IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 23, 2021, accepted May 4, 2021, date of publication May 24, 2021, date of current version June 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3083176

Feature Selection as a Tool to Support the Diagnosis of Cognitive Impairments Through Handwriting Analysis

NICOLE DALIA CILIA[1,2], CLAUDIO DE STEFANO[1], FRANCESCO FONTANELLA[1], AND ALESSANDRA SCOTTO DI FRECA[1]

[1]Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, 03043 Cassino, Italy
[2]Institute for Computing and Information Sciences, Radboud University, 6525 XZ Nijmegen, The Netherlands

Corresponding author: Francesco Fontanella (fontanella@unicas.it)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Cognitive Impairments are cognitive deficits that are greater than expected for a person of a given age and level of education, but which do not significantly interfere with the daily life of the people affected. They range from mild to severe and are seen as a risk factor for Alzheimer's disease, currently the most common neurodegenerative brain disorder worldwide. In a previous study, we presented an experimental protocol comprising different handwriting tasks to be carried out by patients and a healthy control group: the aim was to investigate whether the analysis of the handwriting could be used as a tool to support the diagnosis of this kind of impairment. In the study presented here, we used a well-known and widely-used feature selection approach to determine the most effective features for predicting the symptoms related to cognitive impairments via handwriting analysis. Our intention is to deepen the knowledge about the different cognitive functions affected by the onset of these diseases, as well as to improve the performance of the tools developed to support their diagnosis. The results showed that different sets of highly discriminant features, closely related to the cognitive skills impaired, were selected for the handwriting tasks making up the protocol, thus supporting our hypothesis that their use can be very helpful to support the diagnosis of cognitive impairment.

INDEX TERMS Medical expert systems, cognitive impairments, feature selection.

I. INTRODUCTION

Mild Cognitive Impairment (MCI) (also known as minor neurocognitive disorder), is diagnosed when individuals have cognitive deficits that are greater than those that would be statistically expected for their age and level of education, but which do not significantly interfere with their daily activities. This condition is considered to be the transition state between normal aging and dementia. More generally, cognitive impairments (CI) range from mild to severe. People affected by mild impairment may begin to notice changes in their cognitive functions, but are still able to do their everyday activities. Severe levels affect the understanding of the meaning or importance of events; of things that are said; talking and writing, resulting in the loss of independent living. Although MCI is characterized by a wide variety of symptoms, when memory loss becomes the predominant symptom it is often referred to as "amnestic MCI" (a-MCI) and is frequently seen as a risk factor for Alzheimer's disease (AD) [1]. AD is the most common neurodegenerative brain disorder and typically progresses into severe cognitive impairment and loss of autonomy (i.e. dementia) in old age.

To date, AD is diagnosed by doctors using imaging, blood tests, and lumbar punctures (spinal sampling) amongst others. Unfortunately, there is no remedial cure and early diagnosis would significantly improve the efficacy of available treatments. Recently, researchers have shown that patients affected by AD have altered spatial organization and poor movement control. Therefore, the observation of motor activities should be used in the diagnosis of AD. Handwriting, which is the result of a complex network of cognitive, kinesthetic and perceptive motor skills, can be significantly

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

compromised. For example, in the clinical course of AD, dysgraphia occurs both during the initial phase and progressively [2], [3]. In this framework, many studies have been published in the fields of medicine and psychology, in which standard statistical tools have been used to study the relationship between the disease and the variables used to describe the handwriting of patients [4]–[7]. However, these studies overlook the complex interactions that may occur between multiple features. In many cases, single features weakly correlated to the target class could significantly improve classification accuracy if used together with some complementary features. In contrast, individually relevant features may be redundant when used together with other features.

Currently, researchers in the field of Artificial Intelligence are paying increasing attention to the importance of investigating the characteristics and anomalies of handwriting [8]–[10]. They are committed to modeling the complex interactions between the features extracted from people's handwriting, in order to predict their cognitive state [11]–[17]. The aim is to develop tools that can provide further evidence of cognitive impairment, especially in patients suffering from neurodegenerative diseases (ND). However, to date, there has been no investigation of the effectiveness of the features extracted and the relationship between them and the diseases they may help predict. This study would allow better use of the information that can be extracted from handwriting, and it could be based on the use of well-known and effective feature selection approaches. Such techniques typically use a search strategy to find good solutions (feature subsets) according to a given evaluation function. The methods for defining these functions are generally subdivided into three broad classes; namely filter, wrapper and embedded methods. The first takes into account statistical or geometrical properties of the feature subset space; whereas wrapper ones consider the performance achieved by a given classifier when adopting a feature subset and embedded methods include feature selection as part of the training process.

In [18] we presented an experimental protocol consisting of twenty-five handwriting tasks. These tasks ranged from simple copy tasks, that required a low cognitive load, to more complex ones, that involved high-level cognitive skills. The purpose was to investigate whether and how the wide range of cognitive skills and functions needed to perform the tasks of the protocol are/were affected by ND and CI. We used the protocol to acquire data from about one hundred and seventy people, a much higher number than in the previous studies presented in the literature. From the data we extracted both dynamic and static features such as, for example, the velocity or the total length of the ink trace. In [15], [19] we presented some preliminary results, in which we used a subset of both features and tasks under consideration in the study presented here.

