Pattern Recognition 86 (2019) 112–133

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

From aging to early-stage Alzheimer's: Uncovering handwriting multimodal behaviors by semi-supervised learning and sequential representation learning

Mounîm A. El-Yacoubi[a,*], Sonia Garcia-Salicetti[a], Christian Kahindo[a], Anne-Sophie Rigaud[b,c], Victoria Cristancho-Lacroix[b,c]

[a] SAMOVAR, Telecom SudParis, CNRS, Université Paris Saclay, France
[b] AP-HP, Groupe Hospitalier Cochin Paris Centre, Hôpital Broca, Pôle Gérontologie, Paris, France
[c] Université Paris Descartes, EA 4468 Paris, France

ARTICLE INFO

Article history:
Received 30 November 2017
Revised 3 July 2018
Accepted 31 July 2018
Available online 16 August 2018

Keywords:
Online handwriting
Alzheimer's
Mild Cognitive Impairment
Aging
Unsupervised & semi-supervised learning
Temporal representation learning

ABSTRACT

We present, in this paper, a novel paradigm for assessing *Alzheimer*'s disease and aging by analyzing impairment of handwriting (*HW*) on tablets, a challenging problem that is still in its infancy. The state of the art is dominated by methods that assume a *unique* behavioral trend for each cognitive profile or age group, and that extract *global* kinematic parameters, assessed by standard statistical tests or classification models, for discriminating the neuropathological disorders (*Alzheimer*'s (*AD*), *Mild Cognitive Impairment* (*MCI*)) from *Healthy Controls* (*HC*), or *HC* age groups from each other. Our work tackles these two major limitations as follows. First, instead of considering a unique behavioral pattern for each cognitive profile or age group, we relax this heavy constraint by allowing the emergence of multimodal behavioral patterns. We achieve this by performing semi or unsupervised learning to uncover homogeneous clusters of subjects, and then we analyze how much information these clusters carry on the cognitive profiles (or age groups). Second, instead of relying on global kinematic parameters, mostly consisting of their average, we refine the encoding either by a semi-global parameterization, or by modeling the full dynamics of each parameter, harnessing thereby the rich temporal information inherently characterizing online *HW*. To illustrate the power of our paradigm, we present three studies, one regarding age, and two regarding *Alzheimer*'s. Thanks to our modeling, we obtain new findings that are the first of their kind on this research field. On aging, unlike previous works reporting only one pattern of *HW* change with age, our study, based on a semiglobal parametrization scheme, uncovers three major aging *HW* styles, one specific to aged subjects and two shared with other age groups. On *Alzheimer*'s, a striking finding is revealed: two major clusters are unveiled, one dominated by *HC* and *MCI* subjects, and one by *MCI* and *ES-AD*, thus revealing that *MCI* patients have fine motor skills leaning towards either *HC*'s or *ES-AD*'s. Our paper introduces also a new *temporal representation learning* from *HW* trajectories that uncovers a rich set of features simultaneously like the full velocity profile, size and slant, fluidity, and shakiness, and reveals, in a naturally explainable way, how these *HW* features conjointly characterize, with fine and subtle details, the cognitive profiles.

1. Introduction

1.1. Context and motivation

Alzheimer's disease (*AD*), the most common cause of major neurocognitive disorder (or dementia), is a progressive neurodegenerative disease, characterized by cognitive dysfunction, particularly memory impairment and other cognitive skills, that affect a person's ability to perform everyday activities [4,10]. Given the insidious progression of *AD*, the first troubles are often misinterpreted as due to normal aging. *AD*'s diagnosis criteria are mainly based on clinical markers (i.e. a significant cognitive decline affecting individual's independency), and biological markers. Due to the insidious and slow progress of the disease, research attention has recently focused on *Mild Cognitive Impairment* (*MCI*), a health stage associated with lower performance in one or more cognitive domains, that does not affect, however, a person's independence in carrying out functional activities. About 15–20% of people over 65

* Corresponding author.
E-mail addresses: mounim.el_yacoubi@telecom-sudparis.eu (M.A. El-Yacoubi), sonia.garcia@telecom-sudparis.eu (S. Garcia-Salicetti), christian.kahindo@telecom-sudparis.eu (C. Kahindo), anne-sophie.rigaud@aphp.fr (A.-S. Rigaud), victoria.cristancho-lacroix@aphp.fr (V. Cristancho-Lacroix).

https://doi.org/10.1016/j.patcog.2018.07.029
0031-3203/© 2018 Published by Elsevier Ltd.

have *MCI*, among which those with memory-related *MCI* are more likely to develop *AD* [1,74,46,60].

As persons with *AD* are significantly impacted by episodic memory impairment, loads of studies have been dedicated to language disorders involving spelling, grammatical, syntactic or semantic errors, *etc.* [5,42,53,58,68,73]. A recent review shows, however, that *AD* can be predicted by noncognitive symptoms, in particular by motor impairment occurring during the preclinical phase and before clinical diagnosis [10]. Several studies have assessed gait impairment, mild parkinsonian signs, fatigue and frailty [8,9,33,41], and some works have investigated fine motor impairment, especially *handwriting* (*HW*) changes due to *AD* [16,27,28,32,36,66,71,82,84,85]. Indeed, AD induces cognitive and visuospatial impairment that makes the physical act of writing difficult, which may trigger *HW* impairment [24]. The aim of our study is to characterize handwriting (*HW*), acquired online from tablets, in subjects belonging to three cognitive profiles: early-stage *Alzheimer Disease* (*ES-AD*), *Mild Cognitive Impairment* (*MCI*) and *Healthy Controls* (*HC*), i.e. subjects with a neurotypical cognitive profile, and also to investigate how *HW* changes with aging.

1.2. State of the art

HW recognition [57] is a mature technology with several highly successful commercial applications, whether offline for postal mail sorting and bank check processing [21,22,49], or online for recognizing personal notes on smartphones, tablets and PDA devices [54,48]. *HW* analysis for health assessment [17,56] has been far less studied owing to the difficulty and the cost of acquiring data from patients and the limited datasets obtained as a result, the difficulty of obtaining reliable annotations, and, above all, the unclear understanding of whether *HW* changes may be symptomatic of cognitive decline and the onset of a neurodegenerative disease.

Several studies have investigated the link between *HW* changes and pathologies like *Parkinson* [40,47,76,71,81,77], *Huntington* [52], *Schizophrenia* [11], *Sclerosis* [65], or other health conditions such as *Depression* [66] or *Emotions* [39]. Other works tried to shed light on the link between *HW* deterioration and aging [13,23,35,45,62,70,81]. Such a link is not only fundamental for understanding how fine motor skills evolve with age, but it may be key for distinguishing natural from pathological *HW* changes.

Research on *Alzheimer*'s assessment by *HW* analysis is still in its infancy. The state of the art is dominated by methods that extract *global* kinematic parameters, e.g. their averages, and then consider one of the two following schemes: 1) apply standard tests (e.g. *Anova*) to assess the statistical significance of each parameter for discriminating *AD* and *MCI* from each other, and w.r.t healthy controls (*HC*), or, albeit much less frequently, 2) apply classification techniques to identify a user's cognitive profile. The studies in the first scheme support the tendency of lower velocity, fluidity, and pressure, as well as larger movement duration and number of strokes, observed as the health profile declines from *HC* to *MCI* and later on to *AD* [82,84,66,71,36]. In the second scheme, the approaches proposed recently [82,36,28] essentially gather the global parameters above and provide them as input to simple classifiers. Although they report promising classification rates on some *HW* tasks, these studies are prone to overfitting, as combining *HW* parameters may lead to a curse of dimensionality given the limited training data. Interestingly, the studies on age-related *HW* changes are similar to those in scheme 1. Based on the same global kinematic parameters, they show the general trend that, as age increases, velocity, fluidity and pressure decrease, while in-air time and pen lifts increase [35,45,62,81,23].

Overall, although statistical tests and classification schemes obtain some promising results, showing the potential of *HW* in discriminating *AD, MCI* and *HC* (or age groups) from each other, they suffer from serious limitations. First, they consider, in most cases, only the average values of the kinematic parameters, thus overlooking *HW* dynamics and its potential in detecting subtle changes about the health condition. Such an averaging actually corresponds to a handcrafted feature extraction that converts raw *HW* input into manually-designed features, based on human *a priori* knowledge. This is a clear shortcoming as it implicitly assumes that the handcrafted features are the best way to discriminate the cognitive profiles (or age groups) from each other. Second, these studies assume that each cognitive profile (or age group) is associated with one *HW* pattern that distinguishes it from the others. Such an assumption is limiting and restrictive as it discards, from the outset, the diversity of *HW* patterns that may characterize a single health condition (for instance, all *HC* subjects may not have a fast *HW*, while all *AD* subjects may not write slowly).

1.3. Proposed work

This paper presents a novel paradigm of studying *HW* changes with aging or different cognitive profiles, that addresses the limits above. First, instead of assuming a unique (unimodal) behavior for each cognitive profile or age group, we relax this heavy constraint by allowing, for each, the emergence of a multimodal behavioral pattern. We achieve this by a semi or unsupervised learning to uncover homogeneous groups of subjects, and then we analyze the information these clusters carry on the cognitive profiles (or age groups). Second, instead of relying on average kinematic parameters, we refine the encoding either by a semi-global parameterization, or by modeling the full dynamics of each parameter, harnessing thereby the rich temporal information inherently characterizing online *HW*. The power of our paradigm is illustrated by three studies, one on age, and two on *Alzheimer*'s.

The first study aims to infer different writing styles and their correlation with age, with an emphasis on people over 65 years. Based on a set of words produced by each writer, it first considers a semi-global feature parameterization by encoding the distribution of each spatiotemporal parameter over a fixed number of bins characterizing coarsely its dynamics. Since writing styles are unknown *a priori*, we resort to unsupervised learning to uncover them in an automatic way. In this respect, we propose a novel style categorization model, carried out at two levels, *word*-level (low level) and *writer*-level (high level). At the first, a clustering of words based on their spatiotemporal representation detects the major *writer-independent word styles* (*clusters*). At the second, each writer's set of words is converted into a Bag of Prototypes (*BoP*), associated with the writing styles detected in the 1st stage. This *BoP* is augmented by a descriptor of the writer's variability across words to generate the input to a second clustering algorithm that infers *writer styles* or *categories*. The analysis of the age group distribution over each cluster identifies then the major writing styles that characterize aging. Unlike previous works reporting only one pattern of *HW* change with age, our study unveils three major aging *HW* styles, one specific to aged people and two shared with other age groups.

Our second study seeks to characterize *HW* alterations associated with *ES-AD* and *MCI* w.r.t *HC*. Based on a semi-global feature encoding in a text copying task, it seeks to uncover homogeneous subject groups (clusters), and then analyzes the extent to which these groups are correlated with the cognitive profiles. To enhance the clusters' quality, a semi-supervised learning is proposed where a Normalized Mutual Information feature selection scheme guides a hierarchical clustering algorithm to find the best trade-off between the number of clusters and the discriminative power of each w.r.t the three cognitive profiles. Thanks to this method, a striking finding is revealed: two major clusters are uncovered, one dominated by *HC* and *MCI* subjects, and one dominated by *MCI* and

ES-AD, thus revealing that *MCI* patients have fine motor skills either close to *HC*'s or to *ES-AD*'s.

In the third study, we take a leap further by modeling the full dynamics of *HW* basic units. The key our approach builds on is to harness the online *HW* time ordering to automatically learn, for each raw kinematic parameter, feature representations [6] instead of considering handcrafted global or semi-global features, assumed implicitly to be discriminant. On a task of writing cursive *l* loops, we propose a *temporal* clustering of the loops considered as time series, by a *K-medoids* algorithm taking as similarity measure *DTW* (*Dynamic Time Warping*) that accommodates the sequential aspect of the data. Our scheme allows a *representation learning* from sequences, which is barely addressed in the state of the art [6]. Applied to loop's *velocity* time series, our scheme uncovers a rich set of features simultaneously as a byproduct of the unsupervised learning itself. Indeed, the latter extracts (learns) several loop *medoids* (clusters), each consisting of a different combination of features like the full velocity profile, loop size and slant, fluidity, *etc*. We show that this representation learning can be exploited in several ways. First, by considering a second stage clustering based on the distribution of each user's input over the loop medoids or prototypes (first stage clusters), we uncover new homogeneous groups and study their link with the cognitive profiles. Second, to show the intrinsic information carried by the 1st stage, we consider, in a binary *HC* vs. *ES-AD* classification task, a Bayesian formalism that aggregates probabilistically the contributions of the loop prototypes by leveraging the discriminative power of each. Third, this temporal representation learning offers the advantage of being explainable. It does not only automatically extract new *HW* features for characterizing *ES-AD*, that can be visualized and easily understood, but it also detects clusters and obtains classification results that are naturally explainable to the medical staff and to the layman in general. This is important as a neurologist, for instance, rather than being convinced by mere classification rates, is keen in understanding how the automatic system generates its decision based on the subject's data.

The rest of the paper is as follows. In Section 2, we present our approach on characterizing age-related *HW* changes. Section 3 introduces our work on *HW* changes for subjects with *ES-AD* and *MCI* w.r.t *HC*, composed of two approaches, based on *HW* semi-global parametrization and full dynamics modeling respectively. The details of these two approaches are given in Sections 4 and 5. Section 6 concludes the paper and sketches some future directions of our work.

2. Uncovering writing style changes with aging

Age characterization from HW is fundamental as it may allow distinguishing normal HW change due to aging from abnormal one, potentially related to a pathological cognitive decline. In this section, we address the problem of age characterization from online HW. The goal is to detect HW styles and study their correlation with age, by the analysis of spatiotemporal HW parameters.

2.1. State of the art on aging assessment by HW analysis

Several works have studied *HW* changes as people age. Some were carried out through visual inspection [81,35,72,78]. Automatic studies concerned mostly online *HW* [23,62,55,70,12], although few have addressed the offline case [2,1,7]. Because they rely on a rich set of *temporal* features and not solely on static ones, the former have a much larger potential for uncovering parameters that change with age. This potential is reflected in the state of the art where, based on standard statistical tests or linear regression, several spatiotemporal parameters have been shown to change with aging, such as increasing in-air time and number of pen lifts [62], lower writing velocity, pressure and smoothness [23,35,45].

Although they do show the link between aging and *HW* changes, these studies suffer from two limitations. First, they consider *global* kinematic parameters from the whole writer's text, thus assuming that they are sufficient to assess different writings. In doing so, they overlook another useful information: does a person write different words in a similar way, or does s/he show different spatiotemporal trends from one word to another? this question has not been addressed before. Second, state of the art methods assume that *HW* evolves with age according to a unique pattern. This rules out the possibility of different evolution patterns, or that for some elders, *HW* may not change in any significant way. Different *HW* aging patterns, however, is a sound assumption as they could reflect different biological aging patterns within chronological aging [38].

2.2. Proposed approach for age assessment based on online HW analysis

To tackle these limitations, we have proposed an approach [44] that relaxes the two assumptions above. First, instead of considering average *HW* parameters, we propose a semi-global parameterization scheme that encodes the distribution of each spatiotemporal parameter over a fixed number of bins, characterizing coarsely its dynamics. Second, we propose a two-stage clustering scheme that models the writing style in terms of both the spatiotemporal style and its maintenance/variability across different words, and that makes possible the emergence of several trends within a same age group. Next, we describe the feature extraction phase, the two-stage clustering scheme, the experiments and the results obtained.

2.2.1. Word-based feature extraction

Online *HW* words are comprehensively represented by three temporal functions ($x(t)$, $y(t)$, $p(t)$) encoding the pen trajectory and pressure [26]. For each word, we extract two feature types, dynamic and static. For the first, we extract, for each point n, the horizontal and vertical velocities, $Vx(n)=|\Delta x(n)/\Delta t(n)|$ and $Vy(n)=|\Delta y(n)/\Delta t(n)|$, where $\Delta x(n)=x(n+1)-x(n\text{-}1)$, $\Delta y(n)=y(n+1)-y(n\text{-}1)$ and $\Delta t(n)=t(n+1)-t(n\text{-}1)$. $Vx(n)$ and $Vy(n)$ are then quantized each into a four-bin histogram to encode coarsely their dynamics. We similarly compute local acceleration and jerk (derivative of acceleration). By adding pen pressure, its variation, and in-air time duration ratio (in-air duration/total duration) [62], we obtain 33 dynamic features. For spatial features, we first remove the velocity effect by resampling *HW* trajectories to make constant the distance between consecutive points. Local direction and curvature angles are then extracted and quantized each into a histogram of eight bins in the $0^\circ-180^\circ$ range. We also consider the number of pen-ups, the average horizontal in-air length, the number of strokes (segments between two local minima of velocity) and their average length, as well as the average length of the stroke projection on the X axis, and on the Y axis. This results in 21 spatial features, which combined to dynamic features, yield a feature vector of dimension 54.

