DISCOVERING DISTINGUISHABLE PATTERN OF PARKINSON DISEASE ON PATIENT'S CLOCK DRAWING TEST

Anan Methasate Nisaruj Rattanaaram Pawat Amonpitakpun

Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand

ABSTRACT

The Clock Drawing Test (CDT) is a test for screening for cognitive impairment. In this paper we'll focus on finding any distinct features or patterns regarding Parkinson's Disease. Parkinson's disease is one of many disease caused due to neurodegenerative disorder. According to US Department of Health and Services [1], Parkinson's patient would have the following distinguishable symptoms: shaking, stiffness, and difficulty with walking, balance, and coordination.

1 Introduction

1.1 Motivation

The clock drawing task can check various brain anomalies, including Alzheimer, ADHD, etc. But for PD, we beg to differ. Generally doctors only uses the final clock image for data analyzing but for PD, speed and motion should be the main focus in determining PD, in contrast to other brain diseases. We believe that by taking advantages of time series data instead of using only final image, we can extract many useful information about Parkinson Patients more than sole image could.

1.2 Previous Work

We have found out that Manuel et al.[2] also studied recently how Discrete Fourier Transform extracts useful feature from the time series signals, so we intended to try on our dataset to see if our results reproduce anything similar to theirs.

2 Data Statistics

The data consists of 196 patients record, which 80 of those were diagnosed having Parkinson's disease. The patient's data consists of fundamental demographics as follows: age, gender, educational level, duration of PD disease, handedness, and TMSE score (Thai Mini Mental Status Examination). There are also time-series features which is a recording of patient's clock drawing. The time-series data consists of x-y coordinate, pen pressure, relative time of that point, and action type (start, drag, lift).

	Normal	Abnormal
Control	33	83
PD	26	54

Table 1. Count of patients by diagnosis

3 Data Visualization

We have found numerous visual patterns forming what appears to co-align with what the symptoms of Parkinson Disease describes. Initially, by looking at whole clock image poses challenges due to the reason that we're visually biased to analyze structure of the clock and/or components alignment instead of line patterns. So we tried to analyze by components instead and compared many plots to non-PD patients. We can clearly see the jiggling line pattern for

Fig. 1. Comparison of separated clock components between PD patients (left) and non-PD patients (right). The line density denotes pressure.

PD patient as opposed to non-PD. Second, we mapped each components into derivative and second-order derivative of y with respect to x so we can see the trend of how position of the pen changes. Then, we calculate the curvature of the

stroke and plot the curvature and the logarithm based 10 of the curvature. The equation of the curvature is as below.

$$K = \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}} \quad (1)$$

where K is the curvature, x' is the derivative of x with respect to time ,and y' is the derivative of y with respect to time.

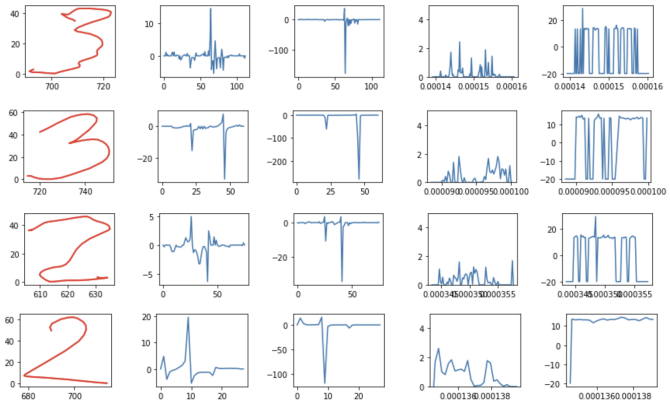

Fig. 2. Components of drawings by PD (1st and 3rd row plots) and non-PD patients (2nd and 4th row plots). Red graph is an x,y co-ordinate of patient's drawing. Each blue graph denotes 1st order, 2nd order of each x, y co-ordinate. We can see higher level of noises for PD patient's drawings.

