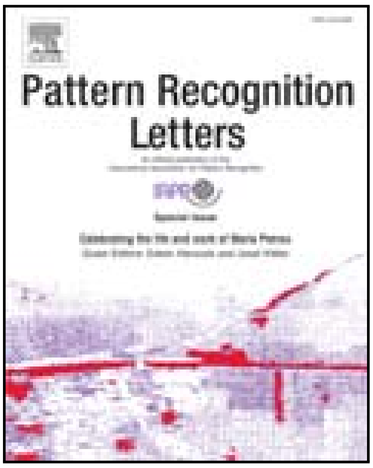
Accepted Manuscript



Assessing Visual Attributes of Handwriting for Prediction of Neurological Disorders - A Case Study on Parkinson's Disease

Momina Moetesum, Imran Siddiqi, Nicole Vincent, Florence Cloppet

PII: S0167-8655(18)30128-4
DOI: 10.1016/j.patrec.2018.04.008
Reference: PATREC 7136

To appear in: *Pattern Recognition Letters*

Received date: 20 November 2017
Revised date: 23 February 2018
Accepted date: 5 April 2018

Please cite this article as: Momina Moetesum, Imran Siddiqi, Nicole Vincent, Florence Cloppet, Assessing Visual Attributes of Handwriting for Prediction of Neurological Disorders - A Case Study on Parkinson's Disease, *Pattern Recognition Letters* (2018), doi: 10.1016/j.patrec.2018.04.008

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. 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.

ACCEPTED MANUSCRIPT

Highlights

- Investigation of visual attributes of handwriting to predict Parkison's Disease
- Use of Convolutional Neural Networks for automatic feature extraction
- Multiple representations of raw data to enhance feature extraction step
- Evaluations on a standard template including drawing and writing tasks
- Fusion of predictions from multiple tasks to enhance performance

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Pattern Recognition Letters

Authorship Confirmation

Please save a copy of this file, complete and upload as the "Confirmation of Authorship" file.

As corresponding author I, Momina Moetesum , hereby confirm on behalf of all authors that:

1. This manuscript, or a large part of it, has not been published, was not, and is not being submitted to any other journal.
2. If presented at or submitted to or published at a conference(s), the conference(s) is (are) identified and substantial justification for re-publication is presented below. A copy of conference paper(s) is(are) uploaded with the manuscript.
3. If the manuscript appears as a preprint anywhere on the web, e.g. arXiv, etc., it is identified below. The preprint should include a statement that the paper is under consideration at Pattern Recognition Letters.
4. All text and graphics, except for those marked with sources, are original works of the authors, and all necessary permissions for publication were secured prior to submission of the manuscript.
5. All authors each made a significant contribution to the research reported and have read and approved the submitted manuscript.

Signature: Momina Moetesum Date: November 20, 2017

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Pattern Recognition Letters
journal homepage: www.elsevier.com

Assessing Visual Attributes of Handwriting for Prediction of Neurological Disorders - A Case Study on Parkinson's Disease

Momina Moetesum[a,**], Imran Siddiqi[a], Nicole Vincent[b], Florence Cloppet[b]

[a]*Bahria University, Islamabad, Pakistan*
[b]*Paris Descartes University, Paris, France*

ABSTRACT

Parkinson's disease (PD) is a degenerative disorder that progressively affects the central nervous system causing muscle rigidity, tremors, slowed movements and impaired balance. Sophisticated diagnostic procedures like SPECT scans can detect changes in the brain caused by PD but are only effective once the disease has advanced considerably. Analysis of subtle variations in handwriting and speech can serve as potential tools for early prediction of the disease. While traditional techniques mostly rely on dynamic (kinematic and spatio-temporal) features of handwriting, in this study, we quantitatively evaluate the visual attributes in characterization of graphomotor samples of PD patients. For this purpose, Convolutional Neural Networks are employed to extract discriminating visual features from multiple representations of various graphomotor samples produced by both control and PD subjects. The extracted features are then fed to a Support Vector Machine (SVM) classifier. Evaluations are carried out on a dataset of 72 subjects using early and late fusion techniques and an overall accuracy of 83% is realized with solely visual information.
Keywords: Handwriting; Parkinson's Disease; Convolutional Neural Networks; Visual Attributes.
© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Parkinson's Disease (PD) is a neurodegenerative disorder that affects the coordinated movements of a person due to loss of dopamine producing neurons in substantia nigra (Campenhausen et al., 2005). According to studies (De Lau and Breteler, 2006; Zhou et al., 2016), it is one of the most prevalent neurological diseases after Alzheimer's (Yan et al., 2008) with an average onset age of 60. Patients with PD experience symptoms like posture deformation, rigidity, tremors and vocal impairments, etc. Traditional diagnostic procedures for determination of the disease include costly, invasive methods like SPECT and CT scans, which are usually effective when the disease has already progressed to a mature stage. Clinical practitioners therefore, first opt for manual, non-invasive screening tests like Unified Parkinson's Disease Rating Scale (UPDRS) (Fish, 2011), for early detection of the disease. While this process is quite established and has been modified over years of experience, it remains relatively subjective.

With the advent of technology, a number of computerized systems have been proposed to identify the early symptoms of Parkinson's and similar neurological diseases. Some of these studies analyze voice or speech patterns to observe subtle but progressive changes which are indicative of PD (Tsanas et al., 2010), while others monitor muscular movements using wearable sensors (Niazmand et al., 2011). Over the period of time, a substantial number of studies (Tucha et al., 2006; Nackaerts et al., 2013; Drotár et al., 2013a) have suggested that handwriting, a product of perceptive, cognitive and fine motor skills (Teulings et al., 1997; Feder and Majnemer, 2007; Caligiuri and Mohammed, 2012), can also be employed as an effective tool for early diagnosis of PD.

Currently most of the research (Van Gemmert and Teulings, 2006; Rosenblum and Livneh-Zirinski, 2008; Johnson et al., 2015; Dirlikov et al., 2017) focuses on analyzing the kine-

**Corresponding author:
e-mail: momina.moetesum@bui.edu.pk (Momina Moetesum), imran.siddiqi@bahria.edu.pk (Imran Siddiqi), nicole.vincent@mi.parisdescartes.fr (Nicole Vincent), florence.cloppet@mi.parisdescartes.fr (Florence Cloppet)

ACCEPTED MANUSCRIPT

matic and pressure aspects of handwriting to determine PD. Although these dynamic features are effective, they mostly require additional temporal information for prediction which can only be acquired by utilizing special equipment like digitizer tablets and customized electronic pens (Ünlü et al., 2006). Recent advancements in image analysis and pattern classification techniques have encouraged researchers to re-investigate static features from offline samples for improved detection of PD as well (Zhi et al., 2015). Hypothetically believing in the importance of visual features, we propose a novel method of assessing their contribution in characterization of PD. We designed an enhanced system that extracts useful visual features from handwriting and drawing samples of subjects and applies various fusion techniques to accurately discriminate between the control and PD groups. Feature learning based classification has previously been applied to sensor based handwriting movement signals (Pereira et al., 2016) for detection of symptomatic signs of PD. Nevertheless, to the best of our knowledge, feature learning of visual attributes has not been explored to its full potential. The main contributions of the paper are listed in the following.