2.2.2. Two-stage semi-supervised learning

As *HW* styles are unknown *a priori*, they are usually inferred by unsupervised learning techniques [80,64,14], that cluster *HW* input into groups, identified as styles. These styles are often inferred at the character, stroke and word levels [14,18]. We believe, however, that writer style inference should rely not only on this raw signal information but also on high-level information associated with the writer's variability across words. This motivated us to propose a two-stage unsupervised approach: the 1st stage

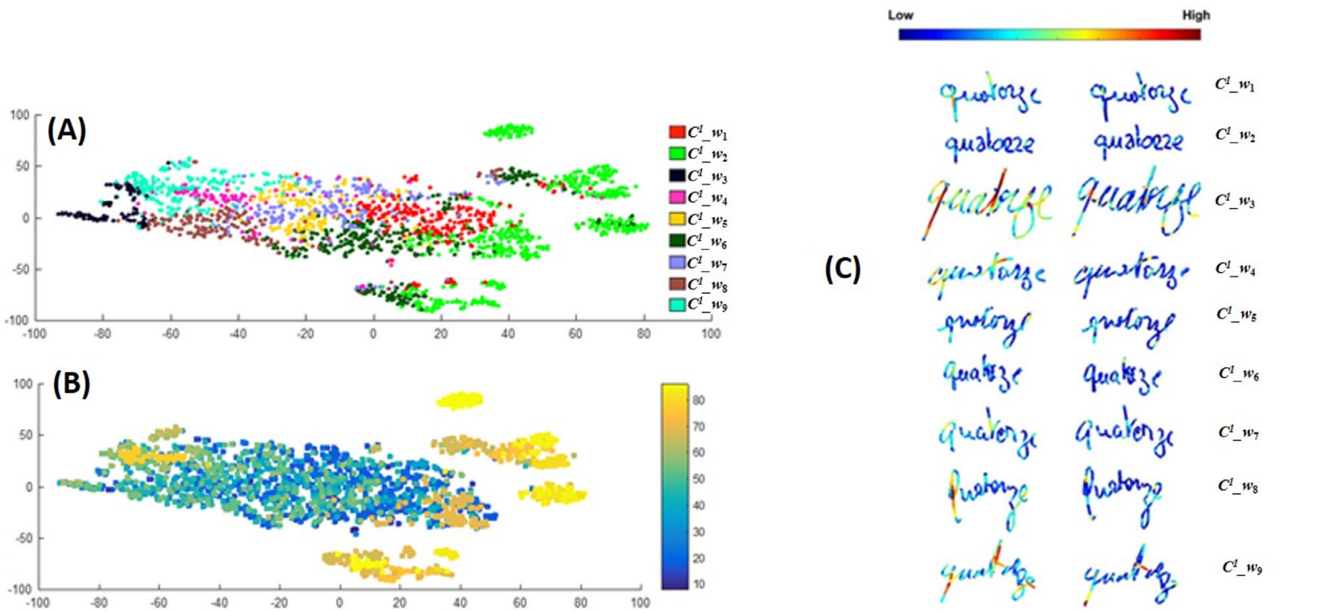

Fig. 1. SNE projections of 1st stage clusters with color encoding (A) cluster labels, and (B) age distribution; (C) *HW* samples in each cluster; the color scale here quantifies the magnitude of Velocity (left column) and Jerk (right).

takes as input the low-level spatiotemporal word representation (encoded by 54 features), and performs a clustering of the set of words regardless of writer identity, generating clusters corresponding to *text-independent* and *writer-independent word* styles. At the 2nd stage, features are computed at the *writer* level. Each writer's set of words is converted into a Bag of Prototype Words (BPW) [69] by assigning each word to its closest 1st stage cluster. This generates a histogram of the writer's word distribution over the 1st stage clusters, that is augmented by the writer pairwise word distance distribution, quantized over five bins. This two-level representation is input to a second clustering, to uncover *writer*-style categories by modeling both the spatiotemporal word style, and its variability across the writer's words. The detected clusters are then analyzed in terms of their correlation with age.

Our two-stage scheme can be seen as a clustering-based deep hierarchical feature representation scheme [15], in which the 1st stage learns *word* writing styles inferred from spatiotemporal information, and the 2nd stage detects the actual *writer* style by learning its words' variability across the 1st stage word styles. Different from [15], nonetheless, our hierarchical learning is performed over two entities, a word in the 1st stage and a writer's set of words in the 2nd stage.

We present the results using *K-means* clustering on both stages (similar results are obtained with other algorithms such as *GMM* or *Hierarchical* clustering), where the number of clusters is automatically determined by the Silhouette criterion. Hereafter, the 1st and 2nd stage clusters will be referred to as *clusters* ($C^1\_w_k$,) and *categories* ($C^2\_A_j$), respectively. $C^1\_w_k$ designates the *k*th cluster of words, obtained at the 1st stage, while $C^2\_A_j$ refers to the age-related 2nd stage *j*th cluster of subjects.

2.3. Experiments

2.3.1. Dataset

For evaluation, we use the Ironoff dataset [79] of online *HW* word samples, acquired by a Wacom tablet at a sampling rate of 100 Hz and a resolution of 300 dpi. Although this set comprises 880 writers, only few are over 60 years. For a more reliable study, we collected, at Broca Hospital in Paris, *HW* samples from 25 elders with no diagnosed pathology, with an average age of 72. The data were acquired on a Wacom Tablet at the same sampling rate but with a higher resolution (5080 dpi), that we decreased to match the 300 dpi of Ironoff. Combining both sets, we obtain 27,683 *HW* words from 905 writers aged from 11 to 86 years old (*y.o.*), among which 772 are between 18 and 50. For the 1st stage unsupervised learning, we use the whole set since the clustering is word-based. For the 2nd stage, we consider the following six Age Groups (*AG*), in a similar way to the state of the art [23]: AG_{11-17} (11–17 *y.o.*), AG_{18-35} (18–35 *y.o.*), AG_{36-50} (36–50 *y.o.*), AG_{51-65} (51–65 *y.o.*), AG_{66-75} (66–75 *y.o.*), and AG_{76-86} (76–86 *y.o.*). To properly evaluate the clustering and its correlation with age, we select, from the whole set, a balanced subset in terms of age groups by retaining 26 writers for each, thus generating a total of 156 writers.

2.3.2. First stage clustering for unsupervised characterization of age-related HW patterns

Based on the *Silhouette* method, the 1st stage uncovers an optimal number of nine clusters. To visualize the clustering quality, we use *Stochastic Neighbor Embedding* (*SNE*) [34], a nonlinear dimensionality reduction technique that optimally maps the points from a high dimensional space onto a lower space by preserving pairwise distances as much as possible. The left of Fig. 1 shows the words projected by *SNE* from the 54-feature dimensional space onto two dimensions; color in Fig. 1.(A) encodes age (from dark blue (youngest) to yellow (eldest)) while it encodes cluster labels in Fig. 1.(B). By Overlaying (A) over (B), remarkable findings are brought to light: we clearly observe a correlation between age and *HW*, as reflected by some groups of aged people emerging automatically from our 1st stage clustering. In particular, cluster $C^1\_w_2$ stands out as it is mostly associated with aged people. We also note that clusters $C^1\_w_3$, $C^1\_w_6$ and $C^1\_w_9$ are partially associated with aged writers.

Fig. 1.(C) shows word samples in each cluster, encoded by the velocity and jerk magnitudes. These words highlight the main *HW* patterns that emerge from *HW* data, summarized in Table 1.

If we focus on clusters $C^1\_w_2$, $C^1\_w_3$, $C^1\_w_6$ and $C^1\_w_9$, i.e. the ones significantly represented by the oldest age groups, AG_{66-75} (26 people) and AG_{76-86} (26 people), two main aging tendencies stand out: $C^1\_w_2$ and $C^1\_w_6$ represent small-size *HW*, with a vertical script style, low velocity and jerk, and medium pressure, while

Table 1
Main characteristics of 1st stage clusters ($C^1\_w_k$). V_x and V_y stand for horizontal and vertical velocity, and V for its magnitude. The same applies for Acceleration and Jerk.

	Dynamics	Slant	Pressure	Curvature	Pen-up frequency
$C^1\_w_1$	Low V, A, and J	Upright	Medium	Round Strokes	Medium
$C^1\_w_2$	Very low V, A, and J	Upright	Low	Round Strokes	High
$C^1\_w_3$	High V, A, and J	Right Slant	Medium	Straight Strokes	High
$C^1\_w_4$	Moderate V, A, and J	Right Slant	High	Straight Strokes	Medium
$C^1\_w_5$	Moderate V, A, and J	Upright	Medium	Medium	High
$C^1\_w_6$	Moderate V_y, low V_x	Upright	Medium	Medium	Medium
$C^1\_w_7$	Moderate V, A, and J	Upright	Medium	Round Strokes	Medium
$C^1\_w_8$	High V_y, moderate V_x	Upright	Medium	Upright Strokes	High
$C^1\_w_9$	Very high V, A, and J	Right Slant	Medium	Upright Strokes	Medium

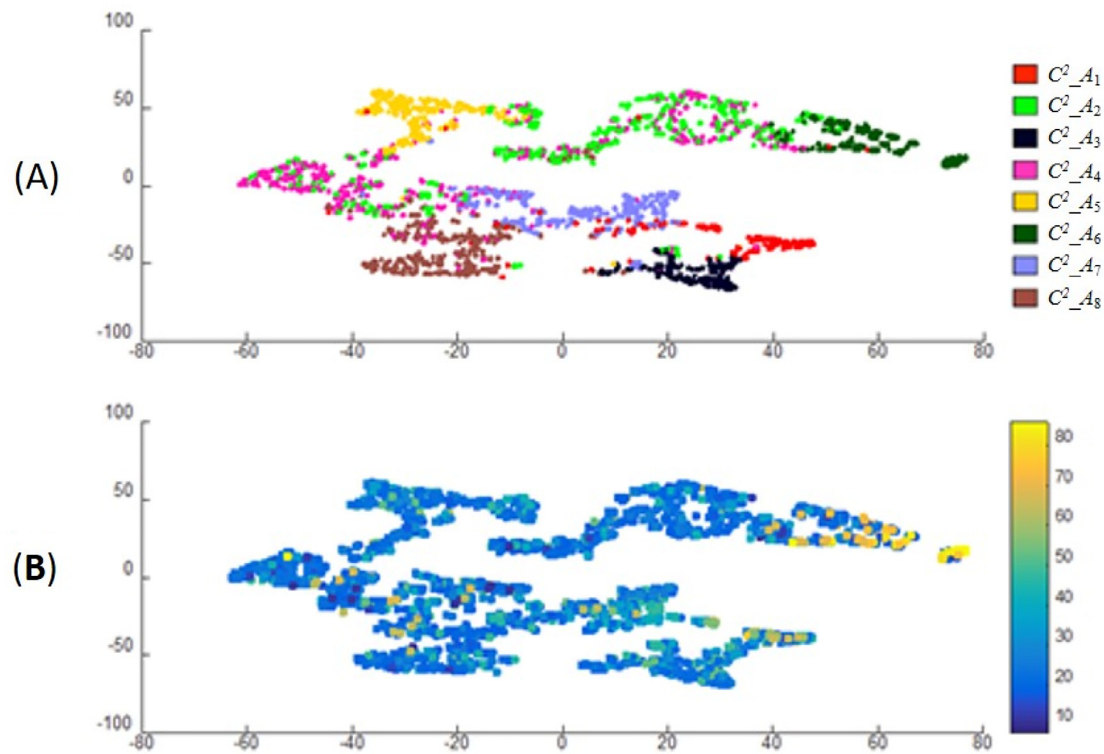

Fig. 2. SNE projections of the subjects from the 14-dimensional space onto two dimensions, labeled by color according to (A) 2nd stage categories ($C^2\_A_1$ to $C^2\_A_8$), and (B) Age, from youngest (blue) to oldest (yellow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$C^1\_w_3$ and $C^1\_w_9$ represent a right slanted cursive style, with very fast dynamics and medium to low pressure, $C^1\_w_3$ characterizing, in addition, large size *HW*.

2.3.3. Second stage clustering for unsupervised characterization of age-related HW patterns

At the 2nd stage, each writer is described by 14 features, nine encoding his/her word distribution over the 1st stage clusters, and five encoding the distribution of his/her word pairwise distances. Our clustering of writers detects eight clusters or *categories* based on the silhouette criterion. Fig. 2 shows the *SNE* projections of the eight categories on the set of writers, where each writer is encoded by age color in Fig. 2.A, and by label color in Fig. 2.B. Again, we observe a striking relationship between the categories and age groups, with particularly category $C^2\_A_6$ standing out, as it is comprised mostly of aged subjects.

To emphasize the link between *HW* changes and aging, we analyze for each category, the sizes of the oldest age groups, i.e. $AG_{66\text{-}75}$ and $AG_{76\text{-}86}$, w.r.t the other age groups. Table 2 reports the size and percentage of $AG_{66–75}$ and $AG_{76–86}$ within each (2nd stage) category, and Fig. 3 shows the age distribution for each category w.r.t to the initial balanced age distribution (1/6 for each group). For instance, age group $AG_{51–65}$'s percentage in $C^2\_A_1$ is two, as it is twice more represented in $C^2\_A_1$ than it was before clustering.

Fig. 3 reveals an important fact: four categories ($C^2\_A_2$, $C^2\_A_3$, $C^2\_A_5$ and $C^2\_A_7$) do not comprise any writer from $AG_{66–75}$ or $AG_{76–86}$, and only three categories ($C^2\_A_1$, $C^2\_A_4$, and $C^2\_A_6$) contain a significant number of aged writers. The distribution of the 1st stage clusters within each 2nd stage category is depicted in Fig. 4, while Fig. 5 reflects this distribution in a visual way, through representative word samples. The major findings follow below.

- $C^2\_A_6$ clearly stands out: it includes virtually only subjects over 65, as the $\{AG_{66–75}+AG_{76–86}\}$ set represents 84% of the subjects (Table 2). As shown in Figs. 4 and 5, $C^2\_A_6$'s subjects write words mostly captured by $C^1\_w_2$, characterized by lowest velocity, acceleration and jerk, as well as round *HW* with the highest number of strokes and smallest stroke length (Fig. 1.(C) and Table 1). As $C^2\_A_6$ contains the highest number of persons (44 writers among 156, i.e. 28%), and 71% of the $\{AG_{66–75}+AG_{76–86}\}$ subjects, this could reflect a major aging trend, characterized by slow and curved *HW*, with medium to high in-air time, probably induced by writing hesitations due

Table 2
Size and percentage of AG_{66-75} and AG_{76-86} within each (2nd stage) category.

	$C^2\_A_1$	$C^2\_A_2$	$C^2\_A_3$	$C^2\_A_4$	$C^2\_A_5$	$C^2\_A_6$	$C^2\_A_7$	$C^2\_A_8$
Size	18	16	10	29	10	44	16	13
$\boldsymbol{AG_{66-75}}$	**11%**	0%	0%	**21%**	0%	**39%**	0%	6%
$\boldsymbol{AG_{76-86}}$	**22%**	0%	0%	**7%**	0%	**45%**	0%	0%

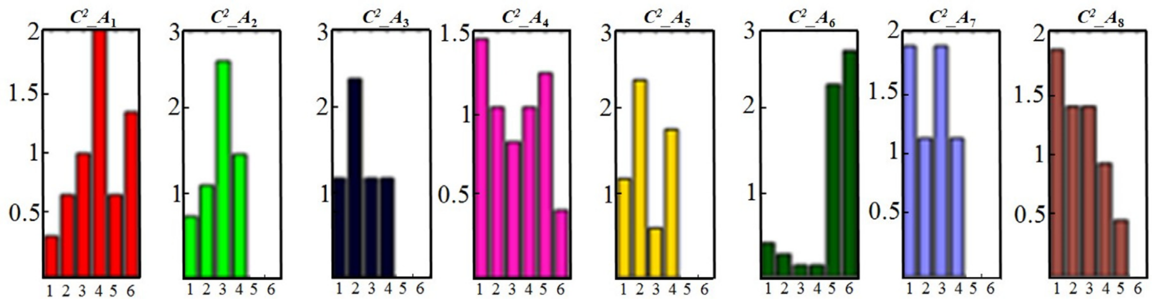

Fig. 3. Age group distribution in each category of the 2nd stage.

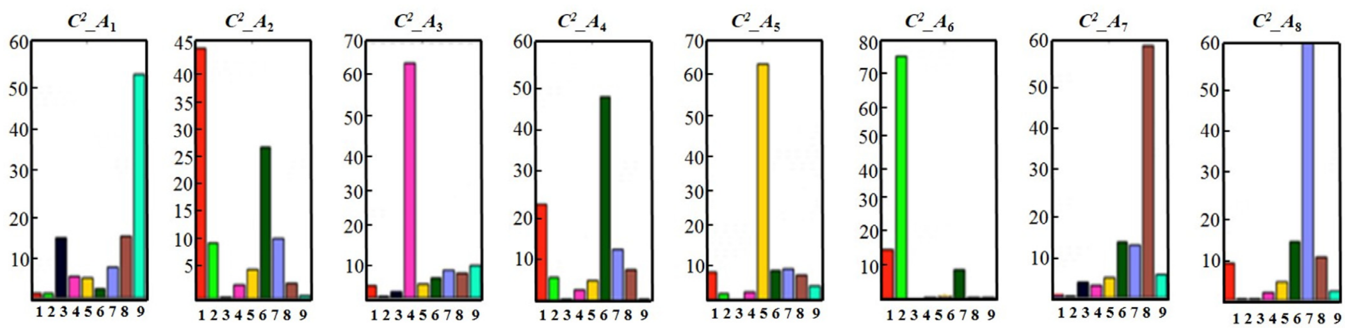

Fig. 4. Distribution of the 1st stage clusters within each 2nd stage category.

Fig. 5. *HW* Samples from each category of the 2nd stage, showing velocity on a color scale.

to mild cognitive decline. These tendencies are a hallmark of a slower and less fluid *HW*. $C^2\_A_6$ is also characterized by words written with small size, as shown by Fig. 5 visually, and by Fig. 4 that indicates a high value for $C^1\_w_2$ in the $C^2\_A_6$ category. $C^1\_w_2$, precisely, corresponds to small size (Fig. 1.(C) and Section 2.3.2).

- $C^2\_A_1$ represents 11.5% of the oldest age groups, AG_{66-75} and AG_{76-86}, and consists of a *HW* style closer to that of AG_{36-50}, in terms of dynamic features. The subjects in this group have the highest velocity, acceleration and jerk, which is the opposite behavior to $C^2\_A_6$'s. $C^2\_A_1$ is also characterized by words with large size as shown by Fig. 5 visually, and by Fig. 4 that indicates a high value for $C^1\_w_3$ in the $C^2\_A_1$ category. $C^1\_w_3$, precisely, corresponds to large size (Fig. 1.(C) and Section 2.3.2).
- $C^2\_A_4$ represents 15.4% of AG_{66-75} and AG_{76-86}, and is characterized by a *HW* with medium velocity, very low horizontal jerk, medium pressure, and low pressure variation.

In summary, unlike previous works reporting a unique HW pattern change with aging, our study unveils three major aging HW styles, one specific to aged people and characterized by slower and less fluid HW, and two, shared with the other age groups, characterized mostly by high dynamics and variability. In the future, it might be interesting to link our findings with the works seeking to study how chronological aging features different biological aging patterns, healthier and unhealthier [38].

