Diagnosing Parkinson's Disease From Gait

Daryl Chang, Marco Alban-Hidalgo, Kevin Hsu
Computer Science Department, Stanford University, Stanford, CA
{dlchang, marcoal, kvnhsu}@stanford.edu

***Abstract*—Machine learning methods were used to diagnose the presence and severity of Parkinson's Disease from gait. Data was drawn from a public Physionet database consisting of force sensors positioned under subjects' feet while walking. Ambulatory features, including swing time variability, center of pressure, and foot strike profile, were used to differentiate normal strides from Parkinsonian strides. An SVM Linear Kernel performed the best for predicting the presence of PD, with an AUC of 92.3% and an accuracy of 85.3%. A Random Forest classifier ranking the severity of Parkinson's disease on a scale of 1-5 achieved an AUC of 76.3% and an accuracy of 75.6%. While past studies have used multiple sensing modalities for this task, we outline a method for achieving similar accuracy with only a single set of force sensors.**

***Index Terms*—Parkinson's Disease, Gait Analysis, Force Sensor**

I. Introduction

There is no standard test to diagnose Parkinson's Disease, a condition that affects up to one million people in the US [1]. Instead, doctors qualitatively assess gait, along with other clinical observations, to determine the presence and severity of Parkinson's Disease; the subjectivity of the assessment opens up the possibility of inconsistent diagnoses[3]. In this paper we present a classifier that will diagnose Parkinson's from quantitative measurements of gait.

Tahrir et al. [2] showed that SVMs can diagnose Parkinson's from a combination of spatiotemporal, kinematic, and kinetic gait data. In that study, spatiotemporal data was collected using infrared sensors attached to the subjects' hips and legs, while kinetic data was collected using force sensors placed on the subjects' feet. Since the former may be cumbersome to collect in practice, we aim to diagnose Parkinson's solely from kinetic gait data.

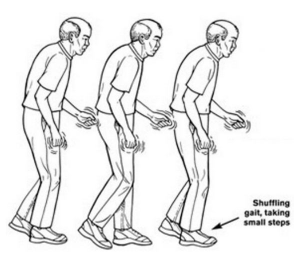

Fig. 1. Abnormal Gait Characteristics of Parkinson's Disease

II. Dataset

Data was drawn from a public dataset maintained by Physionet and originally collected by the Laboratory for Gait & Neurodynamics, Movement Disorders Unit of the Tel-Aviv Sourasky Medical Center. The data consists of 279 gait recordings from 93 patients with idiopathic Parkinson's Disease (PD) and 73 healthy controls. During data collection, 8 sensors were placed underneath each of the subject's feet. For each subject, measurements of the vertical ground reaction force (VGRF) in Newtons were recorded as they walked at their usual, self-selected pace for approximately 2 minutes on level ground. Thus, each datum consists of 16 VGRF time series, as well as an additional 2 time series representing the aggregate force under each foot. Finally, each subject is annotated with demographic information and the presence and severity of Parkinson's Disease.

III. Feature Extraction

A. Overview

To decide which features to extract from the dataset, we conducted a literature survey of medical journal articles describing gait analysis of PD patients. Toledo et al. [3] state that the ability to maintain a steady gait rhythm is impaired in patients with Parkinson's Disease. Amende et al. [4] also show a significant statistical difference in stride length and stride-to-stride frequency in subjects with PD.

After conducting the literature survey and deciding on a feature set, we followed the feature extraction process shown below. First, we segmented gait into stance and swing phases, where stance is the part of a step during which the foot makes contact with the ground, and swing is the part during which the foot is in the air. Then, we calculated the mean and variance for each feature described in the following section, and concatenated them to create the feature vector for each subject.

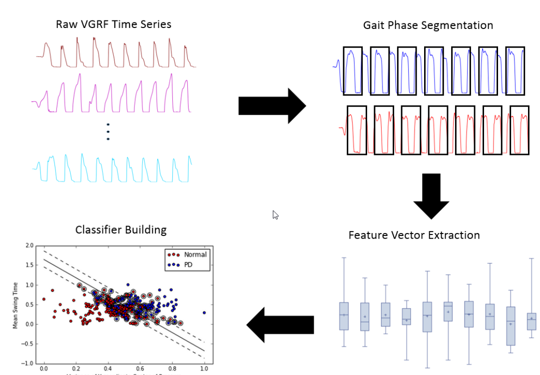

Fig. 2. Summary of Feature Extraction and Modeling

B. Features Extracted

1) Naive Features: As a basic feature set, we included the mean force for each sensor. Though the mean force varies as a function of the subject's weight, it can be used as a naive indicator of the parts of the foot on which the subject tends to exert pressure.