Finally, we looked at underlying structure of the clock by computing T-SNE clustering of every images of clocks final drawings. We converted data into image and resized it into 100 by 100 square pixels. From the results, we can see that there is indeed underlying structure that can distinguish PD by some degree. However, the T-SNE results for abnormal clock results in better cluster than PD label. This is one evidence that just the final image is not enough to completely distinguish PD patient.

4 Baselines

We have felt that the patient's sample sizes were too small, and we have proved that splitting data is enough to have impact on the model's quality. So we will try to redo experiments for a lot of time (like K-fold) to ensure that the metrics that we got truly represents the data. In each following subsection, we split the data and tested accordingly.

4.1 Random baseline

After running 1,000 times of random baseline, we found that accuracy and recall is approximately close to 0.5, while recall precision and F1 score is a bit lower (0.44 to 0.46).

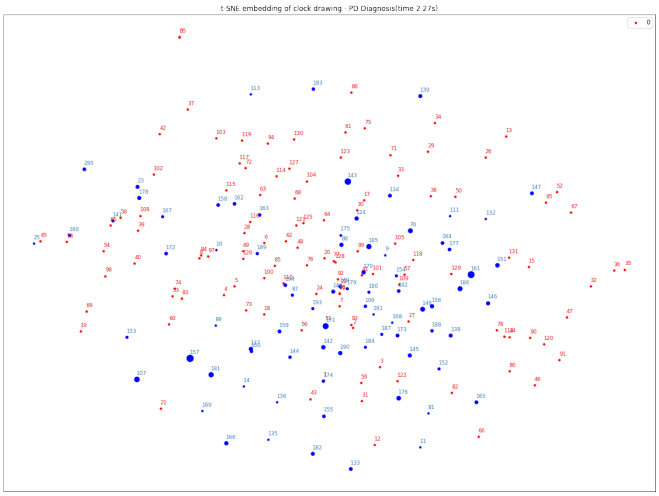

Fig. 3. T-SNE clustering of PD (blue) and non-PD (red) patient's clock images for perplexity 10.

All having std. value between 0.07 to 0.10. This result applied to both label type.

4.2 All 0/1 baseline

We predict either all positive or all negative. For one baseline (all positive), we achieved average metrics of 0.4 to 0.5 for all metrics except recall which was 1. For zero baseline (all negative), we got average accuracy of 0.6. All of the experiments showed some variety of scores just by splitting the data.

4.3 Demographic baseline

We used only demographic data to see if there are any correlation in age, intelligence, etc. While we did not have any time to determine which attribute contributed the most to having PD, we found that those features were useful in determining PD. At first, we only used age and gender as a feature for model prediction. Then we would compare it with model which uses all demographic features. After running for 100 epoch and 30 trials each, we got an average ROC curves which represents that model's performance. We can see that each ROC varies a lot due to small sample size dataset. Dense1 data contained age, gender. Dense2 data contained every demographic data. Dense3 contained all demographic data and average pressure/sample. The results were that Dense3 is not comparably better than Dense2, and Dense1 performed worse than average compared to other curves.

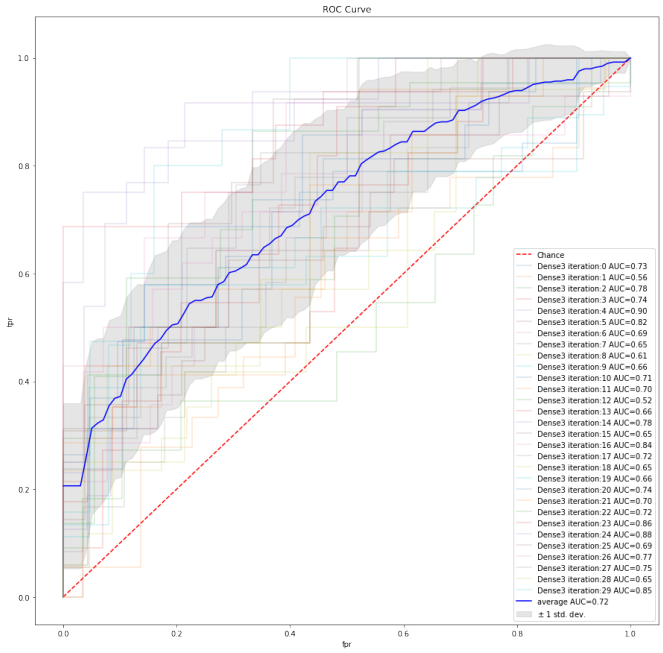

Fig. 4. The plot shows ROC Curve for all demographic features combined and additional average pressure. Blue line indicates average ROC Curve with AUC of 0.72.

