

Journal Pre-proof

Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease

Iqra Kamran, Saeeda Naz, Imran Razzak, Muhammad Imran



PII: S0167-739X(20)33044-2

DOI: <https://doi.org/10.1016/j.future.2020.11.020>

Reference: FUTURE 5928

To appear in: *Future Generation Computer Systems*

Received date: 26 November 2019

Revised date: 6 October 2020

Accepted date: 19 November 2020

Please cite this article as: I. Kamran, S. Naz, I. Razzak et al., Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease, *Future Generation Computer Systems* (2020), doi: <https://doi.org/10.1016/j.future.2020.11.020>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Elsevier B.V. All rights reserved.

Highlights

- In order to improve the identification performance, we combined different Parkinson's dataset and applied deep transfer learning based algorithms.
- We demonstrate its effectiveness by achieving excellent Parkinson identification performance with a value of 99.22% on illuminated task of combined HandPD, NewHandPD and Parkinson's Drawing datasets, which indicates that our method outperforms in comparison to state-of-the-art.
- Furthermore, we analyze different deep transfer learning structure to discover the vital elements in the success of our proposed approach.

Handwriting Dynamics Assessment Using Deep Neural Network for Early Identification of Parkinsons Disease

Iqra Kamran, Saeeda Naz, Imran Razzak, Muhammad Imran

Computer Science Department, Govt Girls Postgraduate College No.1, Abbottabad, Pakistan

School of Information Technology, Deakin University, Geelong, Australia

College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia

Email: {iqrakamran547, saeedanaz292}@gmail.com, {imran.razzak, dr.m.imran}@ieee.org

Abstract—The etiology of Parkinsons disease (PD) remains unclear. Symptoms usually appear after approximately 70% of dopamine-producing cells have stopped working normally. PD cannot be cured, but its symptoms can be managed to delay its progression. Evidence suggests that early diagnosis is important in establishing an effective pathway for management of symptoms. However, PD diagnosis is challenging, particularly in the early stages of the disease. In this paper, we present a method for early diagnosis of PD using patients handwriting samples. To improve performance, we combined multiple PD handwriting datasets and used deep transfer learning-based algorithms to overcome the challenge of high variability in the handwritten material. Our approach achieved excellent PD identification performance with 99.22% accuracy on illuminated task of combined HandPD, NewHandPD and Parkinsons Drawing datasets, demonstrating the superiority of our approach over current state-of-the-art methods.

Keywords: Parkinson's, Neurological disorder, Brain disorder, PD identification, Transfer learning

1. Introduction

Parkinsons disease (PD) is a progressive neurological condition affecting more than 10 million people worldwide, including 100,000 and 145,000 people in Australia and the UK, respectively. It is one of the costliest diseases, with a burden of approximately \$23,000 per patient per patient. The symptoms of PD arise from decreasing levels of the chemical messenger dopamine, occurring due to gradual loss of dopamine-producing nerve cells (neurons) in the brain. Symptoms typically start to appear when the majority of dopamine-producing cells (about 70%) have stopped working normally. PD cannot be cured but its progression can be delayed by management of symptoms, i.e., people with PD can live independent and productive lives if treated appropriately with a combination of medication and multidisciplinary support including speech and occupational therapy, dietetics, and physiotherapy.

Despite its pathological features being well-established, the etiology of PD and related diseases, and appropriate disease-modifying treatments remain elusive. PD is associated with multiple symptoms, with the most prominent

indicators being tremor, muscle stiffness (rigidity), slowness of movement (bradykinesia), and postural instability. Other symptoms can include memory loss, mild cognitive impairment, sleep disruption, anxiety, and depression [1], negatively affecting patients daily lives. PD has a male preponderance and develops gradually, most commonly appearing between the ages of 50 and 60, although so-called young onset PD can occur in people (10-20%) younger than 50 years of age. Currently, diagnosis is based on performance in tasks such as those involving facial expression [2], [3], writing [4], [5] drawing [6], walking [7]–[9], and speaking [10], [11].

PD is a progressive neurological condition whose etiology remains unclear. Various stages of PD have been defined to help determine its progression and associated loss of neurons. Patients often ask what stage of PD they are in, how far their symptoms have progressed, and what they can expect to happen next. The range, order of onset, and intensity of PD symptoms varies significantly between individuals. Stage of PD progression is commonly assessed using two rating scales: The Unified Parkinsons Disease Rating Scale (UPDRS) and the Hoehn and Yahr (H-Y) scale. In first stage, PD symptoms typically affect only one side of the body, progressing to both sides of body in the second stage. In the third stage of PD, movement is affected. In the last two stages, PD patients are unable to perform daily activities without assistance. PD diagnosis is a complex and challenging task, yet early-stage diagnosis is important. Existing diagnostic tools include the MiniMental State Examination [12] questionnaire used to evaluate PD disability, the UPDRS [13], and brain scans [14]. These methods are expensive and require significant expertise to administer. Therefore, research has moved towards development of automated systems that can employ algorithmic methods to differentiate people with PD from healthy individuals.

Currently, diagnosis is based on assessment of performance in activities such facial expression [2], [3], writing [4], [5], drawing [6], walking [7]–[9] and speaking [10], [11]. People with PD have difficulty controlling body movements due to changes in neuronal mechanisms. This can negatively affect fine motor skills, such as those used in writing. Hence, handwriting changes can be an early indicator of PD. In the early stages of the disease, changes

in handwriting are small and barely noticeable. Thus, it is important to detect such small changes in writing patterns as they can signal the presence of subclinical PD. Cramped handwriting (micrographia), or sudden changes in writing size (often becoming smaller) could indicate early PD. The handwriting of PD patients is affected by shaking, stiffness, slow movement, and imbalance. Several previous studies have reported handwriting impairment in the majority of PD patients, hence it is considered an important biomarker for PD identification.

On the basis of such previous research findings, handwriting analysis is known to be a successful diagnostic strategy for PD, compared to more costly and time-consuming neurological tests and brain scans. Analysis of handwriting by experts is a widely used approach for identifying PD. Thus, a current focus of computer science research is automation of this process. However, due to a lack of sufficient data, and the high variability of PD, both traditional machine learning and deep learning methods have been unable to provide accurate results. Recently, the performance improvements offered by transfer learning have made it an essential tool in many applications [15]–[18]. Usually in transfer learning the feature layers of a network pre-trained for a given source task are transplanted into a second target network which is then fine-tuned for a different source task by merely refining the original model using target data for the second task. In this paper, we performed PD identification on handwritten samples from PD patients using deep transfer learning-based techniques. Extensive experimental analysis confirmed that our proposed approach considerably improves diagnostic performance, compared to state-of-the-art anomaly detection methods. The **key contributions** of this work are as follows

- We present an end-to-end deep transfer learning method to transfer pre-learnt knowledge onto handwritten samples from PD patients to develop a tool for early identification of PD.
- We investigate early PD identification through handwriting using different architectures of deep transfer convolutional neural network (CNN) to achieve state-of-the-art performance.
- To improve the overall classification performance and achieve effective results for early identification, we increased the input space by combining different PD handwritten datasets, and applied various data augmentation techniques.
- We investigated different types of handwriting task performed by PD patients to determine the optimal task for disease detection.
- We performed an extensive evaluation study, demonstrating that our proposed system considerably outperforms current state-of-the-art techniques.