In this paper, we are presenting the results of the feature analysis performed on the data acquired using the protocol described above. In particular, for each task, we tried to understand which were the most effective features for the prediction of the symptoms related to CI and ND. The aim of this analysis is twofold. On one hand, we want to improve the performance of the tools used for prediction. On the other hand, since each task involved different cognitive skills and each feature is related to different aspects of these skills, this analysis would provide doctors with further elements to understand the areas of the brain that are damaged in the early phases of CI and ND. The results showed that each task of the protocol was characterized by a different set of relevant features. This confirms our hypothesis that each of the cognitive abilities tested by our protocol is affected by damage to different areas of the brain, caused by CI or ND.

The remainder of the paper is organized as follows. After a brief overview of the research activities related to our study (Section II), in Section III we describe the protocol we used to acquire the handwriting data. Section IV introduces the problem of defining and selecting features as well as the features used. Section V reports the experimental results, and concluding remarks are discussed in Section VI.

II. RELATED WORK

Once the raw data of the handwriting movements have been collected, the features must be extracted. There are two types: function features and parameter features. The difference between function features and parameter features is that the former are time-dependent, while the latter refer to the entire handwriting movement. The most common function features are: (x, y) coordinates, pen pressure, azimuth, altitude, displacement, velocity and acceleration. Some of these features are directly recorded by the acquisition device, e.g. coordinates and pressure, whereas others are numerically derived. Typically, the most used function features are velocity and acceleration: the former contains information related to the slowness of movements, whereas changes in acceleration allow jerk to be revealed.

The most common parameter features are total duration, absolute size, horizontal size, in-air time, stroke number.

Furthermore, handwriting movements can be subdivided into two categories: "on-paper" and "in-air" [20]. The first are recorded when the pen tip touches the surface on which the person is writing, whereas the second records the movements of the pen tip when it is lifted from the surface, but within a maximum distance from the paper.[1] It has been recently demonstrated that features extracted from in-air movements allow a better characterization of the movements of people affected by AD [5], [21]. Indeed, it has been shown that the total in-air time is related to functional decline, as well as to difficulties in activity planning. In [5] and [21] the authors also found that handwriting fluidity anomalies are much more evident during in-air movements than on-paper ones, observing that these anomalies increase with task complexity, whereas other values (e.g. pressure) remain constant. Moreover, they also found that in copy tasks the in-air time

[1]Note that the features mentioned can be separately extracted both for function features and parameter features

reflects the person's hesitations. Some of the parameter features have been specifically designed with the aim of investigating their relationships with ND [11], [13], [22].

An interesting review on the topic was presented in [23]. Handwriting features can be evaluated at the global or local level. The first are obtained from the entire task, whereas the second are obtained from single strokes. The former implies computations covering all movements performed to execute a given task, whereas the latter requires the analysis at the individual ''stroke'' level. A stroke is defined as the single component making up a handwritten movement, and it is represented by sequence of points recorded by the acquisition device. The number of strokes per second can be used as a measure of handwriting frequency: in [24], for example, the authors observed that AD patients had a significantly lower writing frequency than the control group. Jerk, which typically characterizes the handwriting of PD patients, can be measured in terms of changes in acceleration over time and is often used in conjunction with changes in velocity. In the literature, it has been also observed that features like entropy and energy can be used to characterize handwriting ''noise'', i.e. the randomness of movements typically caused by tremor and irregular muscle contractions [25]. In [26], the authors introduced a metric based on the velocity variability: the observations that low-level control of the muscular systems occurs in terms of milliseconds, while the control of conscious movements cannot be at the same frequency. Time-varying (x, y) coordinates can also be seen as a signal to be processed. In [22], for example, the authors decomposed handwriting movements into a small, finite number of components that can be processed using well-known frequency analysis techniques. However, to date, most of these techniques have not been well investigated. A very successful theory for modelling handwriting movements is that developed by Plamondon [27], [28]. This model has found many applications, including, amongst others, the early diagnosis of ND [12], [20], [29].

From the brief literature review outlined, it can be noted that there is no study devoted to the analysis of the features extracted from the handwriting of ND and CI patients, and in particular to the complex interactions that may occur among multiple features. Most of the studies presented only investigated the relationship between these diseases and each of the considered features, overlooking the complex interactions that may occur among multiple features [30].

III. THE ACQUISITION PROTOCOL

As mentioned in the Introduction, we have defined a protocol for the acquisition of data related to handwriting movements, both from patients affected by CI and a healthy control group. The protocol includes twenty five tasks, belonging to the following categories (see Tables 1, 2 and 3):

- Graphic tasks: to test the patient's ability in writing elementary traits, joining some points and drawing geometrical figures (simple or complex);
- Copy and Reverse Copy tasks: to evaluate patient's abilities in repeating complex graphic gestures, which have semantic meaning such as letters, words and numbers;
- Memory tasks: to test the changes in the writing process of words previously memorized or associated with objects shown in a picture;
- Dictation tasks: to investigate how handwriting varies when the working memory is used.

It is worth noting that each task was designed with the aim of testing either function features or parameter features.

Note that task #17, in which the person is asked to write six different words, has been used in two different ways: in the first, we averaged feature values over the entire set

TABLE 1. Graphic tasks.

#	Description
2	Join two points with a horizontal line, continuously for four times
3	Join two points with a vertical line, continuously for four times
4	Retrace a circle (6 cm of diameter) continuously for four times
5	Retrace a circle (3 cm of diameter) continuously for four times
21	Retrace a complex form
24	Draw a clock, with all hours and put hands at 11:05 (Clock Drawing Test)

TABLE 2. Copy and reverse copy tasks.