3. Uncovering writing style alterations with *Alzheimer*'s and *MCI*

3.1. State of the art on Alzheimer's assessment by HW analysis

Over the last decades, loads of research studies have investigated the link between *HW* changes and pathologies like *Parkinson* [76], Huntington [52,71], Schizophrenia [11], Sclerosis [65], or other health conditions such as Depression [66] or Emotions [39]. In particular, *Parkinson* disease (*PD*) has intensively been studied through the analysis of fine movements, acquired on a digitizer. The target tasks for these studies required particular finger and wrist coordination, like the Archimedes spiral, concentric circles, and handwriting input [75,59]. In addition to *Micrographia* (small size writing or drawing), several spatiotemporal parameters as movement duration, velocity, and fluency were reported to be effective in discriminating *PD* patients from *HC* [77].

Although several studies have been proposed for *AD*'s assessment by online *HW* analysis since the late 1990s, this research field is still in its infancy. Characterizing *Alzheimer*'s at an early stage is a challenge, since the onset of the disease is insidious. As there is high heterogeneity of *Alzheimer*'s profiles, and as some *MCI* patients can convert into *Alzheimer*'s, characterizing *AD* requires studying the *MCI* class, and thus developing techniques for discriminating between three classes (*AD* vs. *MCI* vs. *HC*), which brings additional complexity w.r.t *Parkinson*'s (only two classes).

State of the art methods on *Alzheimer*'s assessment by *HW* analysis essentially extract *global* (*average*) kinematic parameters, and then consider one of the two following schemes: 1) apply standard tests (e.g. *Anova*) to assess the statistical significance of each parameter for discriminating a pathological population from a healthy control one, or, albeit less frequently, 2) apply classification techniques to identify a subject's cognitive profile based on a multidimensional description of his/her *HW*.

The studies in the first scheme depend on factors such as the *HW* task (copying a text, sentence, loop series, *etc.*), and the number of cognitive profiles under study (e.g. {*HC* vs. *MCI* vs. *AD*} or {*HC* vs. *AD*}), but they tend to assert overall that lower velocity, fluidity, and pressure, as well as larger movement duration and number of strokes, are observed as the health profile declines from *HC* to *MCI* and later on to *AD* [82,84,66,71,36]. These findings, however, are sometimes disconfirmed or even contradicted [82,85]. This may be explained by the strong implicit assumption in these studies that each cognitive profile features a unique behavioral pattern, which is not realistic, as our study on age in Section 2 has shown for *HC*. Indeed, a discriminant parameter in one study may turn out not discriminant in another if the fine motor skills of *MCI* subjects in the former are statistically more impaired. Such a discrepancy is likely given the small datasets usually considered. Worse, considering early-stage *AD* only as opposed to an *AD* all-inclusive study may heavily impact the results, as it is much easier to detect high significant *HW* impairments in subjects with advanced *AD*.

In the second scheme, the few approaches proposed recently [82,36,28] essentially gather the global kinematic parameters above and provide them as input to a *Linear Discriminant Analysis* (*LDA*) or a *logistic regression classifier* [63]. Although they report promising classification rates on some *HW* tasks, these studies suffer from overfitting as the number of *HW* parameters quickly leads to a curse of dimensionality, given the limited training data. Some reported results are misleading as they are obtained on the very data the classifiers are trained on [28,82].

3.2. Proposed work on Alzheimer's assessment by HW analysis

Assessing HW disorders associated with pathologies like Alzheimer's amounts to detecting pathological HW deteriorations w.r.t writing style changes due to normal aging. The main issue, in this regard, is that there is no agreed-upon definition of deteriorations or changes. Fig. 6 shows same HW samples from six people that underline this issue.

Whether one looks to static or velocity (encoded by color) information, it is hard to identify clues that discriminate the cognitive profiles from one another. Actually, the two HW samples on the left are associated with HC, the two in the middle, with MCI, and the two on the right, with ES-AD, and as the figure shows, mere global assessment of the statistical significance or the discriminative power of kinematic or even distortion-related parameters is doomed to failure in realistic settings. The figure shows that a subject might produce a writing that is slow or fast, large or small, upright or slanted, legible or less so, etc., regardless of the cognitive profile s/he belongs to. The average velocity or distortion-based features, therefore, are unlikely to discriminate the three classes.

To tackle the issues above and the limitations of the state of the art, we propose a novel paradigm for studying HW changes due to ES-AD and MCI w.r.t HC, inspired by our study on HW changes with aging. Instead of considering a unimodal behavioral pattern for each cognitive profile, we relax this restriction by allowing, for each, the emergence of a multimodal behavioral pattern. The key idea is to perform semi-supervised learning with the objective of uncovering clusters of subjects, and then to analyze how these clusters characterize the cognitive profiles. In addition, instead of relying on (global) average spatiotemporal parameters, we refine the encoding either by a semi-global parameterization, or by modeling the full dynamics of each parameter, harnessing thereby the rich temporal information inherently characterizing online HW. We present next the corpus and data acquisition, and then detail our studies with these two types of HW Dynamics' encoding in Section 4 and Section 5 respectively.

3.3. Corpus and data acquisition

Online HW data were acquired at Broca Hospital in Paris from three groups, Healthy Controls (HC), Mild Cognitive Impairment (MCI), and Early-Stage Alzheimer's (ES-AD). All ES-AD were diagnosed on the basis of DSM-5 criteria [3]. To be with Early-Stage AD, a patient was required to have a MMSE over 20, MMSE (Mini Mental State Examination) [3] being a clinical scale based on a questionnaire for assessing cognitive impairment, with a score up to 30 (no impairment). On their side, HC subjects underwent neuropsychological tests to ensure they have a normal cognitive profile. All the subjects from the three cognitive profiles had to be over 60, to read and talk French fluently, and to sign a consent form. Patients with visual impairment or any medical problem, such as stroke and other neurodegenerative diseases, were excluded. The corpus consists of 144 participants, 28 HC , 87 MCI, and 29 ES-AD, with a mean-age of 73.2 ($\pm$5.7), 78.5 ($\pm$7.6), and 79.9 ($\pm$6.4) respectively. HW was acquired on a WACOM Intuos Pro Large tablet with an inking pen. A paper was fixed on the tablet to allow a visual feedback as in natural conditions. The tablet records, with a sampling rate of 125 Hz, the pen's position (x(t), y(t)) and pressure p(t) over time, and the pen's in-air trajectory up to two cm off the table. The participants were asked to perform seven tasks involving copying texts, loop series, and drawings.

4. ***Alzheimer*'s and *MCI* assessment by semi-global parametrization of *HW***

Inspired by our study on age, we propose in this section to characterize *HW* alterations due to *ES-AD* and *MCI*, w.r.t *HC*, based on a semi-global feature encoding. The objective is to uncover homogeneous subject groups (clusters), and then to analyze how these groups are correlated with the cognitive profiles. To this end, we consider the task of copying, by each participant, of the following text in French, extracted from Antoine de Saint-Exupéry's *Le petit prince*: "Tu n'es encore pour moi qu'un petit garçon tout semblable à cent mille petits garçons. Je ne suis pour toi qu'un renard semblable à cent mille renards. Voici mon secret : on ne voit bien qu'avec le cœur. L'essentiel est invisible pour les yeux."

4.1. Text-based feature extraction

On each point n of the pen trajectory, we extract point-wise *kinematic* parameters such as horizontal and vertical velocity $\{V_x(n), V_y(n)\}$ and its first and second derivatives, i.e.

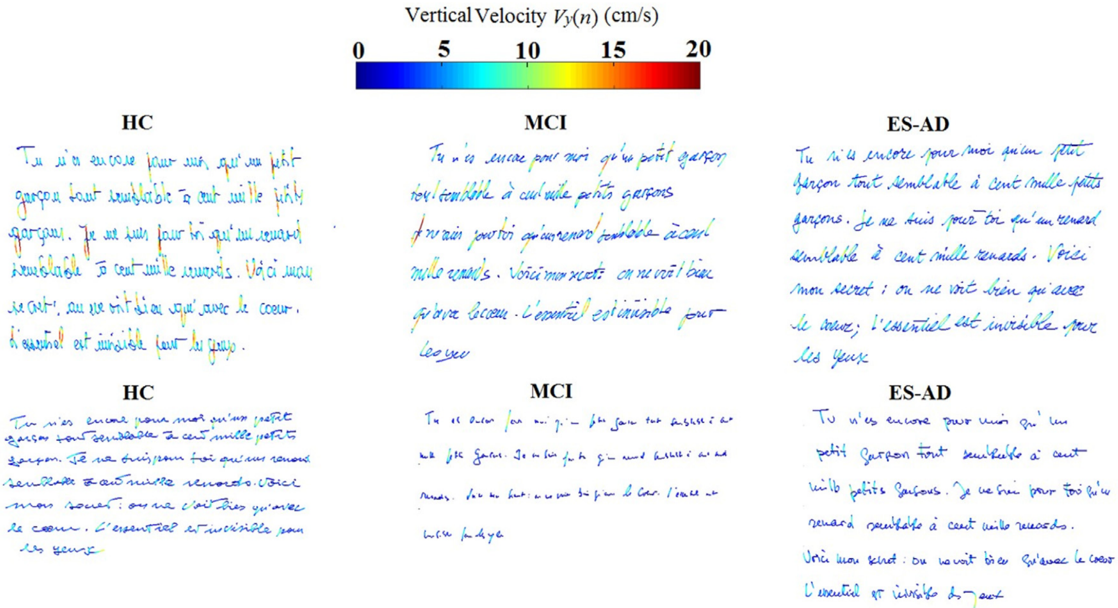

Fig. 6. *HW* samples encoded by velocity, two from *HC* (left), two from *MCI* (center), and two from *ES-AD* (right).

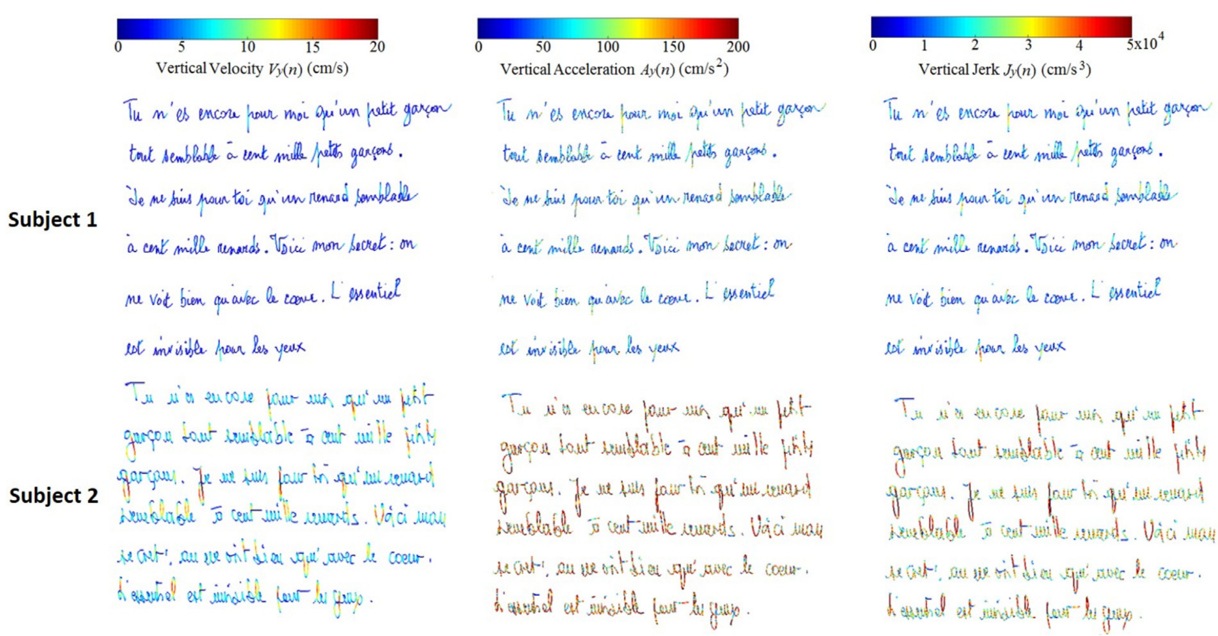

Fig. 7. The evolution of $V_y(n)$, $A_y(n)$ and $J_y(n)$ along the *HW* text, for two subjects. The values of the three parameters for subject 2 are much higher, reflecting a much faster writing.

acceleration $\{A_x(n), A_y(n)\}$, and jerk $\{J_x(n) J_y(n)\}$ (Fig. 7). We also extract pointwise *spatial* parameters related to direction $\theta(n)$ and curvature $\Phi(n)$, and *temporal* parameters such as duration of pen-lifts between consecutive words, and within words. At the *stroke* level, we extract several parameters such as stroke duration and length, and normalized jerk [75], defined as the derivative of acceleration normalized w.r.t stroke length and duration. Other pointwise parameters, such as pen pressure and its variation, are also included. A stroke is defined as the pen trajectory comprised between two consecutive extrema of $y(n)$ (i.e. at $V_y(n) = 0$, where $V_y(n)$ is the pointwise vertical velocity). We obtain, as a result, 46 parameters, 22 pen-down features, and 24 pen-up features. Each feature is then discretized into a histogram of five bins, consisting of the relative frequency of the feature temporal values in each bin. Considering 5 bins allows for a slightly higher level of granularity, w.r.t to the 4 bins used by our age study, for encoding the dynamics in a coarse way.

4.2. Semi-supervised learning

We propose a new approach that generates subject clusters, and analyzes their correlation with the three cognitive profiles. As the optimal number of clusters and the subset of semi-global spatiotemporal features that are discriminant are both unknown, we consider a semi-supervised learning in which a *Normalized Mutual Information* feature selection guides a clustering algorithm to optimize the trade-off between the number of clusters and the discriminative power of each w.r.t the three cognitive profiles.

To analyze the quality of each semi-global feature $\mathbf{F}_l$ (vector of five bins) in discriminating the three classes (*ES-AD, MCI* and *HC*), we perform a *Hierarchical Clustering* [61] of subjects, based on $\mathbf{F}_l$. Then we compute the *Mutual information* (*MI*) between the classes and the obtained clusters as follows:

$$\mathrm{MI(C, A)} = \sum_{k=1}^{N_C}\sum_{i=1}^{N_A} \mathrm{p(C_k, A_i)log_2}\left(\frac{\mathrm{p(C_k, A_i)}}{\mathrm{p(C_k)p(A_i)}}\right) \tag{1}$$

where *C* and N_C are respectively the set and number of clusters, while *A* and N_A are respectively the set and number of classes (*ES-AD, MCI* and *HC*). The better a feature, the greater the associated *MI*. As *MI* increases with the number of clusters, optimization with Eq. (1), leads to obtaining singleton clusters, consisting each of one person. Thus, to determine the optimum number of clusters, we consider, instead, the Normalized Mutual Information (*NMI*) defined as follows:

$$NMI(C, A) = \frac{MI(C, A)}{(H(C) + H(A))/2} \tag{2}$$

where *H(C)* is the cluster entropy:

$$H(C) = -\sum_{k=1}^{N_C} p(C_k)log_2(p(C_k)) \tag{3}$$

and *H*(*A*) is the class entropy [20]:

$$H(A) = -\sum_{i=1}^{N_A} p(A_i)log_2(p(A_i)) \tag{4}$$

The denominator of Eq. (2) is a tight upper bound of *MI*(*C*,*A*), guaranteeing that *NMI* is always between zero and one [43]: one corresponds to highest heterogeneity or disorder, when the persons' cognitive profiles are equally distributed in each cluster, while one reflects complete homogeneity or order, when only one cognitive profile is observed in each cluster. Our feature selection process by semi-supervised learning consists of the following steps:

Step 1: For each feature,

(a) Perform Clustering with different sizes (number of clusters);
(b) Compute the *NMI* for each clustering size;
(c) Select the optimal number of clusters, namely the one maximizing *NMI*;

Step 2: Select the best feature $\mathbf{F}_1 (i = 1)$, namely the one maximizing the *NMI*, based on **Step 1**;

Step 3: Forward Feature Selection:$i = 2$

Repeat

Select the *i*th best feature $\mathbf{F}_i$, i.e. the one that, combined with the previous $(i - 1)$ selected features, maximizes *NMI,* based on the optimal number of clusters$i = i + 1$

Until *NMI* no longer increases.

4.3. Experimental results and discussion

Our algorithm of *NMI*-based conjoint feature selection and clustering detects three clusters and the following three features, selected in a decreasing order: $\mathbf{F}_1$: *Number of extrema in the in-air vertical velocity*, $\mathbf{F}_2$: *Time between words*, $\mathbf{F}_3$: *Vertical pen-down jerk.* The *k*th cluster is here referred to as C_t_k. The letter *t* refers to the *text* the clustering is based on, and superscript 1 or 2 is dropped, as there is a single clustering stage that characterizes both features and subjects. Table 3 shows the distribution of *HC, MCI* and *ES-AD* over these three clusters.

Table 3
The distribution of *HC, MCI* and *ES-AD* over the three clusters (C_t_i), based on the three selected features.

