FSRM588 PROJECT PROPOSAL- Group 4

Parkinson’s Pattern Profiling: A Statistical and Visual Modeling Study

***Team members****: Ananya Chepuri(ac2857), Thomas Stranick(tcs121)*

## **1. Research Problem**

Parkinson’s Disease (PD) is a progressive neurodegenerative disorder affecting over 10 million people worldwide, including 1 million in the U.S. [1]. It primarily impairs motor function, causing tremors, rigidity, and movement difficulties. The diagnosis typically relies on clinical observation, patient history, blood tests, and imaging techniques to rule out other conditions [2]. One of the early observable symptoms of PD is hand tremor, which affects a patient's ability to perform fine motor tasks like handwriting. The HandPD dataset [3] contains spiral and meander tracing images from PD patients and healthy individuals, capturing these tremor-related motor deficits. In earlier studies, image processing techniques were used to extract features such as the Root Mean Squared Error (RMS) between the template and the traced drawings. These features were then used in classification models like Naïve Bayes, Optimum-Path Forest, and Support Vector Machines (SVMs) [4].

Our research problem is derived from the challenge that the diagnosis of Parkinson’s poses and inspired by the purpose of the creation of the HandPD dataset.

***How can we utilize modern machine learning techniques such as Convolutional Neural Networks (CNNs), and Vision Transformers (ViT), to identify whether a given patient has Parkinson’s based on traced templates that observe tremors?***

## **2. Data Preparation**

We will use two publicly available datasets:

* PaHaW Dataset [6]
* HandPD Dataset [2]

To enhance analysis and maximize data utility, we will combine these datasets into a unified corpus of spiral and meander handwriting samples from both Parkinson’s patients and healthy controls. This integration allows for better variability and generalizability in our statistical insights.

Preprocessing Steps:

* Remove low-quality or incomplete handwriting samples
* Segment spiral and meander shapes separately
* Convert all images to grayscale, resize them uniformly, and denoise
* Apply contour detection and skeletonization for clearer analysis
* Extract features such as stroke length, spiral radius variation, line smoothness, stroke density

Data Augmentation and Generation: Since transformer-based methods are data-hungry, we will also explore generating data using Generative Adversarial Networks (GANs) and cycleGANs. GANs and cycleGANs will be trained on spiral and meander samples to generate synthetic handwriting images that mimic Parkinsonian and healthy writing styles. This synthetic data will augment real samples and ensure more robust and diverse feature distributions. All extracted and generated features will be stored in a structured feature matrix for further analysis.

## **3. Methodology**

This project proposes a hybrid feature fusion methodology that strongly incorporates transformer models as a core element alongside interpretable handcrafted features. Vision Transformers (ViTs) will serve as both feature extractors and classification tools in our pipeline.

**3.1 Handcrafted Feature Extraction**

* Handcrafted features refer to manually calculated characteristics derived from handwriting images using classical computer vision techniques. These are interpretable and help provide meaningful statistical insights into motor function irregularities.
* Use OpenCV to extract visual and geometric features such as:
  + Stroke length (total distance traced)
  + Curvature variance (amount of directional change)
  + Radial deviation (distance from ideal spiral/meander path)
  + Line smoothness or jaggedness (level of tremor-induced irregularity)
  + Stroke density (tightness or spread of strokes)

**3.2 Transformer-Based Feature Embedding and Classification (Core Component)**

* Use pre-trained Vision Transformers (ViT) to extract high-dimensional embeddings from spiral and meander handwriting images.
* Optionally fine-tune the ViT architecture on the combined dataset for classification (PD vs Control).
* Integrate Coordinate Attention Modules (CAM) to enhance spatial localization of tremor-related artifacts in strokes
* The ViT will serve both as a deep feature extractor and classifier, making it central to the analysis pipeline.