- Investigation of the effectiveness of visual attributes of handwriting by employing machine learning techniques, in characterizing PD, as opposed to the dynamic online attributes traditionally considered in the literature.
- Use of multiple representations of raw data as input to learn discriminative patterns from handwriting samples.
- Fusion of results from multiple samples acquired from various graphomotor (handwritten & hand drawn) tasks for improved overall classification.

The rest of the paper is organized as follows. Section 2 presents an overview of related works. Section 3 describes the proposed methodology and experimental setup. Section 4 summarizes the results and their analysis. Finally conclusion and future directions are discussed in Section 5.

2. Computerized Analysis of Handwriting for Prediction of Parkinson's Disease

Computerized analysis of handwriting and hand drawn shapes has remained an active area of research in the pattern recognition community for over three decades now. Contrary to popular applications like handwriting recognition, forensic investigation and information retrieval etc., automatic analysis of handwriting and hand drawn shapes for assessment of the mental health of the subject or for prediction of different neurological disorders, still requires further exploration.

2.1. *Correlation Between Handwriting and Parkinson's Disease*

A number of studies (Mavrogiorgou et al., 2001; Rémi et al., 2002; Werner et al., 2006; Yan et al., 2008; Renau-Ferrer and Remi, 2013; Moetesum et al., 2015) highlight strong evidences of correlation between handwriting changes and problems in the nervous system. Neurological disorders like autism (Fuentes et al., 2009; Kushki et al., 2011), Parkinson's (Teulings et al., 2002) and Alzheimer's (Slavin et al., 1999; Schröter et al., 2003) directly impact the graphomotor skills of individuals suffering from them. In addition to these, psychotropic medications (Caligiuri and Mohammed, 2012) and aging (Walton, 1997; Rosenblum et al., 2013a), are also known to influence the handwriting. Visible side effects of neurological disorders like constructional apraxia, dysgraphia and micrographia are directly assessed by analyzing handwriting and hand drawn samples of subjects (Figure 1).

Three most established manifestations of Parkinson's Disease (PD) that can be captured by handwriting, are 'Micrographia', 'Bradykinesia' and 'Tremor' (Smits et al., 2014). In micrographia, it becomes difficult for a patient to maintain the size and alignment of the produced graphomotor impressions (Derkinderen et al., 2002). Bradykinesia or slowness of movement (either due to motor or cognitive dysfunction), causes a potential PD patient to complete a graphomotor task in more time than usually required (Berardelli et al., 2001). 'Tremors' are involuntary to and fro movements that can be visualized by irregular formations of characters and drawings. All these Parkinsonian conditions can either coexist or are present independently depending on the type and progression of the disease. Successful identification of these indicators from the graphomotor samples of patients can assist in early prediction and differential diagnosis of PD.

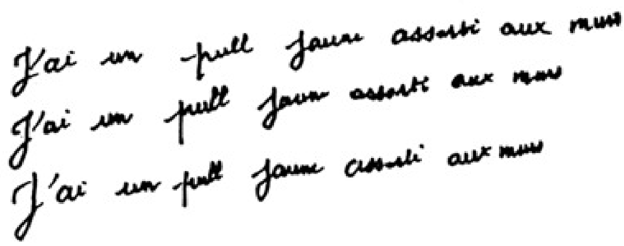


Fig. 1. **Handwriting sample of patient with neurological disorder showing progressive shrinking (Derkinderen et al., 2002)**

2.2. *Templates for Handwriting Samples*

Computerized analysis of 'Archimedean Spiral Drawing Test' (Pullman, 1998) has been widely applied to capture early signs of various neurological dysfunctions (Hsu et al., 2009; Michalec et al., 2014) including PD (Saunders-Pullman et al., 2008; Stanley et al., 2010). Deviations from original template (like loop tightness, loop width variability, drawing speed and acceleration, frequency and amplitude of oscillations and spiral pressure, etc.) are considered as symptomatic indicators of a disorder (Figure 2). Shape modifications (e.g. meander etc.) have also been proposed for improved detection (Aly et al., 2007).

A number of studies (Eichhorn et al., 1996; Teulings et al., 2002; Rosenblum et al., 2013b; Weber et al., 2014) advocate the effectiveness of handwriting analysis in detection of PD as well. Drotár et al. (2013b) presented a template consisting of

ACCEPTED MANUSCRIPT

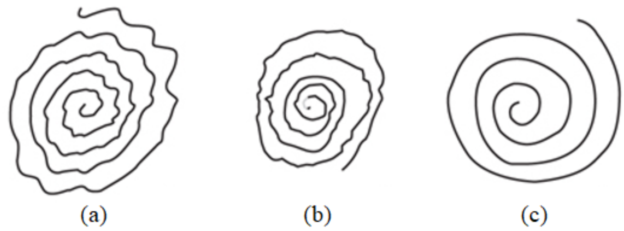


Fig. 2. Archimedean Spiral drawn by (a) Subject with Essential Tremor (b) Subject with Parkinson's Disease (c) Healthy subject (Image Source: (Chung, 2012))

seven different handwriting tasks in addition to conventional spiral drawing task as shown in Figure 3. The study suggested that the choice of template has significant impact on the performance of the proposed features.

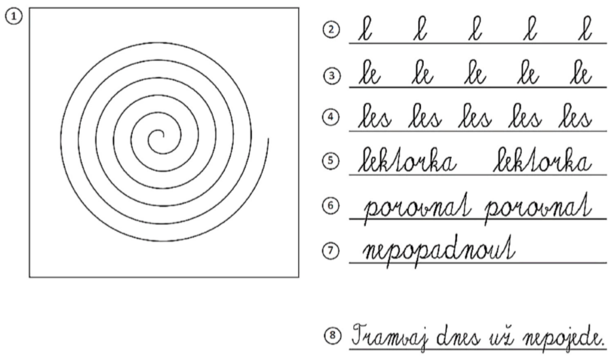


Fig. 3. Template proposed by Drotár et al. (2013b)

2.3. *Handwriting Features for Prediction of Parkinson's Disease*

Various handwriting features have been proposed in the literature for prediction of PD and other neurological disorders. Based on their method of acquisition and measurement, these features can broadly be divided into two categories, *Static* and *Dynamic* features. Static features are usually acquired from offline samples of handwriting and include spatial attributes like shape or character dimensions (e.g. skew and slant etc.) and proportions (e.g. height/width, aspect ratio, etc.), distances, placement on x-y plane, loops, closures and stroke curvature. These features have been employed in literature to indicate signs of micrographia and tremors (Zhi et al., 2015).

On the contrary, dynamic features describe occurrences and thus require additional temporal information, which can only be attained from online samples of handwriting. These mostly include handwriting movements or kinematic measures (like vertical/horizontal stroke speed and acceleration and trajectories etc.) (Van Gemmert and Teulings, 2006; Rosenblum and Livneh-Zirinski, 2008). Popular methods for acquisition of these features include devices like digitizer tablets and smart pens (Palmerini et al., 2011). These devices can also capture in-air/on-surface time intervals which reflect the time a subject is taking to plan the subsequent writing action; more time being indicative of reduced cognitive ability (Johnson et al., 2015; Rosenblum, 2015; Dirlikov et al., 2017).

Pressure or grip on the writing instrument is another useful indicator. With the progression of PD, pen pressure reduces (Ünlü et al., 2006). Although variance in pen pressure can be computed by measuring variance in pixel density from offline images, yet use of specialized equipment in acquisition of online samples, has been frequently applied to quantify pen pressure with more precision.