2) Swing and stance times: As stated by Toledo et al., swing time and stance time variability can be a significant marker for the presence of PD. Swing time is defined as the time from when the foot lifts off the floor to when it lands back on the floor. A healthy person is more likely to have a consistent swing time; conversely, for a patient with PD, the swing time length varies depending on their degree of motor impairment. Thus, a low variance in swing time is associated with a fairly constant stride in a healthy person, whereas in Parkinson's patients, higher variance could signify trouble walking and balancing, or 'freezing of gait,' a phenomenon in which Parkinsons patients feel that their feet are glued to the ground [5].

For each datum, the 16 VGRF time series were processed to extract swing phase and stance intervals for each step. The mean and variance of the swing and stance times were used as classification features. Figure 3 plots the strides of two subjects, in which swing and stance times for the PD subject are markedly higher than those of the non-PD subject.

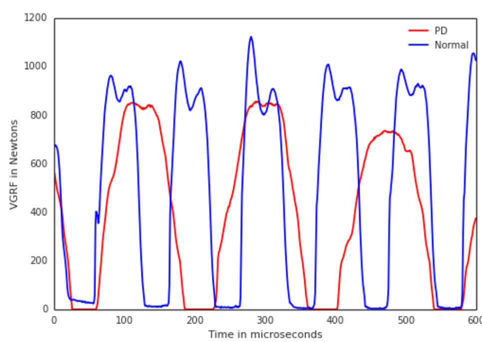

Fig. 3. Stride-to-Stride Time Series of Two Subjects

3) Center of Pressure as a measure of weight imbalance: Most healthy adults have a characteristic weight distribution profile and pattern while walking. In a normal gait cycle, one's center of pressure will shift from the heel to the toe over the course of a step, as shown in Figure 5.

To analyze center of pressure, the 16 VGRF time series were used to produce a 2-coordinate (x and y component) time series of the center of pressure of the subject's foot. The mean and variance of the coordinates were used as features. Note that these features measure shifts in the weight distribution both laterally and longitudinally. We calculate the center of pressure (COP) as follows:

$$COP = \frac{\sum_j^n s_j * f(s_j)}{\sum_j^n f(s_j)}$$

where n is the number of sensors and $f(s_j)$ is the VGRF of sensor j in Newtons.

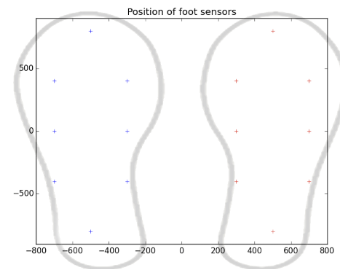

Fig. 4. Sensor Positions

4) Foot Strike Profile: Healthy people tend to walk by lifting their heel and stepping off their toes, then landing on their heel. Parkinsons patients are more likely to have a flat foot strike, where the heel and toe touch the ground at relatively similar time points. Motivated by this, we built a feature set that captures foot strike profile variability. For each datum, we use the 16 VGRF time series to produce time series denoting the point in time were the subject's feet first touched the ground after a swing phase. We then calcluated the (x,y) coordinate of the COP when the foot first contacts the ground, and took the mean and variance as features.

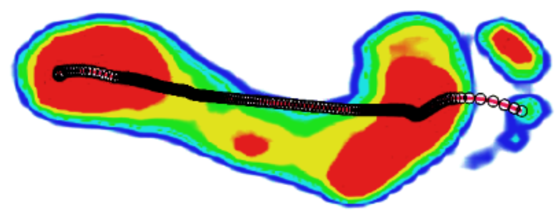

Fig. 5. Transition of Center of Pressure in one gait cycle

C. Normalization Schemes

1) Pre-Feature Extraction: Since each subject's vertical ground reaction force is proportional to the subject's weight, we hypothesized that normalizing each of the 16 time series by the subject's weight in Newtons would refine our features, in particular the summary statistics of the sensors. However, this normalization scheme yielded negligible improvement in performance.

As a data pre-processing step we also tried normalizing the time series so as to have zero mean and unit-variance. However, this also provided negligible improvement as it eliminated much of the variation in force between the different sensors.

2) Post-Feature Extraction: Initially, our logistic regression classifier far outperformed our SVM classifier. After using min-max normalization on the feature vectors we were able to improve the performance of the SVM classifier to be on par with the Logistic Regression classifier. For a more detailed explanation see the Results and Discussion section.

IV. Model Selection and Refinement

A. Classifiers

We selected a basket of classification algorithms to test: logistic regression, random forest, linear kernel SVM with regularization, and RBF kernel SVM. As a baseline algorithm for both classification tasks, we ran Logistic Regression with the naive feature set described above.

B. Feature Selection

To mitigate over-fitting, a subset of the feature set was extracted using forward search, where the subset size k was tuned as as a hyperparameter. The best subset was chosen for the final classifier. The criteria for marginal improvement in the algorithm was the average AUC of an ROC curve in 10-fold cross validation.