#	Description
6	Copy the letters 'l', 'm' and 'p'
7	Copy the letters on the adjacent rows
8	Write exactly four joined lowercase cursive letter 'l', in a single smooth movement
9	Write exactly four joined lowercase cursive bigram 'le', in a single smooth movement
10	Copy the word "foglio"
11	Copy the word "foglio" above a line
12	Copy the word "mamma"
13	Copy the word "mamma" above a line
15	Copy in reverse the word "bottiglia"
16	Copy in reverse the word "casa"
17	Copy six words (regular, non regular, non words) in the appropriate boxes
19	Copy the fields of a postal order
22	Copy a telephone number
25	Copy a paragraph
26	copy the word "pane" in the appropriate boxe
27	copy the word "mela" in the appropriate boxe
28	copy the word "prosciutto" in the appropriate boxe
29	copy the word "ciliegia" in the appropriate boxe
30	copy the word "lonfo" in the appropriate boxe
31	copy the word "taganaccio" in the appropriate boxe

TABLE 3. Memory and dictation tasks.

#	Description
1	Signature drawing
14	Memorize the words "telefono", "cane", and "negozio" and rewrite them
18	Write the name of the object shown in a picture (a chair)
20	Write a simple sentence under dictation
23	Write a telephone number under dictation
32	Memorize the word "telefono" and rewrite it
33	Memorize the word "cane" and rewrite it
34	Memorize the word "negozio" and rewrite it

of words. In the second, we averaged feature values over each single word: this means that we split task #17 into six further tasks (from #26 to #31). The same occurred in task #14, in which the person is asked to memorize and rewrite the words ‘‘telefono’’ (telephone), ‘‘cane’’ (dog) and ‘‘negozio’’ (shop) (tasks from #32 to #34). The rationale behind this choice is to evaluate the effects of fatigue, i.e. to assess whether writing performance degrades more rapidly in subjects affected by neurodegenerative diseases when they have to write several words consecutively.

For recruitment to the study we used standard clinical tests, such as the Mini-Mental State Examination (MMSE) [31], the Frontal Assessment Battery (FAB) [32], the Montreal Cognitive Assessment (MoCA) [33]. In these tests, the cognitive abilities of the subject are assessed using questionnaires covering many areas, ranging from orientation in time and place to registration recall. To avoid any bias, we chose the people forming the control group in such a way as to match the patient group in terms of average age, level of education, gender and type of work (manual or intellectual) as shown in Table 5. Finally, from both groups, we excluded people taking psychotropic medication, or any other drugs influencing their cognitive abilities. Note that patients were referred to the study by medical experts, excluding patients whose cognitive abilities were too compromised.

As acquisition tool we used a Wacom Bamboo Folio smartpad, equipped with a pen that allowed participants to normally write ink traces on A4 white paper sheets placed on it. For each task, the smartpad recorded the x-y coordinates of pen movements (at a frequency of 200Hz) on the plane represented by the paper sheet surface. The smartpad also recorded the pressure exerted by participant when the pen tip was touching the sheet as well as the ‘‘in-air’’ movements, i.e. the pen tip x-y coordinates when it was lifted from the sheet, within a maximum distance of 3cm. During the acquisition the smartpad was positioned about seventy centimeters from the participant. Note that the participants had the same conditions during the acquisition.

IV. FEATURE EXTRACTION AND SELECTION

The following subsections detail the features extracted from the data acquired by using the protocol described above, and the feature selection technique we used to find the most discriminative features for each task.

A. FEATURE EXTRACTION

The features extracted from the raw data available, i.e. (x, y) coordinates, pressure and timestamps, were calculated on the strokes [2] making up the handwritten traits and then averaged over the entire task. Our goal is to describe, for each task, the behavior of a subject taking into account a fixed number of features. Since the number of strokes varies strongly from subject to subject and from task to task, we have averaged the values extracted from each single stroke on the total number of strokes. We extracted both static and dynamic features. The first are computed taking into account the shape or the position of the strokes, whereas the second are related to quantities like velocity and acceleration. Table 4 shows the list of the extracted features.

As many studies in the literature show significant differences in patients’motor performance between in-air and on-paper traits, each feature was calculated separately for the in-air or on-paper traits. In particular, we extracted four groups of features:

- On-paper: the features extracted from the written traits (i.e. during pen-down and the successive pen-up). Note that in this case each sample was represented by twenty six features. These features are described as ’P’ in the following;
- In-air: the features extracted from the in-air traits (i.e. those acquired by the system when the pen is lifted from the sheet, within the maximum acquisition distance of 3 cm). These movements characterize the planning activity for positioning the pen tip between two successive written traits. Note that in this case we extracted twenty five features because pressure (feature #21) is always zero. These features are described as ’A’ in the following;
- All: in this case, each sample is represented by a feature vector containing both in-air and on-paper features, i.e. by adding the vector of in-air features to the vector of on-paper one. The aim was twofold. On one hand, we wanted to perform a direct comparison between in-air and on-paper features. On the other hand, we wanted to investigate the interactions between in-air and on-paper features. Note that in this case the total number of features extracted was forty-seven (personal features and pressure were not repeated). These features are described in the following as ‘‘AL’’;
- In-air and on-paper: these features are computed without distinguishing between in-air and on-paper traits. In practice, for each task, each of the twenty six features listed in Table 4 was extracted averaging the values on both in-air and on-paper traits. This represented a different way of providing all the information to the classification systems. Moreover, they also allowed us to assess the effectiveness of the split between in-air and on-paper features. These features are described in the following as ‘‘AP’’.

In order to take into account the differences due to age, education or work, we have also added the following ‘‘personal’’ features: gender, age, type of work, and level of education.