	HC	*MCI*	*ES-AD*	Total
$\mathbf{C\_t_1}$	2	1	4	7
$\mathbf{C\_t_2}$	3	50	22	75
$\mathbf{C\_t_3}$	23	36	3	62
Total	28	87	29	144

4.3.1. Analysis of the obtained clusters

We observe that the first cluster C_t_1 is very small (comprises 5% of persons) and thus can be ignored when analyzing the main trends (we postpone its analysis to the end of this section). Most people pertain to one of the two major clusters, C_t_2 (52%) and C_t_3 (43%), from which a striking finding is revealed: C_t_2 is dominated by *ES-AD* and *MCI* subjects, while C_t_3 is dominated by *HC* and *MCI*. From these two clusters, two major interpretations can be drawn:

(i) Leaving aside *MCI* subjects, the selected features discriminate *HC* from *ES-AD:* C_t_2 comprises 22 *ES-AD* (76% of *ES-AD* subjects) and only 3 *HC* (11% of *HC* subjects), while C_t_3 comprises 23 *HC* (82% of *HC*) and only 3 *ES-AD* (10% of *ES-AD*). This is remarkable as we include only subjects with *early stage Alzheimer*'s, and this confirms that alterations do show up in the *HW* of *AD* subjects at an earlier stage.
(ii) Despite this, few *HC* are mixed with *ES-AD* in C_t_2 and few *ES-AD* are mixed with *HC* in C_t_3. This confirms our claim that these two cognitive profiles are not homogeneous, but rather may contain subgroups with different behaviors.
(iii) Current state of the art treats *MCI* as a monolithic entity by reporting that some *HW* parameters discriminate *MCI* as a whole from the other cognitive profiles, and that some do not. Our findings, by contrast, reveal that *MCI* patients are split over C_t_2 (57%) and C_t_3 (41%), and this shows that they have fine motor skills shared either by *HC*'s or by *ES-AD*'s. This corroborates the definition of *MCI* as a transitory phase between *HC* and *AD*, and our results are the first of their kind to show two *MCI*'s *HW* behavioral trends, one leaning towards *HC*'s and one towards *ES-AD*'s.

4.3.2. Analysis of the selected features

Among the three selected features, one is a pen-up feature ($\mathbf{F}_1$), one is a pen-down feature ($\mathbf{F}_3$), and one is the time between words ($\mathbf{F}_2$). This shows that these three types of spatiotemporal features are important to detect different writing styles, those characterizing cognitive impairment in particular. Features $\mathbf{F}_1$ and $\mathbf{F}_2$ seem to be relevant as they require visual short-term memory skills when copying the words, one after another, while $\mathbf{F}_3$ characterizes the writing movement fluidity.

As described in Section 4.1, each feature is encoded over five ordered bins, the first and last representing the frequency of the low and high feature values, and the bins in between representing the intermediate values. To characterize the *HW* of each cluster, we show, in Fig. 8, the distribution of each selected feature's bins over the three clusters. The major observations follow below.

- $\mathbf{F}_1$: the 1st bin shows a lower value for C_t_2 w.r.t C_t_3, while the opposite is observed for subsequent bins. This means the number of extrema of pen-up vertical velocity tends to be higher in C_t_2, a finding that reveals that the subjects in C_t_2, dominated by *ES-AD* and partially by *MCI*, have a less fluid *HW*, characterized by a larger number of velocity changes in pen-up trajectories.
- The distribution of $\mathbf{F}_2$ shows roughly the same trend as $\mathbf{F}_1$, meaning that the time between words tends to be higher in

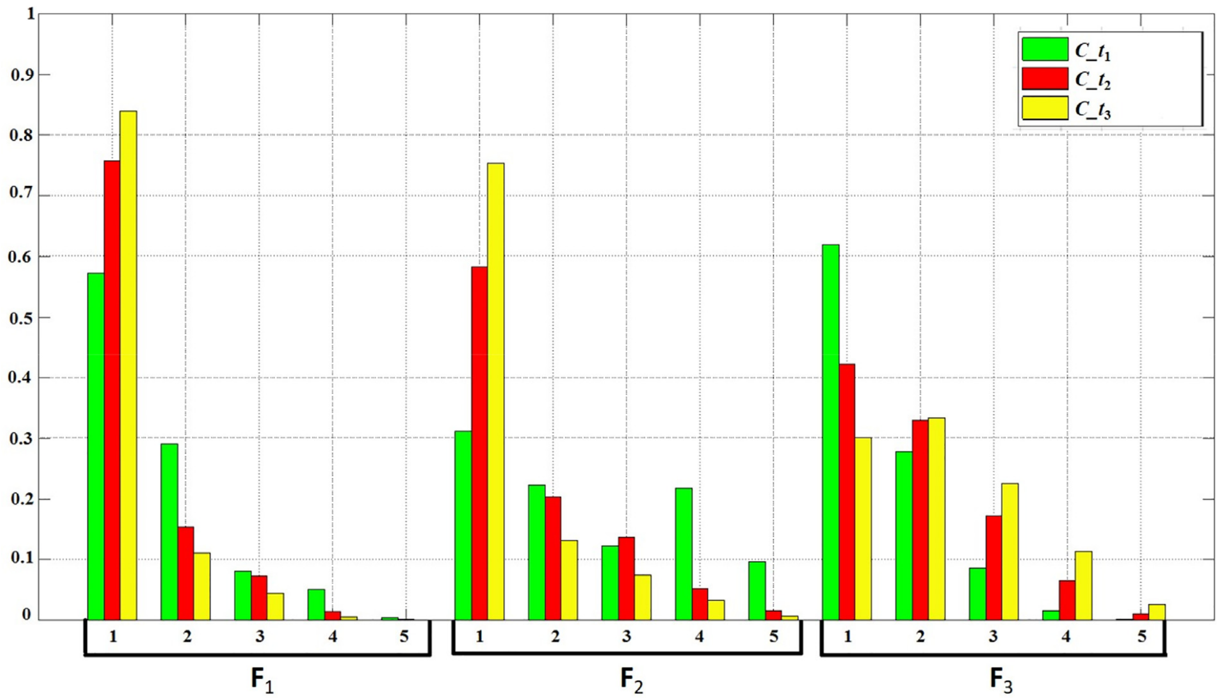

Fig. 8. Distribution of the three selected 5-dimensional features over the clusters (C_t_1 (green), C_t_2 (red), and C_t_3 (yellow)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

C_t_2. This reveals that the subjects of C_t_2 spend more time in copying words one after another, due probably to cognitive impairment inducing hesitations and more back and forth eye movements from the text to be copied to the tablet writing surface.

- The distribution of $\mathbf{F}_3$ shows the opposite trend to that of $\mathbf{F}_1$ and $\mathbf{F}_2$, meaning that the vertical pen-down jerk tends to be lower in C_t_2. This is not surprising as jerk[1] is highly correlated to velocity and acceleration, and thus the subjects in C_t_2 are characterized by lower jerk as they write more slowly.

To summarize this feature comparison, we can conclude that the subjects of C_t_2 write more slowly, less fluidly and with more hesitations. As the subjects in this cluster consist of most *ES-AD* subjects and of about 57% of *MCI*, this means that *fine* motor impairment characterizes not only early stage *AD*, but also, and to a large extent, its preclinical phase. Fig. 9 shows some *HW* samples representing C_t_2 and C_t_3, that highlight the two different behavioral trends: slow *HW* for C_t_2, characterizing most ES-*AD* and a significant part of *MCI*, and fast *HW* for C_t_3, characterizing most *HC* and another significant part of *MCI*.

4.3.3. Link with our study on age

The writing style characterizing C_t_2 shows some similarities with age category 6, $C^2\_A_6$, in our study of age, that uncovered a subgroup of the oldest age groups, $AG_{66–75}$ and $AG_{76–86}$, with a proper writing style, not shared with other subjects from these two groups and from the other age groups, and characterized by the lowest velocity, acceleration, jerk, with a medium to high in-air time. This suggests that the $AG_{66–75}$ and $AG_{76–86}$ subjects in $C^2\_A_6$ might have a cognitive decline that share some fine motor skill impairments with *ES-AD* and a part of *MCI*. This may also mean that the three *HC* subjects pertaining to C_t_2 are similar to those of $C^2\_A_6$, which explains why they end up with most *ES-AD* in the same cluster.

4.3.4. Analysis of tiny cluster C_t_1

This cluster is composed of 2 *HC*, 1 *MCI*, and 4 *ES-AD*. Fig. 8 shows clearly that, w.r.t C_t_2 and C_t_3, C_t_1 is mainly characterized by a much more frequent high number of extrema of in-air vertical velocity, and long time between words (3rd, 4th and 5th bins of $\mathbf{F}_1$ and $\mathbf{F}_2$ are much higher), as well as a much more frequent low vertical pen-down jerk (4th and 5th bins of $\mathbf{F}_3$ are much lower). This corresponds to a neat writing characterized by very slow movement and poor fluidity (Fig. 10), as the subjects resort to frequent stopping, generating thereby more velocity extrema (minima and maxima). The subjects spend also a larger time between words, which again favors a tidy writing.

This peculiar writing style deserves special attention as it characterizes none of the three cognitive profiles. It requires further analysis to rule out potential annotation issues, and to scrutinize other metadata of the subjects, by checking whether additional factors may explain why their writing is so distinct from the rest. That said, however, this cluster indirectly highlights one of the main strengths of our framework, and plays a key role regarding the quality of the obtained clusters. Despite its tiny size, isolating automatically C_t_1 allowed discarding few subjects, but with outlier-like *HW* dynamics, which enabled the model to unveil the two major behavioral trends featured by clusters C_t_2 and C_t_3. Without our automatic detection of three homogeneous clusters, C_t_1 would have "corrupted" its closest cluster, C_t_2 (with slower *HW* than C_t_3) and compelled the clustering algorithm to split the subjects therein in several tiny groups, loosing thereby the emergence of the meaningful and reliable behavioral trend of C_t_2.

4.3.5. Comparison of semi-global with global parameterization

To assess our semi-global feature parameterization w.r.t to the global one, which is adopted by the state of the art, we run the same study as above but by considering this time *global* features. Based on the Normalized Mutual Information (*NMI*) scheme, the

[1] Jerk corresponds here to pointwise jerk over time and is not to be confused with the normalized jerk per stroke, found in some studies to be correlated with tremor.

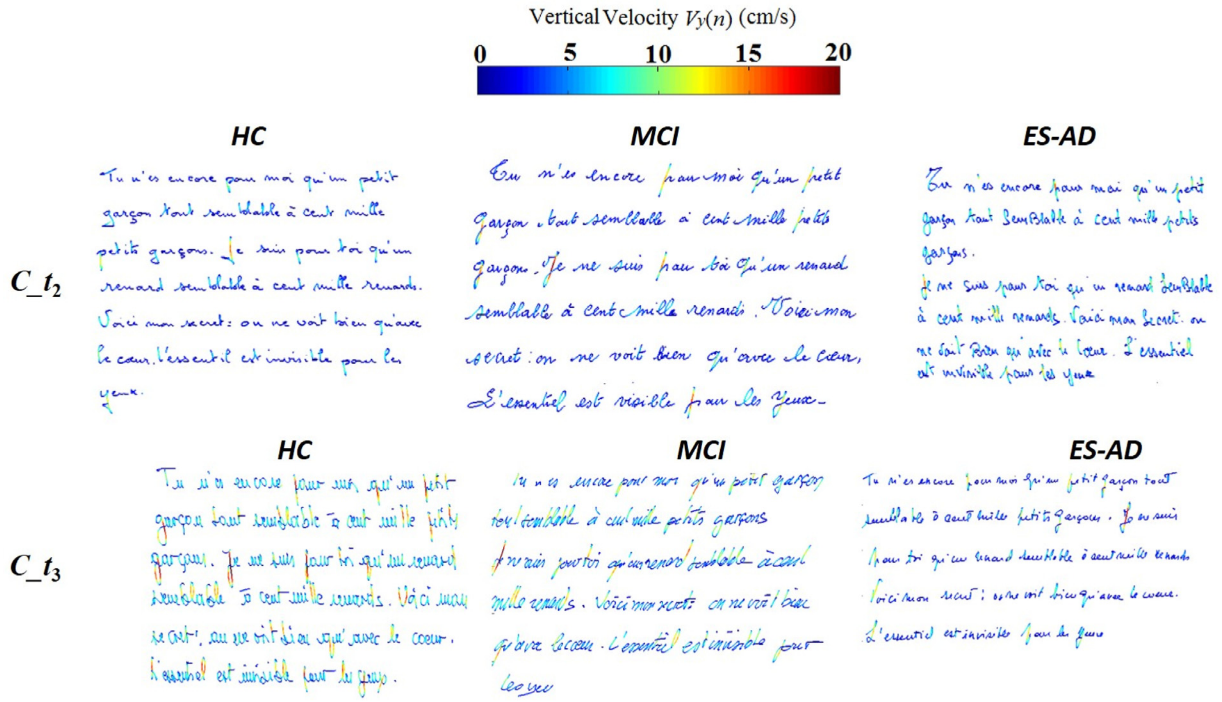

Fig. 9. Text samples from C_t_2 (top) and C_t_3 (bottom), showing two *HW* behavioral trends. slow for C_t_2, and fast for C_t_3.

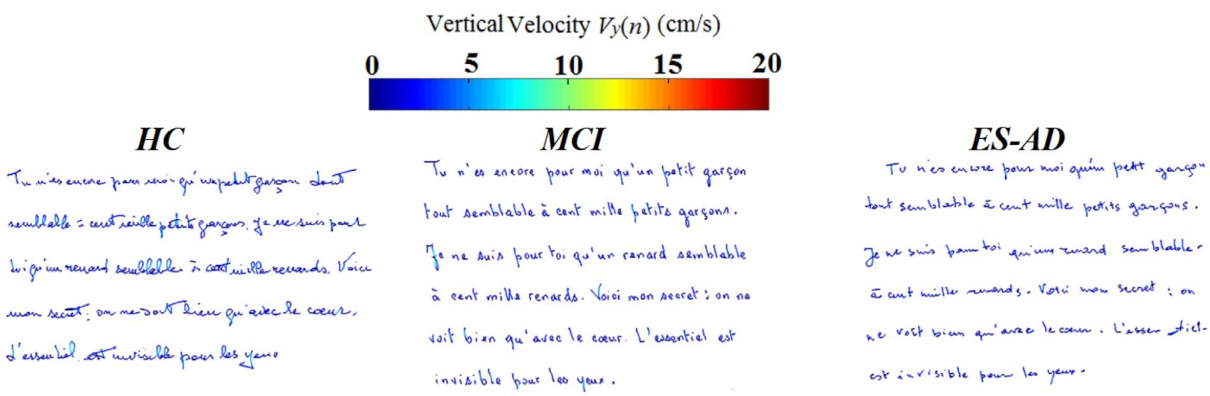

Fig. 10. Text samples from the tiny cluster, pertaining to 1 *HC*, 1 *MCI*, and 1 *ES-AD*.

Table 4
The distribution of *HC, MCI* and *ES-AD* over the three clusters (C_tg_k), based on the nine selected global features.

	HC	*MCI*	*ES-AD*	Total
$\mathbf{C\_tg_1}$	19	37	2	58
$\mathbf{C\_tg_2}$	7	48	21	76
$\mathbf{C\_tg_3}$	2	2	6	10
Total	28	87	29	144

Table 5
NMI values for global and semi-global feature parametrization. For the latter, each feature is encoded by five bins.

	NMI	Number of features
Global parameters	0.10	9
Semi-global parameters	0.14	3×5

detected number of clusters is again 3, shown in Table 4, with an uncovering of a similar behavior of the *MCI* class, split over two clusters (one related to *HC* and one to *ES-AD*), with again a tiny cluster of 10 subjects (2 *HC*; 2 *MCI*; 6 *ES-AD*). The fact that the *MCI* class is split into two parts confirms its bimodal behavioral trend unveiled with semi-global parameterization. Interestingly, the tiny cluster above comprises all the seven subjects (2 *HC*; 1 *MCI*; 4 *ES-AD*) of the analogous one, obtained in Section 4.3.4 with the semi-global setting. This consistency means that these subjects have a HW style so slow and tidy that they are set apart from the rest, regardless of the granularity of the feature encoding adopted, global or semi-global.

Despite these similarities, however, the *NMI* value, as shown in Table 5, is higher for our semi-global parametrization, which proves its better discrimination of the three cognitive profiles (*ES-AD, MCI, HC*). Fig. 8 sheds light on the reason why: a general observation, indeed, is the overall significantly decreasing size from bin 1 to bin 5 regardless of the features. Despite their small size, however, the last bins correspond to infrequent but subtle events, that are important for discriminating different writing styles, as this is shown for $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{F}_3$. Without our semi-global parametrization allowing an automatic detection of such subtle events, the bin

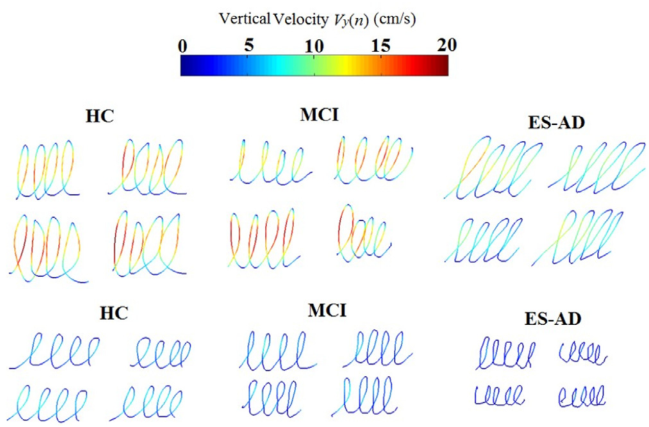

Fig. 11. Four llll series from subjects with different cognitive profiles. Color encodes the velocity dynamics.

values encoding these events would have been diluted into the global values through the averaging process.

In terms of features, nine global parameters are selected: *Task duration, average pen-down velocity magnitude, average horizontal in-air velocity, average pressure variation, average normalized in-air jerk, average number of extrema of pen-down vertical velocity, average vertical in-air jerk, Total pen-down time, Total in-air time*. As in the semi-global parameterization case, the selected global features convey information from both the in-air trajectory and the on-tablet one. Four out of nine are kinematic (velocity and jerk-based), and three are temporal.

If we disregard the dimensionality, the number of selected semi-global parameters is much lower (3 vs. 9). However, as they are encoded over five bins, an additional selection of a semi-global parameter implies adding five dimensions, which limits the number of selected features, given the small size of the training dataset. On a larger dataset, we can expect a larger improvement gap of the semi-global parameter setting over the global one.

A final remark is that, w.r.t our study on age, our semiglobal parametrization scheme for assessing *HC, MCI* and *ES-AD*, is based only on a unique clustering stage. This is because our spatiotemporal parameters are computed over the whole text. An improvement, in this regard, is to consider a two-stage clustering, where the first operates on words instead of the whole text, and the second clusters the subjects based on the distribution of the set of words of each subject over the first stage clusters. To do this, however, a reliable segmentation of the text into words needs first to be performed.

