Final Report: Data Analytics for Movement and Health in Parkinson's Disease

Julienne Charles
December 12, 2023
Computer Science Project
Rutgers University Newark

Table Of Content

Project Overview

The focus of this project is to use data analytics to gain an understanding of movement patterns related to Parkinson's Disease. Dr. Jean-Francois Daneault and his team conducted groundbreaking research from data collected from a 4-day Levodopa Response study on 31 PD subjects. During the study, test subjects wore wrist-worn smartwatches and a waist-worn smartphone to collect accelerometer data.

Team Members:

Sung Huh: ML Engineer | Data Architect

Benny Hernandez: Data Analyst | QA Engineer

Atharv Honap: Subject Matter Expert | Data Analyst

Gregory Hiraldo: Data Analyst | Developer

Julienne Charles: Project Manager | Subject Matter Expert | Analyst

Research Background

A significant effort is being made across various institutions to understand Parkinson's Disease (PD) better and improve treatments, preventative measures, and cures. Dr. Jean-Francois Daneault's team conducted a 4-day Levodopa Response study on 31 PD subjects, using wrist-worn smartwatches and waist-worn smartphones with accelerometer applications.

Our Approach

Objectives

- Analyze severe timer scores.
- Determine tremor duration.
- Identify activities influencing tremors.
- Visualize tremor patterns.

Our main approach involves conducting statistical analysis from the left-sensor wrist data to answer the following questions:

1. What times are the most severe tremor scores?
2. How long do tremors last?
3. What activities affect tremors the most?
4. What do the tremors look like?

Model Development:

Task-Agnostic Model:

- Classification to determine if tremors occur (Decision Trees | Random Forest)
- Regression to measure the severity of tremors (Random Forest)

Deep Learning:

- Utilize deep learning for predictive outcomes on larger datasets.

Data Preparation

- - Converted MATLAB files to Python-compatible data.
- - Utilized TSFEL for feature extraction.
- - Overcame multilevel array challenges by "flattening" data into a compatible data frame.

Data Prep Techniques:

- Used TSFEL for feature extraction to convert multilevel arrays into a usable data frame.

Model Selection and Investigation

- Explored various models for tremor identification.
- Models investigated: Decision Trees, SVM, RNN with LSTM, Random Forest, KNN, Support Vector Regression.

Investigated Models:

- Sung: Decision Trees
- Atharv: Support Vector Machines (SVM)
- Julienne: Recurrent Neural Networks (RNN) with LSTM
- Greg: Random Forest
- Benny: K-Nearest Neighbor (KNN)

Support Vector Regression (SVR):

- - Accuracy: 73%
- - Challenges: Manual cross-validation, suboptimal results.
- - Conclusion: Limited success; not the optimal model.

	precision	recall	f1-score	support
0	0.72	0.86	0.78	757
1	0.69	0.49	0.57	488
accuracy			0.71	1245
macro avg	0.71	0.67	0.68	1245
weighted avg	0.71	0.71	0.70	1245

Random Forest Classifier:

- - Accuracy: 75%
- - True Positive Rate: 77%, True Negative Rate: 72%
- - Challenges: Overfitting, feature importance issues.
- - Conclusion: Promising results but room for improvement.

```
Fitting 5 folds for each of 1 candidates, totalling 5 fits
Corresponding Mean CV accuracy: 0.7886740331491713
Binary Test set accuracy: 0.74
Binary Classification Report:
              precision    recall  f1-score   support

           0       0.78      0.79      0.78       730
           1       0.69      0.68      0.69       515

    accuracy                           0.74      1245
   macro avg       0.74      0.74      0.74      1245
weighted avg       0.74      0.74      0.74      1245

Fitting 6 folds for each of 1 candidates, totalling 6 fits
Corresponding Mean CV accuracy: 0.8630514705882354
Tremor Score Test set accuracy: 0.73
Tremor Score Classification Report:
              precision    recall  f1-score   support

           1       0.86      0.83      0.84       356
           2       0.49      0.49      0.49       112
           3       0.21      0.23      0.22        13
           4       0.00      0.00      0.00         2

    accuracy                           0.73       483
   macro avg       0.39      0.39      0.39       483
weighted avg       0.75      0.73      0.74       483
```

LSTM (Long Short-term Memory):

- Application: Classification of tremor severity based on time series data.
- Metrics: Accuracy, precision, confusion matrix.
- Conclusion: Comprehensive evaluation method, potential for fine-tuning.