Summarizing, we used four groups of features, each represented by thirty-four files (one for each task), each containing one hundred and eighty samples, each made of a number dependent on the feature type. In particular, we extracted twenty-six features (see Table 4) both for the on-paper (P)

[2] A stroke is defined as the single component making up a handwritten trait, and it is represented by the sequence of points between two consecutive segmentation points. We considered as segmentation points: pen-up, pen-down and zero-crossing velocity along the y-axis.

TABLE 4. Feature list. Feature types are: dynamic (D), static (S), and personal (P).

#	Name	Description	Type
1	Duration	Time interval between the first and the last points in a stroke	D
2	Start Vertical Position	Vertical start position relative to the lower edge of the active digitizer area	S
3	Vertical Size	Difference between the highest and lowest y coordinates of the stroke	S
4	Peak vertical velocity	Maximum value of vertical velocity among the points of the stroke	D
5	Peak vertical acceleration	Maximum value of vertical acceleration among the points of the stroke	D
6	Start horizontal position	Horizontal start position relative to the lower edge of the active tablet area	S
7	Horizontal size	Difference between the highest (rightmost) and lowest (leftmost) x coordinates of the stroke	S
8	Straightness error	It is calculated estimating the length of the straight line, fitting the straight line, estimating the (perpendicular) distances of each point to the fitted line, estimating the standard deviation of the distances and dividing it by the length of the line between beginning and end	D
9	Slant	Direction from the beginning point to endpoint of the stroke, in radiant	S
10	Loop Surface	Area of the loop enclosed by the previous and the present stroke	S
11	Relative initial slant	Departure of the direction during the first 80 ms to the slant of the entire stroke.	D
12	Relative time to peak vertical velocity	Ratio of the time duration at which the maximum peak velocity occurs (from the start time) to the total duration	D
13	Absolute size	Calculated from the vertical and horizontal sizes	S
14	Average absolute velocity	Average absolute velocity computed across all the samples of the stroke	D
15	Road length	length of a stroke from beginning to end, dimensionless	S
16	Absolute y jerk	The root mean square (RMS) value of the absolute jerk along the vertical direction, across all points of the stroke	D
17	Normalized y jerk	Dimensionless as it is normalized for stroke duration and size	D
18	Absolute jerk	The Root Mean Square (RMS) value of the absolute jerk across all points of the stroke	D
19	Normalized jerk	Dimensionless as it is normalized for stroke duration and size	D
20	Number of peak acceleration points	Number of acceleration peaks both up-going and down-going in the stroke	S
21	Pen pressure	Average pen pressure computed over the points of the stroke	D
22	#strokes	Total number of strokes of the task	S
23	Sex	Subject's gender	P
24	Age	Subject's age	P
25	Work	Type of work of the subject (intellectual or manual)	P
26	Instruction	Subject's education level, expressed in years	P

and in-air and on-paper (AP) categories, twenty-five in-air features (A) (because pen pressure was not used), and forty-four for the category All (AL).

B. FEATURE SELECTION

To find the features that allow better discrimination between the patients affected by CI and the control group, we used a well-known feature selection technique based on a wrapper evaluation function, named recursive feature elimination (RFE in the following) [34]. RFE performs a greedy search to find the best performing feature subset, based on the backward elimination strategy. Starting from the whole set of available features, the RFE algorithm iteratively creates models and determines the worst-performing feature at each iteration. Then, it builds the subsequent models with the features leftover until all the features are explored. If the data contain N features, in the worst case RFE evaluates N^2 subsets. The algorithm provides as output the feature subset providing the best performance among those tested. As an evaluation function we used the accuracy, computed by using the K-fold cross-validation technique, achieved by using the xgboost classifier. Note that the RFE iterative algorithm tends to select features that are few correlated with each other. To explain the mechanism that allows RFE to select few correlated features let us give an example. Let f_1 and f_2 be two correlated features in S_i, the subset of features left over at the i–th iteration; the algorithm tests both $S' = \{S_i - \{f_1\}\}$ and $S'' = \{S_i - \{f_2\}\}$. Since f_1 and f_2 are correlated, S' and S'' achieve very similar performance. As a consequence, most probably RFE will exclude one of them from the subset S_{i+1}.

TABLE 5. Average demographic data of participants. Standard deviations are shown in parentheses.

	Age	Education	#Women	#Men
Patients	71.5 (9.5)	10.8 (5.1)	46	44
Control group	68.9 (12)	12.9 (4.4)	51	39

The data used for the experiments reported in Section V were obtained from one hundred eighty people. Ninety one of them (forty six females and forty five males) were patients with different levels of cognitive impairments, whereas the remaining eighty nine were healthy control group (fifty females and thirty nine males).

V. EXPERIMENTAL FINDINGS

In order to perform the feature analysis (the objective of this study), we performed three sets of experiments, using the data described in Section III. In the first, we assessed the importance of each feature across the thirty-four tasks of the protocol. In the second, we assessed the effectiveness of the feature selection procedure used and verified if one of the four feature categories outperformed the others. Finally, in the third set, we investigated the relationship between the selected features and the cognitive functions and skills involved.