The rest of this paper is structured as follows. Section 2 reviews the existing literature on PD identification. Section 3 provides a description of the datasets, data augmentation techniques, and the proposed architectures and network parameters selection used in our approach. Section 4 summarizes our experimental setup, and presents our results and a comparative evaluation. Finally, section 5 concludes the paper and provides directions for future work.

2. Related Work

In recent times, use of machine learning and deep neural networks in various clinical applications has gained increasing popularity. There has been significant research interest in development of automated systems for early-stage identification of PD based on voice, gait, and handwriting data. In 2003, Gemmert et al. [19] collected handwriting samples from 13 healthy and 13 non-healthy controls. In this study, individuals were asked to write two patterns, such as “lllllll” and “lililili” without the dots in five different sizes (1.0, 1.5, 2.0, 3.0, and 5.0 cm). Analysis indicated an increase in stroke duration in graphic assignments by PD patients. The authors also estimated stroke amplitude to be reduced when focused on stroke estimate necessities are larger than 1.5 cm. Since 2003, further efforts in this domain have employed traditional machine learning algorithm and deep learning models.

Drotar has presented several works relating to PD identification [20]–[25]. In [20], Drotar et al. classified PD by selecting features using the Mann-Whitney U-test filter and the Relief algorithm. They considered on-surface and in-air movements. A support vector machine (SVM) classifier employed for PD identification yielded an overall accuracy of 80.09%. In another work [22], the authors extracted kinematic features and fed them to an SVM classifier, obtaining 79.4% accuracy using the Parkinson’s Disease Handwriting (PaHaW) dataset. The same authors [21] extracted kinematic features (speed, velocity, acceleration, stroke speed, jerk, etc.) using the PaHaW dataset. An SVM used to identify healthy and non-healthy subjects achieved an overall accuracy of 78% and 84%, respectively. In another work, Drotar et al. extracted handwriting features using the Relief algorithm and then used an SVM model for PD identification. A ten-fold cross-validation approach was used for classification, obtaining an average accuracy of 88.13% with 89.47% sensitivity and 91.89% specificity. In another work [23], the same authors extracted pressure and spatiotemporal features, i.e., stroke height and width, using the PaHaW dataset. They considered on-surface and in-air movements for seven different tasks. The SVM classifier obtained 89% accuracy. In [25], Drotar et al. further applied KNN (K-nearest neighbor), AdaBoost, and an SVM classifier for PD classification, with the SVM classifier yielding the highest recognition rate of 81.3%, compared to AdaBoost (78.9%) and KNN (71.7%). They concluded kinematic and pressure features to be useful for PD diagnosis.

Pereira et al. created the offline HandPD dataset in two sets of studies [26], [27]. The handwritten samples were collected from 55 subjects including 18 healthy and 37 PD subjects. In [27], offline features relating to spatial, tremor, and mean relative measures were extracted from HandPD (90:10 train/test split) and fed to nave Bayes, SVM, and optimum-path forest (OPF) classifiers for identification. A

TABLE 1: Summary of deep learning based methods for Parkinson disease prediction

Study	Features	Classifier	Dataset	Accuracy (%)
Drotar et al. [20]	Kinematics features	SVM	Parkinson's handwriting data	80.09
Drotar et al. [21]	Kinematics features	SVM	PaHaW	79.4
Drotar et al. [22]	Kinematics features	SVM	PaHaW	78 , 84
Drotar et al. [23]	Kinematics and spatio-temporal features	SVM classifier with radial Gaussian kernel	PaHaW	88.13
Drotar et al. [24]	Kinematics features	SVM with rbf kernel	PaHaW	89
Drotar et al. [25]	Kinematics and pressure features	SVM, AdaBoost, KNN	PaHaW	81.3 , 78.9 , 71.7
Pereira et al. [26]	Offline manual features	Naive Bayes, Optimum path forest, SVM	HandPD	78.9, 77.1, 75.8
Pereira et al. [6]	time series based features	CNN	HandPD	93.50
Pereira et al. [27]	Pen based features	CNN	HandPD	80.19
Pereira et al. [28]	Pen based features	CNN	HandPD	90.38
Eskofier et al. [29]	CNN based features	CNN	IMU data	90.9
Caliskan et al. [30]	CNN based features	DNN	Voice data	68.75
Choi et al. [31]	CNN based features	CNN	PPMI and SNUH cohort	90.7
Zhang et al. [32]	Time frequency features	DNN with (KNN)	Speech data	90.53%
Moetesum et al. [33]	CNN based features	SVM	PaHaW	83
Naseer et al. [4]	CNN based features	AlexNet	PaHaW	98.28
Sakar et al. [34]	Offline manual features	SVM , KNN	Speech data	77.50
Gupta et al. [35]	Kinematic, entropic and energetic features	SVM	PaHaW	83.75
Senatore et al. [36]	Offline manual features	Cartesian genetic programing	HandPD	72.36
Impedovo et al. [37]	Velocity based signal features	SVM along with linear kernel	PaHaW	98.44
Afonso et al. [38]	CNN based features	Convolutional Neural Network	HandPD	87
Bernardo et al. [39]	Offline manual features	OPF, SVM and Bayesian classifier	cube, triangle, and an Archimedean spiral	100

maximum average accuracy of up to 78.9% was reported for the nave Bayes classifier. Pereira et al. [27] extracted pen-based features and passed them to a CNN. They performed multiple experiments for each class, and also combined multiple classes such as spiral and meander samples. They achieved an 80.19% recognition rate on the spiral dataset using the CNN. They performed experiments using two splits and two different image resolutions (64 x 64 and 128 x 128 pixels). The first training/test split was 75:25, whereas in the other experiment it was 50:50. In another study, Pereira et al. [28] used the same handwriting dataset with metaheuristic methods, namely the Bat algorithm (BA), firey algorithm (FA), and molecule swarm optimization, with a 30:50:20 training/test/validation split, and employing a CNN for classification. Features extracted by systems fine-tuned by the metaheuristic strategies (BA, PSO, FA) were used for classification. The proposed framework achieved general exactness of 90.385%. Another attempt by the same authors in [6] employed CNN networks using the HandPD dataset and achieved 95% accuracy.

Isenkul and Sakar [40] developed a PD dataset using a graphics tablet to collect handwriting samples at the Department of Neurology, Istanbul University. The static spiral test (SST) and dynamic spiral test (DST) were analyzed based on subjects speed and increasing speed of drawing. Analysis of the dataset illustrated that increasing speed in the SST is statistically closer to that of DST for control

subjects compared to PD patients. It is thought that the two tests can be combined to measure individuals motor performance. Another study was presented by Sakar et al. [34] in 2013, in which they developed a PD voice dataset by collecting different voice samples from 20 subjects with PD and 20 healthy individuals at the Department of Neurology, Istanbul University. They extracted voice features and passed them to SVM and KNN classifiers. The SVM classifier yielded the best performance of up to 77.50%. Caliskan et al. [30] used the two voice dataset, i.e., OPD and PSD. The authors proposed a DNN (deep neural network) classifier for PD detection. They divided the dataset into two parts with a 70:30 training/test split. Creators compared comes about of profound and ordinary models. The DNN classified the OPD and PSD datasets with a precision of 93.79% and 68.05%, respectively, whereas SVM, decision tree, and credulous Bayes had accuracy scores of 85.780%, 84.371%, and 69.64%, respectively.