C. Parameter Search

After pruning our feature vector, for each classifier, we implement a parameter grid search to find the best performing parameters across a set of possible values. For example, for the linear SVM classifier we run a grid search on the possible values of the regularization parameter.

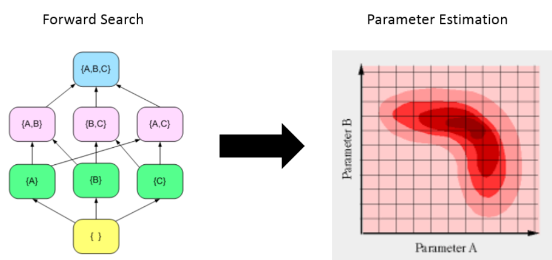

Fig. 6. Forward feature selection and parameter grid search.

V. Evaluation Metric

One of the challenges we faced in building the classifiers was class imbalance and small sample size. As noted in the section detailing the data set, our input classes had a 70% to 30% asymmetry (PD subjects to non-PD subjects). As a result, traditional accuracy measures would not have been an effective evaluation metric. Instead, we used the AUC (Area Under the Curve) of the ROC (Receiver Operator Characteristic Curve) as our evaluation metric. The ROC curve plots the true-positive rate against the false-positive rate for a binary classifier system. Chen et al. [7] have shown that the AUC to be a good evaluation metric in combating class imbalance for small samples. As an added precaution, we sampled training examples with weight inversely proportional to the class frequency, thus downsampling the PD subjects.

VI. Results and Discussion

As shown in Figure 7, the baseline algorithm performed relatively poorly, predicting the presence of PD with 56.7% AUC and the severity with 51.5% AUC. This was expected as the force sensor means contain little information about the abnormal gait features characteristic to Parkinson's such as swing and stance time variability.

	AUC	Accuracy	Precision	Recall
Baseline PD Classifier	56.8%	60.2%	70.6%	70.6%
Baseline Severity Classifier	51.5%	49.8%	30.3%	55.0%

Fig. 7. 10-fold cross validation metrics for baseline classifiers.

To improve classification performance, we extracted the ambulatory features described in Section III, and tested each classification model described in Section V with forward feature selection and parameter grid search. Figure 8 below shows a summary of the performance of the models.

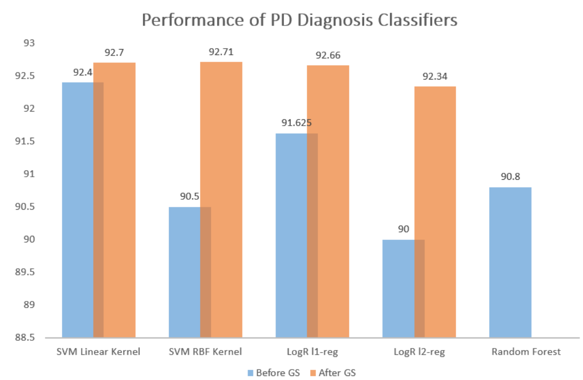

Fig. 8. Average AUC score over 10-fold cross validation for PD diagnosis classifiers. The blue bars show the performance before parameter grid search (GS) for best parameter estimation. The orange bars show the performance after GS.

All models achieved an AUC score higher than 90%, with the linear kernel SVM outperforming the others by a small margin. Examining the cross-validation metrics for the best-performing classifiers in more detail, we see in Figure 9 that the PD classifier and severity classifier perform much better than the baseline. In particular, the PD classifier AUC of 92.3% represents a large improvement over the 56.7% AUC of the baseline algorithm, while the PD severity improved from 51.5% AUC to 76.3%.

	AUC	Accuracy	Precision	Recall
PD Classifier	92.3%	85.3%	99.3%	87.3%
Severity Classifier	76.3%	75.6%	71.8%	61.5%

Fig. 9. Summary of best performance classifiers

As mentioned in section IV, we used a forward search algorithm with the average AUC of an ROC curve in 10-fold cross validation to determine the most crucial features in our model. Table I below shows the top features found by forward selection. Confirming the findings of Toledo et al. and Amende et al., we find that variability in center of pressure and foot strike profile are good predictors of PD [3][4]. The first finding is consistent with the fact that PD patients shift their weight erratically, thus increasing the variance in the center of pressure during the stance phase of gait. The second finding most likely results from PD patients having a flat foot strike rather than a heel strike, which would bring the mean foot strike coordinate closer to the center of the foot.