5. *AD* and *MCI* assessment by representation learning from *HW* trajectories

Encouraged by our findings with semi-global feature encoding, we take a leap forward by modeling the full dynamics of HW strokes, in a task involving writing four series of four cursive concatenated l loops (llll) (Fig. 11). As modeling the HW trajectory for cognitive assessment has not been addressed before in the literature, we have chosen the loops series to study its potential, as they allow a text-dependent study that can reveal clearly the behavioral trends related to the subjects' health conditions, by discarding from the outset any variations due to change of words or characters. Once the potential is confirmed, the approach can be applied in a straightforward manner to any other task.

The key our approach builds on is to harness the online HW time ordering to automatically learn, for each raw kinematic parameter, feature representations [6] in an unsupervised way, instead of considering handcrafted global or semi-global features, assumed implicitly to be discriminant. The modeling of HW dynamics for feature representation has never been considered before, especially, in the context for health assessment.

Our modeling relies first on automatically segmenting the (llll) series into individual loops. The segmentation allows to significantly increase the size of the training data, and accordingly the reliability of the clustering. It also allows generating individual loop-based clusters, that are much more likely to be homogeneous than would be the clusters of entire llll series.

We show next how this representation learning can be exploited either in a semi-supervised setting to uncover the link between homogeneous clusters and the cognitive profiles, or in a supervised one for classification.

5.1. Segmentation into loops and feature extraction

We segment each continuous llll series into isolated instances of letter l, from which we keep only the loop part for subsequent feature extraction. The segmentation process is the following: a low-pass filter is applied to smooth $V_y(n)$, the vertical velocity signal of the llll series, by setting the cutoff frequency to the series' fundamental frequency. We then apply the inverse Fourier transform and segment the trajectory at points n where $V_y(n) = 0$. Each loop is then retrieved by merging its two consecutive strokes, a stroke being the portion between two consecutive points with $V_y(n) = 0$. This process is illustrated on Fig. 12.

As shown by Fig. 12, setting the cutoff frequency to the fundamental frequency allows to segment the loops, irrespective of the irregularities and tremors in shaky handwriting, in a much reliable way than manually-based thresholding techniques would.

We extract, at each loop point, the velocity in x and y directions, $V_x(n)$ and $V_y(n)$. An illustration of these velocities is given in Fig. 13 that shows the temporal velocity magnitude $|V(n)|$ for some loop samples. For the studies below, we use only velocity to encode HW but we will show how the other features can be integrated as well.

5.2. Two-stage clustering

In this task, the number of subjects is 141 (27 HC, 87 MCI, 27 ES-AD), three less than those participating in the text copying task considered by the semi-global approach (three people did not perform the loops' task). As each person writes four llll series, the number of total segmented loops is 2263 (a little more that 16×141 as few subjects produced sometimes more than four l loops).

We consider a two-stage clustering based on the loop's velocity's trajectory. To model HW's full dynamics, we propose a temporal clustering of the loops, considered as time series, by a K-medoids algorithm taking as similarity measure DTW (Dynamic Time Warping) that accommodates the data sequential aspect. This clustering generates a dictionary of prototype medoids, regardless of the cognitive profile, that serves as input to the 2nd stage clustering. The latter then computes for each subject the distribution (histogram) of his/her loops over the medoids (1st stage clusters). Hereafter, the 1st and 2nd stage clusters will be referred to, respectively, as $C^1\_l_k$, and $C^2\_D_j$. Here, $C^1\_l_k$ designates the *k*th cluster of loops (hence the letter l), obtained at the 1st stage, while $C^2\_D_j$ refers the *j*th cluster of subjects with different cognitive profiles (HC, MCI or ES-AD), and obtained at the 2nd stage. Letter D stands for Disease, in order to distinguish the second clusters here from those related to age in Section 2.

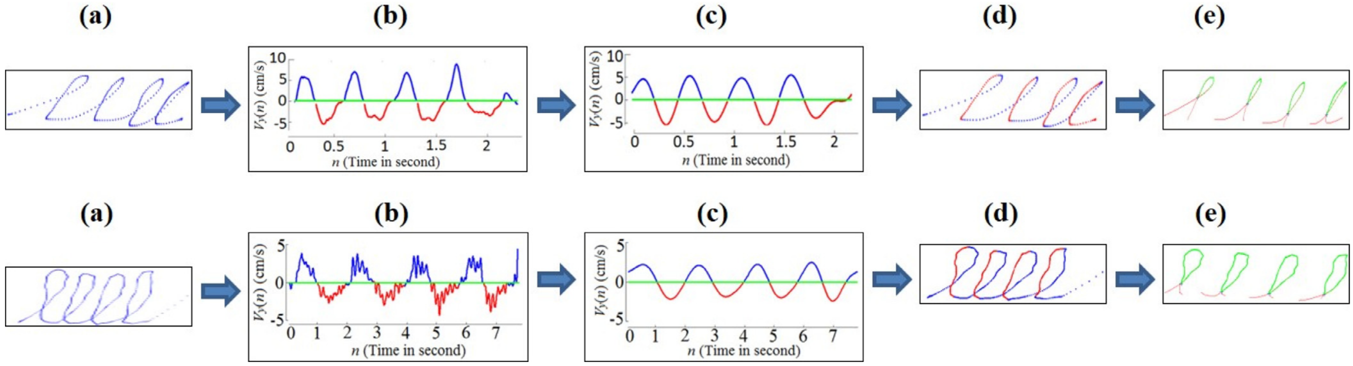

Fig. 12. Loop segmentation: (a) input loops series; (b) the $Vy(n)$ signal, (c) low-pass filtering by the fundamental frequency; (d) segmentation into ascending and descending strokes; (e) extraction of the loops. Top: fluid writing; bottom: shaky writing.

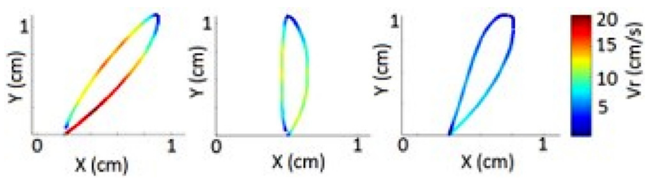

Fig. 13. Loop samples with color encoding velocity magnitude: red stands for high values, and blue for low ones. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2.1. First stage clustering

Fig. 14 shows the results for $K=8$ and $K=30$ medoids. In each case, the K medoids are the major prototypes of the total set of loops produced by *HC, MCI* and *ES-AD*, and as shown, they represent a large diversity in terms of dynamics and shape. Each medoid reflects a different and rich combination of several loop features including full velocity profile, size, slant, fluidity, *etc*.

Fig. 14 also shows the effect of increasing the number of clusters. For $K=8$, the medoids represent the major prototypes in terms of velocity, size and slant, as they attract, each, a relatively large number of loops. A much higher K (e.g. 30 here), by contrast, allows the medoids to capture, each, only the loops they are close to. This allows the algorithm to detect new prototypes with much subtler spatiotemporal dynamics, such as very fast medoids, or those with moderate and mostly low velocity but with shaky writing that induces loss of smoothness and fluidity. High values of K, nonetheless, imply that the obtained clusters have a relatively low size. This underlines the importance of choosing a K value granting a good trade-off level between the depth of details in the medoids and the representativeness of the associated clusters, depending on the data size and the classification task.

Fig. 15.(A) displays samples from four loop clusters when $K=8$ (4 chosen among 8). These clusters were selected as they convey the main tendencies observed for other clustering results with similar but different K values. Each row corresponds to a cluster and includes, for presentation clarity, only the closest loops to their medoid, which appears first. Along with each cluster's index, we display the number of loops it contains, as well as their distribution over the three cognitive profiles. As shown, each cluster is associated with a unique combination of several *HW* features like the full velocity dynamics profile, fluidity, shakiness, slant, and even, to a large extent, size.[2] This is remarkable, as the input to clustering is merely loop raw velocity trajectories, which confirms that our scheme allows an *unsupervised representation learning* from *sequences*, a problem that is not addressed by state of the art representation learning [6]. The rich set of features conjointly uncovered shows the key advantage of modeling the full sequence of each loop instead of simple statistics such as the average of each spatiotemporal parameter, taken separately.

A more in-depth analysis of each cluster gives us insights on the behavioral trends of the three cognitive profiles. If we leave aside, for the moment, the *MCI* class, we observe the following main tendencies (Fig. 15.(A)):

- $C^1\_l_4$ contains loops mostly originating from *HC* (64 *HC;* 4 *ES-AD*), characterized by highly fluid loops with moderate size, and medium to high velocity on their ascending and descending phases. *ES-AD* subjects, therefore, seem to have trouble with maintaining this typical writing style.

[2] In Fig. 13, the loop sizes are mostly homogeneous. The variations that appear on some clusters can be explained by the small number of clusters (medoids) and the reliance only on the velocity signal for the clustering. Consider instead position will generate other clusters, that emphasize more the size information.

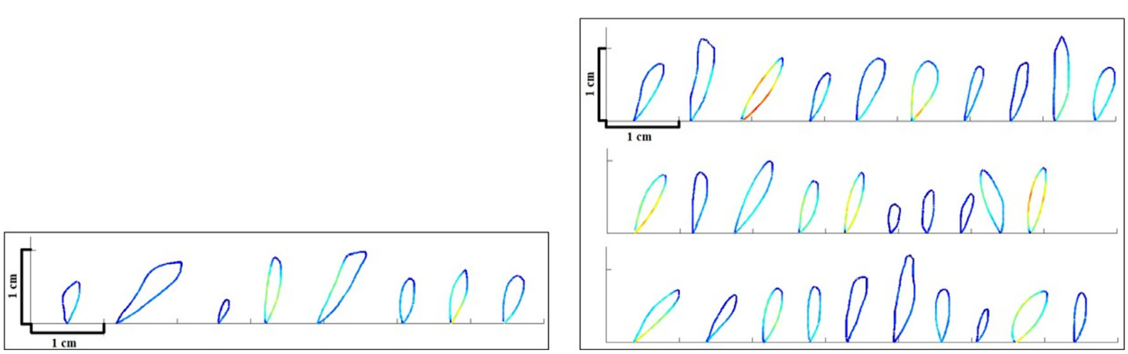

Fig. 14. Medoids obtained on the (V_x,V_y) trajectory for: (A) $K=8$ and (B) $K = 30$ medoids respectively. Color stands for velocity magnitude: red means high local velocity values, and blue, low ones. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

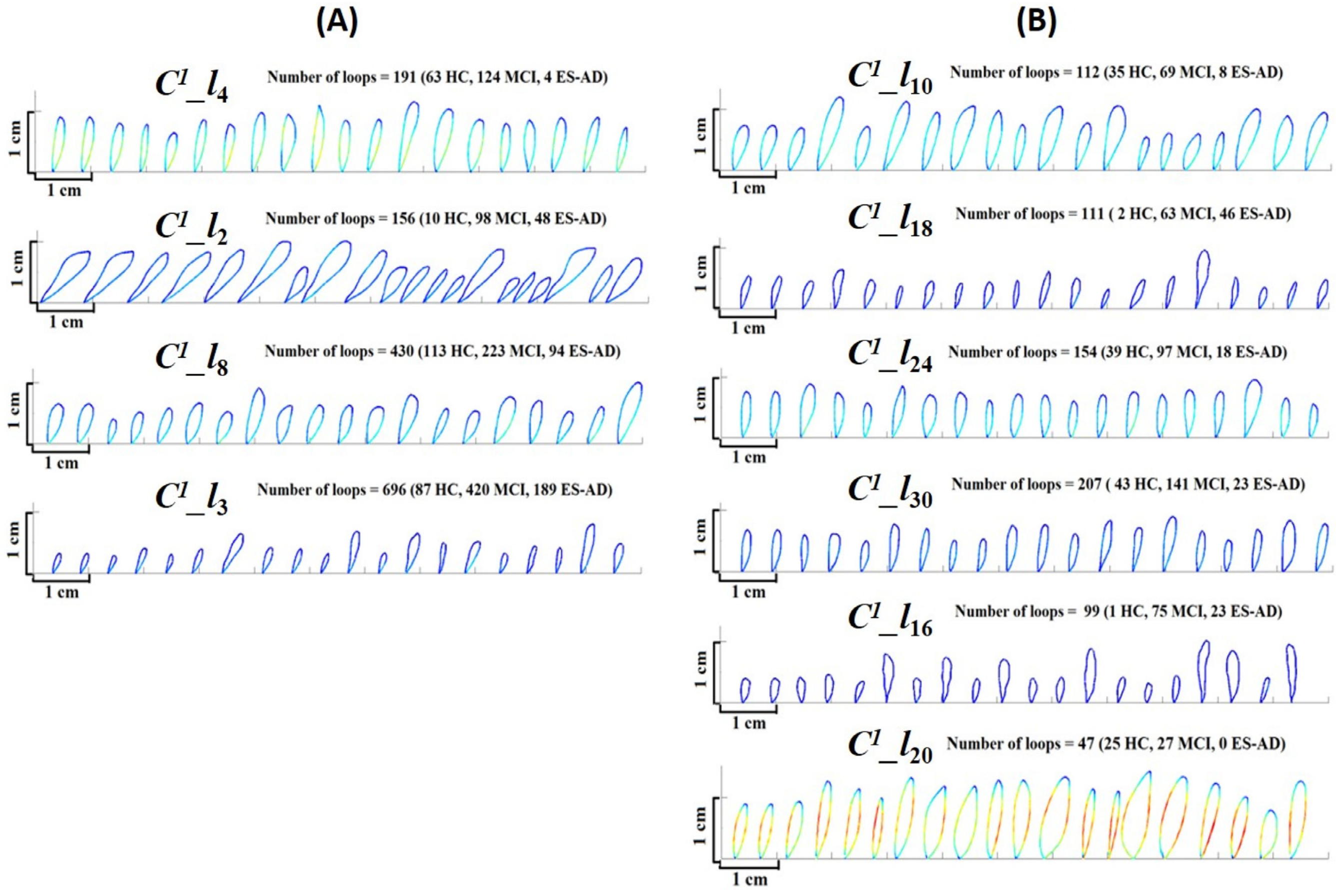

Fig. 15. Samples from some clusters for different numbers of K medoids. (A): $K = 8$, from which four typical clusters are shown; (B) $K = 30$, from which six typical clusters are shown. For each cluster, we report its number of loops, and the number of loops for each class.

- $C^1\_l_2$ shows the opposite trend (10 *HC;* 48 *ES-AD*) as it contains mid-sized to large loops, with mostly low velocity. Here, the subjects actually try to write faster at the onset of the ascending or descending phase, but quickly fail to maintain the rhythm. This results in a loss of fluidity as manifested by the sudden change of loop velocity or slant. This is an example of behavioral handwriting impossible to detect from the global or semi-global spatiotemporal parameters, but which we uncover automatically thanks to our representation learning on the loop's velocity sequences. Given the (*HC; ES-AD*) distribution, the *HW* impairments featured by this cluster seem to appear for *AD* at an early-stage, although they may show-up occasionally for a *HC* subject (one subject usually produces $4 \times 4 = 16$ loops, and $C^1\_l_2$ contains only 10 *HC* loops, i.e. less than one *HC* subject in average).
- $C^1\_l_8$ comprises loops with a moderate velocity in the ascending phase, that decreases in the descending phase, while fluidity is maintained throughout the loop. This is a balanced cluster in terms of *HC* (113) and *ES-AD* (94), and given the fluidity shown, the *ES-AD* subjects here are those who maintained good fine motor skills, contrary to those in $C^1\_l_2$.
- $C^1\_l_3$ consists of a large number of *HC* and *ES-AD* loops, but with a clear skewed distribution in favor of *ES-AD* (87 *HC;* 189 *ES-AD*). It is characterized by mid-sized to very small loops, with low velocity, shakiness and loss of fluidity. Fluidity loss and shakiness show up as the subjects struggle to produce such small and slow loops, which hampers a natural HW rhythm. This style sheds light on the correlation between micrographia and fluidity loss, that can be developed at an early stage of *AD*. This correlation between two types of *HW* impairment is the kind of findings that are not possible with state of the art approaches, but which are brought to light thanks to our framework combining semi-supervised clustering and sequential representation learning. Despite these impairments, an interesting observation, though, is that this style is shared by *HC* and *ES-AD*. The *HC* subjects here might be similar to those in $C^2\_A_6$, the cluster of aged people uncovered by our age study (Section 2), who write very slowly. We may speculate that the *HC* subjects of this cluster and those of $C^2\_A_6$ exhibit a clear behavioral decline induced by either of the two following reasons. The first is that the subjects in $C^2\_A_6$ and $C^1\_l_3$, even if they are clinically healthy, may have an aging cognitive decline that induces similar handwriting alterations to those manifested in *ES-AD*'s *HW*. The second is that these elders may actually already be developing undiagnosed cognitive impairment.

Now, an explanation of *MCI* is in order. Fig. 15.(A) shows that for all the clusters, the *MCI* class is always a significant part. This can be explained by its larger size (87 *MCI* vs. 27 *HC*, and 27 *ES-AD*), and by the fact that *MCI* covers a large cognitive spectrum ranging from the mildest cognitive impairment, when *MCI* is diagnosed at an early stage, to the strongest one, just before *AD* is diagnosed. This explains also why *MCI* appears in clusters that comprise mostly people with no cognitive decline (e.g. $C^1\_l_4$), but also in clusters that comprise people with strong cognitive decline (e.g. $C^1\_l_2$), possibly associated with *AD*.

The results above show the power of our semi-supervised representation learning in uncovering, even with few clusters (e.g. the eight above), the writing styles that define the main behavioral

Table 6
NMI for the optimal number of clusters in the 2nd stage, conditionally on the 1st stage number of medoids.