Choi et al. [31] proposed an automated profound learning-based FP CIT SPECT interpretation system using PPMI and SNUH cohort datasets. The PPMI dataset consists of 431 PD patients and 193 healthy subjects, while SNUH consists of 72 PD patients and 10 healthy subjects. The CNN used for classification achieved overall accuracy of 90.7%. Eskofier et al. [29] proposed a CNN classifier for confirmation of PD using IMU data. The authors compared machine learning techniques with CNN-based methods. CNN perfor-

mance was superior to the machine learning techniques by 4.6%. The CNN achieved an accuracy of 90.9%, whereas the other classifiers including AdaBoost.M1, KNN, PART, and SVM achieved an accuracy of 86.3%, 67.1%, 81.7%, 85.6%, respectively. Zhang et al. [32] applied a KNN classifier to the speech records, having removed noise, jitter, shine, and voice pitch to improve signal quality. The proposed method discriminated subjects with PD from healthy subjects with an accuracy of 90.53%. Gupta et al. [35] proposed a sex-specific and age-dependent classification technique for PD. They employed the PaHaW dataset and considered only seven tasks, extracting kinematic, entropy, and energetic features. SVM ranking techniques were used for classification, achieving an accuracy of 83.75%. Senatore et al. [36] classified PD using a technique known as Cartesian genetic programming (CGP). The authors compared their results with other techniques, concluding that the proposed system performed well than OPF and SVM. They used the HandPD dataset, finding that CGP performed well for meander, achieving a higher global accuracy of 72.36%. Impedovo et al. [37] used the PaHaW dataset consisting of eight different tasks, and extracted velocity-based signal features. An SVM was used with a linear kernel was used for PD diagnosis, achieving 98.44% accuracy.

Moetesum et al. [33] used fusion techniques (early fusion and late fusion) to improve classification. They assessed the visual properties of handwriting to identify PD using an AlexNet fusion with an SVM classifier to obtain 83% accuracy. The authors evaluated their proposed model using the PaHaW dataset. Subsequently, Naseer et al. [4] used the PaHaW dataset with the ImageNet and MNIST datasets, augmenting data from the PaHaW dataset with turns, horizontal and vertical mirror images, and image sharpening. They employed fine-tune and freeze transfer learning approaches and achieved the highest recognition rate of 98.28% using a CNN with AlexNet architecture, with the ImageNet dataset as the source domain and PaHaW dataset as the destination domain. Afonso et al. [38] proposed the application of a recurrence plot to map signals fed into a CNN for learning appropriate data. They used the HandPD dataset with the same splits and image resolution as Pereira et al. [27]. They performed experiments using two splits (75:25 and 50:50) and two different image resolutions (64 x 64 and 128 x 128 pixels) and achieved an average accuracy of 87%. Bernardo et al. [39] classified PD using machine learning methods including OPF, SVM, and a Bayesian classifier. They performed experiments on three different drawings: Cubes (76 examples), triangles (85 examples), and Archimedean spirals (80 examples). The SVM classifier obtained an accuracy of 100%, compared to OPF and the Bayesian classifier. Related works on PD are summarized in Table 1.

3. Deep Parkinson's Identification System

PD cannot be cured, but its symptoms can be managed, helping to delay disease progression. Early diagnosis is important, not only for treatment of fluctuations in motor

function, but also to help establish an optimal medication regimen. Compared to other healthcare applications of deep learning, there is much less data available for PD. Furthermore, PD symptoms are highly variable. Nevertheless, since deep models produce generalizable features, existing out-of-distribution approaches are transferrable to the PD identification problem. Thus, in this work, we employed pre-trained models and fine-tuned them to improve their identification performance

Figure 1 shows the general workflow of our proposed deep PD identification system. The overall proposed system is divided into pre-processing, data augmentation, and CNN architectures for feature learning, classification, and output. We applied several different data augmentation techniques including contrast, illumination, thresholding, flipping, and rotation to increase the sample space of the dataset, followed by convolutional layers and max-pooling layers to extract salient visual features. Finally, the features were fine-tuned on a PD dataset. In the following section, we first describe the datasets, followed by a detailed explanation of the learning parameters and transfer learning framework.

3.1. Dataset and Data-Augmentation Techniques

As discussed in an earlier section, a major challenge in deep learning for PD detection is lack of data. The current approach for early PD identification is based on motor symptoms. Thus, obtaining large datasets is quite challenging. Various handwriting datasets exist, but the number of subjects in each study (shown in Table 2) is small, irrespective of the high variability of PD symptom intensity. In the current work, we implemented a data augmentation strategy (thresholding, flipping, rotation, illumination, and contrast) to enable a significant increase in the diversity of PD handwritten samples. The overall statistics and number of samples after augmentation are shown in Table 4. Furthermore, to overcome the challenge of a small dataset, we combined several well-known PD handwritten datasets, simultaneously increasing the sample size and expanding the range of variation in the different datasets.

3.1.1. Parkinson's Handwriting Datasets. Due to the changes in functionality of neuronal mechanisms, patients with PD have problems controlling body movement, and thus face difficulties in performing motor-based tasks such as writing. In the current work, we used four different handwriting datasets (PaHaW dataset¹, HandPD dataset², NewHandPD dataset and Parkinsons Drawing Dataset³) to differentiate people with PD from healthy subjects.

The **PaHaW dataset** comprises multiple handwriting samples collected in [22], [25] using digitizing tablets from 75 subjects (37/38 PD/healthy subjects). spiral drawing, "l", "le", "les", "lektorka", "porovnat", "nepopadnout" and "Tramvaj dnes uz nepo-jede". Two spiral samples from

1. <https://bdalab.utko.feec.vutbr.cz/>

2. <http://www.fc.unesp.br/papa/pub/datasets/Handpd/>

3. <https://www.kaggle.com/kmader/parkinsons-drawings>

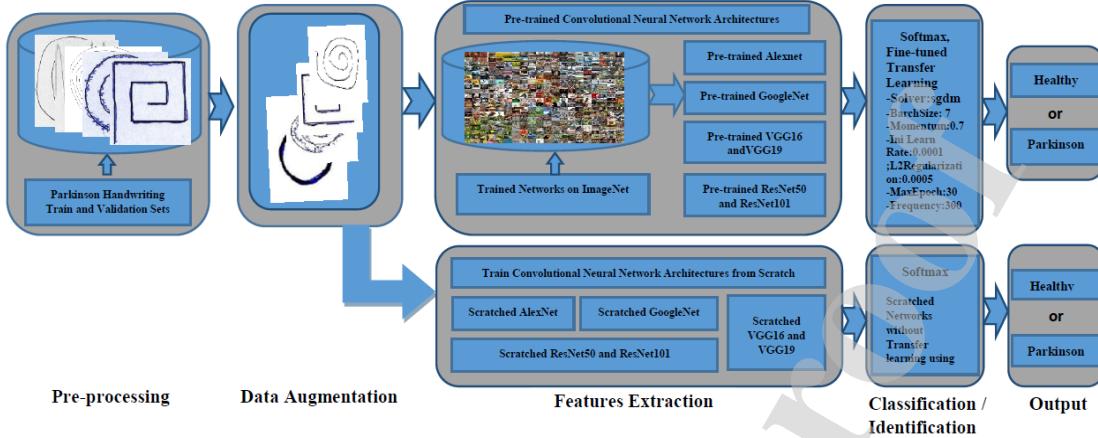


Figure 1: The Proposed Workflow of Automated Identification of Parkinson’s Disease at Early Stage using Fine Tuning Approach of Transfer Learning or Learning through Scratched Weights and Layers using Different CNN Architectures.

healthy individuals and one spiral sample from a PD subject were not completed by those individuals. The PaHaW dataset was collected on digitizing tablets (signals recorded using an Intuos 4M pen of frequency 200 Hz) so does not include images. Statistics for the PaHaW dataset are shown in Table 2. In the current work, we transformed the signal data into image representations. A recent study showed that use of spiral drawing can results in better identification of PD compared to others writing tasks [4]. Thus, we only included the spiral drawing samples from the PaHaW dataset.

