning classifiers like Support Vector Machines (SVM), Decision Trees, Random Forests, and Gradient Boosting Machines will be trained to evaluate the effectiveness of these features in classifying disease and dysarthria severity.

For **Aim 2**, neural network models based on the extracted embeddings will be trained. We will experiment with different ways of combining multiple recordings of the same person to increase classification accuracy, such as averaging and majority voting. We will also attempt training a Transformer that can attend to multiple audio embeddings at once.

**Aim 3** will involve comparing models from Aims 1 and 2 by analyzing their evaluation metrics such as accuracy, F1 score, precision, recall, confusion matrix, and ROC-AUC. Cross-linguistic generalization will also be tested by training on one dataset and evaluating on another.

Our ambitious goal for **Aim 4** is to train our own audio embeddings specific to vowel pronunciation using Convolutional Neural Networks (CNNs) and Transformers. We can utilize healthy control data from clinical studies and check if usage of custom audio embeddings will improve the performance of our ML models.

## 4 Expected Outcomes

- Accurate Machine Learning Classifiers: Development of interpretable machine learning classifiers that reliably distinguish between ALS, PD, and healthy individuals based on hand-crafted acoustic and clinical features. These classifiers will also quantify dysarthria severity (normal, mild, moderate, severe), offering a baseline for future studies and clinical assessments.
- Neural Network Models with Enhanced Performance: Implementation of neural networks leveraging pre-trained audio embeddings to classify disease types and dysarthria severity directly from raw audio recordings. These models are expected to achieve classification accuracy exceeding 80%, demonstrating the potential of deep learning in handling complex speech data.
- Comprehensive Model Comparison: Detailed comparison between machine learning classifiers and neural networks, identifying the most effective models, key acoustic features, and optimal embeddings. This outcome will highlight the strengths and limitations of traditional vs. deep learning approaches for neurodegenerative disease classification and severity assessment.
- Custom Audio Embeddings for Speech Analysis: Development of specialized audio embeddings tailored to vowel phonation data, designed to enhance the detection of dysphonia and dysarthria.

## References


S.B. Appakaya and R. Sankar. Novel unsupervised feature extraction protocol using autoencoders for connected speech: Application in parkinson's disease classification. *IEEE Xplore*, 2021. URL https://ieeexplore.ieee.org/abstract/document/9433683/.

D. Escobar-Grisales, C.D. Ríos-Urrego, and L.M. Mendoza. On the use of a foundation acoustic model to identify highly relevant phonetic information of parkinson's speech. In *Workshop on Advances in Speech Analysis*. Springer, 2024. doi: 10.1007/978-3-031-74595-9_7.

Italy Federico II University Hospital of Naples. Voice signals acquired in amyotrophic lateral sclerosis (voc-als), 2023. URL https://www.synapse.org/Synapse/files/.

A. Illner, A. Kleinschmidt, D. Weise, M. Paulig, L. Leyk, and A. Kirsch. Automated vowel articulation analysis in connected speech among progressive neurological diseases, dysarthria types, and dysarthria severities. *Journal of Speech, Language, and Hearing Research*, 2023. doi: 10.1044/2023_JSLHR-22-00526.

R. Khatwad, S. Tiwari, Y. Tripathi, and P. Sharma. *Parkinson's Disease Detection Using Voice and Speech—Systematic Literature Review*. Wiley, 2024. doi: 10.1002/9781394214211.ch3.

O. Tanglay, C. Cappelen-Smith, and M.W. Parsons. Enhancing stroke recognition: A comparative analysis of balance and eyes–face, arms, speech, time (be-fast) and face, arms, speech, time (fast). *Journal of Clinical Neurology*, 13(19):5912, 2024. doi: 10.3390/jcm13195912.

A. Tsanas, M.A. Little, P.E. McSharry, and J. Spielman. Novel speech signal processing methods for detection and monitoring of parkinson's disease. *PubMed*, 2014. URL https://pubmed.ncbi.nlm.nih.gov/25055311/.

B. Vachhani, C. Bhat, B. Das, and S.K. Kopparapu. Deep autoencoder based speech features for improved dysarthric speech recognition. In *Interspeech*, 2017. URL https://www.researchgate.net/publication/317559243_Deep_Autoencoder_Based_Speech_Features_for_Improved_Dysarthric_Speech_Recognition.

U. Vishniakou and Y. Xia. Technology of neurological disease recognition using gated recurrent unit neural network and internet of things. *Republic of Belarus Research Archive*, 2023. URL https://libeldoc.bsuir.by/handle/123456789/51292.