DEEP CONVOLUTIONAL NEURAL NETWORK FOR PARKINSON'S DISEASE BASED HANDWRITING SCREENING

Mohamed Shaban, Member IEEE
Electrical and Computer Engineering
University of South Alabama, Mobile, AL, USA

ABSTRACT

In this paper, the use of a fine-tuned VGG-19 for screening Parkinson's Disease (PD) based on a Kaggle handwriting dataset is investigated and experimented. The dataset including 102 wave and 102 spiral handwriting patterns was pre-processed where images were resized and a data augmentation based on image rotation was adopted to minimize overfitting. The Convolutional Neural Network (CNN) model was then trained on the pre-processed dataset and validated using both 4-fold and 10-fold cross validation techniques. The CNN model achieved an accuracy of 88%, 89%, and a sensitivity of 89%, 87% on the wave and spiral patterns respectively when a 10-fold cross validation was used. The proposed approach offers a promising solution for assessing and screening PD based on handwriting drawings and achieves a comparable high performance on the two different handwriting patterns as compared with the-state-of-the-art architecture that adopted a fine-tuned AlexNet.

Index Terms— Parkinson's Disease, Computer Aided Diagnosis, Deep Learning, Convolutional Neural Networks.

1. INTRODUCTION

Parkinson's disease (PD) is a neuro-degenerative disorder that is characterized by motor symptoms such as slowness of movement, and tremors, and non-motor symptoms such as cognitive changes, anxiety, depression, and sleep problems [1]. Early diagnosis of PD is critical to maintain an acceptable quality of life for PD patients using appropriate medical treatments. Diagnosis, and evaluation of the disease is currently done using the Unified Parkinson Disease Rating Scale (UPDRS) which consists of 42 items in 4 subscales making the diagnosis very challenging and subjective [2].

Machine, and deep learning methods have been recently introduced for the detection, and classification of PD based on speech, and handwriting patterns [3-8]. In [3], support vector machine was used to classify speech signals of 33 PD patients, and 10 controls based on 132 dysphonia measures with an accuracy of 99%. In [4], a Convolutional Neural Network (CNN) was used to classify raw speech on a relatively small training set. In [5], feature selection, and support vector machine methods were used to distinguish between 37 PD patients, and 38 controls based on the handwriting movements. In [6] [7], a handwriting dataset (i.e. "HandPD") was introduced, and a fine-tuned AlexNet was adopted to classify subjects into PD, and Controls with a best-case accuracy of 79.6% to 83.1% and 89.6% to 90.4% on meander, and spiral patterns respectively. Further, in [7] few images were used for training and testing the model (i.e. 132 images) providing a less model generalization. In [8], several CNN architectures were developed for classifying handwriting dynamics obtained from a smart pen equipped with a series of sensors for 224 PD patients, and 84 controls.

In this paper, a pre-processing stage followed by a fine-tuned VGG-19 was introduced to analyze, and classify spiral, and wave handwriting datasets into controls, and PD in an automated fashion. The CNN model provides a consistent performance for PD screening based on the two types of handwriting drawings as compared with the-state-of-the-art architecture [7] where the performance of the model works better on the spiral pattern with respect to the meander pattern.

2. MATERIALS AND METHODS

2.1 Dataset Description

Dataset used in the current study was analyzed by Zham et al. in [9], and published on Kaggle [10].

The dataset is composed of 102 spiral, and 102 wave handwriting drawings for fifty-five age, and gender matched control subjects, and PD patients. PD patients representing various stages of PD, and including severe PD where UPDRS > 24 were recruited from PD outpatient clinic at Dandenong Neurology, Melbourne, Australia. Controls (UPDRS = 0) were recruited from multiple age-care facilities using word-of-mouth, and poster ads. All subjects are right handed. Examples of the dataset samples are shown in fig. 1.

978-1-7281-7401-3/20/$31.00 ©2020 IEEE

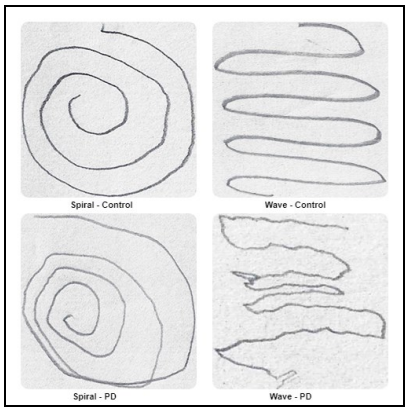

Fig. 1. Spiral, and Wave Handwriting Drawings for Control, and PD Subjects

2.2 CNN Model Description

In this paper, a fine-tuned VGG-19 architecture was adopted to classify spiral, and wave handwriting images into control, and PD as shown in fig. 2. A set of 100 images was selected from each of the spiral and wave datasets.