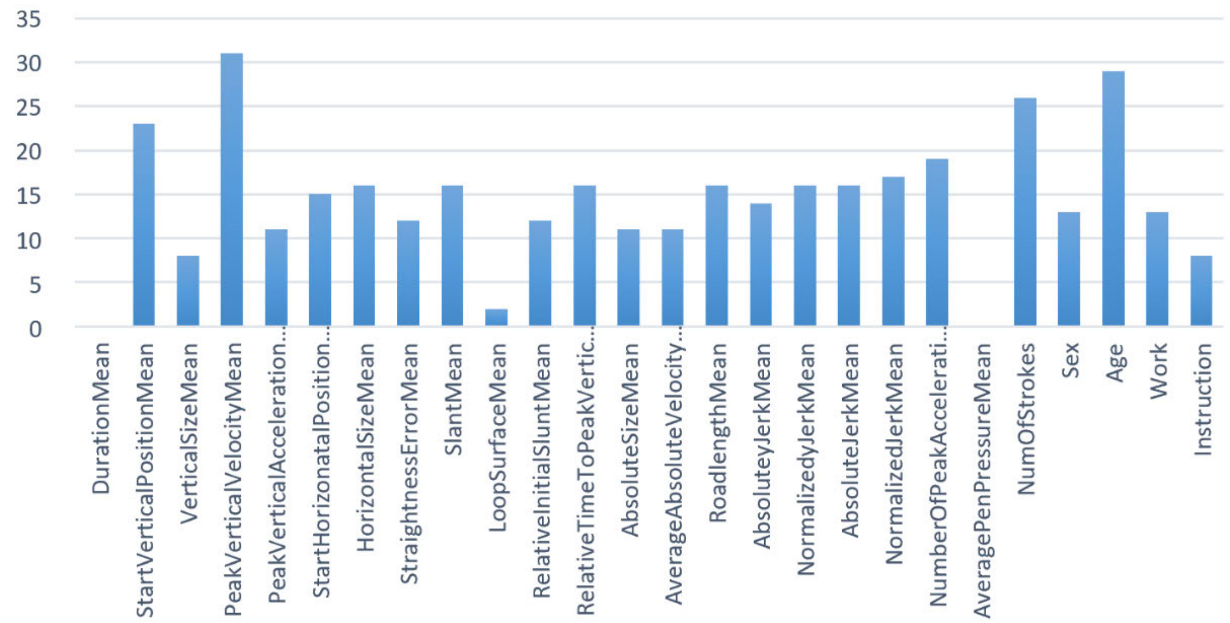

FIGURE 1. Histogram for the in-air features selected using the RFE algorithm. The average number of features selected, computed on the thirty four tasks, is 10.41, with a standard deviation equal to 6.81.

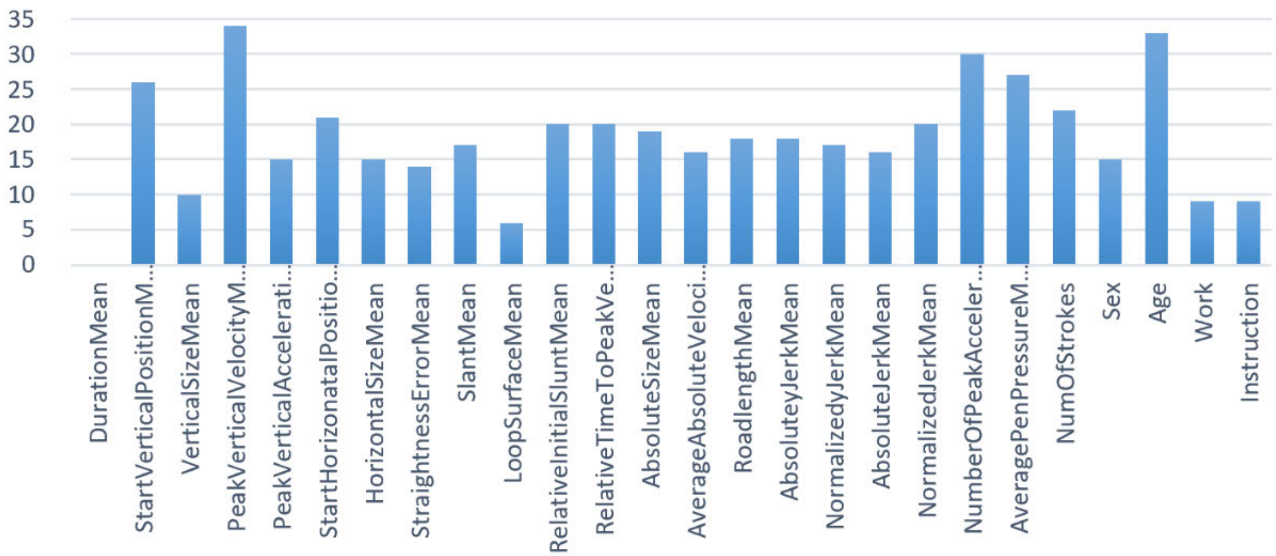

FIGURE 2. Histogram for the on-paper features selected using the RFE algorithm. The average number of features selected, computed on the thirty four tasks, is 13.15, with a standard deviation equal to 6.57.

For all the experiments reported in the following we used Scikit-learn, an open-source machine learning library [35]. The results reported have been achieved after a hyper-parameter optimization step. We used the grid search procedure provided by Scikit-learn. Once a set of values has been defined for each of the parameters to be tuned, this procedure exhaustively tests all parameter combinations. The set of values tested for each of the hyper-parameter tuned is shown in Table 6.

The experiments performed are detailed in the following subsections.

