Deep Learning Approaches for Automated Diagnosis of Parkinson's Disease from Handwriting Samples

Anthony Chen
43408687
UC Irvine
anthonc9@uci.edu

Shreya Chippagiri
53365562
UC Irvine
schippag@uci.edu

Robert Logan
87482533
UC Irvine
rlogan@uci.edu

Pratik Shetty
39844769
UC Irvine
ppshetty@uci.edu

Abstract

Parkinson's disease is a neurodegenerative disease whose symptoms are characterized by a loss of motor skills. One way which these symptoms commonly manifest is through the deterioration of handwriting quality. In this paper, we explore different applications of deep neural networks to diagnosing Parkinson's disease. This is done using handwriting data collected from a population of 37 patients diagnosed with Parkinson's disease and 38 healthy controls. Our final model leads to accurate diagnoses of 74.3% of subjects.

1 Introduction

Parkinson's disease is a neurodegenerative disease which mainly affects the motor system. It is the second most common neurodegenerative disease worldwide, and typically affects elderly patients. Symptoms include tremors, rigidity and loss of balance at early stages, as well as dementia and other thinking and behavioural problems at later stages (Sveinbjornsdottir, 2016).

While there is no cure for Parkinson's disease, a variety of medications exist which alleviate the symptoms and/or slow down its progression (National Institute of Neurological Disorders and Stroke, 2017). Accordingly, the ability to diagnose the disease at early stages is essential to the success of these treatments. Unfortunately, diagnosis is currently a difficult task as there is no definitive test which identifies the disease, and so diagnosis must be done on a case by case basis using clinical criteria (Jankovic, 2008).

In this paper, we seek to solve this problem by using neural networks to diagnose Parkinson's disease from handwriting samples. Previous work on this problem by Drotár et al. (2014; 2016) demonstrates that accurate diagnoses can bo obtained using an SVM to classify handwriting samples. Seperate research by (Periera et al., 2016) has shown that convolutional neural networks are also successful at performing this task. Our goal in this study is to further investigate the effectiveness of different neural network architectures towards solving this problem.

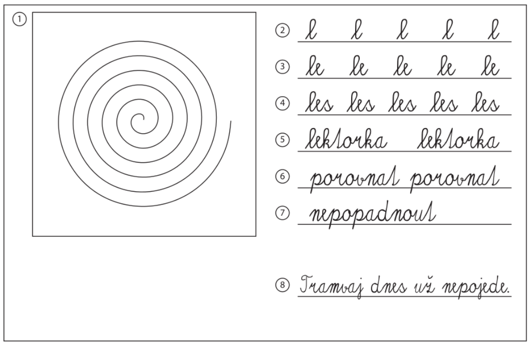


Figure 1: PaHaW template

2 Overview

2.1 PaHaW Database

We train and evaluate our model on the Parkinson's Disease Handwriting Database (PaHaW) (Drotár et al., 2014, 2016). This dataset consists of multiple handwriting samples from 37 parkinsonian patients (19 men/18 women) and 38 gender and age matched controls (20 men/18 women). Each subject was asked to copy the drawings/sentences in the template displayed in Figure 1 on a blank piece of paper containing only reference lines/boxes using a conventional ink pen.

Pen movements were recorded using an Intuos 4M tablet with a 200 Hz sampling frequency. Each sample contains the following data: *x coordinate, y coordinate, time stamp, button status, pressure,*

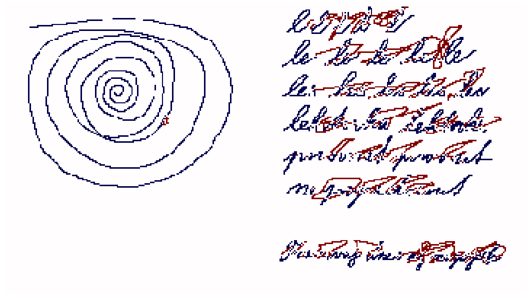


Figure 2: Example graphical representation. On-paper image is drawn in blue. In-air image is drawn in red.

tilt, and *elevation*. The *button status* variable indicates whether or not the pen is in the air.

2.2 Derived Features

The position features described in the previous section can cause issues when used in classification due to their global coordinates. That is, the classifier output can differ depending on where the subject draws on the paper, even if the shape they draw is exactly the same. To counteract this, we replace the the x and y *coordinate* features with x and y *velocity, acceleration,* and *jerk* which are translation invariant.