The proposed architecture starts with a pre-processing stage where images were resized to a standard size of 224×224×3. Since the size of the datasets is relatively small, a data augmentation approach was then deployed in order to enhance the proposed classifier performance where each image was rotated by 90 degrees. The approach was then repeated for three times creating datasets of a larger size of 400 each. The 90 degrees' image rotation operation was selected for expanding and augmenting the dataset as no details were lost using such operation as compared to other operations including scaling, cropping or translation.

Images were then processed by a set of five consecutive stages where each stage is composed of two, and four convolutional layers for the first two, and last three stages respectively. If *x(i,j,k)* is the input to a convolutional layer, then the output or feature map generated can be represented as follows:

$$y(l,m,n)=\sum_{k=1}^{K}\sum_{i=1}^{I}\sum_{j=1}^{J}w(l,i,j,k)x(i+m-1,j+n-1,k)+b(l) \quad (1)$$

where *w*, and *b* are the weights and biases. *I, J,* and *K* are the width, height, and the depth (*K = 1* for gray scale handwriting patterns) of the input respectively. The number of convolutions or filters used in each stage are 64, 128, 256, 512, and 512 respectively where each filter has a size of 3×3×3.

A single 2×2 max pooling layer was deployed between consecutive stages in order to subsample the feature maps by 2 reducing the computational complexity of the model. The output of the five stages was then fed to a set of two consecutive fully connected layers with 4096 neurons each. The last fully connected layer of the original pre-trained VGG-19 was replaced with a two-neurons layer, and the final model with parameters that require fine-tuning was trained on the Kaggle dataset [10].

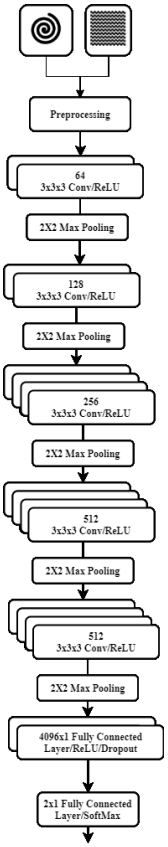

Fig. 2. Fine-Tuned VGG-19 Model

Rectified Linear Unit (ReLU) activation function was used for convolutional, and early fully connected layers while the SoftMax function was deployed at the last fully connected layer for non-linear classification. Both ReLU and SoftMax activations can be represented as follows:

$$a_i=\begin{cases}z_i & z_i>0\\ 0 & z_i<0\end{cases} \quad (2)$$

$$d_i=\frac{e^{-c_i}}{\sum_{j=0}^{2}e^{-c_j}} \quad (3)$$

where z_i is an input to the ReLU, and a_i is the corresponding activation generated by the ReLU. Also, c_i is the i^{th} output of the last fully connected layer, and d_i is the corresponding SoftMax activation. Almost 50% of the outputs of the first two fully connected layers were randomly dropped to further minimize overfitting and improve the robustness of the architecture.

The cross entropy loss *e* representing the difference between the predicted and desired activations of the

SoftMax was then measured. The loss can be defined as follows:

$$e = -\sum_{j=0}^{L} \hat{d}_j \log(d_j) \quad (4)$$

where $\hat{d}_j$ is the expected desired probability. The cross entropy loss was then minimized using the Stochastic Gradient Descent (SGD). The previously described optimization approach is widely known as the backpropagation algorithm.

2.3 Evaluation Metrics

In this paper, the confusion matrix, the sensitivity, specificity, validation accuracy, Receiver Operating Characteristic Curve (ROC), and Area Under the Curve (AUC) were used to evaluate the performance of the CNN model. Both 4-fold and 10-fold cross validation methods were used to test the classifier performance.

First, sensitivity and specificity of the proposed architecture are defined as follows:

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

where *TP, FP, TN* and *FN* are the true positive, false positive, true negative, and false negative counts respectively. Further, accuracy is defined as the number of the correctly classified images divided by the total number of images of the validation dataset.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

ROC was also represented to visualize the ability of the classifier to distinguish between controls and PD. ROC represents the relationship between the true positive rate (i.e. detection probability) and the false positive rate (i.e. false alarm probability) at various threshold settings. Based on ROC, AUC was measured to determine the separability of the classifier to better evaluate the ability of the CNN architecture to differentiate between controls, and PD.