A. FEATURE EVALUATION

In this first set of experiments we tried to answer the following question: among the features we extracted, are there some that are more important than others (useful for many or even most of the tasks of the protocol)? To answer this, for each feature category, we plotted a histogram reporting how many times each feature was selected by RFE across the thirty-four tasks. These histograms are shown in Fig. 1-4. From the figures we can observe that the four histograms show the same trend, with few peaks and valleys, and most of the features selected in a range between ten and

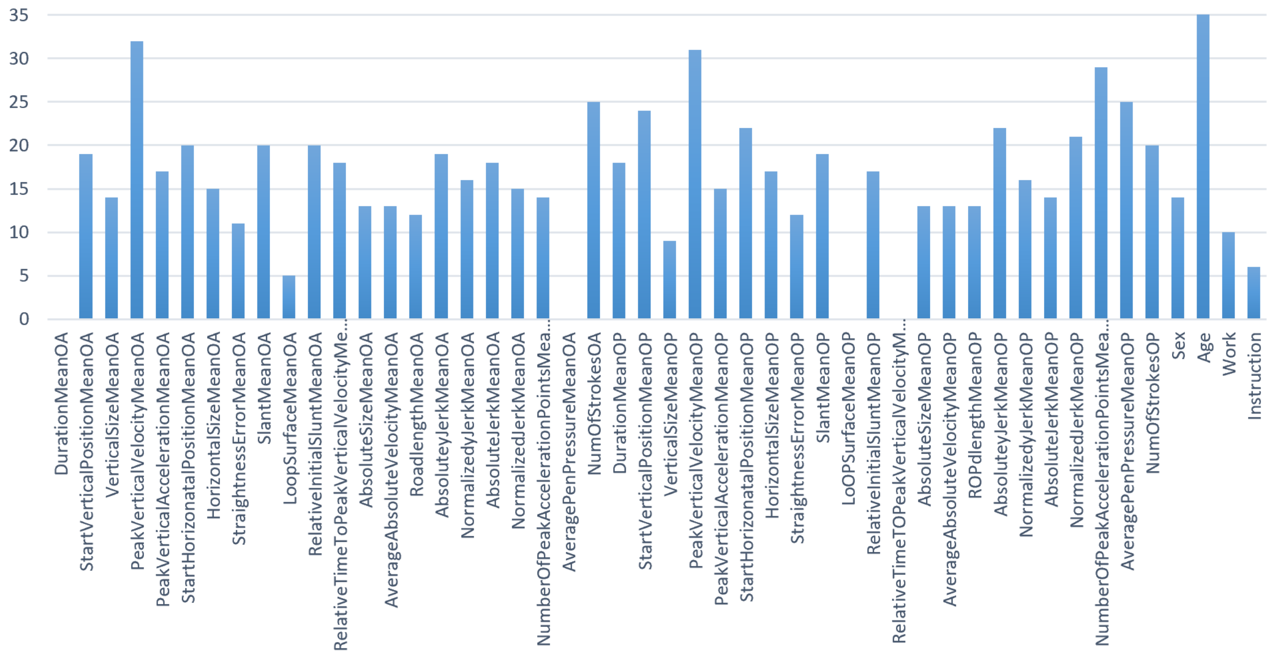

FIGURE 3. Histogram for the all features selected using the RFE algorithm. The average number of features selected, computed on the thirty four tasks, is 23.50, with a standard deviation equal to 7.53.

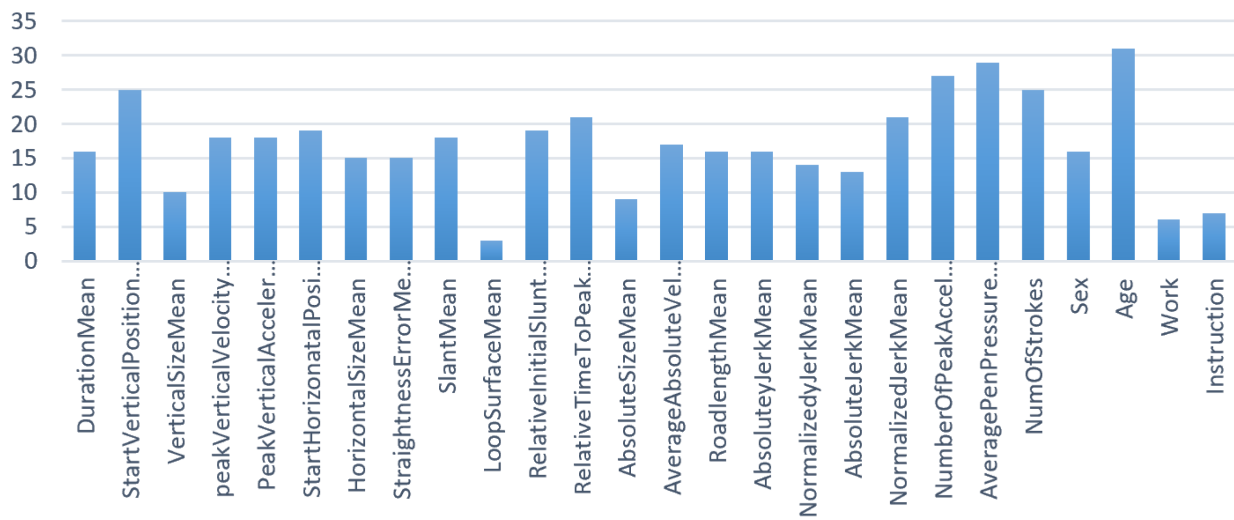

FIGURE 4. Histogram for the in-air and on-paper features selected using the RFE algorithm. The average number of features selected, computed on the thirty four tasks, is 16.65, with a standard deviation equal to 5.42.

twenty. This trend confirms our hypothesis that the features that best distinguish the handwriting of the CI patients from that of healthy people vary across the tasks. In the peaks, age is common among the four categories of features, confirming, in this case, that age affects handwriting processes and must be taken into account. A further feature mostly selected is the peak vertical velocity, confirming, in this case, the importance of velocity as a discriminating feature, both for the in-air and on-paper traits. Duration and loop surface, instead, were the least selected features. Most probably, the first was not selected because it is correlated with peak velocity, which is more discriminating. Therefore, according to the rationale of the RFE iterative algorithm (see Section IV), a feature was not selected if it correlated with previously selected ones. The non-selection of the loop surface, its non-selection confirmed that handwriting size is not affected by CI.