2.4. *Related Works*

In a number of related studies (Drotár et al., 2013a,b, 2014), effectiveness of various online features for prediction of PD is evaluated using the customized template illustrated in Figure 3. Most of these features included kinematic measurements (like stroke speed, vertical/horizontal acceleration and jerk etc.), computed from in-air as well as on-surface time intervals. Classification was carried out by applying Support Vector Machine (SVM) and maximum accuracy of 85.16% was achieved on a database of 75 subjects (37 control and 38 PD patients). Authors also suggested that classification accuracy depends on the choice of template. Some tasks are better representative of symptomatic signs than others.

In another study (Rosenblum et al., 2013b), a combination of various spatio-temporal and pressure measures of on-surface strokes was computed from online samples of 20 PD and 20 control subjects using a digitizer tablet. By employing discriminant analysis, a classification rate of 97.5% was reported. Graça et al. (2014) presented an Android application that employs gait analysis in addition to handwriting analysis for prediction of PD. Three classifiers including C4.5, RipperK and Bayesian Network were employed reporting accuracies of 86.67%, 80.83% and 87.50% respectively on 35 subjects.

In another set of related studies (Pereira et al., 2015, 2016), authors assessed both offline and signal-based online features for prediction of PD using various machine learning and graph based classifiers. A dataset comprising of hand drawn spiral samples of 55 individuals was used to assess the offline features. Approximately 75% accuracy was achieved using a combination of different offline features (i.e. Mean Relative Tremor and spatial measures) and SVM as classifier. A dataset consisting of spiral and meander drawing samples of 35 subjects was used to assess performance of signal-based features. Accuracies achieved on signal-based features ranged from 79% to 87%. It was, however, observed that the realized results are sensitive to the choice of classifier used. An interesting aspect of their work is use of pre-trained Convolutional Neural Networks (CNNs) (LeCun et al., 1989, 1998) for classification. Nevertheless, classification was performed on a combination of signals captured by smart pen and not on visual attributes.

It is observed that due to ease of acquisition and depth of information present, researchers are becoming more inclined to-

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

wards using online dynamic features in their studies. Nevertheless, the potential of static visual information cannot be undermined (Moetesum et al., 2015; Zhi et al., 2015). It is however important to comprehend, that the performance of a computerized prediction system is a function of a number of parameters including acquisition template, discriminating features and the classification techniques employed.

3. Proposed Methodology and Experimental Setup

As mentioned in the introductory section, the objective of this study is to evaluate the effectiveness of visual attributes as discriminators between graphomotor impressions created by PD and control subjects. To achieve this objective, we propose a system consisting of multiple networks trained on samples of various graphomotor tasks. To enrich feature learning, multiple representations of input data are used to train these networks. The features learned by different networks are then combined together (early fusion). Later, predictions from different tasks are pooled in using a voting scheme (late fusion) to enhance classification results. The overall system architecture is illustrated in Figure 4.

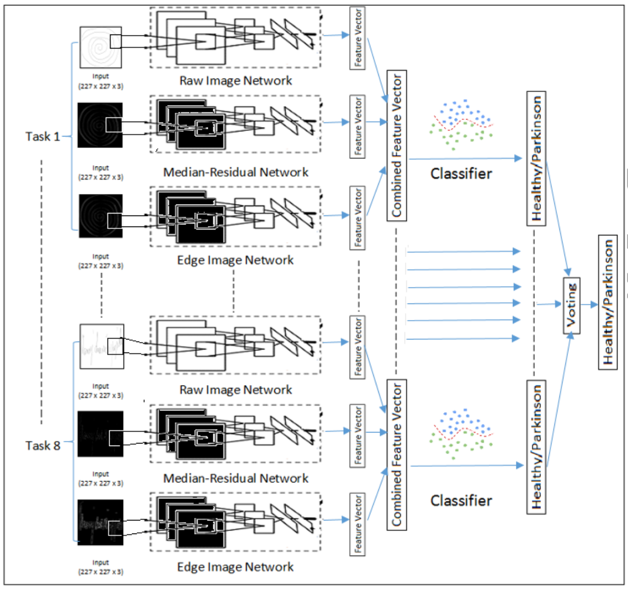
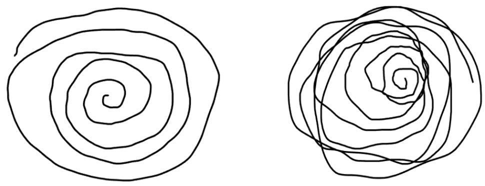


Fig. 4. System Overview

3.1. Dataset

To assess the performance of our proposed methodology, we used the 'Parkinson's Disease Handwriting Database' (PaHaW), compiled by authors in (Drotár et al., 2014). The employed dataset comprises of samples collected from 75 subjects (37 PD patients and 38 (age and gender comparable) control subjects). The mean UPDRS-Part V score for PD patients is 2.27 ± 0.84. The tasks were based on the template proposed in (Drotár et al., 2013b) and described in previous section (Figure 3). However, some subjects did not complete all tasks in the template, hence their samples were not considered in our study. A total of 576 samples from 72 (36 PD and 36 control) subjects were used. The idea of capturing more information from different samples produced by the same subject was the prime incentive for using PaHaW.

The original PaHaW database does not include images, instead it consists of various online attributes such as the (x,y) coordinates of the pen trajectory as well as the pen status (whether touching the writing surface or in air). A visual image of the drawing produced by the subject is generated by plotting the normalized (x,y) coordinates corresponding to all positions where the pen is touching the writing surface. The x coordinate is normalized to 0 (by subtracting the minimum value from every coordinate) while the y coordinate is normalized by subtracting the mean from each value. The sampling frequency in acquisition of online data is sufficiently high (200Hz), therefore connecting the coordinates of pen trajectories produces smooth drawing traces that can be considered very much similar to the real offline images, an example is illustrated in Figure 5.



(a) (b)

Fig. 5. Generated image of (a) Archimedean spiral drawn by Control Subject, (b) Archimedean spiral drawn by PD Patient

3.2. Convolutional Neural Network (CNN) based Feature Extraction

While domain specific 'hand-crafted' feature representations have been investigated for decades to solve similar classification problems, a number of recent studies (Szegedy et al., 2015; Mollahosseini et al., 2016; Herath et al., 2017), advocate the superiority of machine learned features over traditional ones. Deep feature learning approaches like Convolutional Neural Networks (CNNs), can easily be applied to various datasets and are known to capture useful information from diverse samples within a dataset. Pre-trained CNNs can also be employed off-the-shelf both as feature extractors and as classifiers (Sharif Razavian et al., 2014). In this study, we have employed a pre-trained CNN (AlexNet, (Krizhevsky et al., 2012)) for feature extraction purposes. This technique is known as 'Transfer Learning' and is generally applied in scenarios where training data is limited, like the case under consideration.

AlexNet architecture comprises of 5 convolution layers, max-pooling layers, dropout layers, and 3 fully connected layers. It is trained on 1.2 Million images (with 1000 different classes) of the ImageNet dataset. The network constructs a hierarchical representation of input images. Deeper layers contain higher-level features, constructed using the lower-level features of earlier layers. Together, the convolutional and down sampling layers serve as feature extractors while the fully connected layers represent a trainable classifier similar

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

to a standard multi-layer neural network. In transfer learning that we have chosen, the fully connected layers (performing classification) are removed and the output of the feature extractor layers is fed to another classifier. In this study, we have employed transfer learning by extracting features at $fc7$ layer.