1st stage medoids	4	5	6	7	8	9	30	50	100
2nd stage clusters	5	5	4	6	**6**	3	4	5	3
NMI	0.04	0.02	0.03	0.04	**0.06**	0.03	0.04	0.03	0.03

trends of the cognitive profiles. If we increase the number of clusters (medoids), this capability increases accordingly. Fig. 15.(B) displays samples from six loop clusters when $K=30$ (6 chosen among 30), as they are typical also of the kind of clusters obtained at such higher values of *K*. The analysis of the six clusters shows an interesting evolution of the writing styles as *K* increases. Concretely, the top four clusters of Fig. 15.(B) can be considered as similar to the four clusters of Fig. 15.(A). In each set, (A) or (B), indeed, the top cluster underlines a typical *HW* style of *HC* (barely any *ES-AD*), the second from the top characterizes a typical style of *ES-AD* (barely any *HC*) with degraded style, the third detects the *ES-AD* subjects still maintaining good fine motor skills, and the 4th the *HC* subjects with significant cognitive decline, the *HW* symptoms of which are shared with *ES-AD* and the *MCI* subjects with more pronounced cognitive impairment. Notwithstanding, if we scrutinize both sets, it becomes clear that the ones on the right (obtained with $K=30$) are much more homogeneous: the top cluster on the right, for instance, includes only loops with consistently moderate velocity in their ascending and their descending phases, while its left counterpart comprises much more variations, reflected in the loops' moderate to high velocity appearing inconsistently on the ascending or descending phase. This difference is maintained overall, as the clusters on the right side have higher homogeneity in their velocity profile. Besides, with $K=30$, new clusters emerge, like the one at the bottom ($C^1\_l_{20}$), which includes very fast loops, or the second ($C^1\_l_{18}$) and the fifth ($C^1\_l_{16}$) consisting of very slow and non-fluent loops, with different levels of shakiness. It comes as no surprise that these three clusters are highly discriminative, as reflected by their cognitive profile distributions which are very sharp ($C^1\_l_{18}$ and $C^1\_l_{16}$ contain respectively only two and one loops of *HC*, while $C^1\_l_{20}$ does not contain any *ES-AD*'s).

The higher homogeneity of fine motor skills observed in the clusters, as *K* increases, underlines the fact that a high *K* value allows the medoids to attract, each, only their closest loops, in terms of the *DTW* distance, used in our unsupervised learning. However, as *K* continues to increase, the clusters become even more homogeneous but with small sizes, which in turn, decreases their reliability of characterizing an actual behavioral trend that is not peculiar to the data at hand, but rather generalizable to unseen data. To overcome this issue, we resort again to the Normalized Mutual Information (*NMI*) criterion, to minimize both the number of medoids (loop clusters) and the number of clusters at the second (subject-based) clustering stage, as detailed next.

5.2.2. Second stage clustering

As the optimal number of clusters (groups of subjects) in the 2nd stage depends on the size of the dictionary of loop prototypes (medoids), we perform a joint optimization of the two sizes (respectively K_1 and K_2), based on NMI, to maximize the mutual information between the 2nd stage and the cognitive profiles, while penalizing the increase of both K_1 and K_2. Table 6 shows some values of NMI for different combinations of the two sizes. The optimum is obtained for ($K_1=8$, $K_2=6$). Note that much higher values of K_1 were not selected even if they show a finer motor skill characterization of the cognitive profiles (as shown for $K_1=30$ in Fig. 15.(B)). The reason is that we select the first local NMI maximum, instead of the global one, to minimize, as much as possible, the number of clusters, thus ensuring that their sizes are sufficiently large to allow reliable interpretation.

Table 7
Distribution of the cognitive profiles over the 2nd stage clusters $C^2\_D_j$, based on nine medoids (1st stage).

	HC	*MCI*	*ES-AD*	Size
$C^2\_D_1$	4	15	6	25
$C^2\_D_2$	18	46	7	71
$C^2\_D_3$	5	26	14	45
Size	27	87	27	141

5.2.2.1. Analysis of the ($K_1=9$, $K_2=3$) clustering pair.

Before delving in the analysis of the optimal clustering pair ($K_1=8$, $K_2=6$), we start first by analyzing the ($K_1=9$, $K_2=3$) combination as it consists of the fewest number of subject clusters, which allows focusing first on the major behavioral trends on the data.

Table 7 shows the distribution of the subject's cognitive profiles over the three clusters. The results are strikingly like those obtained with our semi-global parametrization, shown in Table 3. This confirms our findings that *HW* alterations do appear for *AD* at an early stage, and discriminate *HC* from *ES-AD* in most cases, and that *MCI*'s *HW* is subject to two behavioral trends, one leaning toward *HC*'s and one towards *ES-AD*'s. The new findings, nonetheless, are obtained only with the raw velocity signal on the loop's writing task, consisting usually of 16 'l' instances, while those obtained by the semi-global parametrization relied on a feature selection operating on 46 spatiotemporal parameters extracted from a rich text of 44 words, made up of over 200 characters.

The distribution of the three clusters over the nine medoids (Fig. 16) confirms our interpretations at the first level, as we see that the cluster with mostly *HC* and *MCI* ($C^2\_D_2$, in blue), comprises people mostly producing fluid loops with moderate to high velocity, while $C^2\_D_3$ (in red), with mostly *ES-AD* and *MCI*, comprises mainly people producing shaky loops with lower velocity and size. Again, a small group of subjects ($C^2\_D_1$, in yellow) appears, similar to that observed with the semi-global scheme.

Given the consistency of the two behavioral trends of *MCI*, we have studied their correlation with two metadata, age and *MMSE* (Mini Mental State Examination). Table 8 shows the same results in Table 7, but enriched by the mean and standard deviation of *MMSE* and age, for each cognitive profile and in each cluster. If we focus again on the two largest clusters, $C^2\_D_2$ and $C^2\_D_3$, we find that the *MMSE* and age information sources give new insights on our results, summarized below:

- If we compare the *MMSE* mean values (yellow cells), we observe that *ES-AD* subjects in $C^2\_D_2$ have a much higher *MMSE* than those in $C^2\_D_3$ (24.4 vs 21.9). This confirms that higher *MMSE* is correlated with maintaining fine motor skills, like those shown in the *ES-AD*'s writing in cluster $C^1\_l_8$. Not surprisingly, $C^2\_D_2$ is the only 2nd stage cluster represented by $C^1\_l_8$ (Medoid 8) (Fig. 16). Likewise, the averagely higher *MMSE* for *MCI* subjects in $C^2\_D_2$ w.r.t $C^2\_D_3$ (28.13 vs. 26.9) may be one of the explanations why the former lean towards a *HC* behavior ($C^2\_D_2$) while the second lean to an *ES-AD*'s ($C^2\_D_3$). Note

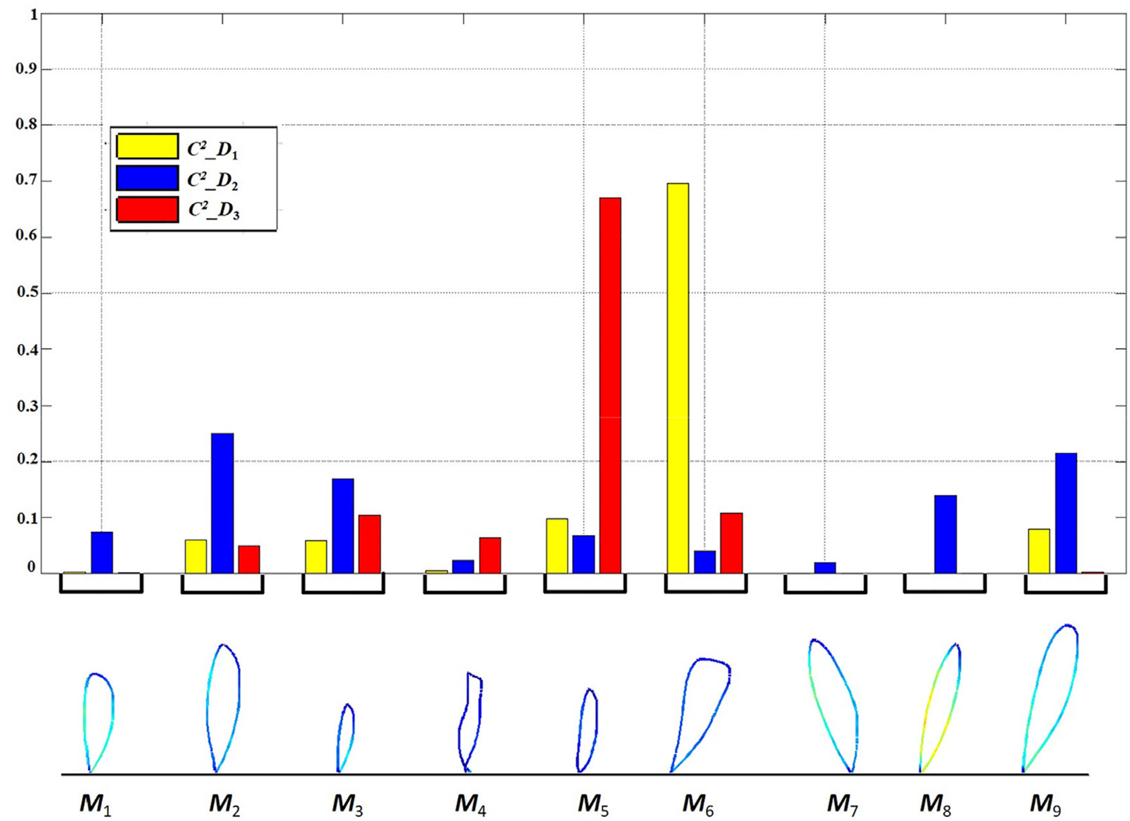

Fig. 16. Distribution of the three optimal clusters obtained in the 2nd stage, based on nine medoids (M_k) in the 1st stage.

Table 8
Distribution of *HC*, *MCI* and *ES-AD*, based on nine medoids (1st stage), enriched by the mean and standard deviation of *MMSE* and age, for each cognitive profile and in each cluster ($C^2\_D_j$).

	HC			*MC*			*ES-AD*			Total Size
	MMSE	Age	Size	*MMSE*	Age	Size	*MMSE*	Age	Size	
$C^2\_D_1$	28.8 ± 1.3	75.5 ± 3.7	4	28.26 ± 1.9	79.4 ± 6.4	15	22.6 ± 5.0	79.5 ± 6.2	6	25
$C^2\_D_2$	28.4 ± 1.5	72.6 ± 6.2	18	28.1 ± 1.7	75.7 ± 8.4	46	24.4 ± 2.5	78.4 ± 6.9	7	71
$C^2\_D_3$	29.2 ± 0.4	73.6 ± 5.8	5	26.9 ± 2.4	82.0 ± 4.8	26	21.9 ± 3.4	80.0 ± 6.7	14	45

that this difference is not maintained for *HC* where the *MMSE* is slightly lower in $C^2\_D_2$.

- A similar trend is observed for the mean age (orange cells), which is consistently lower in $C^2\_D_2$ w.r.t $C^2\_D_3$, for the three cognitive profiles, especially for *MCI* where the margin is much wider. Age advancement, therefore, may explain why *MCI* patients in $C^2\_D_3$ fail to maintain their fine motor skills.
- For $C^2\_D_1$, no clear trend seems to emerge. This cluster is mainly covered by Medoid 6 (Fig. 16) that characterizes a somewhat neat writing. This cluster has a larger size than the similar one detected with the semi-global scheme (25 vs. 7), but this is because we are not considering here the best clustering configuration, but one with the lowest number of clusters in the second stage ($K_1 = 9$, $K_2 = 3$); $C^2\_D_1$, as a result, attracts a relatively larger number of subjects.

Overall, the *MMSE* and age show a correlation with the maintenance of fine motor skills in *ES-AD* and *MCI* subjects. This effect, nevertheless, is not systematic, as the standard deviation values show that, for both *MMSE* and age, an overlapping is observed between their distributions in $C^2\_D_2$ and $C^2\_D_3$. This was expected since, besides *MMSE* and age, other key factors may explain the two major *MCI* behavioral trends we observe, chief among them, the type of *MCI* that is diagnosed. Indeed, subjects with *MCI* are usually classified into amnestic *MCI* or non-amnestic *MCI* subtypes, based on standard neuropsychological tests. The former suffer from clinically significant memory deficits, while the latter demonstrate impairment in non-memory cognitive domains including language, executive functions, or visuospatial functions. These subtypes can be further classified into single domain or multiple domain *MCI*s, based on the involvement of a single domain or multiple different cognitive domains [74]. These MCI annotations, unfortunately, are still not available to us, at this point of our study.

5.2.2.2. Analysis of the ($K_1 = 8$, $K_2 = 6$) *optimal clustering pair.* If we now select the actual optimal clustering pair ($K_1 = 8$, $K_2 = 6$), we obtain, as Table 9 shows, a much higher discrimination of the cognitive profiles, which are split into more homogeneous groups with smaller sizes. This is reflected by the emergence of two clusters with no *ES-AD* ($C^2\_D_1$ and $C^2\_D_5$), and of a cluster with no *HC* ($C^2\_D_2$).

This higher discrimination is also reflected by sharper distributions of the 2nd stage clusters over the medoids (Fig. 17), which underlines the capture of more homogeneous writing styles (fine motor skills). It is also reflected by sharper *MMSE* and age distributions over the subject clusters (Table 10), which highlights the overall higher homogeneity of these writing styles in terms of the two metadata as well. An in-depth analysis of each 2nd stage

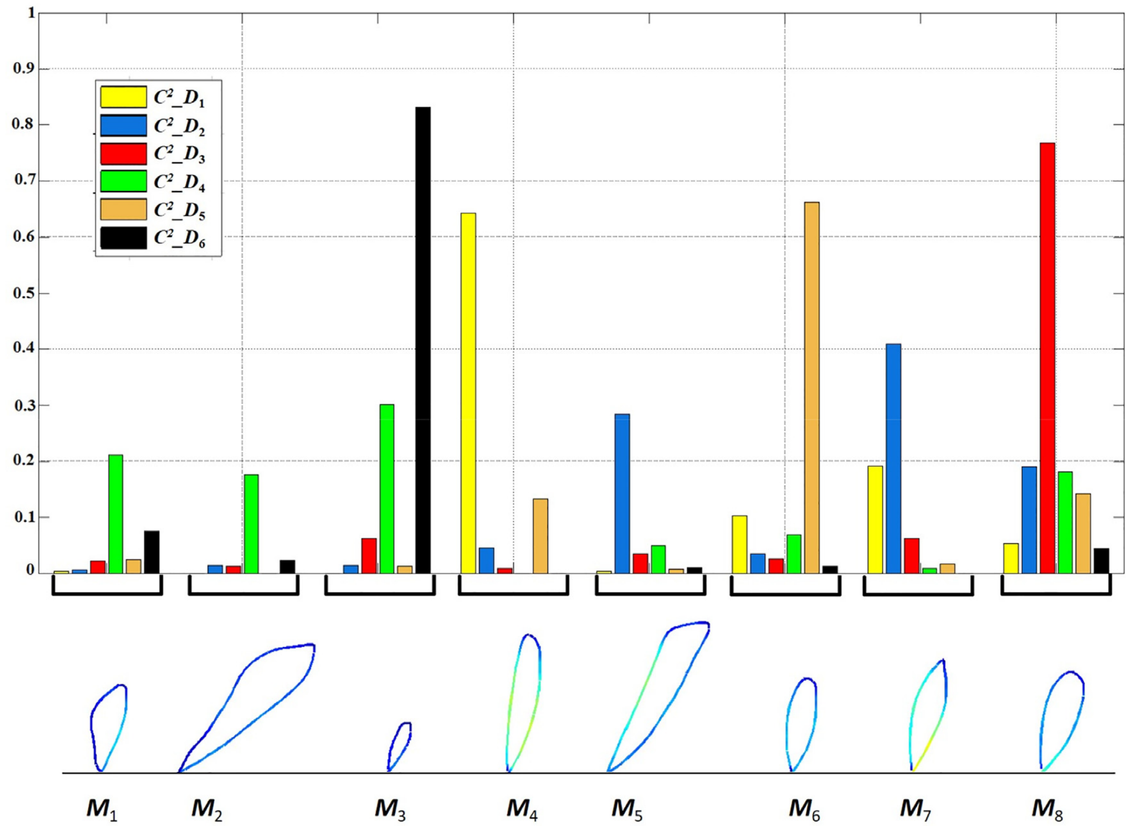

Fig. 17. Distribution of the six optimal clusters obtained in the 2nd stage, based on eight medoids (M_k) in the 1st stage.

Table 9
Distribution of the cognitive profiles over the 2nd stage clusters ($C^2\_D_j$), based on eight medoids (1st stage).

	HC	*MCI*	*ES-AD*	Total
$C^2\_D_1$	4	10	0	14
$C^2\_D_2$	0	16	2	18
$C^2\_D_3$	5	4	5	14
$C^2\_D_4$	9	25	12	46
$C^2\_D_5$	6	9	0	15
$C^2\_D_6$	3	23	8	34
Total	27	87	27	141

cluster in terms of *HW* features (velocity, fluidity, shakiness, etc.) can be done as before, based on Fig. 17, and a careful observation of the loops pertaining to each medoid-based cluster, but we drop this analysis due to the amount of space required to describe the details of six clusters ($C^2\_D_j$), and also because such an analysis would be less reliable given the smaller sizes of the groups of people in each cluster (C^2_D sizes).

5.2.3. Comparison with the global parametrization in the loops' task

To complete our analysis, we run another clustering of the subjects based on the average velocity computed on the whole loops task, as commonly adopted in the state of the art, instead of our full dynamics modeling. Based on the mean (V_x,V_y) of each subject, the *NMI*-based semi-supervised scheme detects seven clusters with a *NMI* value of 0.03, which is much lower (half) than the best *NMI* values observed for the 2nd stage clustering (Table 6). This shows the huge improvement brought by modeling the full trajectory dynamics over considering mere global parameters, and by breaking writing style modeling into two stages, the first detecting the spatiotemporal writing styles at a *HW* unit level (here loops, but it can be words as shown in our age study), and the second detecting the writer's variability over these unit-based handwriting styles.

If we now analyze the seven clusters, denoted by C_Dg_k, and shown in Table 11, it becomes clear, that the clusters with higher discrimination of *HC* vs. *ES-AD*, for instance, become smaller. This can be explained by the poor discriminative capabilities of the average velocities, which compels the clustering algorithm to detect smaller groups for which this average is discriminant. This comes

Table 10
MMSE and age of the cognitive profiles in each 2nd stage cluster ($C^2\_D_j$) (8 medoids in the 1st stage).