HandPD is a fairly large dataset, containing samples from 92 subjects (74/18 PD/healthy subjects) [41]⁴ performing spiral and meander tasks. It includes a total of 736 images, comprising 368 spirals (296/72 PD/healthy subjects) and 368 meanders (296/72 PD/healthy subjects). Statistics for the HandPD dataset are shown in Table 2. PD patients were right-handed (69 subjects) and left-handed (5 subjects), and ranged from 38 to 78 years of age. Participants were asked to fill out a form that included both the spiral and meander drawings. The dataset includes noise, which affects classification performance.

The **NewHandPD** dataset consists of data from 66 individuals comprising 31 PD patients and 35 healthy individuals [42]⁵. Of the PD patients, 21 were male and 10 were female. Of the healthy subjects, 18 were male and 17 were female. Healthy individuals ranged in age from 14 to 79 years old; 5 were left-handed and 30 were right-handed. The PD patients were between 38 and 78 years old; 2 were left-handed and 29 were right-handed. The NewHandPD dataset was collected while subjects performed four different writing tasks, including circle, spiral, meander, and signals. Each subject was asked to produce 1 circle sample, 4 meander and spiral samples, and 12 signal samples. The dataset thus contains 394 images and 792 signals. In this paper, we evaluated our proposed method using the image

data. Statistics for the NewHandPD dataset are shown in Table 2.

The **Parkinsons Drawings dataset** consists of handwriting samples⁶. The dataset consists of training and testing groups containing two different types of image [43], specifically, 102 spiral samples and 102 wave samples. Of the 102 images of each type, 51 samples were provided by subjects with PD, and 51 by healthy subjects.

3.1.2. Data Augmentation. As discussed in an earlier section, a major challenge in PD diagnosis is lack of suitable data. In the current work, we applied data augmentation methods to significantly increase the diversity of the PD handwritten samples. Data augmentation was organized into the following categories: Threshold, flipping, rotation, illumination, and contrast, as illustrated in Table 3. The overall statistics and number of samples after augmentation are shown in Table 4.

- Flipping: Flipping produces a mirror image; we performed both horizontal and vertical flips.
- Rotation: Images are rotated by a given angle, such as 90, 180, or 270 degrees.
- Illumination: Illumination can be implemented by adjusting image intensity values or by transforming the colormap (RGB values). We added random values to the R, G, and B channels and enhanced the resulting image.
- Contrast: Contrast is performed by converting images from RGB to HSV, and then adding values ranging from 0 to 2 to the ‘v’ channel of an image and concatenating all the channels. Finally, we convert the HSV images back to RGB. The resulting image is enhanced.
- Thresholding: We performed thresholding to remove noise from the images, using a method that creates a binary image from the input image by replacing all values above a determined threshold level with 1s, and setting the remaining values to 0.

4. <http://wwwp.fc.unesp.br/papa/pub/datasets/Handpd/>

5. <http://wwwp.fc.unesp.br/papa/pub/datasets/Handpd/>

TABLE 2: Statistics of Each Dataset Included in Study for identification of PD

Dataset	Class	Individuals	Samples	Total	Spiral	Meander	Circle	Waves
PaHaW [22], [25]	Healthy Control	38	1 (2)	36		-	-	-
	PD Patients	37	1 (1)	36		-	-	-
	Total	-	-	72				
HandPD [41]	Healthy controls	18	1	18			-	-
	PD Patients	74	1	74			-	-
	Total	-	-	92				
NewHandPD dataset [42]	Healthy controls	35	4+4+1=9	315				-
	PD Patients	31	4+4+1=9	179				-
	Total	66		394				
Parkinson's Drawings [43]	Healthy controls			102		-	-	
	PD Patients			102		-	-	
	Total			204				

When capturing the PaHaW dataset, individuals were asked to complete eight tasks, but not all tasks were completed successfully. In the current study, when using the spiral samples we omitted two samples from healthy individuals and one from a subject with PD. In the above table, the number of samples omitted is shown in brackets.

3.2. Deep Transfer Learning

As shown in Figure 1, our deep transfer learning framework consist of two main parts. The first part utilizes pre-trained knowledge (pre-learnt features), followed by setting of the network (shared network) architecture parameters and retraining the weights of the output layer for PD identification in handwritten samples. The choice of source affects performance on the target task, and is thus one of the main considerations in transfer learning. A common assumption in many machine learning methods is that the training and test datasets are drawn from the same feature space with the same distribution, i.e., the source data comes from the same distribution as the target data. However, this assumption is revealed to be false as soon as we apply

our models in the real world, where most of the realistic data sources we encounter will be very different from our original training data, significantly impacting performance. Selecting an optimal source can avoid negative transfer occurring when information from our source training data is not only unhelpful but actually counter-productive to good performance in our target domain.

In the current work, we conducted a thorough analysis of PD and investigated the six most commonly used transfer learning architectures, namely AlexNet [44], GoogleNet [45], VGGNet-16/19 [46], and ResNet-50/101 [47]. Each of these networks is trained on over a million images and can classify images into 1000 different object categories of the ImageNet dataset. The transfer learning network architecture

TABLE 3: The Sample Images After Applying Different Techniques of Data Augmentation

Dataset	Sample	Original	Threshold	Contrast	Illumination	Rotation	Flipping
PaHaW	Spiral						
HandPD	Meander						
HandPD	Spiral						
NewHandPD	Circle						
NewHandPD	Spiral						
NewHandPD	Meander						
Parkinson's Drawings	Wave						
Parkinson's Drawings	Spiral						

TABLE 4: Total Number of Images in Augmented Dataset.

Dataset	Healthy	PD	Total
Original	561	973	1534
Contrast	561	973	1534
Illumination	561	973	1534
Flipping	1122	1946	3068
Rotation	1683	2919	4602
Combined	5049	8757	13806
Total	9537	16541	26078

is composed of multi-scale CNN architectures (AlexNet, GoogleNet, VGGNet, and ResNet). Multiscale CNNs are used to detect patterns and extract different scale features from handwritten images, followed by batch normalization to control the distributions of the output vector, reducing internal covariate shift. After the shared network architec-

ture, there is a fully connected layer with ReLU activation function to predict the probability of a subject having PD. In the current work, we divided the PD dataset into training and test sets. We first retrained (fine-tuned) the previously learned features on ImageNet to converge the model to handwritten samples using the training set. Then, we used the test set to evaluate PD identification performance. Thus, our model was able to deliver improved performance in early PD identification through fine tuning of previously trained parameters.