Although internal representations of CNN layers are hard to decode, an intuitive guess can be made by visualizing the output of various layers of the network. Figure 6, shows outputs of a random activation channel after convolutional layer 3 on two input images of a spiral drawn by a healthy subject and a PD patient. It is clearly seen that neurons of the same activation channel react differently to smoothly drawn spiral edges and to irregular ones. This gives some indication that the network is capable of learning discriminating features required for this problem.

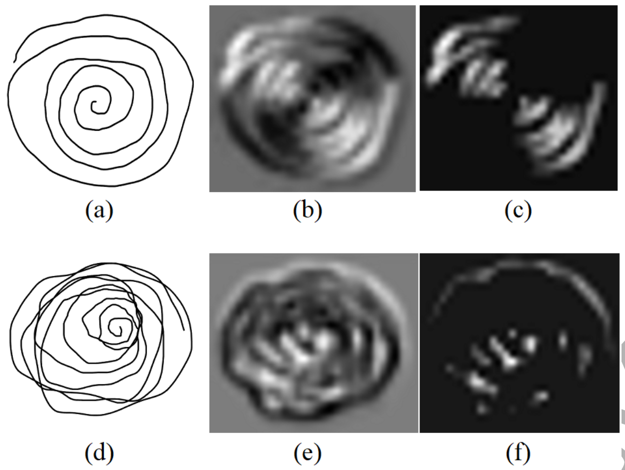


Fig. 6. (a) Archimedean spiral drawn by Control subject, (b) Random image after convolutional layer 3, (c) Channel with maximum neuron activity, (d) Archimedean spiral drawn by PD patient (e) Random image after convolutional layer 3 (f) Same channel with very less neuron activity

3.3. Fusion Techniques for Performance Enhancement

Till now, one of the limiting factors for application of feature learning for this problem, is the unavailability of large amount of training data. Nevertheless, literature shows that complementing methods like synthetic data generation (Huang et al., 2017), data augmentation (Ding et al., 2016) and fusion techniques (Park et al., 2016), can significantly improve the performance of a CNN-based system in such scenarios. In this study, we have used two light weight fusion techniques to enhance both feature learning and classification.

3.3.1. Early Fusion Technique

One limitation of a CNN is the computation of only linear characteristics. Hence to enhance feature learning, we propose to present to the CNN the initial data and the result of the transformation of this initial data through different non linear transforms. Three representations (Figure 7) of the input data are used to train three independent networks for each of the 8 tasks performed by a subject. AlexNet computes 4096-dimensional features from each input image. The features extracted by the 3 networks are then fused into a combined feature vector. The combined feature vector is then fed to a dedicated classifier (i.e. SVM) which takes a decision. Same is repeated for each of the 8 tasks. Benefits of using this technique are two-fold. Multiple representations of the input image not only increase training data but training smaller networks also reduces computational overheads. Furthermore, appropriate selection of representation can help the network in learning better features. Brief details of the three representations used are given below.

- Raw data (D_r): Conventionally raw images are used as input data for CNNs as they contain different frequency components that can be extracted and used for image classification. Raw images of the 8 tasks, completed by subjects, are used to train the first network.

- Median filter residual data (D_m): The second network is trained using median filter residuals of the same raw images. To compute the median filter residual, we applied a 3×3 median filter on the raw image and then subtracted the raw image from the resultant filtered image. The idea is to preserve high frequency imperfections. Median filter residual of a raw image (Figure 7-a) is shown in Figure 7-b.

- Edge data (D_e): The third network is trained using images containing only the edge information from the raw images (Figure 7-c). Edges are known to contain useful information in most cases. By applying linear convolution filters in vertical and horizontal directions, the magnitude of the gradient is computed in a non linear way. As a result, we obtained emphasized edge information of the shape and used it to train our network.

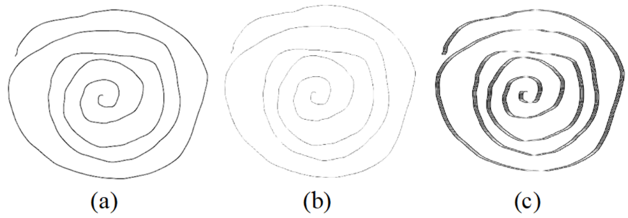


Fig. 7. (a) Raw image, (b) Median filter residual image (Pixel values inverted for better visualization), (c) Edge image (Inverted for better visualization)

3.3.2. Late Fusion Technique

The resultant combined feature vector of each task, extracted by CNNs, is fed to Support Vector Machine (SVM) for classification. A number of studies (Szarvas et al., 2005; Mori et al., 2005; Lauer et al., 2007; Niu and Suen, 2012) have investigated the combination of CNN-SVM, and reported improved classification results. The objective at hand is to take task level decisions from multiple samples of the same subject. The outputs of the 8 classifiers in our system form the decision vector d defined as $d = [d_1, d_2, d_3, ..., d_8]^T$, where $d_i \in \{c_1, c_2\}$ and c_i denotes label of either of the class (i.e. Healthy/PD). In the next

ACCEPTED MANUSCRIPT

step, we applied voting based late fusion. This strategy is suitable for a multiple classifier system (Ruta and Gabrys, 2000), where each classifier gives a single class label as an output, as considered in our proposed system. In general voting, the output class is decided only when all classifiers produce the same output. However, for our system, we have employed 'Majority Voting'. By varying parameters, we can adjust the weightage given to number of tasks used for final diagnosis.

4. Analysis of Results

In this section, we evaluate the performance of our proposed scheme in light of the results of the experiments conducted.

4.1. Evaluation Metrics

The effectiveness of the proposed image representations is evaluated by computing the system accuracy for each of the tasks separately and against each representation. Accuracy is also reported by combining the feature vectors of the three representations (early fusion) as well as by combining the predictions of the eight modalities through majority vote (late fusion). Furthermore, class-wise precision, specificity and sensitivity values are also reported. For completeness, each of these metrics is described briefly in the following in terms of True Positives (t_p), False Positives (f_p), True Negatives (t_n) and False Negatives (f_n).

- **Accuracy** measures the overall ability of the system to correctly classify PD patients and healthy subjects: $\frac{t_p+f_n}{t_p+t_n+f_p+f_n}$.
- **Sensitivity** measures the ability to correctly classify the PD patients and is calculated as the proportion of True Positives in the diseased cases and is defined by the ratio $\frac{t_p}{t_p+f_n}$.
- **Specificity** measures the ability of the system to correctly classify the healthy subjects and is defined as $\frac{t_n}{t_n+f_p}$
- **Precision** is the true positive relevance rate and is defined as $\frac{t_p}{t_p+f_p}$.

The experimental protocol is same as that of (Drotár et al., 2013a, 2014, 2015, 2016), i.e. 10-fold cross validation; the reported performances hence representing the average of 10 runs.

4.2. Experimental Results

We first present the accuracy of the system against different representations for each of the tasks in Table 1. The accuracies represent average accuracies of ten runs. Comparing the performance of various tasks, it can be observed from Table 1 that Task-1 ('Archimedean Spiral') reports highest accuracies across all three data representations (Accuracy of 57% on raw images and 65% on median residual and edge images). Comparing the performance of various image representations, median residual images (D_m) and edge images (D_e) outperform the raw image representations (D_r). Combining the predictions of all tasks using majority voting results in increasing the accuracies, the enhancement, however, seems to be marginal.