	HC			*MC*			*ES-AD*			Total
	Size	*MMSE*	Age	Size	*MMSE*	Age	Size	*MMSE*	Age	
$C^2\_D_1$	4	29.2 ± 1.0	69.0 ± 6.7	10	28.6 ± 1.2	75.0 ± 6.8	0	N/A	N/A	14
$C^2\_D_2$	0	N/A	N/A	16	28.2 ± 2.0	75.6 ± 10.2	2	22.5 ± 0.7	77.0 ± 9.9	18
$C^2\_D_3$	5	29.4 ± 0.6	71.0 ± 3.7	4	28.5 ± 1.7	74.2 ± 4.5	5	24.8 ± 2.9	78.8 ± 7.2	14
$C^2\_D_4$	9	28.8 ± 1.0	77.0 ± 4.1	25	27.6 ± 2.1	80.7 ± 4.6	12	22.3 ± 4.2	78.0 ± 6.8	46
$C^2\_D_5$	6	27.0 ± 1.4	71.6 ± 5.6	9	28.6 ± 1.5	73.4 ± 9.4	0	N/A	N/A	15
$C^2\_D_6$	3	29.3 ± 0.6	73.6 ± 8.1	23	27.0 ± 2.4	81.7 ± 4.3	8	22.1 ± 3.5	83.3 ± 4.0	34

Table 11
Distribution of the cognitive profiles over the clusters (C_Dg_k), based on global parametrization.

	HC	MCI	ES-AD	Total
C_Dg_1	12	29	9	50
C_Dg_2	2	10	6	18
C_Dg_3	1	8	0	9
C_Dg_4	9	20	5	34
C_Dg_5	1	12	6	19
C_Dg_6	2	2	0	4
C_Dg_7	0	6	1	7
Total	27	87	27	141

with a price, though, as the small size of these clusters (clusters C_Dg_2, C_Dg_3, C_Dg_5, C_Dg_6 and C_Dg_7) makes them unreliable for drawing meaningful conclusions; these clusters are likely to be overfitting the data.

An additional and important shortcoming with global parametrization of the spatiotemporal features is the much poorer visualization and interpretability properties they offer. By relying only on average parameters, they are unable to explain and to show the subtle local dynamic changes differentiating different cognitive profiles, and groups within each cognitive profile.

5.3. Classification

5.3.1. Bayesian scheme for the two-class (HC vs. ES-AD) discrimination problem

So far, we have proposed semi-supervised learning techniques in which the label information was used to guide the clustering algorithms to select the optimal number of clusters, whether at the feature level or the subject level. This was motivated by our goal of automatically discovering the most relevant features for characterizing ES-AD and MCI, w.r.t HC, as they are unknown a priori. Our representation learning, however, can be harnessed in a classification setting as the 1st stage medoid-based clusters are, each, discriminant to some degree, given their unbalanced distribution in terms of the cognitive profiles associated with the loops they contain. For classification, we consider only the two-class (HC vs. ES-AD) classification setting. The reason is that MCI is overly represented, and its inclusion would entail an unbalanced data distribution that is not suitable for supervised learning. Selecting a subset of MCI, instead, is not viable as this health condition includes a large diversity of types (amnestic, executive, multidomain, etc.) that are still unavailable in our dataset, but which are important for annotating the MCI subjects, prior to include MCI in a supervised classification task.

To merge the intrinsic information carried by the clusters, we consider a Bayesian formalism for classifying a writer as *AD* or *HC*, that aggregates probabilistically the contributions of the loop clusters (medoids), by leveraging the discriminative power of each. We use Bayes' rule to compute, for each subject, the posterior probability to be *ES-AD* or *HC* given his/her respective data (loops). Let us assume that the ith subject, s_i, produces N_i ($\sim 4\times 4=16$) loops, distributed over the clusters obtained by *K-medoids*, performed on the loops' training set. The posterior probability of class C_k (*ES-AD* or *HC*), given data D_i (loops from s_i), is:

$$P(C_k/D_i) = \frac{P(D_i/C_k)\ \times P(C_k)}{P(D_i)} \tag{5}$$

where $P(D_i) = \sum_{k=HC,\ ES-AD} P(D_i/C_k)\ \times P(C_k)$ and $P(C_k)$ is the *a priori* probability of class k (50% in our dataset). Assuming the data D_i (loops) from subject s_i are class-conditionally independent, we have:

$$P(D_i/C_k) = \prod_{j=1}^{N_i} P\left(M_j^i/C_k\right) \tag{6}$$

where M_j^i is the closest cluster (Medoid) to the jth loop of subject s_i. Thus,

$$P(M_j^i/C_k) = \frac{P(C_k/M_j^i)\ \times P\left(M_j^i\right)}{P(C_k)} \tag{7}$$

$P(M_j^i)$ being the *a priori* probability of cluster M_j^i, estimated by:

$$P\left(M_j^i\right) = \frac{N_{B_j^i}}{N_{Total}} \tag{8}$$

here $N_{M_j^i}$ is the number of loops in cluster M_j^i and N_{Total} is the total number of loops (~16 loops x 54 participants). Likewise,

$$P\left(C_k/M_j^i\right) = \frac{N_k^j}{N_{B_j^i}} \tag{9}$$

N_k^j being the number of loops in cluster M_j^i from class k (*ES-AD* or *HC*). Each subject i is then classified by selecting the class (*HC* or *ES-AD*) with the maximum *a posteriori* probability:

$$\mathbf{C^*} = \underset{k=HC;ES-AD}{\boldsymbol{arg\ max}}\ \mathbf{P(C_k/D_i)} \tag{10}$$

5.3.2. Experiments

For experiments, we consider the two-class dataset consisting of 27 *ES-AD* and 27 *HC*. We use the Leave-one-person-out procedure for performance evaluation. We did not use the *NMI* criterion here, as this would have entailed to consider it only on the training data, which is heavy given our leave-one-out scheme involving 54 different training datasets. To get meaningful clusters w.r.t the size of data, we have tried several numbers of clusters, by varying K between 10 and 50, and obtained similar optimal performance for K between 30 and 50. Here, we report the results for $K=30$.

For comparison, we assess the main classification approach used in the literature, namely *Linear Discriminant Analysis* (*LDA*). We implement *LDA* as in [82,36,28], by extracting global kinematic features, and we carry out two experiments: in the first, *LDA* takes as input the mean velocity ($\overline{V_x}$,$\overline{V_y}$) computed over each writer's loops, and in the second, a combination of global kinematic features computed in the same way: ($\overline{V_x}$,$\overline{V_y}$), mean acceleration ($\overline{A_x}$,$\overline{A_y}$), and mean jerk ($\overline{J_x}$,$\overline{J_y}$). Table 12 reports the classification rates of these three experiments on the training and validation sets. Note that we do not report confidence intervals for *LDA*, since it is not subject to a random parameter initialization, but we do so for our Bayesian approach since it relies on the clustering of loops, obtained from an initialization of the medoids (cluster centers); we perform then 10 independent classification runs and report the classification mean and its standard deviation.

As shown in Table 12, on the validation set, *LDA*, with ($\overline{V_x}$,$\overline{V_y}$) as input, obtains a classification rate of 51.9%, and of 50%, when ($\overline{V_x}$,$\overline{V_y}$), ($\overline{A_x}$,$\overline{A_y}$), and ($\overline{J_x}$,$\overline{J_y}$) are combined. These rates correspond essentially to chance, as a blind classifier, choosing systematically *HC* for output, gets a 50% classification rate. Incidentally, these results confirm those in [85], obtained on a similar task (3×8 loops), that report no significant difference between *AD* and *HC*, with *Anova*, based on the mean stroke velocity, despite including all AD subjects, and not only those at an early stage. This underscores the poor discrimination capabilities of the global parameters, even when they are combined. By contrast, thanks to our modeling of the full dynamics of (V_x,V_y), our approach obtains, on

Table 12
Classification rates obtained with global parameters, and with full dynamics, encoded by the temporal clusters (Medoids).

Features		Classifier	Learning set	Validation Set
Global (Average)	$(\overline{V_x}, \overline{V_y})$	*LDA*	55.9%	51.9%
Global (Average)	$\{\overline{V_x}, \overline{V_y}\} + \{\overline{A_x}, \overline{A_y}\} + \{\overline{J_x}, \overline{J_y}\}$	*LDA*	52%	50%
Full dynamics (Trajectory)	$\{V_x(n), V_y(n)\}$	Bayes' Classifier	$83.2 \pm 0.7\%$	$74 \pm 3\%$

validation, a classification rate of 74.3%, which brings an improvement margin of 50% over these global schemes. This is remarkable given that we consider only velocity in comparison to the combination above of velocity, acceleration, and jerk. This shows that the velocity *full dynamics* is a good parameter for discriminating *ES-AD* from *HC*, as it considers the changes of the velocity trajectory throughout the loop. In sharp contrast to global parametrization, this enables the discovery of subtle changes in the writing styles, occurring at different movement (and location) phases.

Although our approach outperforms the state of the art by a high margin, it is in its promising phase only, as there still remains a gap of 25% to perfect classification. This gap, however, can be significantly narrowed if the data increase. Our dataset of 54 persons is still very small, compared to the ones used for handwriting recognition, consisting of thousands of samples, or if we take into account the heterogeneity of each cognitive profile. We have shown extensively in this paper, that HC and ES-AD comprise, each, several subgroups of subjects with clearly different fine motor skills. To ensure robust classification, therefore, each subgroup needs to be represented by a sufficient number of subjects. As an example, in our results above, the ES-AD subjects still maintaining their fine motor skills, and the HC subjects failing to maintain theirs, are likely to be misclassified given their respective small number. The same applies to other subgroups not sufficiently represented in the training data. It is to be expected that, by enrolling new subjects in the study –the acquisition campaign in the Broca hospital is continuing to this date–, the classification rates would go up accordingly.

It is worth to stress that our Bayesian approach, coupled to the K-medoids temporal clustering of the loops, remains fully explainable. The classifier decision can be understood in a top-down manner by first comparing the a posteriori probabilities of the two classes, which can be broken, each, into the product of the class-conditional probabilities of the subject's loops. The values of these loop-based probabilities can, in turn, be easily understood by checking the frequency of the loops from each class in each medoid-based cluster. Finally, the visualization of the clusters – Fig. 15 illustrates an example for the three-class scenario– gives insights on the types of writing styles shared by all the cognitive profiles, and on those specific to cognitive profile declines. Such interpretability is of utmost importance to the medical staff. For instance, a neurologist that understands how the automatic classification system generates its decision, based on the subject's data, is likely to be convinced by the usefulness of such a system, and to be interested in integrating it as an aid-to-decision tool. Moreover, the medical staff can also provide an informed feedback on how to potentially improve the decision system, based on its expertise.

6. Conclusion & perspectives

We proposed in this paper a novel paradigm for studying handwriting changes due to cognitive decline associated with MCI and early-stage Alzheimer, or to aging. Our work has addressed two major limitations of the state of the art, the assumption of a unique behavioral trend for each cognitive profile, and the encoding of the HW spatiotemporal dynamics by simple global parameters. First, we relax the one per-class behavioral pattern restriction by allowing, for each, the emergence of a multimodal behavioral pattern reflecting the diversity of behaviors within a given health condition. We achieve this by performing unsupervised or semi-supervised learning to uncover homogeneous groups of subjects, and then we analyze how much information these clusters carry about the cognitive profiles (or age groups). Second, instead of relying on global (mostly average) kinematic parameters, we refine the coarse encoding, first by a semi-global parameterization, and then by modeling the full dynamics of each parameter. To illustrate the power of our paradigm, we presented three studies, one regarding age, and two regarding Alzheimer's disease.

Regarding our age study, unlike previous works reporting only one pattern of *HW* change with aging, our first study, based on a semiglobal feature parametrization scheme unveils, in an unsupervised way, three major aging *HW* styles, one specific to aged people and two shared with other age groups. In our second study, through a semi-supervised learning based on the same semiglobal parametrization, a striking finding is revealed: two major clusters are uncovered, one dominated by *HC* and *MCI* subjects, and one dominated by *MCI* and *ES-AD*, thus highlighting that *MCI* patients have fine motor skills leaning towards either *HC*'s or *ES-AD*'s.

In the third approach, our novel modeling of the full dynamics of *HW* units allowed to harness the rich temporal information inherently characterizing online *HW*. For each raw kinematic parameter, our approach can learn feature representations [6] instead of considering handcrafted global or semi-global features, assumed implicitly to be discriminant. Our scheme allows a *representation learning* from sequences, which is barely addressed in the state of the art, as it is suitable for sequential data from which temporal feature representations are to be uncovered. As a comparison, current sequential deep learning models [25,31], including end-to-end versions like *CNN*/*MLP*→*LSTM* [83,30], leave the task of *static* feature learning to *CNN* or *MLP*, *LSTM* (*RNN*) taking charge of the sequential modeling. Such an approach would not be applied in our case, as it is fully supervised, and second because *temporal*, not static, features are to be uncovered from the sequences themselves.

Applied to loops represented by their *velocity* time series, our temporal representation learning uncovers a rich set of features simultaneously as a byproduct of the unsupervised learning itself, by automatically extracting several loop prototypes, each consisting of a different combination of features like the full velocity profile, size, slant, fluidity, and shakiness. By considering a two-stage clustering based on the distribution of each user's input over the loop prototypes, we uncover again two major clusters, one leaning towards *HC* and one to *ES-AD*, with *MCI* subjects distributed over the two clusters in comparable proportions. We have shown that this bimodal behavioral trend of *MCI* is coherent with age and *MMSE* metadata, found to be higher and lower respectively in the second cluster, which strongly suggests that the *MCI* subjects gathered with *ES-AD*'s are likely to be more cognitively impaired than those with *HC*'s. Although this finding has also been unveiled by our second study as well, it was discovered here based only the velocity trajectory, instead of the large set of spatiotemporal features considered in the semi-global parametrization scheme. We also have shown that our sequential representation learning can be harnessed for classification through a Bayesian formalism aggregating probabilistically the contributions of the loop prototypes; this approach outperforms with a large margin state of the art methods based on discriminant classifiers, taking as input a set of global features.

A key advantage of our temporal representation learning is that it is fully explainable. It does not only automatically extract new *HW* features for characterizing *ES-AD*, that can be visualized and easily understood, but it also detects clusters and obtains classification results that are naturally explainable to the medical staff. This is a highly desirable property for health professionals, as they can better exploit and integrate such a system with other aid–to–decision tools.

In terms of perspectives, our work opens the door for several future directions, whether short term or mid to long term. At the short term, we have considered, in our modeling of the spatiotemporal full dynamics, only velocity trajectory. A straightforward improvement is to consider this modeling also for the other spatiotemporal parameters like, acceleration, jerk, pressure, etc., and to fuse the results from these streams. In the same spirit, the fusion can take place at the task level. Combining the input the writer produces for different tasks (loops, text to copy, free text, and drawings) will uncover potential writing impairments under different contexts, thus increasing discrimination between ES-AD, MCI and HC subjects.

On another side, although our results are already promising, and are expected to improve based on the two fusion levels mentioned above, we should assess additional metadata that were not considered as non-inclusion factors, such as the subject's education level and frequency of handwriting in daily life. Our dataset can be seen as a snapshot at a particular time for each person, and although our assumption is that these factors are expected to be statistically similar for the three cognitive profiles, they may actually induce bias in our study. To circumvent this problem, two strategies can be conjointly adopted. The first is to increase significantly the size of the dataset to ensure that each cognitive profile covers sufficiently all the factors such as the two above. Acquiring a large dataset in our health context, however, is extremely difficult, as explained in the introduction, and requires a large timeline duration. The second strategy is to consider a longitudinal study where the different methods proposed in this paper can be assessed for the subjects at two different sessions, separated by a time period between 12 to 24 months, for instance. Such a study will focus on the changes of the writing style of each subject irrespective of his/her education level, frequency of handwriting, and other metadata of this kind. In doing so, the longitudinal study will implicitly remove the possible bias introduced by these factors. Moreover, it may help assessing the predicting power of our approach by investigating *HC* subjects that may convert into *MCI* or *ES-AD*, or *MCI* patients that become *ES-AD*.

As our approach is generic and fully data-driven, it can be applied for characterizing other pathologies. This is because it automatically uncovers the features associated with different health conditions by an automatic learning of online handwriting data. The clusters resulting from such learning implicitly encode several spatiotemporal features like velocity, jerk, shape, and above all, subtle irregularities possibly associated with pathologies, like Parkinson's or Huntington's, as long as the data are acquired from patients with these pathologies and from healthy controls. Finally, the genericity of our approach makes it also applicable in a straightforward manner to non-Latin languages as well [19,29,37,50,51,67,86].

Acknowledgments

This work was funded by Institut Mines Telecom and MAIF Foundation. We would like to thank Hospital Broca for making possible the data acquisition campaign.

References

[1] M.S. Albert, S.T. DeKosky, D. Dickson, B. Dubois, H.H. Feldman, N.C. Fox, et al., The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimers. Dement 7 (2011) 270–279.

[2] S. Al Maadeed, A. Hassaine, Automatic prediction of age, gender, and nationality in offline handwriting, EURASIP J. Image Video Process. 2014 (10) (2014) 1–10.

[3] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, fifth ed., American Psychiatric Association, Arlington, VA, 2013 (DSM-5).

[4] Alzheimer's Association Report, 2017 Alzheimer's disease facts and figures, Alzheimer Dement. 13 (2017) 325–373.

[5] G.M.L Bencini, L. Pozzan, R. Biundo, W.J. McGeown, V.V. Valian, A. Venneri, C. Semenza, Language-specific effects in Alzheimer's disease: Subject omission in Italian and English, J. Neurolinguist. 24 (2011) 25–40.

[6] Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.

[7] N. Bouadjenek, H. Nemmour, Y. Chibani, Age, gender and handedness prediction from handwriting using gradient features, in: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR), 2015, pp. 1116–1120.