3.3. Network Parameters

In our experimental analysis, we conducted various experiments to determine the optimal network for identification of PD. Table 5 presents a comparative evaluation of different network parameters, with bold text indicating parameters that provide best performance. In the current

TABLE 5: Experiments for Selection of Parameters for Proposed Networks

Exp	Initial learn rate	Momentum	L2 Regularization	Accuracy
1	0.0001	0.1	0.0005	30.2
2	0.0001	0.2	0.0005	35.4
3	0.0001	0.7	0.0005	62.50
4	0.0001	0.7	0.5	45.1
5	0.0001	0.7	0.05	50.7
6	0.0001	0.7	0.005	56.2
7	0.01	0.3	0.0005	50.6
8	0.001	0.4	0.0005	55.9
9	0.0001	0.5	0.0005	45.7
10	0.000001	0.6	0.0005	43

work, we used stochastic gradient descent with the momentum optimizer to accelerate gradient vectors in the right direction, leading to faster convergence. In the hidden layers, the ReLU unit is used as an activation function for different input blocks. In the current study, we considered 15 to be the maximum number of epochs for the training network to learn thoroughly. Since transfer learning approaches do not require a large number of maximum epoch time, thus, the network needs fewer epochs to learn and converge. We observed that the network started to converge and learn after epoch 8 in all experiments. In each iteration, we passed seven images, so the mini batch size was 7. We set the initial learn rate, momentum, and L2 regularization to 0.0001, 0.7, and 0.0005, respectively.

We divided the data into training and test sets with a 90:10 split. The validation frequency was specified as 300, to calculate and test the accuracy after every epoch during training. Like all other classification-based models, CNN architectures also input class labels along with the input images to learn visual patterns. The input size, and the number and size of convolutional layers and pooling layers are discussed in detailed in section 3.2.

4. Experimental Results

A series of experiments were performed to analyze the handwriting dynamics of PD patients in four benchmark handwriting datasets. Selection of an appropriate source is crucial to transferring and exploiting the learned information to achieve optimal performance. Negative transfer may occur if the information from our source training data is not only unhelpful but actually counter-productive for good performance in our target domain. MNIST is a handwriting digit dataset that resembles a PD handwriting dataset. However, studies have shown that semantic relevance between source (i.e., the MNIST handwriting dataset) and target tasks (i.e., PD handwriting) is not always necessary; performance improvement has been observed even when the source and target tasks are superficially unrelated [4], [48], [49]. Thus, in the current work, we used ImageNet as the source due to its improved performance compared to MNIST based iden-

tification [4]. We learned the features using different well-known CNN architectures such as AlexNet, GoogLeNet, VGG, and ResNet. In the current paper, we considered only fine tuning based on performance rather than the freezing approach presented in our earlier study [4]. To evaluate performance in early PD identification, we performed several experiments by fine-tuning different CNN architectures as well as the same network without transfer learning on the original and data-augmented datasets, as depicted in tables 6, 7 and 9. In each experiment, a 90:10 training/test split was used.

We performed 84 experiments to assess various hand-drawn patterns, including spirals, meanders, circles, waves, and their combinations in the original samples from the different datasets, i.e., PaHaW (spiral), HandPD (spiral and meander), NewHandPD (spiral, meander, and circle) and Parkinsons Drawing (spiral and wave). Our earlier study [49] showed that original offline handwriting samples, such as spiral, meander, and wave, play a vital role in detecting PD. The results were reported as the average accuracy of five runs using hold-out cross validation.

4.1. Original Dataset Evaluation

We analyzed performance on the original dataset as follows: First, we performed experiments on each individual dataset followed by the combined dataset. Table 6 shows the results. In our first experiment, we explored six CNN architectures on the spiral task data from the PaHaW dataset. Results showed that the original PaHaW spiral sample achieved poor recognition rates (with a maximum of approx. 62.50% using AlexNet and VGG16). One reason for this could be that the images are synthetically generated from time series data, as the patients wrote using a digital tablet. Furthermore, the number of samples from PD and healthy individuals is extremely limited. Next, in cases 2 and 3, we performed several experiments on the HandPD (spiral, meander, and combined) and NewHandPD (circle, spiral and meander) datasets. Results for case 3 in Table 6 show that we achieved better performance (up to 89.19%) on the meander task using AlexNet, VGG16, and VGG19. We observed a slight increase in performance to 90.41% on the combined spiral and meander tasks using VGG16 in case 4. A similar trend can be seen in the NewHandPD dataset (92.31% and 88.46% for meander and spiral tasks, respectively, and 97.73% for the combined spiral and meander task). Notice that circle displayed poor performance, indicating that it demands less motor control from the subject compared with meander and spiral production. Results for the combination of circle, spiral, and meander achieved 98.31% using ResNet, compared to 90.41% on the spiral and meander data from HandPD. Furthermore, we evaluated the performance of PD identification on the Parkinsons Drawing dataset. Mixed trends were observed for both the spiral and wave tasks with the different CNN networks.

As discussed in earlier sections, each person with PD will have different symptoms, disease progression, lifestyle effects, and physical tolerances, i.e. no two people with PD

TABLE 6: Results of Parkinson's Original Dataset.

Dataset	Cases	Samples	AlexNet	GoogLeNet	VGG16	VGG19	ResNet50	ResNet101
(a) PaHaW	Case-1	Spiral	62.50	37.50	62.50	50	37.50	50
	Case-2	Spiral	86.49	81.08	81.08	62.50	81.08	81.08
(b) HandPD	Case-3	Meander	89.19	83.78	89.19	89.19	81.08	81.08
	Case-4	Spiral + Meander	84.93	83.56	90.41	82.19	84.93	80.83
(c) NewHandPD	Case-5	Circle	83.33	66.67	66.67	83.33	83.33	83.33
	Case-6	Spiral	84.62	84.62	88.46	88.46	88.46	69.23
	Case-7	Meander	92.31	88.46	84.62	80.77	84.62	76.92
	Case-8	Spiral + Meander	97.37	90.91	97.73	93.18	97.73	84.09
	Case-9	Circle + Spiral + Meander	84.75	84.75	88.14	81.36	98.31	71.19
(d) Parkinson's Drawing	Case-10	Spiral	60	60	90	80	80	70
	Case-11	Waves	80.00	80.00	70	80	90	50
	Case-12	Spiral+Waves	70	90	75	75	80	55
(e) Combined a, b, c, d	Case-13	PaHaW + HandPD + NewHandPD + Parkinsons's Drawing	90.12	81.48	80.25	77.78	88.89	79.01
(f) Combined b, c, d	Case-14	HandPD, NewHandPD & Parkinsons's Drawing	90.85	95.42	96.08	95.42	90.20	90.85

TABLE 7: Performance of pre-trained CNNs with fine-tuned approach using different Data Augmentation schemes for the Parkinson's datasets (HandPD, NewHandPD, Parkinson's Disease Drawing) excluding PaHaW dataset.