We also carry out experiments by combining the feature vectors of different image representations (D_r, D_m and D_e). The accuracies realized in these experiments are summarized in Table 2 where it can be seen that combining different image representations (early fusion) serves to improve the accuracies on each of the tasks. The performance of individual tasks when using combined feature vectors is similar to those reported in Table 1; Task 1 outperforms all other tasks while accuracies across different tasks exhibit similar trends in Table 1 and Table 2. The highest accuracy realized is 83% when combining feature vectors of all three representations (early fusion) as well as the predictions of all tasks (late fusion).

System performance in terms of precision, specificity and sensitivity, on all tasks (late fusion) using the three image representations and their combination is summarized in Table 3. Similar to accuracies, the highest values of these metrics are realized using a combination of the features from the three representations which are comparable to those reported in (Drotár et al., 2016).

4.3. Statistical Analysis

To statistically compare the effectiveness of different representations and tasks, we investigated the following.

- Whether the performance of different representations is statistically different and whether the performance enhancement of combining multiple representations is statistically significant.
- Whether the difference in performance of different tasks is statistically significant, which tasks are statistically better than which other tasks.

For the aforementioned statistical comparisons, we employed the non-parametric Friedman Test (Friedman, 1940). The test ranks different scenarios based on their performance, for instance the best performance is ranked 1 and so on. In case of a tie, the two scenarios are assigned an average rank. The Friedman statistic is then computed as a function of the average rank of all scenarios (Equation 1).

$$X_F^2 = \frac{12N}{k(k+1)}\left(\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right) \tag{1}$$

Where R_j is the average rank of scenario j, N is the number of experiments (i.e. 10 for 10-cross validation in our case) and k is the number of scenarios (tasks/representations). The following improved version (Iman and Davenport, 1980) of the Friedman statistic is generally employed.

$$F_F = \frac{(N-1)X_F^2}{N(k-1) - X_F^2} \tag{2}$$

The null hypothesis of Friedman pre-test assumes that all scenarios are performance-wise equivalent. If rejected a post-hoc test is employed to determine the statistical difference

ACCEPTED MANUSCRIPT

Table 1. Task-wise System Accuracies for Different Data Representations:(*D_r*: **Raw Image,** *D_m*: **Median Residual Image,** *D_e*: **Edge Image)**

	Data Representation		
Task	D_r	D_m	D_e
1 (Archimedean Spiral)	0.57 ± 0.05	0.65 ± 0.06	0.65 ± 0.09
2 (Letter ‘l’)	0.53 ± 0.09	0.55 ± 0.10	0.57 ± 0.09
3 (Bigram ‘le’)	0.48 ± 0.09	0.51 ± 0.09	0.54 ± 0.08
4 (Word ‘les’)	0.50 ± 0.11	0.57 ± 0.09	0.55 ± 0.07
5 (Word ‘lektorka’)	0.49 ± 0.10	0.58 ± 0.07	0.52 ± 0.11
6 (Word ‘porovnat’)	0.46 ± 0.08	0.49 ± 0.09	0.48 ± 0.08
7 (Word ‘nepopadnout’)	0.54 ± 0.07	0.64 ± 0.07	0.60 ± 0.05
8 (Sentence)	0.48 ± 0.08	0.49 ± 0.09	0.48 ± 0.09
All Tasks	0.58 ± 0.07	0.68 ± 0.07	0.66 ± 0.07

Table 2. Task-wise System Accuracies for Different Combinations of Data Representations: (*D_r*: **Raw Image,** *D_m*: **Median Residual Image,** *D_e*: **Edge Image)**

	Data Representation				
Task	$D_r + D_m$	$D_r + D_e$	$D_m + D_e$	$D_r + D_m + D_e$	(Drotár et al., 2016)
1 (Archimedean Spiral)	0.67 ± 0.08	0.70 ± 0.05	0.65 ± 0.08	0.76 ± 0.08	0.62
2 (letter ‘l’)	0.55 ± 0.12	0.52 ± 0.08	0.50 ± 0.09	0.62 ± 0.08	0.72
3 (Bigram ‘le’)	0.51 ± 0.09	0.52 ± 0.11	0.55 ± 0.07	0.57 ± 0.09	0.71
4 (Word ‘les’)	0.54 ± 0.07	0.52 ± 0.08	0.57 ± 0.05	0.60 ± 0.08	0.66
5 (Word ‘lektorka’)	0.54 ± 0.08	0.52 ± 0.11	0.51 ± 0.09	0.60 ± 0.07	0.65
6 (Word ‘porovnat’)	0.50 ± 0.09	0.49 ± 0.09	0.47 ± 0.06	0.51 ± 0.09	0.73
7 (Word ‘nepopadnout’)	0.65 ± 0.06	0.59 ± 0.06	0.63 ± 0.08	0.68 ± 0.07	0.67
8 (Sentence)	0.50 ± 0.09	0.49 ± 0.10	0.49 ± 0.06	0.51 ± 0.08	0.76
All Tasks	0.73 ± 0.08	0.76 ± 0.07	0.79 ± 0.07	0.83 ± 0.09	0.81

between scenarios. In our study, we employed the popular post-hoc Nemenyi test (Nemenyi, 1963), which performs a pairwise comparison of scenarios under consideration. The test computes a Critical Difference (CD) that is used to determine whether the distance between the average ranks of a pair of scenarios is statistically significant. For instance if the difference between the mean ranks of two scenarios is greater than CD, there exists a significant performance difference between the two.

Our first investigation was to validate the use of multiple representation of input data instead of one particular representation. The results of Friedman pre-test rejected the null hypothesis that all representations are equally effective. As a consequence, Nemenyi test was performed, results of which are presented in Figure 8. It is evident from the results that combined representation outperforms all individual representations thus validating the use of early fusion in the given scenario. Another interesting observation is that using various representations instead of raw data representation yields significantly improved classification results. The performance difference between the median-residual and edge images, however, is not significant. Nevertheless, combination of all three individual representations significantly enhances overall system performance.

We also carried out statistical investigations to evaluate the performance of our scheme on various tasks. The statistics computed by the Friedman pre-test state that performance of our proposed scheme on individual tasks is significantly different. As a result Nemenyi post-hoc test was performed. The results of Nemenyi post-hoc test are summarized in Figure 9. It is clearly seen that fusing decisions of all tasks based on majority voting significantly outperforms individual task-wise classification (with an exception of Task 1). While considering effectiveness of individual tasks, it is seen that classification performance of Task 1 (Archmidean Spiral) and Task 7 (word ‘nepopadnout’) is statistically better than the rest of the handwriting tasks. Same observations were made while conducting the experiments suggesting that static visual features can capture effective information from templates which support on-surface continuity. Task 8 and Task 6 have the least impact on the performance of the system. Although Tasks (2, 4 & 5) perform significantly better than Task 8 and Task 6, there is no significant difference between their own performances.

