Unsupervised learning for characterization of Arabic online handwriting of Parkinson's disease patients

Ibtissame Aouraghe¹ · Alae Ammour¹ · Ghizlane Khaissidi¹ · Mostafa Mrabti¹ · Ghita Aboulem² · Faouzi Belahsen²

Received: 29 October 2019 / Accepted: 18 December 2019 / Published online: 1 January 2020
© Springer Nature Switzerland AG 2020

Abstract

In this paper, we propose to characterize the on-line handwriting for the early detection of Parkinson's disease. Thus, using kinematics, mechanical, and spatial features of handwriting, we are looking for the characterization of Parkinson's disease. This paper describes the phase of the data acquisition which is currently carried out with in the Neurological department of UHC Hassan II of Fez. Following this paper, we have proposed an approach based on unsupervised learning techniques for analyzing on-line handwriting of 34 Parkinson's disease patients and 34 Healthy Controls according to quantitative and qualitative features. Based on 230 computed features for each participant, our study has uncovered three different types of writers. The results show that the complications of fine motor abilities in Parkinson's disease patients is especially characterized by a significant degradation in handwriting kinematic features.

Keywords Online handwriting · Parkinson's disease · Principal component analysis · K-means clustering

1 Introduction

Handwriting (HW) is a complex activity requiring fine motor control, it is characterized by a major involvement of muscles and brain, and touches the deepest mechanisms of cognition and sensorimotor. The HW alteration is mostly due to neurodegenerative diseases such as Alzheimer [1–3], and Parkinson's Disease (PD) [4–7].

In fact, HW is a conventional and fine activity that requires learning, which can be different from one country to another, from one culture to another, and obviously from one language to another. The handwriting can be degraded with the age, the physical and mental state of the writer and according to the linguistic context or the task performed: Drafts, personal notes, private or official letters, thus the handwritten trace takes several different aspects.

The analysis of this handwritten trace is necessary and important to see how it changes or deteriorates through age, culture, language, and pathology. Moreover, several studies have been done on HW to develop a simple and early diagnosis of neurodegenerative pathologies. As a result, HW analysis has become a new and promising field of exploration for health researchers, especially in neurosciences. Starting from the fact that some changes in HW resulting from an irregular movement can be evocative of a disease. A degraded handwritten text could bring considerable diagnostic information on its author.

The real time acquisition of the HW from a digital tablet is called online handwriting. It allows to extract kinematic and spatio-temporal features while writing.

Based on a set of wisely elaborate HW exercises, the majority of the research studies conducted on this subject, have statistically proven that a strong relationship exists between the degradation of HW and neurodegenerative

✉ Ibtissame Aouraghe, aouraghe.ibtissame@gmail.com; Alae Ammour, alae.am19@gmail.com; Ghizlane Khaissidi, ghizlane.derkaoui1@hotmail.com; Mostafa Mrabti, mostafa.mrabti@yahoo.fr; Ghita Aboulem, aboulemghita@gmail.com; Faouzi Belahsen, belahsenfaouzi@gmail.com | ¹Laboratory LIPI ENS, Sidi Mohamed Ben Abdellah University, Fez, Morocco. ²Laboratory ERMSC, FMPF, University Hospital Center Hassan II, Fez, Morocco.

pathologies [8–15]. Thus, the early detection of these diseases based on online HW could be useful for helping to detect and control the progression of these pathologies.

In this paper we present a study that was conducted on online HW of PD patients. In fact, PD is a progressive nervous system disorder that affects movement. This disease results in a slow and progressive destruction of dopaminergic neurons [16, 17]. The motor symptoms of this pathology are tremors, muscle rigidity, impossibility or slowing of movements, as well as the intellectual impairment symptoms such as deterioration of memory, hesitation, lack of attention [18]. The average age of onset of the disease is between 55 and 65 years, but there are early forms that begin around age 40, and late forms that begin after age 75.

Following this context, our research team of ENS Fez, in collaboration with the neurology department of the UHC, launched a project called ENEMAR (Etude Neurologique de l'Écriture des MARocains). The project is approved by the ethical agreement of the committee HOSPITALO-UNIVERSITAIRE FES (CEHUF) filed under number "N° 03/15; July 10, 2015; Fez, Morocco".