3. EXPERIMENTAL RESULTS

A total of 200 labelled spiral, and wave images were acquired from the Kaggle dataset [10]. Images were resized to a standard dimension of 224×224×3, and augmented such that a total of 400 images are available in each of the spiral, and wave datasets. Both 4-fold, and 10-fold cross validation techniques were deployed. Datasets were further split into batches of 50 images, and fed to the proposed architecture. The backpropagation algorithm was then executed with a learning rate of 10^{-3} for 10 epochs, and 8 epochs when 4-fold, and 10-fold cross validation approaches were used respectively. Both training loss, and training accuracy were monitored to optimize the model parameters where the model training was terminated once the desired validation accuracy was achieved. Table 1 shows the confusion matrix for both datasets when 4-fold, and 10-fold cross validation were used.

Table 1. Confusion Matrix of the CNN Model for 4-Fold, and 10-Fold Cross Validation on (a) Spiral Dataset (b) Wave Dataset

(a)

4-Fold	No PD	PD	10-Fold	No PD	PD
No PD	57	1	**No PD**	22	0
PD	7	35	**PD**	3	15

(b)

4-Fold	No PD	PD	10-Fold	No PD	PD
No PD	51	2	**No PD**	18	0
PD	6	41	**PD**	0	22

The ROC curves for the 4-fold, and 10-fold cross validation of the CNN model when applied on spiral, and wave datasets are shown in fig. 3. It is generally clear that a low false alarm probability (< 0.1) is achieved with a relatively high detection probability (> 0.8) at a certain threshold setting.

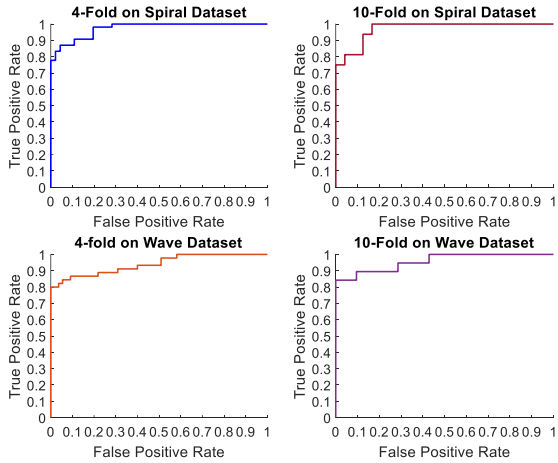

Fig. 3 Receiver Operating Characteristic Curve for 4-fold, and 10-fold Cross Validation on Spiral, and Wave Datasets

Table 2 shows the validation accuracy, sensitivity, specificity, and AUC of the proposed architecture when 4-fold, and 10-fold cross validation were applied on both the spiral, and wave datasets. The AUC was almost 93%, and 94% (4-fold), 89%, and 92% (10-fold) for the wave, and spiral datasets respectively. Further, the CNN model showed an elevated sensitivity of 87% when applied on the spiral dataset while achieved 89% sensitivity on the wave dataset. The proposed classifier is then able to accurately detect both control and PD subjects to a great extent.

A validation accuracy of almost 87% (4-fold), and 88% (10-fold) was reached. Since the-state-of-the-art architectures [5-8] use different handwriting datasets with respect to the proposed architecture, the obtained results

may not be comparable. However, the CNN model shows a promising advantage and comparable consistent performance when applied on both datasets collected from PD patients as compared with the fine-tuned AlexNet used in [7] which shows a relatively higher performance on the spiral dataset as compared with the meanders dataset.

Table 2. Performance Measure of the CNN Model for 4-Fold, and 10-Fold Cross Validation on Spiral, and Wave Datasets

	Spiral (4-Fold)	Spiral (10-Fold)	Wave (4-Fold)	Wave (10-Fold)
Accuracy	86.5%	88.5%	87.3%	88%
Sensitivity	86.7%	86.5%	88.7%	89.2%
Specificity	85.1%	92.2%	85.8%	87.9%
AUC	94.1%	91.6%	92.6%	88.6%

Although the aforementioned results are promising, they basically rely on handwriting exams that may not be helpful for early stage PD diagnosis where tremors, and bradykinesia, are usually diagnosed when most of the nerve cells are degenerated [1].