Case No	Augmentation methods	AlexNet	GoogLeNet	VGG16	VGG19	ResNet50	ResNet101
Case-1	Contrast	94.05	87.32	92.81	93.46	91.50	91.21
Case-2	Flipping	90.99	91.86	88.23	95.77	89.90	88.26
Case-3	Rotation	89.64	91.09	89.89	93.04	90.22	90.85
Case-4	Threshold	91.50	87.58	91.72	92.81	85.62	93.46
Case-5	Illumination	99.22	98.04	97.78	98.56	97.52	95.42
Case-6	Contrast + Illumination	96.29	97.72	98.50	98.69	97.25	97.39
Case-7	Flipping + Illumination	92.61	94.11	94.13	95.05	93.26	90.22
Case-8	Contrast + Flipping + Illumination	94.94	95.11	94.45	92.17	94.78	94.29
Case-9	Case-1 to Case-5	92.58	92.34	90.38	89.65	90.14	88.26
Case-10	Case-1 to Case-5 and original datasets excluding PaHaW	93.85	95.29	95.37	96.31	92.76	92.90

will experience the condition in the same way. Thus, due to the high variability of PD symptoms and progression of motor symptoms, and the lack of adequate quantities of data, both traditional machine learning methods and deep learning approaches have displayed poor performance for early identification of PD. To overcome such challenges, we combined all the datasets to increase the training set size for deep neural networks, and performed several experiments using different combinations of datasets. Results indicated that we achieved 96.08% with VGG-16 using HandPD, NewHandPD, and the Parkinsons Drawing datasets. We observed that inclusion of the PaHaW dataset resulted in a decrease in performance to 80.25% using VGG-16, and 90.12% (highest) using AlexNet. This may be due to the different collection method (digital tablet) used to obtaining the dataset, i.e., the images in PaHaw are synthesized from online coordinated, and thus differ from those in the other three datasets. From the above experiments, we can conclude that the proposed transfer learning-based PD identification method achieved considerably better performance when the

dataset size was increased by combining different datasets.

4.2. Augmented Dataset Evaluation

To further improve performance, we performed additional experiments on the combined datasets modified with five different data augmentation techniques: Contrast, flipping, rotation, thresholding, and illumination. In the first phase, we conducted several experiments with each individual augmentation technique on each dataset using the different CNN models, followed by various combinations of data augmentation methods to determine the best technique. Table 7 shows a comparative analysis of the different augmentation techniques using the HandPD, NewHandPD, and Parkinsons Drawing datasets, excluding the PaHAW dataset. Contrast, flipping, and illumination achieved the highest average accuracy of up to 94.04%, 95.77%, and 99.22%, respectively, using AlexNet and VGG19. Note that the other CNN networks, GoogleNet and ResNet, showed comparable performance. We observed that rotation and thresholding

TABLE 8: Performance of Pre-trained CNNs with fine-tuned approach using different Data Augmentation Schemes for Parkinson’s Datasets (HandPD, NewHandPD, Parkinsons Disease Drawing, PaHaw).

Case No	Augmentation methods	AlexNet	GoogLeNet	VGG16	VGG19	ResNet50	ResNet101
Case-1	Contrast	97.27	90	81.48	93.69	97.27	87.96
Case-2	Flipping	80.86	82.72	88.27	86.42	88.33	79.01
Case-3	Rotation	75.76	78.40	78.03	81.82	78.41	77.27
Case-4	Threshold	81.48	82.71	83.95	83.95	81.48	83.95
Case-5	Illumination	88.89	88.89	88.89	85.18	85.19	90.12
Case-6	Contrast + Flipping	86.76	86.76	88.97	84.56	87.13	90.07
Case-7	Contrast + Illumination	92.12	90.05	87.92	92.23	93.02	89.55
Case-8	Case-1 to case-5 and four original datasets	85.49	83.06	84.47	84.72	85.62	84.47

TABLE 9: Scratch Networks using combined augmented data from three datasets (HandPD, NewHandPD, Parkinson’s Disease Drawing)

Exp	Accuracy
Scratched AlexNet	91.24
Scratched GoogLeNet	63.85
Scratched VGG	63.43
Scratched ResNet	75.38

had a negative impact on performance, and that illumination showed significantly better performance in comparison. Table 3 shows the handwritten text images after application of illumination and contrast techniques. Illumination showed comparatively better performance than flipping and contrast. Similarly, we performed experiments on combinations of all four datasets (including the PaHaW dataset). Results are shown in Table 8. Note that transfer learning (results shown in Table 3 and Table 8) achieved better performance in comparison to the learning-from-scratch network (results shown in Table 9).

4.3. Discussion

We performed 5 runs of cross validation and compared the performance of our best network with state-of-the-art methods. In section 4.2, the fine-tuned CNN with illumination augmentation displayed best performance. In general, deep learning approaches have benefited from the increasing availability of large datasets, but in our case, the available datasets are not that large. We applied a variety of data augmentation techniques to address this, as previously discussed, finding illumination and contrast to provide better performance compared with flipping, rotation, and thresholding. We implemented illumination and contrast for image enhancement by adding random values to the pixels of an image. Image enhancement is a method of altering advanced pictures so that the outcomes are more suitable for further image investigation. In the current work, we performed experiments using different network sizes, observing that smaller networks showed much better performance than larger networks. This may be due to the one-class problem or small dataset size, i.e., the network

TABLE 10: Comparison of Deep Convolutional Neural Networks on Handwriting Datasets

Reference	Classification Technique	Dataset	Accuracy (%)
Pereira et al. [27]	CNN	HandPD	80.19
Pereira et al. [28]	CNN	HandPD	90.38
Priera et al. [6]	CNN	HandPD	95
Moetesum et al. [33]	AlexNet based fusion	PaHaW	83
Naseer et al. [4]	AlexNet	PaHaW	98.28
Afonso et al. [38]	CNN	HandPD	87
Proposed system	AlexNet fine-tuned	HandPD, NewHandPD and Parkinson’s Drawing	99.22

has learnt the handwritten sample well, but it has not learnt enough to generalize to variations.

Table 10 presents a comparison of various CNNs for PD classification. In the current work, we compared the performance of our approach with that of state-of-the-art frameworks such as those of Naseer et al. [4] Pereira et al. [27], [28] and [6], bernardo et al. [39], Afonso et al. [38] and Moetesum et al. [33]. Handcrafted features showed poor performance, achieving a 78.9% recognition rate for spiral and meander tasks. Recent applications of deep learning have demonstrated considerable improvement over traditional machine learning methods. Pereira et al. used CNN and metaheuristic methods on the HandPD dataset and achieved recognition rates of 80.19% and 90.385%, respectively [27], [28]. Recently, Pereira et al. [6] implemented a unique approach in which time series images are passed to a CNN, achieving 93.50% accuracy. Afonso et al. [38] also used the HandPD dataset but they performed experiments using two different splits and achieved an accuracy of 87% for PD classification. In comparison with Pereira et al. [27], [28] and Naseer et al. [4], our system displayed an impressive increase in performance, i.e., 80.19% [27], 90.38% [28], 98.28% [4] to 99.22%. Our key observations

are summarized as follows:

- Illumination resulted in significantly better performance compared to other augmentation methods.
- Even though the MNIST dataset and Parkinsons handwriting have semantic relevance, which suggests using MNIST as source, our investigation found ImageNet to have better transferable knowledge in comparison with MNIST.
- We notice that smaller the network, better the performance, this showed that number of classes and dataset size impact on the size of the network.
- We observed that complex handwriting tasks such as meander, spiral, and wave are best for identifying early PD accurately, as these are more difficult for subjects with PD to perform.
- Experimental results are based on only handwritten samples, whereas clinicians have the advantage of additional patient information. Performance may thus be improved by inclusion of additional clinical features.
- We observed that complex handwritten tasks such as meander, spiral, and wave showed better performance, thus, we suggest developing a complex task dataset.