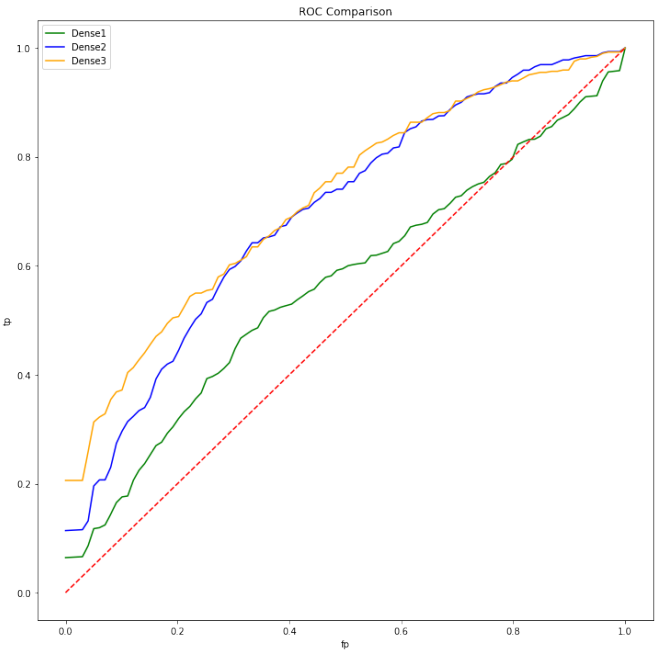

Fig. 5. The plot shows ROC Curve for all demographic features combined and additional average pressure. The model was trained using fully connected layers

5 Feature Extraction

5.1 Angle between two points

Arctan2 function returns the angle between the ray to point (x, y) to x-axis. We use arctan2 between current point and previous point to determine how jiggling the line is. Figure 6 represents the difference between arctan2 values of Non-PD and PD class. From Fig. 6, we can visually notice higher amount of noise pattern.

5.2 Drawing order

We hypothesized that we can gain information on PD symptoms by observing their drawing clock components order. For example, a healthy person tends to write 12 6 3 and 9 before writing other numbers. This metric would be able to capture the planning capability of patients however, the quality of the drawing were not captured such as curvature, etc. So we extract the order of numbers on the clock (1 to 12), the clock's center (0) and the clock's hands (13 and 14) by hand. We programmed animation tools to help visualizing which helped labelling each components easier. Then we will be able to use these sequence as time-series feature. The main problem was that many drawings were ambiguous, and was hard to tell whether it is a number or not, sometimes the clock's hand length could not be distinguished, etc. Another problem is that since the feature was hand engineered, we also have to label again. Person who label also affect the label as well.

5.3 Components

We divide the overall time series into a set of components measured by pointer type. Each component begin with pressing the pen and end with lifting the pen. By divide it into components, we can calculate sum of pressure, mean of pressure and average drawing speed of each stroke more precisely.

5.4 Drawing Speed

There are some research that was wrote about the correlation between PD and writing speed of the patient. We can extract mean of drawing speed and variance of mean of drawing speed of the components by calculate the drawing speed from ratio between euclidean distant of two points and time duration between these points as an equation below.

$$v_i = \frac{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{t_{i+1} - t_i} \tag{2}$$

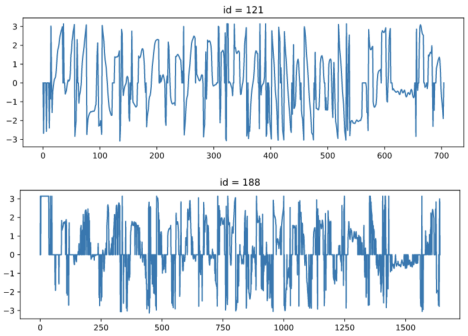

Fig. 6. The difference between Non-PD (Above) and PD (Below). Y-axis represents arctan2 value. The value is in range ($-\pi$, π).