For the in-air features, from the histogram (Fig. 1) it can be noted that the feature reporting the number of strokes was selected twenty-six times. This value confirms that the in-air movements of CI patients are longer and typically

TABLE 6. Values of the classifier hyper-parameters used in the experiments.

Classifier	Parameter	Values tested
XGB		
	min child weight	1, 5, 10
	gamma	0.5, 1, 1.5, 2, 5
	subsample	0.6, 0.8, 1
	colsample bytree	0.6, 0.8, 1
	max depth	3, 4, 5
RF		
	max depth	10, 20, 50, 100, None
	min samples leaf	1, 2, 4
	min samples split	2, 5, 10
	#tree	100, 200, 500, 1000, 2000
DT		
	criterion	gini,entropy
	min samples split	2, 10, 20
	max depth	None, 2, 5, 10
	min samples leaf	1, 5, 10
	max leaf nodes	0, 5, 10, 20
SVM		
	C	0.1, 1, 10, 100, 1000
	gamma	1, 0.1, 0.01, 0.001, 0.0001
	kernel	rbf
MLP		
	hidden layer sizes	50, 100, 200
	activation	tanh, relu
	solver	lbfgs,sgd
	alpha	0.0001, 0.05
	learning rate	constant,adaptive

more unsteady than those of the control group, as reported in previous studies [5], [11].

Regarding the on-paper features, from the histogram (Fig. 2) it can be seen that the most selected features, apart from the age and the peak of vertical velocity discussed above, are: start vertical position, the number of peaks of acceleration, and pressure. The first most probably reveals the spatial orientation difficulties of CI patients, whereas the second and the third features indicate that the handwriting of CI patients is also unsteady during the on-paper movements.

As for the all features, from the histogram (Fig. 3), first of all we can observe that there is no "predominance" of one category over the others; this confirms that in-air and on-paper features contribute equally in distinguishing CI patients from the control group. It is also worth noting that, in this case, there was a different interaction among the whole set of features, both in-air and on-paper. The effect of this interaction is that features rarely or never selected in the previous two cases were selected several times. For example, duration on-paper which was never selected in the previous case, was selected eighteen times. This indicates that this feature, interacting with one or more in-air features, allows the achievement of good classification performance. The histogram also suggests the comparison between each couple of in-air and on-paper features, to understand which kind of feature is more useful. For example, for the number of strokes, in-air are more selected than on-paper, confirming that in-air movements of CI patients are more unsteady than those of the control group.

Finally, in the case of in-air and on-paper features, i.e. those computed without distinguishing between in-air and on-paper traits, from the histogram shown in Fig. 4 we can observe the phenomenon mentioned above: a feature never previously selected was selected a significant number of times. This is the case of the duration mean feature, selected here in sixteen tasks. It is worth noting that most probably this feature is correlated with the peak vertical velocity. Since this last feature was selected many times for the in-air and on-paper cases, this excludes the selection of the duration feature. In this last case we can observe that the peak velocity was selected far less than the previous cases. Therefore we can conclude that when there is no distinction between the in-air and on-paper features, stroke duration can be more discriminating than the peak velocity feature.

From the histogram, we can also note that, apart from age, the most selected features (more than twenty times) were: start vertical position, number of strokes (twenty-five), number of peak acceleration points, pressure, and number of strokes. This confirms that the handwriting of CI patients is unsteady (number of strokes and of peak acceleration points) they have spatial orientation difficulties (start vertical position) and tend to exert more pressure on the surface used to write on (paper sheet in our case).

B. FEATURE SELECTION EVALUATION

In this second set of experiments we tried to answer the following questions: does the feature selection procedure used (RFE) allow us to distinguish CI patients better from the control group? Among the four categories of features considered, is there one that outperforms the others? We therefore compared the performance achieved with and without the RFE algorithm, by using three well-known and widely used classifiers, namely xgboost (XGB in the following) [36], random forest (RF) [37] and decision tree (DT) [38]. It is worth noting that the first two techniques build ensembles of classifiers and typically allow the achievement of good performance, whereas the third provides models that allow the explainability of the results. In order to provide an overview of the classification performance obtained, Tables 7, 8 and 9 show the accuracies (computed by using the 10-fold cross-validation strategy) achieved by the classifiers used, with the various feature categories considered, with (RFE columns) or without feature selection (NO columns). For each table, the last rows report the average (avg) and standard deviation (std), computed over the thirty-four tasks whereas for each task the best result is in bold. From the tables we can observe that, for a given task, results may vary a lot across the different feature types and/or classifiers, confirming that the features used represent the handwriting data in different ways, which can be more or less effective for each classifier. However, taking into account the entire set of tasks, they achieved similar results, as can be seen from the last two rows.

With the aim of summarizing these results, Table 10 reports the best results achieved for each task. In particular, the table

TABLE 7. Classification results achieved by the XGB classifier.