Representation	Combined	Dm	De	Dr	Mean Rank
Combined	0	1	1	1	1.1
Dm	-1	0	0	1	2.5
De	-1	0	0	1	2.7
Dr	-1	-1	-1	0	3.6

0	Row representation is same as column representation
-1	Row representation is worse than column representation
1	Row representation is better than column representation

Fig. 8. Nemenyi Pairwise Statistical Test (CD = 0.49) for Performance Comparison of Individual Representations (*D_r*: Raw Image, *D_m*: Median Residual Image, *D_e*: Edge Image) & Combined Representation

4.4. *Comparative Analysis*

We also present an overview of the performance of the notable studies on handwriting based prediction of PD in Table 4. Although a meaningful comparison of our system is possible with the works of Drotár et al. (2013a, 2014, 2015, 2016), results of other studies are also listed to provide an idea on the current accuracies on this problem across multiple datasets. It is interesting to note that most of the studies listed in Table 4 have exploited dynamic features of handwriting while the visual attributes have not been explored extensively. Although CNNs have been employed by Pereira et al. (2016), the features are learned using the signals extracted through

ACCEPTED MANUSCRIPT

Table 3. Overall System Performance using Individual & Combined Representations: (D_r: Raw Image, D_m: Median Residual Image, D_e: Edge Image)

Metric	Features				
	D_r	D_m	D_e	Combined	Drotár et al. (2016)
Precision	0.64 ± 0.13	0.67 ± 0.05	0.75 ± 0.19	0.89 ± 0.12	-
Sensitivity	0.55 ± 0.13	0.69 ± 0.14	0.72 ± 0.14	0.84 ± 0.14	0.87
Specificity	0.64 ± 0.07	0.65 ± 0.13	0.63 ± 0.24	0.82 ± 0.15	0.80

Task	All Tasks	Task1	Task7	Task2	Task4	Task5	Task3	Task8	Task6	Mean Rank
All Tasks	0	0	1	1	1	1	1	1	1	1.4
Task1	0	0	1	1	1	1	1	1	1	2.3
Task7	-1	-1	0	1	1	1	1	1	1	3.8
Task2	-1	-1	-1	0	0	0	0	1	1	5.2
Task4	-1	-1	-1	0	0	0	0	1	1	5.5
Task5	-1	-1	-1	0	0	0	0	1	1	5.8
Task3	-1	-1	-1	0	0	0	0	0	0	6.3
Task8	-1	-1	-1	-1	-1	-1	0	0	0	7.2
Task6	-1	-1	-1	-1	-1	-1	0	0	0	7.3

0	Row task is same as column task
-1	Row task is worse than column task
1	Row task is better than column task

Fig. 9. **Nemenyi Pairwise Statistical Test (CD = 1.28) for Performance Comparison of Tasks**

a smart pen with multiple sensors capturing handwriting dynamics; the visual attributes of writing are not considered. Comparing the accuracy of 83% reported by the proposed image representations and fusion techniques, with those realized in Drotár et al. (2013a, 2014, 2015, 2016) on the same dataset[1] (and same experimental protocol), it can be observed that visual features of handwriting report comparable performance to those realized using dynamic features.

Some interesting observations can be made while comparing performance results of static visual features with other dynamic features. For instance, in (Drotár et al., 2016), Task 1 (Archimedean Spiral) is reported to be the least effective amongst the 8 tasks (Table 2). On the contrary, in our proposed scheme, Task 1 significantly outperforms rest of the tasks. However, Drotár et al. (2016) attributed the relatively poor performance of spiral test in their study to the features under consideration, suggesting that different features can perform differently on different tasks. Authors did not find significantly effective kinematic features for both Task 1 (Archimedean Spiral) and Task 4 (Trigram 'les') and consequently relied only on pressure features, thus resulting in a lower classification accuracy for both tasks. On the contrary, the experimental as well as the statistical analysis carried out in our study showed that visual attributes extracted from both these tasks (especially Task 1) yield good results. While comparing the performance of the proposed scheme with other techniques (Pereira et al., 2015; Graça et al., 2014) considering the Archimedean spiral only, it is seen that employing only the static visual features measured from online drawing samples yield comparable results to schemes which employed a combination of different dynamic features. Further investigation in using the proposed features from such drawings (e.g. spiral and meander, etc.) can be potentially useful for improving accuracy in prediction or differential diagnosis of Parkinson's disease.

[1] We have used 72 subjects from 75 as data for three subjects was incomplete

Contrary to the results reported in (Drotár et al., 2016) where Task 8 (sentence writing task) performed best classification (with accuracy 76.5%), the same task performed poorly in our proposed scheme. Drotár et al. (2016) attributed the performance of Task 8 to the in-air time interval which subjects required to complete the task. Such temporal information cannot be employed in our proposed scheme, nevertheless tasks which employed more on-surface time or required continuity (i.e. Task 1 and Task 7) performed best when visual features were extracted from them. Such observations strongly indicate the correlation between features and tasks and support the notion that template selection or design must be kept in view while designing a decision support system for early prediction and differential diagnosis of PD or any other neurological disease.

5. Conclusion

This study investigated the potential of visual attributes of handwriting to predict Parkinson's disease. While the existing literature primarily targets kinematic, pressure and spatio-temporal features, we exploit the static visual attributes of handwriting extracted using Convolutional Neural Networks. The idea is not to deny the effectiveness of the rich online features but to manifest the fact that visual information in handwriting can still be effectively employed for this problem. Indeed, combining the two types of features (online and offline) can lead to some interesting findings. In order to enhance learning, we used median residual and edge images in addition to raw images to enrich the feature set (early fusion). Classification is carried out using Support Vector Machine (SVM) and predictions of different tasks are combined using majority voting (late fusion). Evaluations on a standard dataset of 72 subjects reported an overall accuracy of 83%. Comparable performance of the proposed technique with those based on online features validate the idea that visual attributes extracted from images of handwriting and hand-drawn shapes can be effectively employed for this problem. In our further study on this problem, we intend to combine these offline features with other online features to evaluate their effect on the prediction accuracies of PD and other neurological diseases. It would also be interesting to explore other drawing tasks and whether similar results are acquired on actual offline images. Performance comparison across multiple datasets (diverse templates) will also make the subject of our further study.

ACCEPTED MANUSCRIPT

Table 4. Performance Comparison of Handwriting based Parkinson Prediction Systems

Study	Features	Classifier	Dataset	Template	Results
Drotár et al. (2013a)	Online in-air movement based Features	SVM	75 subjects	8 Task Template	80.09%
Drotár et al. (2013b)	Online Kinematic Features	SVM	75 subjects	8 Task Template	80%
Rosenblum et al. (2013b)	Online Spatio-Temporal & Pressure Features	Discriminant Analysis	40 subjects	Signature	97.5%
Drotár et al. (2014)	Online in-air & on-surface movement based Features	SVM	75 subjects	8 Task Template	85.16%
Graça et al. (2014)	Online Spatio-Temporal & Pressure Features	C4.5	35 subjects	Archimedean Spiral	86.67%
		RipperK			80.83%
		Bayesian Network			87.50%
Pereira et al. (2015)	Mean Relative Tremor and Spatial Features	Naive Bayes	55 subjects	Archimedean Spiral	78.9%
		Optimum-Path Forest			77.1%
		SVM			75.8%
Drotár et al. (2015)	Online Spatio-Temporal & Kinematic Features, Entropy, Signal Energy, Empirical Mode Decomposition	SVM	75 subjects	7 Task Template	88.13%
Drotár et al. (2016)	Online Kinematic & Pressure Features	SVM	75 subjects	8 Task Template	81.3%
		AdaBoost			78.9%
		K-NN			71.7%
Pereira et al. (2016)	Pen-based Features	CNN	35 subjects	Archimedean Spiral	80.19%
				Meander	87.14%
		Optimum-Path Forest		Archimedean Spiral	79.2%
				Meander	84.42%
Proposed Technique	**CNN based Visual Features**	**SVM**	**72 Subjects**	**8 Task Template**	**83%**

References

Aly, N.M., Playfer, J.R., Smith, S.L., Halliday, D.M., 2007. A novel computer-based technique for the assessment of tremor in parkinson's disease. Age and Ageing 36, 395–399.