In order to experiment with wider classes of models, we also created two alternative fixed-width representations of the data which we will refer to as the summary representation, and the graphical representation.

The summary representation is created by extracting predefined set of summary statistics for each feature in the sequence. The summary statistics used were: *mean, standard deviation, minimum value, maximum value, median, 25th-percentile,* and *75th-percentile*. We also extracted these summary statistics for subsequences corresponding to when the pen was on paper, and when it was in air. This closely follows the data preprocessing used by Drotár et al. (2014).

The graphical representation consists of two 300×500 binary arrays. The first binary array replicates the image traced out by the subject, and the second plots path traced out by the pen while in the air. An example is plotted in Figure 2.

Data was normalized for all representations.

2.3 Model Architecture

Different network architectures exist which best suit each of the different data representations mentioned in the previous section. For instance, the raw handwriting sequence data is best modeled using recurrent neural networks since it has variable length. Similarly, since the summary representation has a fixed-width and no meaningful structure in how the summary statistics are ordered, it is best modeled using a simple feed-forward neural network. Lastly, convolutional neural networks are best applied to the graphical representation since they are translation invariant.

Since the problem of Parkinson's detection is a binary classification problem, all networks have a single hidden unit sigmoid activation layer and are trained to minimize binary cross-entropy. In our experiments we found the following architectures had the best performance:

Recurrent Neural Network

- 16 unit LSTM layer - activation: ReLU
- 50% dropout
- 32 unit dense layer - activation: ReLU
- 50% dropout
- 32 unit dense layer - activation: ReLU
- 50% dropout
- sigmoid output

An L2 regularization penalty with weight 0.1 was added to both dense layers.

Feed-Forward Neural Network

- 35 unit dense layer - activation: ReLU
- 50% dropout
- sigmoid output

An L2 regularization penalty with weight 0.2 was added to the dense layer.

Convolutional Neural Network

- 4 unit 3×3 convolution layer
- 3×3 max pooling layer
- 4 unit 3×3 convolution layer
- 3×3 max pooling layer
- 4 unit 3×3 convolution layer
- 3×3 max pooling layer
- 95% dropout
- sigmoid output

An L2 regularization penalty with weight 5.0 was added to the dense layer. Seperate convolutional neural networks were trained for the in-air and on-paper images.

Combined Network

We also experimented with creating a final 'combined' neural network by concatenating together the final hidden layers of the previously described feed-forward and CNN architectures. The concatenated layer is connected to the sigmoid output with random weights.

3 Results

Following Drotár et al. (2014; 2016) we evaluate our models by looking at the average classification accuracy using 10-fold cross-validation. Training and test loss plots are provided in the appendix.

We observe that loss for the LSTM model randomly fluctuates around a constant values of 0.7, indicating that the model is underfitting the data. We experimented with many other recurrent architectures and ways of altering the sequence (e.g. subsampling, truncation, keeping in-air values, etc.) and our results were the same across all of these experiments.

In contrast, we see that the loss plots for the remaining models are quite well-behaved - the loss decreases sharply early on and then levels out, and also the train and validation loss curves move in parralel. This suggests that these models are neither over- or under-fitting the data.

The results for our models are presented alongside the results from (Drotár et al., 2016) in Table 1. As is expected from the loss plot, the LSTM model performs quite poorly - it is essentially the same as randomly guessing. The feed-forward network also does not perform very well, which is surprising given that the features are very similar to those used in (Drotár et al., 2016).

The convolutional models have by far the best performance, beating out one of the k-NN from (Drotár et al., 2016). Consistent with their findings, the in-air data gives slightly better performance than the on-paper data.

We observe that the combined model has exactly the same accuracy as the in-air CNN. This implies that the model merely learned to replicate the output from the in-air CNN network and did not learn anything from the other networks.

Unfortunately, we were unable to beat the benchmark accuracy on this dataset. We expect that this is largely due to the small amount of training data.

