An End-to-End Deep Learning Approach for Diagnosis of Parkinson's Disease Based on Hand Drawing

Yifan Wang1,a,*,†, Yifan Zhang2,b,*,†

1Department of Computer Science, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY, USA
2College of Science and Engineering, University of Glasgow, University Avenue, Glasgow G12 8QQ, UK

*Corresponding email: a2594230Z@student.gla.ac.uk, bwangy56@rpi.edu
†These authors contributed equally.

Abstract. At present, Parkinson's disease is mainly subject to the doctor's subjective assessment of the patient's symptoms through the unified Parkinson's disease score scale, which easily leads to a high rate of misdiagnosis. Some wearable devices have been developed to assist in providing accurate diagnostic results. However, this process requires the cooperation of patients and doctors, which is very inconvenient for Parkinson's patients who have limited mobility. According to reports, Parkinson's patients will show some writing disabilities in the early stages. Therefore, to provide a non-invasive and convenient diagnosis program, this research is devoted to developing a Parkinson's disease diagnosis program based on hand-drawn data. This search provides diagnostic solutions for Convolutional Neural Network (CNN) and Residual Neural Network (ResNet). Unlike the traditional hand drawing preprocessing scheme that only retains the trajectory, the color information of the image for the hand drawing was reserved, which allows the subsequent model to learn the stroke pressure, speed, stroke time, and other based on the color information. Compared with CNN, the ResNet can better solve the semantic gap problem, achieve faster convergence speed, and get higher diagnostic accuracy (accuracy = 88.9%). In general, the two diagnostic schemes proposed in this study only require the patient to provide a hand drawing to confirm whether the patient has Parkinson's disease, which is of great significance for the convenient diagnosis of Parkinson's disease.

1. Introduction

Parkinson's disease (PD) is a brain disorder that leads to shaking, stiffness, and difficulty walking balance [1]. As the disease deteriorates, patients will encounter walking and talking troubles which seriously reduce the PD patients' quality of life. Therefore, patients must detect PD at its early phase and get appropriate treatment as soon as possible. Nowadays, most diagnoses are made by reviewing signs of symptoms combined with both neurological and physical tests [2,3]. These tests and checking usually are time-consuming and costly. Another approach is using wearable devices to detect real-time physical abnormalities to assist diagnosis [4]. However, this approach can also be considered as an invasion of privacy.

Since research shows that the early signs of PD include dysgraphia, which can also be utilized to diagnose the disease, this model will provide a subjective and efficient means of diagnosing PD's basic
diagnosis. Because it is far less sophisticated than the traditional tests and can even be done remotely from home [5]. In 2016, Pereira et al. created the dataset containing 74 PD patient hand drawings and 18 non-PD-patient drawings to diagnose [6]. Nevertheless, the accuracy was only 67%. In 2019, Ali applied Cartesian genetic programming and the random undersampling method to enhance the accuracy to 76.44% [7]. Also, in 2019, Liu et al. used Convolutional Neural Network (CNN) to do the training, and the accuracy improved to 82.68% [8]. This research designed the Residual Network (ResNet), which can be considered an advanced version of CNN, to diagnose PD. This article will provide a clear comparison between the use of CNN and ResNet in diagnostic training models and quantify the accuracy, recall rate, and AUC (Area Under Curve) of these models regarding diagnosing PD symptoms. Unlike the binarizing graph processing, the original color of the hand-drawing graph was maintained during the pretreatment process. By doing so, more information and details of the graph can be obtained and analyzed. Simultaneously, these colored graphs can be beneficial for future diagnosing models, such as learning based on stroke pressure and stroke speed.