Berardelli, A., Rothwell, J.C., Thompson, P.D., Hallett, M., 2001. Pathophysiology of bradykinesia in parkinson's disease. Brain 124, 2131–2146.

Caligiuri, M.P., Mohammed, L.A., 2012. The neuroscience of handwriting: Applications for forensic document examination. CRC Press.

Campenhausen, S., Bornschein, B., Wick, R., Bötzel, K., Sampaio, C., Poewe, W., Oertel, W., Siebert, U., Berger, K., Dodel, R., 2005. Prevalence and incidence of parkinson's disease in europe. European Neuropsychopharmacology 15, 473–490.

Chung, S.J., 2012. Diagnosis and treatment of hand tremor. Journal of the Korean Medical Association/Taehan Uisa Hyophoe Chi 55.

De Lau, L.M., Breteler, M.M., 2006. Epidemiology of parkinson's disease. The Lancet Neurology 5, 525–535.

Derkinderen, P., Dupont, S., Vidal, J.S., Chedru, F., Vidailhet, M., 2002. Micrographia secondary to lenticular lesions. Movement disorders 17, 835–837.

Ding, J., Chen, B., Liu, H., Huang, M., 2016. Convolutional neural network with data augmentation for sar target recognition. IEEE Geoscience and Remote Sensing Letters 13, 364–368.

Dirlikov, B., Younes, L., Nebel, M.B., Martinelli, M.K., Tiedemann, A.N., Koch, C.A., Fiorilli, D., Bastian, A.J., Denckla, M.B., Miller, M.I., et al., 2017. Novel automated morphometric and kinematic handwriting assessment: A validity study in children with asd and adhd. Journal of Occupational Therapy, Schools, & Early Intervention 10, 185–201.

Drotár, P., Mekyska, J., Rektorova, I., Masarova, L., Smékal, Z., Faundez-Zanuy, M., 2013a. A new modality for quantitative evaluation of parkinson's disease: In-air movement, in: IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE), 2013, IEEE. pp. 1–4.

Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., Faundez-Zanuy, M., 2014. Analysis of in-air movement in handwriting: A novel marker for parkinson's disease. Computer methods and programs in biomedicine 117, 405–411.

Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., Faundez-Zanuy, M., 2015. Decision support framework for parkinsons disease based on novel handwriting markers. IEEE Transactions on Neural Systems and Rehabilitation Engineering 23, 508–516.

Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., Faundez-Zanuy, M., 2016. Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease. Artificial intelligence in medicine 67, 39–46.

Drotár, P., Mekyska, J., Smekal, Z., Rektorova, I., Masarova, L., Faundez-Zanuy, M., 2013b. Prediction potential of different handwriting tasks for diagnosis of parkinson's, in: E-Health and Bioengineering Conference (EHB), 2013, IEEE. pp. 1–4.

Eichhorn, T., Gasser, T., Mai, N., Marquardt, C., Arnold, G., Schwarz, J., Oertel, W., 1996. Computational analysis of open loop handwriting movements in parkinson's disease: a rapid method to detect dopamimetic effects. Movement disorders 11, 289–297.

Feder, K.P., Majnemer, A., 2007. Handwriting development, competency, and intervention. Developmental Medicine & Child Neurology 49, 312–317.

Fish, J., 2011. Unified Parkinson's Disease Rating Scale. Springer New York, New York, NY. pp. 2576–2577.

Friedman, M., 1940. A comparison of alternative tests of significance for the problem of m rankings. The Annals of Mathematical Statistics 11, 86–92.

Fuentes, C.T., Mostofsky, S.H., Bastian, A.J., 2009. Children with autism show specific handwriting impairments. Neurology 73, 1532–1537.

Graça, R., e Castro, R.S., Cevada, J., 2014. Parkdetect: Early diagnosing parkinson's disease, in: IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2014, IEEE. pp. 1–6.

Herath, S., Harandi, M., Porikli, F., 2017. Going deeper into action recognition: A survey. Image and Vision Computing 60, 4–21.

Hsu, A.W., Piboolnurak, P.A., Floyd, A.G., Yu, Q.P., Wraith, J.E., Patterson, M.C., Pullman, S.L., 2009. Spiral analysis in niemann-pick disease type c. Movement Disorders 24, 1984–1990.

Huang, H., Kalogerakis, E., Yumer, E., Mech, R., 2017. Shape synthesis from sketches via procedural models and convolutional networks. IEEE Transactions on Visualization and Computer Graphics 23, 2003–2013.

Iman, R.L., Davenport, J.M., 1980. Approximations of the critical region of the fbietkan statistic. Communications in Statistics-Theory and Methods 9, 571–595.

Johnson, B.P., Phillips, J.G., Papadopoulos, N., Fielding, J., Tonge, B., Rinehart, N.J., 2015. Do children with autism and asperger's disorder have difficulty controlling handwriting size? a kinematic evaluation. Research in Autism Spectrum Disorders 11, 20–26.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097–1105.

Kushki, A., Chau, T., Anagnostou, E., 2011. Handwriting difficulties in children with autism spectrum disorders: A scoping review. Journal of autism and developmental disorders 41, 1706–1716.

Lauer, F., Suen, C.Y., Bloch, G., 2007. A trainable feature extractor for handwritten digit recognition. Pattern Recognition 40, 1816–1824.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L., 1989. Handwritten digit recognition with a back-propagation network, in: Neural Information Processing Systems (NIPS).

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278–2324.

Mavrogiorgou, P., Mergl, R., Tigges, P., El Husseini, J., Schröter, A., Juckel, G., Zaudig, M., Hegerl, U., 2001. Kinematic analysis of handwriting movements in patients with obsessive-compulsive disorder. Journal of Neurology, Neurosurgery & Psychiatry 70, 605–612.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Michalec, M., Hernandez, N., Clark, L.N., Louis, E.D., 2014. The spiral axis as a clinical tool to distinguish essential tremor from dystonia cases. Parkinsonism & related disorders 20, 541–544.

Moetesum, M., Siddiqi, I., Masroor, U., Djeddi, C., 2015. Automated scoring of bender gestalt test using image analysis techniques, in: 13th International Conference on Document Analysis and Recognition (ICDAR), 2015, IEEE. pp. 666–670.

Mollahosseini, A., Chan, D., Mahoor, M.H., 2016. Going deeper in facial expression recognition using deep neural networks, in: IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, IEEE. pp. 1–10.

Mori, K., Matsugu, M., Suzuki, T., 2005. Face recognition using svm fed with intermediate output of cnn for face detection., in: MVA, pp. 410–413.

Nackaerts, E., Vervoort, G., Heremans, E., Smits-Engelsman, B.C., Swinnen, S.P., Nieuwboer, A., 2013. Relearning of writing skills in parkinson's disease: a literature review on influential factors and optimal strategies. Neuroscience & Biobehavioral Reviews 37, 349–357.

Nemenyi, P., 1963. Distribution-free Multiple Comparisons.