Model	Mean Accuracy [%]
LSTM	50.7
Feed-Forward	62.9
Drotár k-NN	71.7
CNN (On-Paper)	72.9
CNN (In-Air)	74.3
Combined	74.3
Drotár SVM	81.3

Table 1: 10-Fold Cross Validation Results

4 Division of Labor

Anthony created the summary representation of the data as well as trained the final feed-forward and LSTM networks.
Shreya did research on previous work on this dataset, and helped with creating feed-forward networks.
Robert processed the data, created the graphical representation, and trained the final convolutional network and combined network.
Pratik created the presentation, and performed background research on Parksinson's disease.

Acknowledgments

We would like to thank the Brain Disease Analysis Lab at Brno University of Technology for providing access to the Parkinson's Disease Handwriting Database.

References

Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdenek Smékal, and Marcos Faundez-Zanuy. 2014. Analysis of in-air movement in handwriting: A novel marker for parkinson's disease. *Computer Methods and Programs in Biomedicine* 117(3):405 – 411.

Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, and Marcos Faundez-Zanuy. 2016. Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease. *Artificial Intelligence in Medicine* 67:39 – 46.

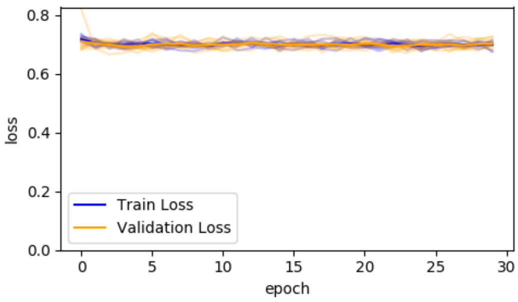
J Jankovic. 2008. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry* 79(4):368–376.

National Institute of Neurological Disorders and Stroke. 2017. Parkinson's disease information page. https://www.ninds.nih.gov/Disorders/All-Disorders/Parkinsons-Disease-Information-Page.

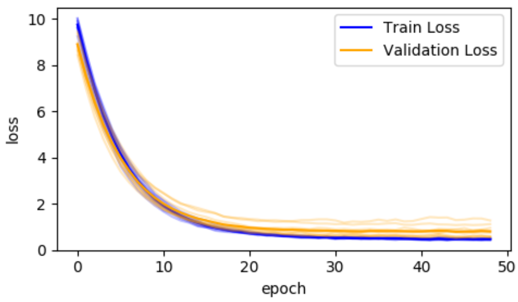
Clayton R Periera, Danillo R Periera, Joao P Papa, Gustavo H Rosa, and Xin-She Yang. 2016. Convolutional neural networks applied for parkinson's disease identification. *Machine Learning for Health Informatics* pages 377–390.

Sigurlaug Sveinbjornsdottir. 2016. The clinical symptoms of parkinson's disease. *Journal of Neurochemistry* 139:318–324.

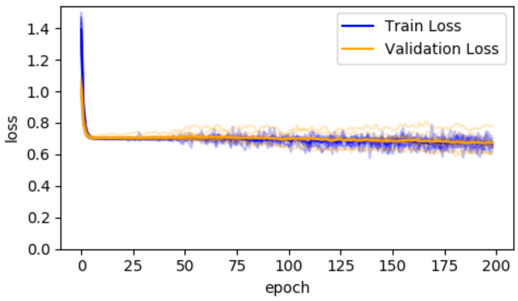
Appendix



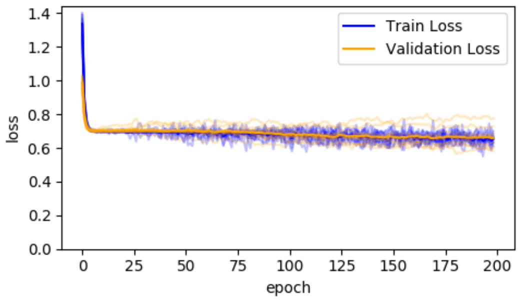
LSTM



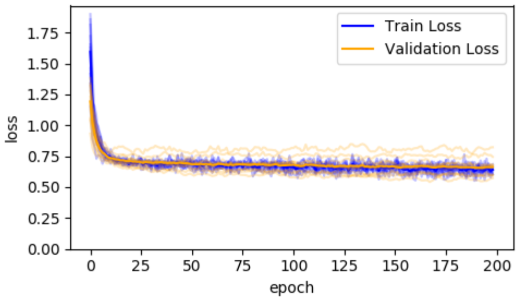
Feed-Forward



CNN (On-Paper)



CNN (In-Air)



Combined