2. Method

2.1. Hand-drawn image preprocessing

2.1.1. Hand-drawn trajectory extraction. The data set used in the study is the spiral images (Figure 1). Compared with ordinary pictures, most of these images are backgrounds, and they also contain a black template line and a blue hand-drawn trajectory line. The difference between non-Parkinson's patients and Parkinson's patients is in these blue hand-drawn lines. To extract this part of the feature, most studies used binarization to extract the hand-drawn trajectory. However, this feature extraction method has lost some information about the color change of handwriting. To obtain more feature information, only remove the background to process the image.

\[
IMG_i^j(p_i^j) = \begin{cases} 
0, & \text{if } R_i^j < 50 \cup R_i^j > 202 \cup G_i^j < 50 \\
\cup G_i^j > 220 \cup B_i^j < 50 \cup B_i^j > 240 \\
p_i^j, & \text{others}
\end{cases}
\]  

(1)

\(IMG_i^j\) represents the \(i\)-th updated image, \(R_i^j\), \(G_i^j\), \(B_i^j\) represent the R, G, and B values for three channels corresponding to pixel \(j\) of the input image \(IMG_i^j\), and \(p_i^j\) represents the original pixel \(j\) value of the corresponding image. Use the threshold segmentation method to remove the background part by removing the value of R, G, B close to 255 and remove the black template line part by removing the value close to 0. The processed image is shown in Figure 2. Although the template lines and
background are not completely separated, more image features are retained through reasonable thresholds.

2) The image after threshold segmentation leaves obvious white template line edges and background noise points. The background noise is similar to salt-and-pepper noise, and the white edge and the blue trajectory line have a large color difference. Therefore, use a 3*3 median filter to remove background noise and weaken the edges of the white template lines.

3) After the previous step of processing, it can be found that the background noise points are almost eliminated, but the edges of the white template lines are still obvious. Considering that the hand-drawn trajectory accounts for a small proportion of the overall image, increase the median filter to 7*7 to obtain the final preprocessed image.

![Figure 2. The preprocessing process of the spiral graph, 1), 2), and 3) represent the images after steps 1), 2), and 3) respectively.](image)

2.1.2. Image standardization. The normalization process can convert the pixels of the image into a standard model, avoiding the influence caused by the affine transformation. In addition, normalization can facilitate data processing and accelerate the speed of gradient descent for neural network training. Therefore, normalize the obtained hand-drawn trajectory feature image.

The processed hand-drawn trajectory map has been converted into R, G, B storage mode. Since the maximum pixel value is 255, normalize all pixels as follows:

\[
\begin{align*}
R_{i,j}' &= \frac{R_{i,j}}{255} \\
G_{i,j}' &= \frac{G_{i,j}}{255} \\
B_{i,j}' &= \frac{B_{i,j}}{255}
\end{align*}
\]  

The abscissa is represented by \( i \), and the ordinate is represented by \( j \). Figure 3 shows the normalized image. It can be found that the normalized processing image will not affect the characteristics of the image.

![Figure 3. Normalized image.](image)

After normalization, compress the size of the image. The size of the original images in the image set is different, and most of the sizes are more than 600*600, which is a huge amount of data for the parameter training of the neural network. Therefore, to reduce the computational complexity of the neural network training process, use the nearest neighbor interpolation method to compress the image to a uniform size of 64*64.
2.2. Parkinson's disease diagnosis based on hand drawing

2.2.1. Construction of Convolutional Neural Network (CNN). The traditional fully connected deep neural network connects each neuron to the neuron in the adjacent layer. For images, the traditional artificial neural network inputs all pixels equally to neurons. This type of neural network will treat these pixels equally, ignoring the relationship between pixels and pixels. To cope with this problem, CNN proposed three ideas: local receptive fields, shared weights, and pooling.

First, the CNN scans the image in the "field of view" through the local receptive field. Each movement scans a different area and assigns corresponding parameters to learn the local features of the image. This is usually called the Convolution process.

In addition, in the task of image learning, the target object will usually be pixels with similar characteristics. Therefore, CNN uses weight sharing to share the weights between the convolution kernels to obtain the regional features of the image. However, an image usually contains more than one feature, so different windows (different filters or convolution kernels) are designed to scan the features.

Finally, according to the specific situation of the image task, average pooling or maximum pooling is used to retain the background information or contour information of the image. This step can significantly reduce the number of features of network learning and reduce the complexity of network learning. This operation is also called the down-sampling process.