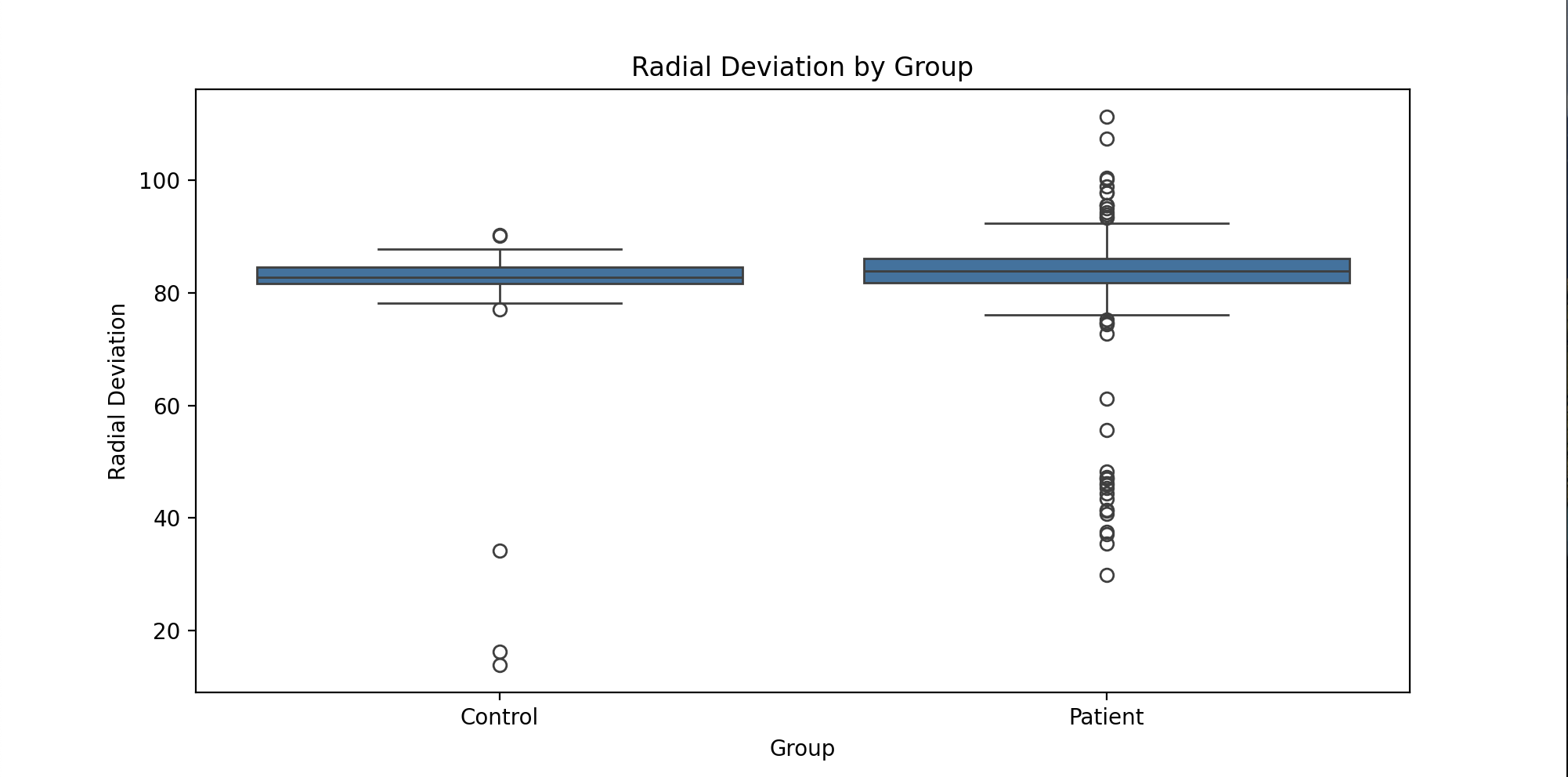
**3.3 Feature Fusion and Statistical Analysis**