The aim of this collaboration is to develop a diagnosis aid system of neurodegenerative pathologies based on the automatic analysis of online HW for a Moroccan bilingual population.

2 Methods

2.1 Handwriting acquisition protocol

The HW acquisition is done in the neurology department at the UHC Fez. First, the participant sign freely a consent form before the execution of handwriting's tasks. For participants' recruitment, an inclusion and non-inclusion criteria were defined, among these criteria:

Inclusion criteria:

- Signing a consent form agreeing to participate freely in the study.
- The patient must be followed by neurological consultation at UHC HASSAN II Fez.
- A cognitive assessment is passed for all the participants, and are classed by:
- Class 1: Healthy controls with normal cognitive functioning.
- Class 2: PD patients, with Mini Mental State Examination between 20 and 25; UPDRS motor scale; and complete neuropsychological test.
- Class 3: Alzheimer's patients (MMSE between 20 and 25).
- Class 4: Mild Cognitive Impairment patients.

Exclusion criteria:

- Participants suffering from visual or hearing problems.
- Each participant suffering from stroke or head trauma.
- Subjects under trusteeship or guardianship.

Participants are encouraged to reproduce a handwritten trace following the protocol in Arabic and French as well [19, 20]. Data is acquired confidentially and anonymously. The Volunteers have indeed, the right to withdraw from the study at any moment.

2.2 Materials

A graphical tablet *WACOM Intuos Pro Large* with Inking Pen is used for collecting the HW data. This tablet is capable to detect 2048 of pressure levels, with a 125 Hz of sampling frequency. It records over time the coordinates ($x(t)$, $y(t)$), pressure, as well as pen's inclination. The tablet does not allow a visual feedback of the HW but allows to capture the signals using a sheet of paper placed on its surface. This tablet detect the pen's coordinates when it is in the air (up to 1.5 cm above the tablet).

2.3 Used database

Matched according to level of education and age, altogether, 68 subjects (34 PD patients and 34 healthy controls (HCs)) participated in this work. The volunteers are all right-handed, completed 6 years of education, and native speakers of Arabic. All PD patients were on medication and took the treatment 30 to 60 min before the HW task. The metadata such as age, HW frequency per week, treatment, profession, etc. as well as the HW data are collected anonymously. In Table 1, we present the demographic data of the studied populations.

2.4 Handwriting task

The analysis presented in this article concerns the first Arabic exercise of the protocol [19, 20]. The task consist of copying a provided Arabic text consisting of six lines. An example of the imposed text is shown in Fig. 1. The blue color represents the low speeds and the red color represents the high speeds in ranges from 0 to 20 cm/s.

Table 1 Demographic data

	Mean age (std)	MMSE score	Mean UPDRS	Hoehn and Yahr stages
PD	55.10 (± 9.18)	28.5 (± 1.2)	12 (± 7.9)	2.16 (± 0.3)
HCs	53.10 (± 10.3)	30	–	–

Fig. 1 Visualization of the HW velocity

3 Feature extraction

The Wacom tablet generates, for each point (n) of the pen trajectory, $X(n)$, $Y(n)$, $P(n)$. From these vectors we compute new features that can be arranged in three categories:

- Kinematics: Horizontal and vertical velocity, acceleration, jerk, normalized jerk, speed, changes in velocity number NCV, changes in acceleration number NCA, etc.
- Spatial: Words height, length of intra and inter word spaces, complexity of the trajectory pen (when the stylus is in the air).
- Mechanical: Pressure, pressure variation.

Thus, some of statistical functions are computed to extract information contained in HW features, such as: mean, standard deviation, maximum, minimum, quantiles Q1 and Q3, entropy, median, Kurtosis, and Skewness. Based on these three categories, we computed 230 statistical features for each participant.

4 Visualization and clustering results

4.1 PCA visualization

Principal Component Analysis (PCA) [21, 22] is one of the most commonly used multivariate data analysis methods. It makes it possible to explore multidimensional data sets made up of quantitative variables. PCA was used to visualize the disposition of writers based on the 230 features. The first two PCA axes retrieve a 98.73% of inertia.