In the PD diagnosis research based on hand drawing, this research set up a seven-layer network for CNN, as shown in figure 4.

![Figure 4. CNN structure based on PD hand drawing.](image)

First, design a layer of convolutional layer Conv1 with 80 10*10 two-dimensional convolution kernels to initially learn the local features of hand-drawn drawings. Since the hand-drawn drawing used is a color drawing, the channel is set to 3. It is worth noting that the image to be classified contains a large amount of background, so pass the image to the down-sampling layer Maxpooling1 for maximum pooling operation to preserve the contour features of the image. This model set the step size for convolution and pooling to 1 and 3, respectively. Considering that the image becomes smaller after the first layer of convolution and pooling operations, set up 40 5*5 convolutional layers Conv2 to further perform feature abstraction. Similarly, the maximum pooling MaxPooling2 is used to preserve the contour information of the image.

Furthermore, to learn more non-linear relationships between images and labels, set up two fully connected layers F1 and F2, to capture this relationship. In the meantime, the Flatten layer is set to connect convolution operation and fully connected operation. It is worth noting that the tanh activation function is beneficial to CNN convergence and achieves higher accuracy, so tanh is used to activate each layer of the CNN network.

2.2.2. Construction of residual neural network (ResNet). Due to too few data sets available for training and the large amount of feature information in the picture, ordinary CNNs are prone to overfitting, resulting in poor robustness. In addition, due to the process of backpropagation, the CNN is prone to vanishing gradients or network degradation. In order to solve this problem, ResNet was used for training. The ResNet introduces a residual network structure. The problems of gradient
disappearance and network degradation can be solved through this residual network structure, and a better classification effect can be achieved [9,10].

The residual network is composed of a series of residual blocks. A residual block can be expressed as:

\[ x_{t+1} = x_t + F(x_t, W_t) \]  \hspace{1cm} (3)

The residual block is divided into two parts, a direct mapping part and a residual part. \( x_{t+1} \) is the direct mapping. \( F(x_t, W_t) \) is the residual part, which is generally composed of two or three convolution operations.

In a convolutional network, the number of Feature Maps of \( x \) may differ from that of \( x \). At this time, use 1*1 convolution for dimensionality increase or dimensionality reduction. The residual block is expressed as:

\[ x_{t+1} = h(x_t) + F(x_t, W_t) \]
\[ h(x_t) = W'_t x_t \]  \hspace{1cm} (4)

Where \( W'_t \) is a 1*1 convolution operation. Since 1*1 convolution has a limited improvement in model performance, it is generally used when the dimensionality is increased or reduced.

Therefore, a building block of ResNet was construct, as shown in Figure 5.

![Figure 5. ResNet: a building block.](image)

As shown in figure 6, according to a structure similar to CNN, first, design a layer of convolutional layer Conv1 with 64*7*7 two-dimensional convolution kernels to initially learn the local features of the hand-drawn drawing. Then pass the image to the down-sampling layer Maxpooling1 for maximum pooling operation to preserve the contour features of the image. Next, construct three sets of residual blocks, as shown in Figure 5, to further perform feature abstraction. Similarly, the maximum pooling MaxPooling2 is used to preserve the contour information of the image. Finally, F's fully connected layer is designed to establish the relationship between the image and the label.

![Figure 6. ResNet structure model based on PD hand drawing.](image)

2.2.3. Neural network training. In this study, the protocol used to train the network is the backpropagation algorithm. The traditional stochastic gradient descent optimizer (SGD) is prone to frequent oscillations due to frequent updates. In addition, as one of the most commonly used optimizers in neural network training, SGD is easily limited to the saddle point, which makes the
model unable to achieve optimal performance. To solve this problem, this research adopted an adaptive moment estimation optimizer (Adam). The optimizer is not susceptible to learning speed and gradient size interference and can adjust the learning rate adaptively. In general, Adam can achieve more stable and faster convergence. Therefore, Adam is used in the neural network training of this study to update the weight \( W \) and bias \( b \) based on the deep learning classifier.