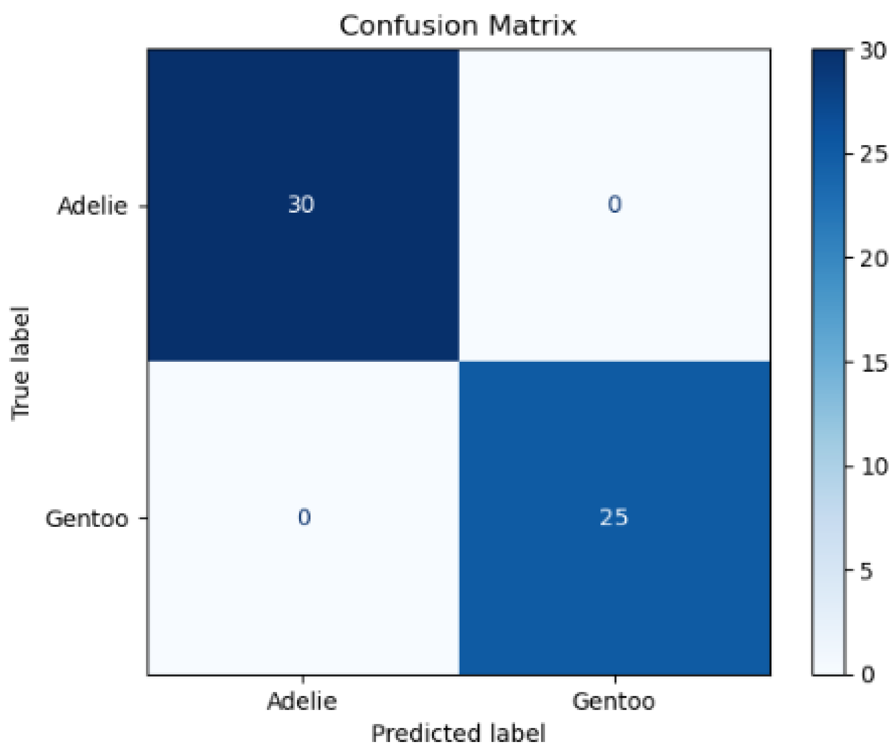
```
Classification Report:
              precision    recall  f1-score   support

           0       0.72      0.79      0.75       465
           1       0.63      0.54      0.59       316

    accuracy                           0.69       781
   macro avg       0.68      0.67      0.67       781
weighted avg       0.68      0.69      0.68       781
```

```
Mean Accuracy: 0.7627 +/- 0.0305
```

The mean accuracy after applying kfolds is 76%, with time I can make additional adjustments and reevaluate to see which parameters ar e optimal.



SVM (Support Vector Machines):

- Tremor Identification: Near 100% accuracy.
- Tremor Severity: Near 100% accuracy, except for severity classification.
- Conclusion: High accuracy in identifying tremors; suitability for severity classification needs improvement.

	precision	recall	f1-score	support
0	0.77	0.84	0.80	500
1	0.72	0.62	0.66	330
accuracy			0.75	830
macro avg	0.74	0.73	0.73	830
weighted avg	0.75	0.75	0.75	830

Final Results

Tremor Identification:

- Best Model: Atharv's SVM with near 100% accuracy.

Tremor Severity Classification:

- Best Model: Sung's Random Forest with 87% accuracy.

Challenges and Room for Improvement

Challenges Faced:

- Difficulty in increasing accuracy.
- Imbalance in research time and implementation.
- Balancing imbalanced data.

Potential Improvements:

- Implement SMOTE hyperparameter tuning.
- Consider using RandomizedSearchCV for more efficient hyperparameter tuning.
- Enhance code reusability.

Lessons Learned and Reflections

- Acknowledged the learning curve with new models.
- Identified the need for better initial questions.
- Recognized the importance of seeking clarification on dataset formats.
- Educate ourselves thoroughly on selected models.
- Efficiently communicate with domain experts to save time and avoid unnecessary challenges.

Future Considerations:

- Collaborate closely with domain experts.
- Formulate precise initial questions.
- Continuous learning for optimal project execution.

Conclusion

In conclusion, our team successfully developed models for identifying tremors and classifying their severity in Parkinson's Disease patients. Especially in Tremor identification and severity. Although we faced many challenges and learning curves, we achieved significant results and identified areas for improvement. Our teamwork efforts showcased the importance of effective communication and continuous learning in the field of computer science. Given more time, we aim to further refine our models and enhance their applicability.

References

Lauzé, Martine, et al. "The Effects of Physical Activity in Parkinson's Disease: A Review." *Journal of Parkinson's Disease*, IOS Press, 1 Jan. 2016, content.iospress.com/articles/journal-of-parkinsons-disease/jpd160790.

"Carbidopa-Levodopa Oral: Uses, Side Effects, Interactions, Pictures, Warnings & Dosing." *WebMD*, WebMD, www.webmd.com/drugs/2/drug-3394-41/carbidopa-levodopa-oral/carbidopa-levodopa-oral/details. Accessed 17 Oct. 2023.