The Fig. 2 represents the visualization PCA. Writers are colored according to their group. Red dots correspond to PD patients and the blue ones correspond to HCs.

4.2 K-means clustering

According to Fig. 2, the majority of the HCs are practically grouped in the same area, and the PD patients are scattered throughout the rest of the figure. Visually, we can distinguish two or three classes. For more precision, we applied the K-means clustering [23] method on the first two principal components corresponding to the first two axes. Using Silhouette criterion, we generated an optimal number of three clusters, which are determined very clear in the Fig. 3.

In this figure, we have 3 clusters, the dots in red are mainly HCs, the points in blue and green are both consisting only of PD patients.

4.3 Characterization of obtained clusters

The aim of this step is to find the HW parameters characterizing each obtained cluster. Thus, a one-way analysis of variance [24] is applied between the HW quantitative variable which is the response, and the class variable which has the role of the explanatory variable.

- Cluster 1: The writers of this cluster are mostly HCs, they have a very high velocity, acceleration and jerk. They have also an average pressure. These writers have a medium to high educational level with a high HW frequency.

Fig. 2 PCA projection of the computed features. Red color: PD patients. Blue color: HCs

Fig. 3 PCA projection of the three classes generated by K-means over the first two principal components

- Cluster 2: Consisting of only PD patients, this cluster is characterized by an average velocity, acceleration and jerk. The writers of this cluster have also an average pressure. All these writers have a high educational level with a high HW frequency.
- Cluster 3: This cluster is characterized by the lowest kinematic features, as well as a very low pressure. These writers spend a lot of time in air, they also have an advanced stage of PD compared to PD patients in cluster 2, according to their UPDRS score.

5 Discussion

Parkinson's Disease is the most widespread neurological pathology after Alzheimer. It's mainly characterized by a gradual and persistent loss of dopaminergic neurons, which affects some specific areas in the brain. This brain's deterioration causes an important decline of cognitive, functional and behavioral abilities of the affected subject. Thus, handwriting, which is a combination of motor and cognitive abilities, is also affected [25]. According to the obtained results, PD patients undergo alterations

mostly in kinematic parameters. On the contrary of healthy controls, PD patients are characterized by a significant slowness in velocity, acceleration and jerk. We are aware that our research may have some limitations such as the size of the used database. For this purpose, the data acquisition is still carried out within the neurological department in order to enlarge our database.