Specifically, the Adam algorithm calculates value correction based on the first and second momentum and only requires one step, which reduces the memory requirement \([11]\). This is computationally effective and is especially suitable for random optimization of deep learning algorithms. The update strategy of the network parameter \( W \) in the ith iteration can be expressed as:

\[
\begin{align*}
\hat{g}_i &= \frac{dL(W_{i-1})}{dW} \\
\omega_i &= \mu_1 \omega_{i-1} + (1 - \mu_1) g_i \\
\nu_i &= \mu_2 \nu_{i-1} + (1 - \mu_2) g_i^2 \\
\hat{W}_i &= \frac{\omega_i}{1 - \mu_1^i}, \hat{V}_i = \frac{\nu_i}{1 - \mu_2^i} \\
W_i &= W_{i-1} - \frac{\alpha \hat{W}_i}{\sqrt{\hat{V}_i} + \varepsilon}
\end{align*}
\]

Among them, \( W_i \) represents the bias parameter or network weight; \( L(W_{i-1}) \) corresponds to the loss function; \( \omega_i \) and \( \nu_i \) are first-order and second-order vectors, and \( \hat{W}_i \) and \( \hat{V}_i \) are correction values for moment estimation. \( \mu_1 \) and \( \mu_2 \in [0, 1) \) are the coefficients that determine the exponential decay rate of the moment estimate, and \( \alpha \) is the learning rate.

3. Results and Discussion

This chapter presents and analyzes the results of the research and the indicators used to evaluate the model's performance. To present more objective and fair results, use the average result of 10 times of 5-fold cross-validation to present the final results \([12]\).

3.1. Experimental results presentation indicators

In this study, the designed question is a binary classification problem, whether the diagnosis is PD. Firstly, based on this problem, accuracy (ACC) is an important measurement index, which is usually expressed by the percentage of the number of samples correctly classified to the number of all samples for diagnosis. In addition, for the diagnosis of diseases, missed detection is often more costly than misjudgment. Therefore, recall is also an important evaluation index for the performance of this type of classification model. It is defined as the probability that a sample suffering from PD is predicted to be PD. In addition, AUC (Area under the curve) as a model performance evaluation indicator is usually used to measure the model's generalization ability. It is calculated from the area under the ROC curve, and its meaning is shown in Table 1\([13]\).

| AUC       | Classifier performance |
|-----------|------------------------|
| 100%      | Perfect                |
| 85%~95%   | Very good              |
| 70%~85%   | General                |
| 50%~70%   | Lower                  |
| 50%       | Random guess           |
| 0~50%     | Lower than random guess|

The stability of deep learning has always been a concern because different initializations or gradient descent directions will lead to different evaluation results. Therefore, introduce the standard deviation (STD) to present the ACC stability performance of the developed deep learning scheme, which is calculated as follows:
\[
STD = \frac{1}{m} \sum_{i=1}^{m} \sqrt{\frac{1}{k} \sum_{j=1}^{k} (ACC_j - \overline{ACC}_i)^2}
\]  
(6)

Among them, \(m\) and \(k\) represent the number of experiments performed and the fold number of the cross-validation experiment.

### 3.2. Diagnosis results based on hand-drawn drawings

Based on the hand-drawn data set, this research provides diagnosis solutions of CNN and ResNet. Figure 7 shows the ACC changes during the training process of the two models. First of all, on the whole, CNN and ResNet's Acc both showed a steady upward trend. In terms of convergence speed, ResNet is better than CNN. And the ACC of CNN oscillates between 5 to 15 iterations, which indicates that the network has a certain degree of instability. The model has a risk of overfitting, while ResNet is relatively stable.

![Figure 7. Training Acc changes based on hand-drawn drawings.](image)

Furthermore, compare the Acc, AUC, RS, and STD of the model, as shown in Figure 8. It can be seen from the figure that ResNet is better than CNN in terms of the performance of ACC, AUC, and RS, while STD does not show obvious advantages. Table 2 presents a further comparison of numerical results. It can be found that for almost all indicators, ResNet has achieved more superior performance. It is worth noting that compared to CNN, ResNet's AUC is significantly better than CNN, which shows that ResNet can achieve more robust diagnostic performance. In general, ResNet achieves the comprehensive optimal performance based on the hand-drawing diagnosis of PD.