[8] P.A. Boyle, A.S. Buchman, R.S. Wilson, S.E. Leurgans, D.A. Bennett, Physical frailty is associated with incident mild cognitive impairment in community-based older persons J. Am. Geriatrics Soc., 58 (2) (2010), 248–255.

[9] A.S. Buchman, P.A. Boyle, R.S. Wilson, Y. Tang, D.A. Bennett, Frailty is associated with incident Alzheimer's disease and cognitive decline in the elderly, Psychosomatic Med. 69 (5) (2007) 483–489.

[10] A.S. Buchman, D.A. Bennett, Loss of motor function in preclinical Alzheimer's disease, Expert Rev. Neurotherapeutics 11 (5) (2011) 665–676.

[11] M.P. Caligiuri, H-L. Teulings, C.E. Dean, A.B. Niculescu, J. Lohr, Handwriting movement analyses for monitoring drug-induced motor side effects in schizophrenia patients treated with risperidone, Hum. Mov. Sci. 28 (5) (2009) 633–642.

[12] R. Camicioli, Handwriting and pre-frailty in the Lausanne cohort 65+(Lc65+) study, Arch. Gerontol. Geriatrics 61 (1) (2015) 8–13.

[13] C. Carmona-Duarte, M.A. Ferrer, A. Parziale, A. Marcelli, Temporal evolution in synthetic handwriting, Pattern Recognit. 68 (2017) 233–244.

[14] S.K. Chan, Y.H. Tay, C. Viard-Gaudin, Online text independent writer identification using character prototypes distribution, SPIE Electronic Imaging 6815 (2008) 1–9.

[15] A. Coates, A.Y. Ng, Learning feature representations with K-means, in: Neural Networks: Tricks of the Trade, 7700, 2012, pp. 561–580. Lecture Notes in Computer Science book series.

[16] B. Croisile, Ecriture, vieillissement, Alzheimer, Psychologie et NeuroPsychiatrie du Vieillissement 3 (3) (2005) 183–197.

[17] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, A.S. di Freca, Handwriting analysis to support neurodegenerative diseases diagnosis: a review, Pattern Recognit. Lett. (2018).

[18] B.A. Deepu V. S. Madhvanath, An approach to identify unique styles in online handwriting recognition, in: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 2005, pp. 775–778.

[19] Y. Elarian, I. Ahmad, S. Awaida, W.G. Al-Khatib, A. Zidouri, An Arabic handwriting synthesis system, Pattern Recognit. 48 (3) (2015) 849–861.

[20] M.A. El-Yacoubi, M. Gilloux, R. Sabourin, C. Suen, Objective evaluation of the discriminant power of features in an HMM-based word recognition system, in: Proceedings of the Brazilian Symposium on Advances in Document Image Analysis, 1997, pp. 60–73.

[21] M.A. El-Yacoubi, M. Gilloux, J-M Bertille, A statistical approach for phrase location and recognition within a text line: an application to street name recognition, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2) (2002) 172–188.

[22] M.A. El-Yacoubi, R. Sabourin, M. Gilloux, C.Y. Suen, Improved model architecture and training phase in an offline hmm-based word recognition system, in: Proceedings of the 13th I C2_DR, 1998, pp. 1521–1525.

[23] B. Engel-Yeger, S. Hus, S. Rosenblum, Age effects on sensory-processing abilities and their impact on handwriting, Can. J. Occup. Therapy 79 (5) (2012) 264–274.

[24] Je. Firger, Handwriting Changes Can Indicate Alzheimer's Progression, 2013. Available at https://www.everydayhealth.com/alzheimers/handwriting-changes-can-indicate-alzheimers-progression-8042.aspx.

[25] V. Frinken, A. Fischer, M. Baumgartner, H. Bunke, Keyword spotting for self–training of BLSTM NN based handwriting recognition systems, Pattern Recognit. 47 (3) (2014) 1073–1082.

[26] S. Garcia-Salicetti, N. Houmani, Digitizing tablet, in: Z. Li Stan (Ed.), Encyclopedia of Biometrics, Publisher Springer, 2009 XXXII, 1433 p., ISBN: 978-0-387-73004-2.

[27] K.E. Forbes, M.F. Shanks, A. Venneri, The evolution of dysgraphia in Alzheimer's disease, Brain Res. Bull. 63 (2004) 19–24.

[28] J. Garre-Olmo, M. Faúndez-Zanuy, K. López-de-Ipiña, L. Calvó-Perxas, O. Turró-Garriga, Kinematic and pressure features of handwriting and drawing: preliminary results between patients with mild cognitive impairment, Alzheimer disease and healthy controls, Current Alzheimer Res. 14 (9) (2017) 960–968.

[29] J. Gao, B. Zhu, M. Nakagawa, Building compact recognizer with recognition rate maintained for on-line handwritten Japanese text recognition, Pattern Recognit. Lett. 35 (2014) 169–177.
[30] N. Granger, M.A. El Yacoubi, Comparing hybrid NN-HMM and RNN for temporal modeling in gesture recognition, in: Proceedings of the International Conference on Neural Information Processing (ICONIP), 2017.
[31] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, J. Schmidhuber, A novel connectionist system for improved unconstrained handwriting recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (5) (2009) 960–968.
[32] A. Hayashi, H. Nomura, R. Mochizuki, A. Ohnuma, T. Kimpara, K. Ootomo, Y. Hosokai, T. Ishioka, K. Suzuki, E. Mori, Neural substrates for writing impairments in Japanese patients with mild Alzheimer's disease: A SPECT study, Neuropsychologia 49 (7) (2011) 1962–1968.
[33] L.E. Hebert, J.L. Bienias, J.J. Mccann, P.A. Scherr, R.S. Wilson, D.A. Evans, Upper and lower extremity motor performance and functional impairment in Alzheimer's disease, Am. J. Alzheimer's Dis. Dementias 25 (5) (2010) 425–431.
[34] G. Hinton, S. Roweis, Stochastic neighbor embedding, NIPS 15 (2002) 833–840.
[35] F. Holekian, Handwriting Analysis, The role of age and education, Int. J. Modern Manag. Foresight 1 (6) (2014) 208–221.
[36] J. Kawa, A. Bednorz, P. Stepien, J. Derejczyk, M. Bugdol, Spatial and dynamical handwriting analysis in mild cognitive impairment, Comput. Biol. Med. 82 (2017) 21–28.
[37] M. Khayyat, L. Lam, C.Y. Suen, Learning-based word spotting system for Arabic handwritten documents, Pattern Recognit. 47 (3) (2014) 1021–1030.
[38] S. Kim, S.M. Jazwinski, Quantitative measures of healthy aging and biological age, Healthy Aging Res. 4 (2015) 4–26.
[39] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clémençon, G. Cordasco, EMOTHAW, A novel database for emotional state recognition from handwriting and drawing, IEEE Trans. Hum. Mach. Syst. 47 (2) (2017) 273–284.
[40] C. Loconsole, G.D. Cascarano, A. Brunetti, G.F. Trotta, G. Losavio, V. Bevilacqua, E. Di Sciascio, A model-free technique based on computer vision and sEMG for classification in Parkinson's disease by using computer-assisted handwriting analysis, Pattern Recognit. Lett. (2018).
[41] E.D. Louis, N. Schupf, J. Manly, K. Marder, M.X. Tang, R. Mayeux, Association between mild parkinsonian signs and mild cognitive impairment in a community, Neurology 64 (7) (2005) 1157–1161.
[42] C. Luzzatti, M. Laiacona, D. Agazzi, Multiple patterns of writing disorders in dementia of the Alzheimer type and their evolution, Neuropsychologia 41 (7) (2003) 759–772.
[43] C.D. Manning, P. Raghavan, H. Schütze, Introduction to Information Retrieval, Cambridge University Press, New York, NY, USA, 2008 ISBN:0521865719 9780521865715.
[44] G. Marzinotto, J.C. Rosales., M.A. El-Yacoubi, S. Garcia-Salicetti, C. Kahindo, H. Kerhervé, V. Cristancho-Lacroix, A.S. Rigaud, Age-related evolution patterns in online handwriting, J. Comput. Math. Methods Med. 16 (2016) 1–15.
[45] R. Mergl, P. Tigges, A. Schröter, H.J. Möller, U. Hegerl, Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives, J. Neurosci. Methods 90 (2) (1999) 157–169.
[46] A.J. Mitchell, M. Shiri-Feshki, Rate of progression of mild cognitive impairment to dementia: meta-analysis of 41 robust inception cohort studies, Acta Psychiatrica Scand. 119 (4) (2009) 252–265.
[47] M. Moetesum, I. Siddiqi, N. Vincent, F. Cloppet, Assessing visual attributes of handwriting for prediction of neurological disorders—A case study on Parkinson's disease, Pattern Recognit. Lett. (2018).
[48] H. Mouchère, C. Viard-Gaudin, R. Zanibbi, U. Garain, ICFHR2016 CROHME: competition on recognition of online handwritten mathematical expressions, in: Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition, 2016, pp. 607–612.
[49] B. Moysset, T. Bluche, M. Knibbe, M.F. Benzeghiba, R. Messina, J. Louradour, C. Kermorvant, The A2iA multi-lingual text recognition system at the Maurdor evaluation, in: Proceedings of the International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 297–302.
[50] S. Naz, K. Hayat, M.I. Razzak, M.W. Anwar, S.A. Madani, S.U. Khan, The optical character recognition of Urdu-like cursive scripts, Pattern Recognit. 47 (3) (2014) 1229–1248.
[51] H.T. Nguyen, C.T. Nguyen, P.T. Bao, M. Nakagawa, A database of unconstrained Vietnamese online handwriting and recognition experiments by recurrent neural networks, Pattern Recognit. 78 (2018) 291–306.
[52] M.J.U. Novak, S.J. Tabrizi, Huntington's disease, BMJ. 340 (2010) 30–34.
[53] S. Pekkala, Lexical retrieval in discourse: An early indicator of Alzheimer's dementia, Clin. Linguist. Phon. 27 (12) (2013) 905–921.
[54] J.A. Pittman, Handwriting recognition: tablet PC text input, Computer 40 (9) (2007) 49–54.
[55] R. Plamondon, C. O'Reilly, C. Rémi, T. Duval, The lognormal handwriter: learning, performing, and declining, Front. Psychol. 4 (2013) 248–255.
[56] R. Plamondon, G. Pirlo, É. Anquetil, C. Rémi, H.L. Teulings, M. Nakagawa, Personal digital bodyguards for e-security, e-learning and e-health: a prospective survey, Pattern Recognit. 81 (2018) 633–659.
[57] R. Plamondon, S.N. Srihari, Online and off-line handwriting recognition: a comprehensive survey, IEEE Trans. Pattern Anal. Mach. Intel. 22 (1) (2000) 63–84.
[58] H. Platel, J. Lambert, F. Eustache, B. Cadet, M. Dary, F. Viader, B. Lechevalier, Characteristics and evolution of writing impairment in Alzheimer's disease, Neuropsychologia 31 (1993) 1147–1158.
[59] S.L. Pullman, Spiral analysis: a new technique for measuring tremor with a digitizing tablet, Mov. Disorders 13 (1998) 85–89.
[60] R. Roberts, D.S. Knopman, Classification and epidemiology of *MCI*, Clin. Geriatric Med. 29 (2013) 753–772.
[61] L. Rokach, O. Maimon, Clustering methods, in: Data Mining Knowledge Discovery Handbook, 2005, pp. 321–352.
[62] S. Rosenblum, B. Engel-Yeger, Y. Fogel, Age-related changes in executive control and their relationships with activity performance in handwriting, Hum. Mov. Sci. 32 (2) (2013) 1056–1069.
[63] N. Sakano, M. Notoya, S. Tanabe, N. Sunahara, T. Fujita, K. Nakatani, K. Inoue, A characteristic and the assessment of the writing impairments in subjects with Alzheimer's disease, J. Tsuruma Health Sci. Soc. Kanazawa Univ. 37 (2) (2013) 35–43.
[64] P. Sarkar, G. Nagy, Style consistent classification of isogenous patterns, IEEE Trans. Pattern Anal. Mach. Intel. 27 (1) (2005) 88–98.
[65] T. Schenk, E.U. Walther, N. Mai, Closed- and open-loop handwriting performance in patients with multiple sclerosis, Eur. J. Neurol. 7 (3) (2000) 269–279.
[66] A. Schröter, R. Mergl, K. Bürger, H. Hampel, H.J. Möller, U. Hegerl, Kinematic analysis of handwriting movements in patients with alzheimer's disease, mild cognitive impairment, depression and healthy subjects, Dementia Geriatric Cognit. Disorders 15 (3) (2003) 132–142.
[67] S. Sen, S. Chowdhury, M. Mitra, F. Schwenker, R. Sarkar, K. Roy, A novel segmentation technique for online handwritten Bangla words, Pattern Recognit. Lett. (2018).
[68] MC. Silveri, F. Corda, M. Di Nardo, Central and peripheral aspects of writing disorders in Alzheimer's disease, J. Clin. Exp. Neuropsychol. 29 (2) (2007) 179–186.
[69] J. Sivic, Efficient visual search of videos cast as text retrieval, IEEE Trans. PAMI 31 (4) (2009) 591–605.
[70] M.J. Slavin, J.G. Phillips, J.L. Bradshaw, Visual cues and the handwriting of older adults: a kinematic analysis, Psychol. Aging 11 (3) (1996) 521–526.
[71] M.J. Slavin, J.G. Phillips, J.L. Bradshaw, K.A. Hall, I. Presnell, Consistency of handwriting movements in dementia of the Alzheimer's type: a comparison with Huntington's and Parkinson's diseases, J. Int. Neuropsychol. Soc. 5 (1) (1999) 20–25.
[72] N.M. T, M.M.A. Sultan, K.Y. Wong, A study on the age related retention of individual characteristics in hand writings and signatures for application during forensic investigation, Malaysian J. Forensic Sci. 1 (1) (2010) 54–60.
[73] V. Taler, E. Klepousniotou, N.A. Phillips, Comprehension of lexical ambiguity in healthy aging, mild cognitive impairment, and mild Alzheimer's disease, Neuropsychologia 47 (5) (2009) 1332–1343.
[74] R.R. Tampi, D.J. Tampi, S. Chandran, A. Ghori, M. Durning, Mild cognitive impairment: a comprehensive review, Healthy Aging Res. 4 (39) (2015) 1–11.
[75] H.L. Teulings, JL. Contreras-Vidal, G.E. Stelmach, C.H. Adler, Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control, Exper. Neurol. 146 (1) (1997) 159–170.
[76] H.L. Teulings, G.E Stelmach, Control of stroke size, peak acceleration, and stroke duration in Parkinsonian handwriting, Hum. Mov. Sci. 10 (2) (1991) 315–334.
[77] M. Thomas, A. Lenka, P. Kumar Pal, Handwriting Analysis in Parkinson's Disease: Current Status and Future Directions, Movement Disorders Clinical Practice, 2017.
[78] N. Van Drempt, A. McCluskey, and N.A. Lannin, Handwriting in healthy people aged 65 years and over, Aust. Occup. Therapy J., 58 (4), (2011), 276–286.
[79] C. Viard-Gaudin, P.M. Lallican, S. Knerr, P. Binter, The IRONOFF dual handwriting database, in: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 1999, pp. 455–458.
[80] V. Vuori, Clustering writing styles with a self-organizing map, in: Proceedings of the International Workshop on Frontiers in Handwriting Recognition, 2002, pp. 345–350.
[81] J. Walton, Handwriting changes due to aging and Parkinson's syndrome, Forensic Sci. Int. 88 (3) (1997) 197–214.
[82] P. Werner, S. Rosenblum, G. Bar-On, J. Heinik, A. Korczyn, Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment, J. Gerontol. Psychol. Sci. 61 (4) (2006) 228–236.
[83] Z. Xie, Z. Sun, L. Jin, Z. Feng, S. Zhang, Fully Convolutional Recurrent Network for Handwritten Chinese Text Recognition, 2016. available at https://arxiv.org/pdf/1604.04953.pdf.
[84] J.H. Yan, S. Rountree, P. Massman, R. Smith Doody, H. Li, Alzheimer's disease and mild cognitive impairment deteriorate fine movement control, J. Psychiatric Res. 42 (14) (2008) 1203–1212.
[85] N.Y. Yu, S.H. Chang, Kinematic analyses of graphomotor functions in individuals with Alzheimer's Disease and amnestic mild cognitive impairment, J. Med. Biol. Eng. 36 (3) (2016) 334–343.
[86] X-Y. Zhang, Y. Bengio, C-L. Liu, Online and offline handwritten Chinese character recognition: a comprehensive study and new benchmark, Pattern Recognit. 61 (2017) 348–360.

Mounîm A. EL-YACOUBI is a Professor at Telecom SudParis, University Paris Saclay. He holds a Ph.D. in Signal Processing & Telecommunications from Rennes University. He is specialized in modeling human user data, especially behavioral signals like Handwriting and Gestures, based on Machine Learning, with applications in e-health, and human-computer interaction.

Sonia Garcia-Salicetti is a Professor at Telecom SudParis, University Paris Saclay. She received her Ph.D. in Computer Science from the University of Paris 6, on a Neuro-markovian hybrid model for On-line Handwriting Recognition. Her research interests are Machine Learning, Online Handwriting Analysis, Statistical quality measures, and e-health applications.

Anne-Sophie Rigaud (MD, Ph.D. in Neurosciences) is a geriatrician and psychiatrist. She is a Professor in Geriatrics in University Paris 5. She heads the Geriatric Department at Broca Hospital and the Research Unit EA 4468 including LUSAGE Lab. She has expertise in dementia and gerontechnology.

Christian Kahindo is a Ph.D. student at Telecom SudParis, University Paris Saclay. He got his Master Degree in Data processing and Machine Learning at University of Versailles Saint-Quentin in June 2015, after an Engineering degree at the National School of Applied Sciences (ENSAF) at Fès, Morocco.

Victoria CRISTANCHO LACROIX is a research neuropsychologist and psychologist at hospital Broca. She is Ph.D. in Psychology and M.Sc. in Cognitive Neurosciences, and is Lecturer at University Paris Descartes. She is specialized in Alzheimer's disease cognitive assessment.