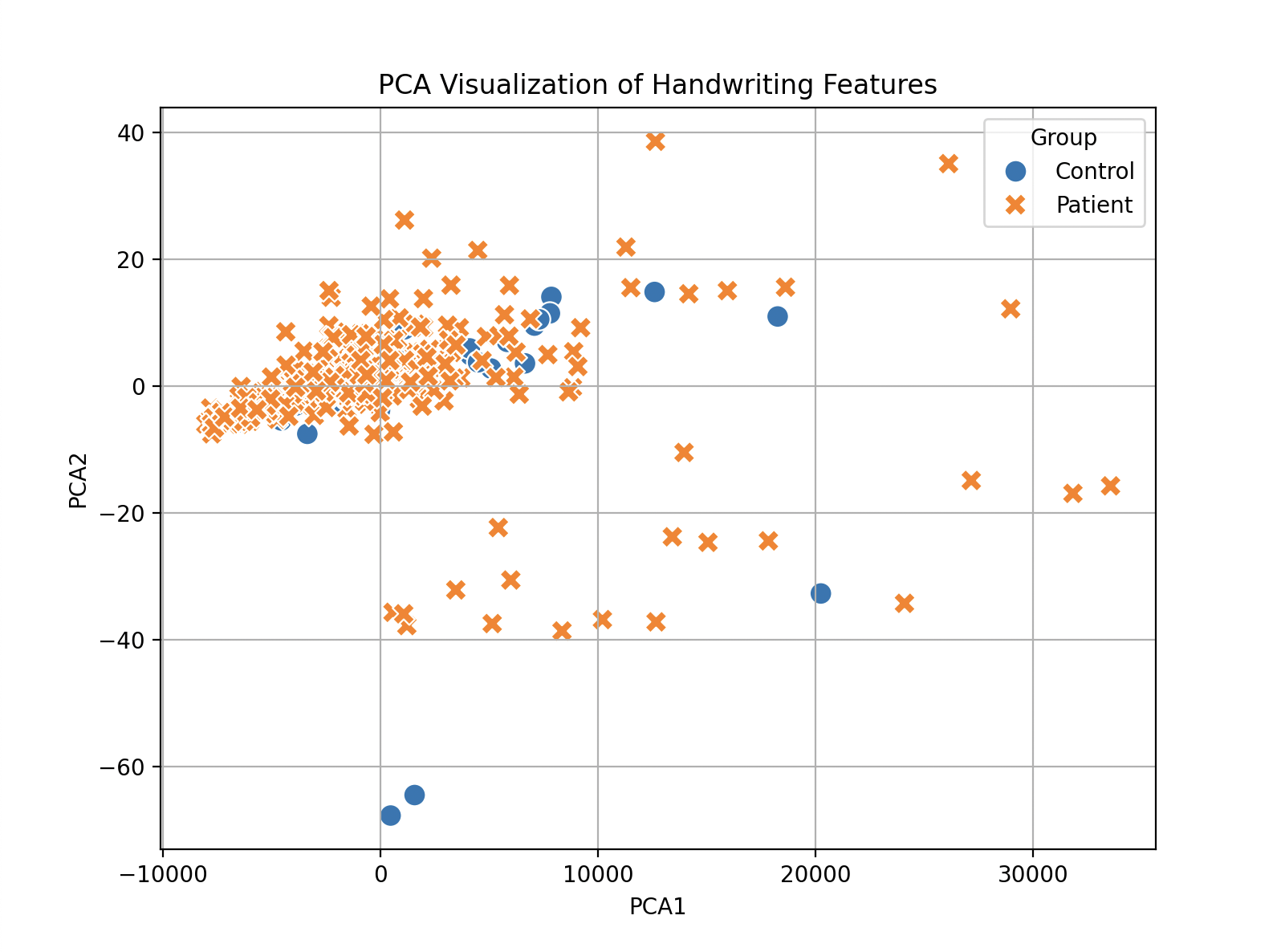
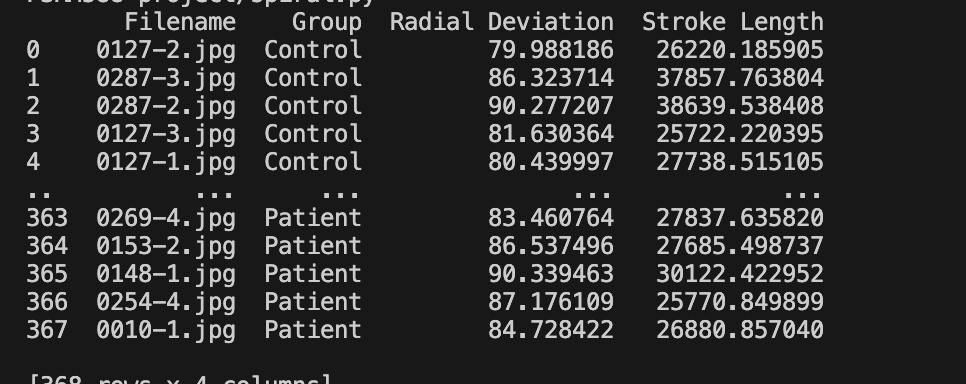
* Combine handcrafted and transformer embeddings into a joint feature matrix
* Perform descriptive statistics and statistical tests (like T-tests, ANOVA) to compare PD and control groups
* Apply PCA for visualization and correlation analysis for feature relationships

**3.4 Classification**

* Inspired by Wang et al., we will apply shallow classifiers such as:
  + Logistic Regression (LR)
  + Decision Tree (DT)
  + K-Nearest Neighbors (KNN)
* These classifiers will operate on the fused feature set (handcrafted + transformer embeddings) to assess classification performance, while keeping interpretability and model complexity low.

## **4. Preliminary results**

Preliminary testing using spiral samples from the HandPD dataset indicates:



* Boxplot visualizations revealed visibly higher variability in both radial deviation and stroke length among Parkinson’s patients compared to the control group, indicating irregularity in stroke smoothness and tracing consistency — consistent with known symptoms such as tremor and loss of fine motor control.
* PCA (Principal Component Analysis) demonstrated partial clustering between the two groups, suggesting that even basic handcrafted features may hold valuable separability patterns for early-stage screening.

## 

## **5. References**

1. <https://www.parkinson.org/understanding-parkinsons/statistics>
2. [https://www.ninds.nih.gov/health-information/disorders/parkinsons-disease](https://www.ninds.nih.gov/health-information/disorders/parkinsons-disease#:~:text=Parkinson's%20disease%20is%20a%20progressive,%2C%20stiffness%2C%20and%20impaired%20balance)
3. <https://www.kaggle.com/datasets/claytonteybauru/spiral-handpd?resource=download>
4. <https://pubmed.ncbi.nlm.nih.gov/27686705/>
5. <https://en.wikipedia.org/wiki/Convolutional_neural_network>
6. <https://en.wikipedia.org/wiki/Vision_transformer>
7. <https://bdalab.utko.fekt.vut.cz/>