5. Conclusion

In this paper, we presented an approach for early diagnosis of PD using handwritten samples produced by individuals with PD. To increase the dataset size, we combined multiple PD handwriting datasets and applied different data augmentation techniques, resulting in improved diagnostic performance. We achieved promising results using an end-to-end deep transfer learning method to transfer pre-learned knowledge onto the domain of handwriting samples. Our results demonstrate that the proposed approach outperforms current state-of-the-art methods. Our experimental results showed that the use of the data-augmented images provides best performance when combined with CNN fine-tuned architectures. Furthermore, these results show that all CNN models significantly outperform the classical approaches with or without data augmentation. The learn-from-scratch CNNs delivered worse performance in terms of classification accuracy than the CNN fine-tuned approach combined with data-augmented images, which achieved an accuracy of 99.22%. Further improvements could be obtained by using additional clinical features alongside handwritten samples. Furthermore, we observed that drawing of complex patterns can improve early-stage identification of the disease. In future, we plan to collect a dataset of complex tasks, with additional clinical features that will help to not only identify PD at early stage, but also to investigate disease severity and the effects of levodopa and other medications.

Acknowledgement: The authors thank the Deanship of Scientific Research and RSSU at King Saud University for their technical support.

References

- [1] R. G. Ramani and G. Sivagami, "Parkinson disease classification using data mining algorithms," *International journal of computer applications*, vol. 32, no. 9, pp. 17–22, 2011.
- [2] S. Argaud, M. Vérin, P. Sauleau, and D. Grandjean, "Facial emotion recognition in parkinson's disease: a review and new hypotheses," *Movement Disorders*, vol. 33, no. 4, pp. 554–567, 2018.
- [3] M. Hirai, T. Sakurada, and S.-i. Muramatsu, "Face-to-trait inferences in patients with parkinsons disease," *Journal of clinical and experimental neuropsychology*, vol. 41, no. 2, pp. 170–178, 2019.
- [4] A. Naseer, M. Rani, S. Naz, M. I. Razzak, M. Imran, and G. Xu, "Refining parkinsons neurological disorder identification through deep transfer learning," *Neural Computing and Applications*, pp. 1–16, 2019.
- [5] L. A. Passos, C. R. Pereira, E. R. Rezende, T. J. Carvalho, S. A. Weber, C. Hook, and J. P. Papa, "Parkinson disease identification using residual networks and optimum-path forest," in *2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI)*. IEEE, 2018, pp. 000 325–000 330.
- [6] C. R. Pereira, D. R. Pereira, G. H. Rosa, V. H. Albuquerque, S. A. Weber, C. Hook, and J. 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.
- [7] S. Mehrang, M. Jauhainen, J. Pietil, J. Puustinen, J. Ruokolainen, and H. Nieminen, "Identification of parkinsons disease utilizing a single self-recorded 20-step walking test acquired by smartphones inertial measurement unit," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 2913–2916.
- [8] E. Stack, V. Agarwal, R. King, M. Burnett, F. Tahavori, B. Janko, W. Harwin, A. Ashburn, and D. Kunkel, "Identifying balance impairments in people with parkinsons disease using video and wearable sensors," *Gait & posture*, vol. 62, pp. 321–326, 2018.
- [9] D. Govindaraju, G. Nagarajan, and P. Alagumariappan, "Assessment of gait disorder in parkinson's disease," in *Early Detection of Neurological Disorders Using Machine Learning Systems*. IGI Global, 2019, pp. 108–127.
- [10] A. Antonini, P. Odin, Y. J. Jalundhwala, P. N. Schmidt, A. M. Skalicky, L. Kleinman, J. Zamudio, K. Onuk, P. Kukreja, Y. Bao *et al.*, "Manage-pd: A clinician-reported tool to identify patients with parkinsons disease inadequately controlled on oral medications—results from vignette-based validation (p5. 8-039)," 2019.
- [11] J. A. Whitfield and D. D. Mehta, "Examination of clear speech in parkinson disease using measures of working vowel space," *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 7, pp. 2082–2098, 2019.
- [12] L. Kurlowicz, M. Wallace *et al.*, "The mini-mental state examination (mmse)," *Journal of gerontological nursing*, vol. 25, no. 5, pp. 8–9, 1999.
- [13] P. Martínez-Martín, A. Gil-Nagel, L. M. Gracia, J. B. Gómez, J. Martínez-Sarries, F. Bermejo, and C. M. Group, "Unified parkinson's disease rating scale characteristics and structure," *Movement disorders*, vol. 9, no. 1, pp. 76–83, 1994.
- [14] C. Taleb, M. Khachab, C. Mokbel, and L. Likforman-Sulem, "Feature selection for an improved parkinson's disease identification based on handwriting," in *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*. IEEE, 2017, pp. 52–56.
- [15] R. Ahmad, S. Naz, M. Afzal, S. Rashid, M. Liwicki, and A. D., "A deep learning based arabic script recognition," *International Arab Journal of Information Technology (IAJIT)*, 2020.
- [16] S. Naz, A. I. Umar, R. Ahmad, I. Siddiqi, S. B. Ahmed, M. I. Razzak, and F. Shafait, "Urdu nastaliq recognition using convolutional-recursive deep learning," *Neurocomputing*, vol. 243, pp. 80–87, 2017.