6 Proposed Models

6.1 Using non-time series features

We start from using Fully connected layer with non-time series features. We use 3 layers of Dense with age, gender, handedness (side), education (Ed), and TMSE.

6.2 Combine with time-series features

We also try to create a model using both non-time series and time-series features. We split the model into two parts. We handle time-series features by using two Gated Recurrent Unit (GRU) layers and two Dense layers for non-time series features. We combine these two parts by using Concatenate and Dense layer.

6.3 CNN with Fast Fourier's Transform

We try to apply Manuel et al. [2] method with our dataset. First, we need to preprocess time-series signal such as coordinates or pen pressure by sampling long sequence into many windows of size 30. Then we uses the sliding windows to move and repeatedly apply Fast Fourier's Transform to each window and label all windows from PD as 1, and non-PD as 0. The model is organized in two parts. In the first part, we use Convolutional Neural Network (CNN), MaxPooling and to extract features from the inputs. The second part includes fully connected layers and dropout layers for classification. Window size and striding value are also hyperparameters. However, we did not have much time to try out many possible hyperparameters and pick out the best one.

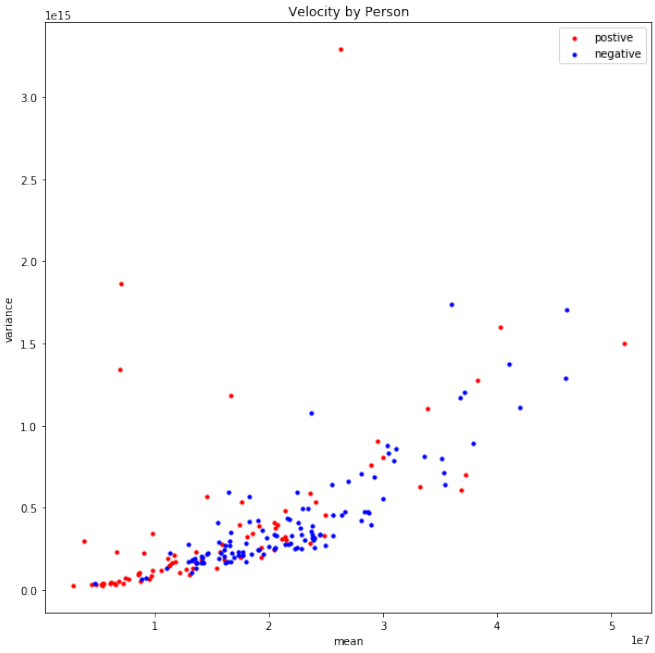

Fig. 7. Relation between mean (x axis) and variance (y axis) of each patient. (red is PD).

7 Tuning hyperparameter

7.1 Threshold

We try to optimize the Model 6.2 by finding the threshold that gives the best in accuracy, recall, precision and F1. We have The training involved using time series data became very challenging to train due to the few samples, causing it to overfit very easily. The trick was to find a way to sample a lot of data from the whole signals using sliding windows in order to upsample the data. After trying to predict the output with different thresholds, we found that 0.5 is the best threshold for this model.

8 Experimental Result

We use accuracy, recall, precision and F1-score as metrics. Table 2 shows the result of the experiments with baselines.

	Accuracy	Recall	Precision	F1
Random	0.5	0.5	0.41	0.44
All PD	0.4	1	0.4	0.57
Mixed (6.2)	0.64	0.68	0.69	0.66
CNN + FFT	0.68	0.69	0.69	0.68

Table 2. Result of the models compared with baselines.