6 Conclusion

Following this paper, we propose an unsupervised learning technique for characterizing the on-line handwriting of 34 Parkinson's Disease patients and 34 Healthy Controls according to quantitative and qualitative features. Based on 230 computed features, our study has uncovered three different types of writers. The complications of handwriting skills in Parkinson's Disease patients are reflected by a considerable deterioration in kinematic aspects.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Onofri E, Mercuri M, Salesi M, Ferrara S, Troili GM, Simeone C, Ricciardi MR, Ricci S, Archer T (2013) Dysgraphia in relation to cognitive performance in patients with Alzheimer's disease. *J Intellect Disabil Diagn Treat* 1(2):113–124
- Garre-Olmo J, Faúndez-Zanuy M, López-de-Ipiña K, Calvó-Perxas L, Turró-Garriga O (2017) Kinematic and pressure features of handwriting and drawing: preliminary results between patients with mild cognitive impairment, Alzheimer disease and healthy controls. *Curr Alzheimer Res* 14(9):960–968
- Schröter A, Mergl R, Bürger K, Hampel H, Möller HJ, Hegerl U (2003) Kinematic analysis of handwriting movements in patients with Alzheimer's disease, mild cognitive impairment, depression and healthy subjects. *Dement Geriatr Cognit Disord* 15(3):132–142
- Werner P, Rosenblum S, Bar-On G, Heinik J, Korczyn A (2006) Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment. *J Gerontol Ser B Psychol Sci Soc Sci* 61(4):P228–P236
- Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M (2014) Decision support framework for Parkinson's disease based on novel handwriting markers. *IEEE Trans Neural Syst Rehabil Eng* 23(3):508–516
- Drotár P, Mekyska J, Smékal Z, Rektorová I, Masarová L, Faundez-Zanuy M (2015) Contribution of different handwriting modalities to differential diagnosis of Parkinson's disease. In: 2015 IEEE international symposium on medical measurements and applications (MeMeA) proceedings, pp 344–348. IEEE
- Aouraghe I, Ammour A, Khaissidi G, Mrabti M, Aboulem G, Belahsen F (2019) Automatic analysis of Arabic online handwriting of patients with parkinson's disease: statistical study and classification. In: Proceedings of the new challenges in data sciences: acts of the second conference of the moroccan classification society, p 24. ACM
- Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M (2013) A new modality for quantitative evaluation of Parkinson's disease: in-air movement. In: 13th IEEE international conference on bioinformatics and bioengineering, pp 1–4. IEEE
- Smekal Z, Mekyska J, Rektorova I, Faundez-Zanuy M (2013) Analysis of neurological disorders based on digital processing of speech and handwritten text. In: International symposium on signals, circuits and systems ISSCS2013, pp 1–6. IEEE
- Rosenblum S, Samuel M, Zlotnik S, Erikh I, Schlesinger I (2013) Handwriting as an objective tool for Parkinson's disease diagnosis. *J Neurol* 260(9):2357–2361
- Drotar 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. *Comput Methods Progr Biomed* 117(3):405–411
- Akyol K (2017) A study on the diagnosis of Parkinson's disease using digitized wacom graphics tablet dataset. *Int J Inf Technol Comput Sci* 9:45–51
- Kawa J, Bednorz A, Stępień P, Derejczyk J, Bugdol M (2017) Spatial and dynamical handwriting analysis in mild cognitive impairment. *Comput Biol Med* 82:21–28
- Müller S, Preische O, Heymann P, Elbing U, Laske C (2017) Diagnostic value of a tablet-based drawing task for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. *J Alzheimer's Dis* 55(4):1463–1469
- Müller S, Preische O, Heymann P, Elbing U, Laske C (2017) Increased diagnostic accuracy of digital versus conventional clock drawing test for discrimination of patients in the early course of Alzheimer's disease from cognitively healthy individuals. *Front Aging Neurosci* 9:101
- Man JH, Groenink L, Caiazzo M (2018) Cell reprogramming approaches in gene-and cell-based therapies for Parkinson's disease. *J Control Release* 286:114–124
- Lozano AM, Dostrovsky J, Chen R, Ashby P (2002) Deep brain stimulation for Parkinson's disease: disrupting the disruption. *Lancet Neurol* 1(4):225–231
- Kalia SK, Sankar T, Lozano AM (2013) Deep brain stimulation for Parkinson's disease and other movement disorders. *Curr Opin Neurol* 26(4):374–380
- Ibtissame A, Alae A, Ghita A, Ghizlane K, Mostafa M, Faouzi B, El-Yacoubi MA, Garcia-Salicetti S (2017) Analyse automatique de l'écriture manuscrite en ligne pour la détection précoce des pathologies neurodégénératives
- Ammour A, Aouraghe I, Khaissidi G, Mrabti M, Aboulem G, Belahsen F (2020) A new semi-supervised approach for characterizing the Arabic on-line handwriting of Parkinson's disease patients. *Comput Methods Progr Biomed* 183:104979
- Jolliffe I (2011) Principal component analysis. *International Encyclopedia of Statistical Science*. Springer, Berlin, pp 22–32
- Campbell MC, Markham J, Flores H, Hartlein JM, Goate AM, Cairns NJ, Videen TO, Perlmuter JS (2013) Principal component analysis of PiB distribution in Parkinson and Alzheimer diseases. *Neurology* 81(6):520–527
- Hartigan JA, Wong MA (1979) Algorithm AS 136: a k-means clustering algorithm. *J R Stat Soc Ser C Appl Stat* 28(1):100–108
- Cuevas A, Febrero M, Fraiman R (2004) An ANOVA test for functional data. *Comput Stat Data Anal* 47:111–122
- McLennan JE et al (1972) Micrographia in Parkinson's disease. *J Neurol Sci* 15(2):141–152

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.