TABLE I
MOST SIGNIFICANT FEATURES IN CLASSIFIER PERFORMANCE

PD Classifier	center of pressure variance, foot strike coordinate means
Severity Classifier	sensor means, center of pressure variance, foot strike coordinate means

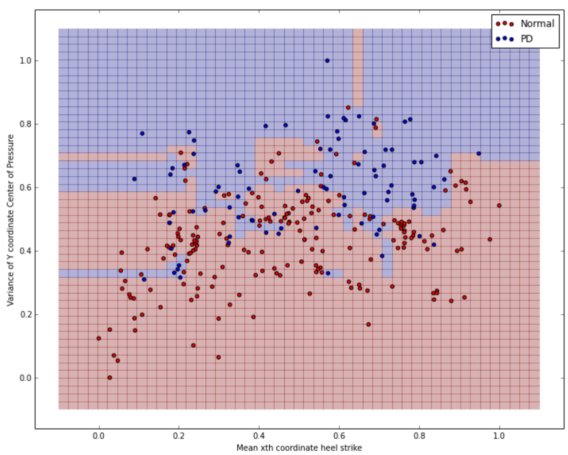

Fig. 10. Random Forest Decision boundary using features: Mean xth coordinate foot strike and variance of y coordinate of center of pressure

Indeed, plotting the subjects by foot strike mean and center of pressure variance, as in Figure 10, clearly separates the data into PD and non-PD clusters by center of pressure variance, and to a lesser degree by the foot strike mean. The plot also helps to explain the relatively high AUC score of the Random Forest classifier, as it is able to modify the decision boundary to handle the noise in the dataset.

Interestingly, both SVM classifiers performed poorly ($\sim$ 60% AUC) before min-max normalization. The literature about SVMs suggests that large margin classifiers are sensitive to the way features are scaled, resulting in severe accuracy degradation if the data is not normalized. In the case of the linear kernel SVM, a prediction is calculated as follows:

$$\Sigma_{i=1}^{m} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b$$

If a particular dimension of x is much larger in magnitude than the rest, it will dominate the dot product and thus affect the separating hyperplane. Normalizing adjusts for this magnitude discrepancy, explaining the performance boost that we observed upon normalization.

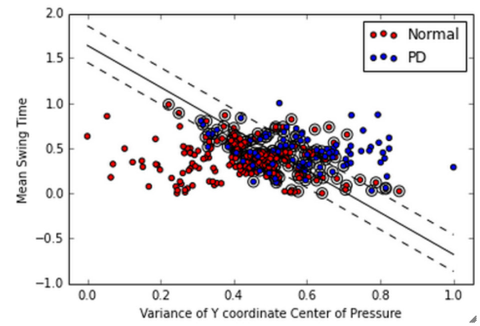

Fig. 11. Linear SVM Decision Boundary for swing time and variance of yth coordinate of the centre of pressure (min-max normalized). Data normalized using min-max normalization. Regularization C= 5.2

VII. CONCLUSION AND FUTURE WORK

This paper presents several successful classifiers for the diagnosis of the presence and severity of Parkinson's Disease from gait data. We provide an improvement on previous work by Tahrir et al. [2] by presenting a classifier that only requires kinematic gait data (as opposed to kinematic and spatiotemporal). We also successfully extract gait abnormalities previously shown to be associated with Parkinson's Disease and use them for classification [3][4].

For future work, we would like to improve the gait segmentation algorithm to be more resistant against noise and to segment the stride into further subphases of the gait cycle. We would also like to extract a richer feature set that incorporates these subphase features to improve the accuracy of the severity classifier.

ACKNOWLEDGMENT

The authors would like to thank Andrew Ng and the TAs of CS229 for their guidance.

REFERENCES

[1] *Understanding Parkinson's Disease*, Parkinson's Disease Foundation, 2014.
[2] N. Tahir, H. Manap. *Parkinson Disease Gait Classification based on Machine Learning Approach*, Journal of Applied Sciences 2012, 12(2), 180-185.
[3] S. Frenkel-Toledo, N. Giladi, C. Peretz, et at. *Effect of gait speed on gait rhythmicity in Parkinson's disease*, Journal of Neuroengineering and Rehabilitation, 2005, 2(23).
[4] I. Amende, A. Kale2, S. McCue, et al. *Gait dynamics in mouse models of Parkinson's disease and Huntington's disease*, Journal of NeuroEngineering and Rehabilitation, 2005, vol. 20, p2.
[5] S. Perez-Lloret, L. Negre-Pages, P. Damier, et al. *Prevalence, Determinants, and Effect on Quality of Life of Freezing of Gait in Parkinson Disease*, The Journal Of the American Medical Association, 2014.
[6] Andre Ng. *Advice on Applying Machine Learning*, Stanford University, http://cs229.stanford.edu/materials/ML-advice.pdf.
[7] M. Wasikowski, X. Chen, *Combating the Small Sample Class Imbalance Problem Using Feature Selection* IEEE Transactions on Knowledge & Data Engineering, vol.22, 2010.