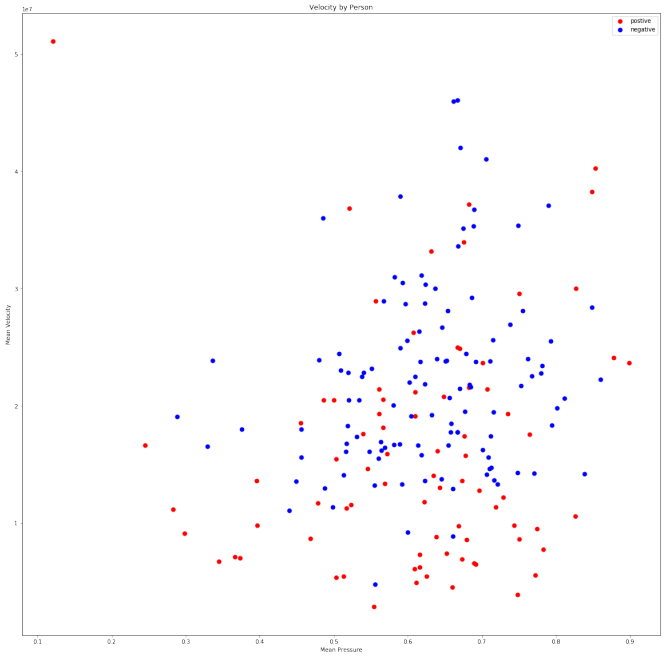

Fig. 8. Relation between mean pressure (x axis) and mean speed (y axis) of drawing. (red is PD)

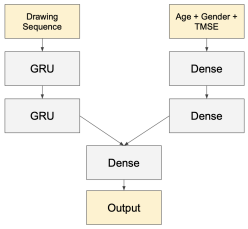

Fig. 9. The model combines time series features and non-time series features using concatenate layer.

9 Conclusions

According to experimental result, the CNN with FFT model gave better result than using normal time-series features. We can improve the FFT model by using RNN layers instead of CNN. The main problem was we did not have enough time to try many variant of models architecture and tuning hyperparameters. Our processed data, benchmarking tools and visualizing data will be publicly available for those interested. There were many interesting approach to take, and many opportunity to improve beyond our model.

10 References

[1] National Institute on Aging, "Parkinson's disease," *Electronics*, May 2017.

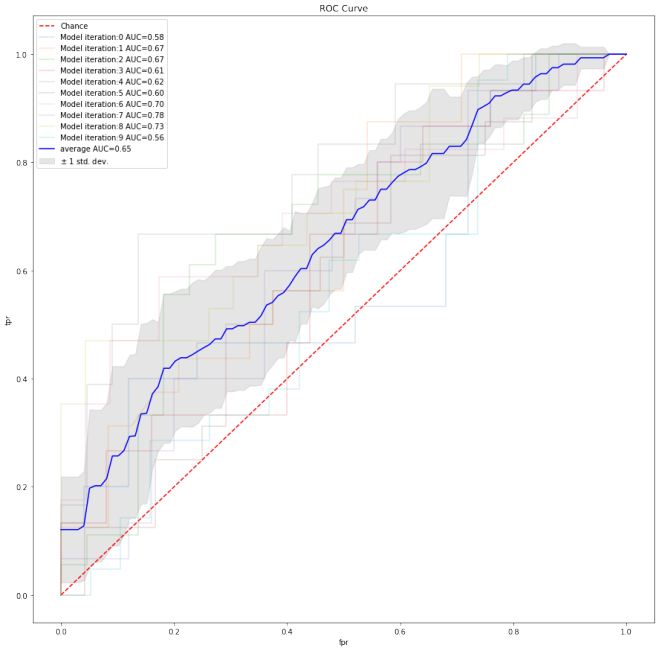

Fig. 10. ROC Curve for method 6.2, acheived average AUC of 0.65

[2] Manuel Gil-Martín, Juan Manuel Montero, and Rubén San-Segundo, "Parkinson's disease detection from drawing movements using convolutional neural networks," *Electronics*, vol. 8, no. 8, pp. 907, Aug 2019.