Electroencephalography (EEG) which measures the brain activity within a certain period of time is often used in the diagnosis of brain diseases such as epilepsy, tumors, stroke as well as study of the cognitive dysfunction in neuro-degenerative disorders such as PD. Frequency, and power analysis of EEG signals may contain useful information about motor, and non-motor dysfunction of PD by exploiting delta [1-4 Hz], theta [4-8 Hz], alpha [8-12 Hz], beta [12-35 Hz], and gamma [35-45 Hz] waves [13-16]. Accordingly, as a future plan, a deep learning approach will be introduced to automate EEG analysis, and extract various unique high-level features that will be further visualized, and analyzed to help in identifying the early biomarkers of the disease.

4. CONCLUSIONS

A fine-tuned VGG-19 was used to classify public Kaggle spiral, and wave handwriting datasets [10] into control or PD. Handwriting images were preprocessed, and then fed to a fine-tuned VGG-19 where the last fully connected layer with associated pretrained parameters was replaced with a two-neurons layer.

Performance of the CNN model was studied using validation accuracy, sensitivity, and specificity where the proposed architecture attained an accuracy of 87% (4-fold), and 88% (10-fold) with a robust sensitivity of 87% (spiral dataset), and 89% (wave dataset). ROC curves were provided showing an elevated AUC of 93%, and 94% (4-fold), and 89%, and 92% (10-fold) when applied on both the wave, and spiral datasets respectively. While the datasets used in this study and [7] are different, the proposed approach provides a comparable accuracy and performance when applied on both datasets of interest as compared with [7] that used a fine-tuned AlexNet with a remarkable deviation in performance on the respective datasets.

REFERENCES

[1] William Dauer, and Serge Przedborski, Parkinson's disease: mechanisms, and models, *Neuron,* Vol. 39, Issue 6, PP. 889-909, 2003.

[2] Joel S. Perlmutter, Assessment of Parkinson Disease Manifestations, *Current Protocols in Neuroscience*, Vol. 49, Issue 1, PP. 10.1.1-10.1.14, 2009.

[3] Athanasios Tsanas, Max A. Little, Patrick E. McSharry, and Lorraine O. Ramig, Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests, *IEEE Transactions on Biomedical Engineering,* Vol. 57, Issue 4, PP. 884-893, 2010.

[4] Alex Frid, Ariel Kantor, Dimitri Svechin, and Larry M. Manevitz, Diagnosis of Parkinson's Disease from Continuous Speech using Deep Convolutional Networks without Manual Selection of Features, *IEEE International Conference on the Science of Electrical Engineering*, Eilat, Israel, 2016.

[5] Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdenek Smékal, and Marcos Faundez-Zanuy, Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease, *Computer Methods, and Programs in Biomedicine*, Vol. 117, Issue 3, PP. 405-411, 2014.

[6] Clayton R. Pereira, Danillo R. Pereira, Francisco A. da Silva, Christian Hook, Silke A.T. Weber, Luis A.M. Pereira, and Joao P. Papa, A Step Towards the Automated Diagnosis of Parkinson's Disease: Analyzing Handwriting Movements, *Annual IEEE Symposium on Computer-Based Medical Systems*, Sao Carlos, Brazil, 2015.

[7] Clayton R. Pereira, Danillo R. Pereira, Joao P. Papa, Gustavo H. Rosa, and Xin-She Yang, Convolutional Neural Networks Applied for Parkinson's Disease Identification, *Machine Learning for Health Informatics*, PP. 377-390, 2016.

[8] Clayton R. Pereira, Danillo R. Pereira, Gustavo H. Rosa, Victor H.C. Albuquerque, Silke A.T. Weber, Christian Hook, and Joao P. Papa, Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification, *Artificial Intelligence in Medicine*, Vol. 87, PP. 67-77, 2018.

[9] Zham P, Kumar DK, Dabnichki P, Poosapadi Arjunan S, and Raghav S, Distinguishing Different Stages of Parkinson's Disease Using Composite Index of Speed and Pen-Pressure of Sketching a Spiral, *Frontiers in Neurology,* Vol. 8, 2007.

[10] Kevin Mader, "Parkinson's Drawing", 2019. [Online]. Available: https://www.kaggle.com/kmader/parkinsons-drawings.