![Figure 8. Cross-validation performance comparison of training and testing based on hand-drawn drawings.](image)
Table 2. Comparison of numerical results between CNN and ResNet based on hand-drawn drawings.

|        | ACC  | AUC  | RS   | STD  |
|--------|------|------|------|------|
| ResNet | 88.9%| 94.5%| 87.4%| 3.8% |
| CNN    | 85.6%| 90.0%| 84.9%| 4.3% |

4. Conclusion

Accurate and convenient diagnosis of PD has always been a problem of concern. This study proposed a scheme for early diagnosis of PD based on hand-drawn drawings, in which patients only need to provide their hand-drawn drawings to diagnose whether they have PD. This is of great significance for remote, long-term, convenient, and low-cost monitoring of PD.

Based on the presented experimental results, the ResNet achieves the best overall classification performance based on hand drawings, and this performance is significantly superior to CNN. Besides, the convergence speed is also faster than CNN.

In general, this research has implemented a PD diagnosis scheme based on deep learning and has achieved certain research results. To assist doctors in diagnosing PD, the proposed scheme can be further developed into a real-time online learning system to learn the latest collected data to improve the model's generalization ability to new data. This can also further enrich the hand drawing database. In the case of sufficient data to participate in training, the number of layers of ResNet can be further increased to improve the diagnosis accuracy of PD.

References

[1] J. A. Obeso et al., "Past, present, and future of Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy," Mov Disord, vol. 32, no. 9, pp. 1264-1310, Sep 2017.
[2] A. Castrillo, C. Gilpolo, R. Gutiérrezríos, and A. Mendozarodríguez, "Neuroimaging in parkinsonisms: magnetic resonance," Sensors, vol. 11, no. 11, pp. 2972-81, 2011.
[3] J. F. Morley et al., "Optimizing olfactory testing for the diagnosis of Parkinson's disease: item analysis of the university of Pennsylvania smell identification test," npj Parkinson's Disease, vol. 4, no. 1, p. 2, 2018.
[4] R. Erika, M. Carlo, and C. Filippo, "How Wearable Sensors Can Support Parkinson's Disease Diagnosis and Treatment: A Systematic Review," Frontiers in Neuroscience, vol. 11, no. 11, 2017.
[5] G. M. Clark and J. A. G. Lum, "Procedural learning in Parkinson's disease, specific language impairment, dyslexia, schizophrenia, developmental coordination disorder, and autism spectrum disorders: A second-order meta-analysis," Brain Cogn, vol. 117, pp. 41-48, Oct 2017.
[6] C. R. Pereira et al., "A new computer vision-based approach to aid the diagnosis of Parkinson's disease," Comput Methods Programs Biomed, vol. 136, pp. 79-88, Nov 2016.
[7] L. Ali, C. Zhu, N. A. Golilarz, A. Javeed, M. Zhou, and Y. Liu, "Reliable Parkinson's Disease Detection by Analyzing Handwritten Drawings: Construction of an Unbiased Cascaded Learning System Based on Feature Selection and Adaptive Boosting Model," IEEE Access, vol. 7, pp. 116480-116489, 2019.
[8] H. Li et al., "Diagnosis of Parkinson's Disease with a hybrid feature selection algorithm based on a discrete artificial bee colony," 2020, p. 97.
[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.
[10] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," ArXiv, vol. abs/1603.05027, 2016.
[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
[12] T. Wong and N. Yang, "Dependency Analysis of Accuracy Estimates in k-Fold Cross Validation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 11, pp. 2417-2427, 2017.

[13] M. Luque-Fernandez, D. Redondo-Sánchez, and C. Maringe, "evauroc: Command to compute cross-validated area under the curve for ROC analysis after predictive modeling for binary outcomes," *The Stata Journal: Promoting communications on statistics and Stata*, vol. 19, pp. 615-625, 09/01 2019.