Niazmand, K., Tonn, K., Kalaras, A., Fietzek, U.M., Mehrkens, J.H., Lueth, T.C., 2011. Quantitative evaluation of parkinson's disease using sensor based smart glove, in: 24th International Symposium on Computer-Based Medical Systems (CBMS), 2011, IEEE. pp. 1–8.

Niu, X.X., Suen, C.Y., 2012. A novel hybrid cnn–svm classifier for recognizing handwritten digits. Pattern Recognition 45, 1318–1325.

Palmerini, L., Rocchi, L., Mellone, S., Valzania, F., Chiari, L., 2011. Feature selection for accelerometer-based posture analysis in parkinson's disease. IEEE Transactions on Information Technology in Biomedicine 15, 481–490.

Park, E., Han, X., Berg, T.L., Berg, A.C., 2016. Combining multiple sources of knowledge in deep cnns for action recognition, in: IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, IEEE. pp. 1–8.

Pereira, C.R., Pereira, D.R., da Silva, F.A., Hook, C., Weber, S.A., Pereira, L.A., Papa, J.P., 2015. A step towards the automated diagnosis of parkinson's disease: Analyzing handwriting movements, in: IEEE 28th International Symposium on Computer-Based Medical Systems (CBMS), 2015, IEEE. pp. 171–176.

Pereira, C.R., Weber, S.A., Hook, C., Rosa, G.H., Papa, J.P., 2016. Deep learning-aided parkinson's disease diagnosis from handwritten dynamics, in: 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2016, IEEE. pp. 340–346.

Pullman, S.L., 1998. Spiral analysis: a new technique for measuring tremor with a digitizing tablet. Movement Disorders 13, 85–89.

Rémi, C., Frélicot, C., Courtellemont, P., 2002. Automatic analysis of the structuring of children's drawings and writing. Pattern Recognition 35, 1059–1069.

Renau-Ferrer, N., Remi, C., 2013. A method for visuo-spatial classification of freehand shapes freely sketched. arXiv preprint arXiv:1305.1520 .

Rosenblum, S., 2015. Do motor ability and handwriting kinematic measures predict organizational ability among children with developmental coordination disorders? Human movement science 43, 201–215.

Rosenblum, S., Engel-Yeger, B., Fogel, Y., 2013a. Age-related changes in executive control and their relationships with activity performance in handwriting. Human movement science 32, 363–376.

Rosenblum, S., Livneh-Zirinski, M., 2008. Handwriting process and product characteristics of children diagnosed with developmental coordination disorder. Human Movement Science 27, 200–214.

Rosenblum, S., Samuel, M., Zlotnik, S., Erikh, I., Schlesinger, I., 2013b. Handwriting as an objective tool for parkinsons disease diagnosis. Journal of neurology 260, 2357–2361.

Ruta, D., Gabrys, B., 2000. An overview of classifier fusion methods. Computing and Information systems 7, 1–10.

Saunders-Pullman, R., Derby, C., Stanley, K., Floyd, A., Bressman, S., Lipton, R.B., Deligtisch, A., Severt, L., Yu, Q., Kurtis, M., et al., 2008. Validity of spiral analysis in early parkinson's disease. Movement disorders 23, 531–537.

Schröter, A., Mergl, R., Bürger, K., Hampel, H., Möller, H.J., Hegerl, U., 2003. Kinematic analysis of handwriting movements in patients with alzheimers disease, mild cognitive impairment, depression and healthy subjects. Dementia and Geriatric Cognitive Disorders 15, 132–142.

Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. Cnn features off-the-shelf: an astounding baseline for recognition, in: Proceedings of IEEE conference on computer vision and pattern recognition workshops, pp. 806–813.

Slavin, M.J., Phillips, J.G., Bradshaw, J.L., Hall, K.A., Presnell, I., 1999. Consistency of handwriting movements in dementia of the alzheimer's type: a comparison with huntington's and parkinson's diseases. Journal of the International Neuropsychological Society 5, 20–25.

Smits, E.J., Tolonen, A.J., Cluitmans, L., van Gils, M., Conway, B.A., Zietsma, R.C., Leenders, K.L., Maurits, N.M., 2014. Standardized handwriting to assess bradykinesia, micrographia and tremor in parkinson's disease. PloS one 9, e97614.

Stanley, K., Hagenah, J., Brüggemann, N., Reetz, K., Severt, L., Klein, C., Yu, Q., Derby, C., Pullman, S., Saunders-Pullman, R., 2010. Digitized spiral analysis is a promising early motor marker for parkinson disease. Parkinsonism & related disorders 16, 233.

Szarvas, M., Yoshizawa, A., Yamamoto, M., Ogata, J., 2005. Pedestrian detection with convolutional neural networks, in: IEEE Proceedings of Intelligent vehicles symposium, 2005., IEEE. pp. 224–229.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of IEEE conference on computer vision and pattern recognition, pp. 1–9.

Teulings, H., Contreras-Vidal, J.L., Stelmach, G., Adler, C.H., 2002. Adaptation of handwriting size under distorted visual feedback in patients with parkinson's disease and elderly and young controls. Journal of Neurology, Neurosurgery & Psychiatry 72, 315–324.

Teulings, H.L., Contreras-Vidal, J.L., Stelmach, G.E., Adler, C.H., 1997. Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control. Experimental neurology 146, 159–170.

Tsanas, A., Little, M.A., McSharry, P.E., Ramig, L.O., 2010. Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests. IEEE transactions on Biomedical Engineering 57, 884–893.

Tucha, O., Mecklinger, L., Thome, J., Reiter, A., Alders, G., Sartor, H., Naumann, M., Lange, K., 2006. Kinematic analysis of dopaminergic effects on skilled handwriting movements in parkinsons disease. Journal of neural transmission 113, 609–623.

Ünlü, A., Brause, R., Krakow, K., 2006. Handwriting analysis for diagnosis and prognosis of parkinsons disease, in: International Symposium on Biological and Medical Data Analysis, Springer. pp. 441–450.

Van Gemmert, A.W., Teulings, H.L., 2006. Advances in graphonomics: Studies on fine motor control, its development and disorders. Human Movement Science 25, 447–453.

Walton, J., 1997. Handwriting changes due to aging and parkinson's syndrome. Forensic science international 88, 197–214.

Weber, S.A.T., Santos Filho, C.A.d., Shelp, A.O., Resende, L.A.L., Papa, J.P., Hook, C., et al., 2014. Classification of handwriting patterns in patients with parkinson´ s disease, using a biometric sensor. Global Advanced Research Journal of Medicine and Medical Sciences , 362–366.

Werner, P., Rosenblum, S., Bar-On, G., Heinik, J., Korczyn, A., 2006. Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 61, P228–P236.

Yan, J.H., Rountree, S., Massman, P., Doody, R.S., Li, H., 2008. Alzheimers disease and mild cognitive impairment deteriorate fine movement control. Journal of Psychiatric Research 42, 1203–1212.

Zhi, N., Jaeger, B.K., Gouldstone, A., Frank, S., Sipahi, R., 2015. Objective quantitative assessment of movement disorders through analysis of static handwritten characters, in: ASME 2015 Dynamic Systems and Control Conference, American Society of Mechanical Engineers. pp. V001T16A006–V001T16A006.

Zhou, Y., Jenkins, M., Naish, M.D., Trejos, A.L., 2016. The measurement and analysis of parkinsonian hand tremor, in: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), 2016, IEEE. pp. 414–417.

ACCEPTED MANUSCRIPT