- [17] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, "A deep learning-based framework for automatic brain tumors classification using transfer learning," *Circuits, Systems, and Signal Processing*, pp. 1–19, 2019.
- [18] S. Naz, A. I. Umar, R. Ahmad, S. B. Ahmed, S. H. Shirazi, and M. I. Razzak, "Urdu nastaliq text recognition system based on multi-dimensional recurrent neural network and statistical features," *Neural computing and applications*, vol. 28, no. 2, pp. 219–231, 2017.
- [19] A. Van Gemmert, C. H. Adler, and G. Stelmach, "Parkinsons disease patients undershoot target size in handwriting and similar tasks," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, no. 11, pp. 1502–1508, 2003.
- [20] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "A new modality for quantitative evaluation of parkinson's disease: In-air movement," in *13th IEEE International Conference on BioInformatics and BioEngineering*. IEEE, 2013, pp. 1–4.
- [21] P. Drotár, J. Mekyska, Z. Smékal, I. Rektorová, L. Masarová, and M. Faundez-Zanuy, "Prediction potential of different handwriting tasks for diagnosis of parkinson's," in *2013 E-Health and Bioengineering Conference (EHB)*. IEEE, 2013, pp. 1–4.
- [22] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Analysis of in-air movement in handwriting: A novel marker for parkinson's disease," *Computer methods and programs in biomedicine*, vol. 117, no. 3, pp. 405–411, 2014.
- [23] P. Drotár, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Decision support framework for parkinsons disease based on novel handwriting markers," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 508–516, 2014.
- [24] P. Drotár, J. Mekyska, Z. Smékal, I. Rektorová, L. Masarová, and M. Faundez-Zanuy, "Contribution of different handwriting modalities to differential diagnosis of parkinson's disease," in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*. IEEE, 2015, pp. 344–348.
- [25] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease," *Artificial intelligence in Medicine*, vol. 67, pp. 39–46, 2016.
- [26] C. R. Pereira, D. R. Pereira, F. A. da Silva, C. Hook, S. A. Weber, L. A. Pereira, and J. P. Papa, "A step towards the automated diagnosis of parkinson's disease: Analyzing handwriting movements," in *2015 IEEE 28th international symposium on computer-based medical systems*. IEEE, 2015, pp. 171–176.
- [27] C. R. Pereira, S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson's disease diagnosis from handwritten dynamics," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. Ieee, 2016, pp. 340–346.
- [28] C. R. Pereira, D. R. Pereira, J. P. Papa, G. H. Rosa, and X.-S. Yang, "Convolutional neural networks applied for parkinsons disease identification," in *Machine Learning for Health Informatics*. Springer, 2016, pp. 377–390.
- [29] B. M. Eskofier, S. I. Lee, J.-F. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz *et al.*, "Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson's disease assessment," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 655–658.
- [30] A. Caliskan, H. Badem, A. Basturk, and M. E. Yuksel, "Diagnosis of the parkinson disease by using deep neural network classifier," *Istanbul University-Journal of Electrical & Electronics Engineering*, vol. 17, no. 2, pp. 3311–3318, 2017.
- [31] H. Choi, S. Ha, H. J. Im, S. H. Paek, and D. S. Lee, "Refining diagnosis of parkinson's disease with deep learning-based interpretation of dopamine transporter imaging," *NeuroImage: Clinical*, vol. 16, pp. 586–594, 2017.
- [32] Y. Zhang, "Can a smartphone diagnose parkinson disease? a deep neural network method and telediagnosis system implementation," *Parkinsons Disease*, vol. 2017, 2017.
- [33] M. Moetesum, I. Siddiqi, N. Vincent, and F. Cloppet, "Assessing visual attributes of handwriting for prediction of neurological disorders case study on parkinsons disease," *Pattern Recognition Letters*, vol. 121, pp. 19–27, 2019.
- [34] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.
- [35] U. Gupta, H. Bansal, and D. Joshi, "Gender specific and age dependent classification model for improved diagnosis in parkinson's disease," *arXiv preprint arXiv:1904.09651*, 2019.
- [36] R. Senatore, A. Della Cioppa, and A. Marcelli, "Automatic diagnosis of neurodegenerative diseases: An evolutionary approach for facing the interpretability problem," *Information*, vol. 10, no. 1, p. 30, 2019.
- [37] D. Impedovo, "Velocity-based signal features for the assessment of parkinsonian handwriting," *IEEE Signal Processing Letters*, vol. 26, no. 4, pp. 632–636, 2019.
- [38] L. C. Afonso, G. H. Rosa, C. R. Pereira, S. A. Weber, C. Hook, V. H. C. Albuquerque, and J. P. Papa, "A recurrence plot-based approach for parkinsons disease identification," *Future Generation Computer Systems*, vol. 94, pp. 282–292, 2019.
- [39] L. S. Bernardo, A. Quezada, R. Munoz, F. M. Maia, C. R. Pereira, W. Wu, and V. H. C. de Albuquerque, "Handwritten pattern recognition for early parkinsons disease diagnosis," *Pattern Recognition Letters*, vol. 125, pp. 78–84, 2019.
- [40] B. K. O. senkul, M.E.; Sakar, "Improved spiral test using digitized graphics tablet for monitoring parkinson's disease," in *The 2nd International Conference on e-Health and Telemedicine (ICEHTM-2014)*, 2014, pp. 171–175.
- [41] C. R. Pereira, D. R. Pereira, F. A. Silva, J. P. Masieiro, S. A. T. Weber, C. Hook, and J. P. Papa, "A new computer vision-based approach to aid the diagnosis of parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 136, pp. 79–88, 2016.
- [42] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson's disease diagnosis from handwritten dynamics," in *Proceedings of the SIBGRAPI - Conference on Graphics, Patterns and Images*, 2016, pp. 340–346.
- [43] P. Zham, D. K. Kumar, P. Dabnichki, S. Poosapadi Arjunan, and S. Raghav, "Distinguishing different stages of parkinsons disease using composite index of speed and pen-pressure of sketching a spiral," *Frontiers in neurology*, vol. 8, p. 435, 2017.
- [44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [45] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [48] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps*. Springer, 2018, pp. 323–350.
- [49] I. Razzak, I. Kamran, and S. Naz, "Deep analysis of handwritten notes for early diagnosis of neurological disorders," in *2020 International Joint Conference on Neural Network (IJCNN)*. IEEE, 2020.

Saeeda Naz: SAEEDA NAZ received the B.S. degree from the University of Peshawar (UoP), Peshawar, Pakistan, in 2006, the M.S. degree in computer science from COMSATS Institute of Information Technology (CIIT), Pakistan, in 2012, and the Ph.D. degree (Hons.) in computer science from Hazara University, Mansehra, Pakistan, in 2016. She has been an Assistant Professor and the Head of the Computer Science Department, GGPSC No.1 Abbottabad, Higher Education Department of Government of Khyber Pakhtunkhwa, Pakistan, since 2008. She has published two book chapters and more than 50 articles in peer-reviewed national and international conferences and journals. Her current research interests include optical character recognition, pattern recognition, machine learning, medical imaging, and natural language processing

Imran Razzak: Imran Razzak is currently Sr. Lecturer at School of Information Technology at Deakin University, Australia. Imran's research interest is machine learning and data analytics in general, particularly in healthcare industry. He is a passionate health informatician who wants to make the healthcare industry a better place through informatics. He has published more than 70 refereed research publications in international journals and conferences. He is an editorial board member of many reputable international journals as well as session co-chair, session chair and TPC member of dozens of conferences.

MUHAMMAD IMRAN received the Ph.D. degree in information technology from University Teknologi PETRONAS, Malaysia, in 2011. He is currently an Associate Professor with the College of Applied Computer Science, King Saud University, Saudi Arabia. His research interests include the Internet of Things, mobile and wireless networks, big data analytics, cloud computing, and information security. His research is financially supported by several grants. He has completed a number of international collaborative research projects with reputable universities. He has published more than 150 research articles in top conferences and journals. He has been involved in more than seventy conferences and workshops in various capacities, such as the Chair, the Co-Chair, and a Technical Program Committee Member. European Alliance for Innovation (EAI) has appointed him as the Editor in Chief of EAI Transactions on Pervasive Health and Technology. He serves as an Associate Editor for reputable international journals, such as IEEE COMMUNICATIONS MAGAZINE, Future Generation Computer Systems, IEEE ACCESS, Ad Hoc & Sensor Wireless Networks (Journal) (SCIE), IET Wireless Sensor Systems, the International Journal of Autonomous and Adaptive Communication Systems (Inderscience). He served/serving as a Guest Editor for more than a dozen special issues in journals, such as IEEE COMMUNICATIONS MAGAZINE, Computer Networks (Elsevier), Future Generation Computer Systems (Elsevier), Sensors (MDPI), the International Journal of Distributed Sensor Networks (Hindawi), the Journal of Internet Technology, and the International Journal of Autonomous and Adaptive Communications Systems



Saeeda Naz



Imran Razzak:



Muhammad Imran:

Authors declare no conflict of interest.

Journal Pre-proof