	A		P		AP		AL	
Task	NO	RFE	NO	RFE	NO	RFE	NO	RFE
1	**88.7**	83.0	84.2	87.1	85.2	83.1	83.8	84.9
2	84.3	**90.3**	85.3	83.1	84.6	81.6	87.5	80.7
3	86.9	85.3	83.8	87.5	81.0	**87.6**	83.9	86.8
4	82.5	84.5	83.0	83.0	82.2	**86.9**	83.1	83.3
5	83.6	**87.5**	85.3	83.8	79.6	83.9	81.8	84.6
6	85.1	**87.4**	83.8	83.1	82.5	80.3	82.5	83.1
7	84.8	87.0	83.9	**88.3**	87.7	83.3	83.3	80.4
8	79.3	85.4	**86.0**	81.6	85.3	83.8	83.1	**86.0**
9	**88.4**	79.7	86.7	85.2	85.3	87.5	80.9	83.8
10	86.6	**90.2**	86.0	83.8	80.9	83.1	81.6	78.7
11	80.7	84.7	**86.8**	84.6	77.9	84.6	84.6	84.6
12	81.2	**86.8**	85.9	83.7	83.0	83.3	85.2	83.0
13	**87.3**	85.2	**87.3**	81.3	86.6	83.3	85.1	85.8
14	83.1	82.3	82.3	83.9	79.6	**86.4**	84.9	80.8
15	82.0	**87.2**	86.3	85.5	81.5	85.9	83.0	85.4
16	85.3	84.5	82.4	84.6	85.3	84.6	83.1	**87.6**
17	82.2	84.7	85.9	83.0	83.0	**89.6**	88.9	85.9
18	83.1	84.8	84.6	86.8	84.6	86.0	84.6	**89.0**
19	89.7	89.7	87.0	85.2	88.8	91.4	87.9	**91.7**
20	85.3	83.1	80.8	82.3	84.7	84.7	**85.5**	83.9
21	80.0	**88.2**	81.3	85.4	82.0	82.8	84.4	85.9
22	84.0	**89.6**	87.8	87.8	82.4	86.3	80.9	86.3
23	85.4	83.1	84.7	**88.6**	85.5	84.7	83.2	**88.6**
24	**85.4**	82.3	83.1	83.9	83.9	81.5	80.8	83.1
25	83.6	84.0	80.5	83.6	**85.2**	84.7	76.0	80.5
26	81.9	81.9	78.6	**85.7**	83.3	84.9	79.4	83.3
27	79.3	**87.9**	82.8	85.9	80.5	81.3	77.3	82.0
28	84.1	**87.9**	85.4	87.5	82.3	86.3	83.9	**87.9**
29	86.3	81.4	85.9	83.7	84.4	**91.0**	80.0	89.2
30	85.1	80.9	84.6	82.4	**85.3**	78.7	80.9	84.6
31	83.9	85.4	84.4	87.4	**88.2**	86.7	82.2	83.9
32	82.5	84.1	89.0	89.0	83.1	86.8	**91.2**	82.5
33	85.0	83.3	86.0	**86.8**	80.9	80.2	**86.8**	82.4
34	82.7	**87.3**	86.0	86.1	83.8	84.0	78.7	79.4
Avg	84.1	85.3	84.6	85.0	83.5	84.7	83.2	84.4
Std	2.6	2.8	2.3	2.1	2.5	2.9	3.2	3.0

TABLE 8. Classification results achieved by the RF classifier.

	A		P		AP		AL	
Task	NO	RFE	NO	RFE	NO	RFE	NO	RFE
1	80.9	**87.2**	82.0	82.0	81.0	82.4	78.2	79.1
2	85.1	**87.3**	82.4	83.1	80.9	80.2	82.4	83.0
3	82.0	83.6	81.6	80.9	83.2	83.2	81.0	**86.0**
4	78.6	77.7	**82.0**	**82.0**	78.0	78.7	72.9	75.0
5	82.8	**87.5**	81.6	79.4	79.6	81.0	79.6	81.6
6	84.3	**87.4**	84.6	**87.5**	82.5	81.8	80.3	82.4
7	84.8	82.6	83.2	86.9	**87.7**	**87.7**	80.4	81.9
8	78.4	82.6	81.6	**87.5**	85.3	81.6	80.9	83.8
9	78.6	77.0	83.0	**85.2**	82.4	83.8	80.9	80.2
10	87.5	**88.4**	77.2	80.2	79.4	86.8	77.9	79.4
11	**86.3**	81.9	80.9	83.8	83.1	81.6	76.5	86.0
12	84.6	83.3	84.4	83.0	80.7	**85.4**	80.7	82.2
13	84.1	79.7	82.1	**87.3**	83.6	85.4	82.1	85.1
14	83.1	80.8	84.6	84.6	84.1	83.3	80.3	**87.7**
15	87.2	**88.7**	82.4	82.4	85.2	83.7	85.2	85.4
16	**89.2**	84.5	76.5	83.8	82.4	86.0	81.6	84.5
17	82.2	84.7	85.9	85.2	85.9	81.5	83.7	**86.7**
18	84.8	83.1	86.8	**87.5**	84.6	83.1	82.4	82.4
19	84.5	88.8	81.5	87.0	86.2	83.6	85.3	**88.9**
20	**85.3**	82.3	83.9	80.8	83.2	81.7	80.9	81.5
21	**86.7**	86.1	82.8	86.1	82.0	83.6	78.1	82.8
22	84.0	86.1	82.4	**87.8**	84.7	84.7	81.7	87.0
23	77.7	84.6	79.4	**87.0**	83.2	80.9	84.7	84.0
24	**85.4**	81.5	80.8	82.3	83.1	80.0	**85.4**	84.6
25	82.8	81.3	83.6	81.3	82.8	**85.4**	84.5	82.8
26	81.0	**85.3**	81.0	80.2	81.0	84.9	80.2	81.0
27	**83.6**	81.0	82.8	80.5	76.6	80.5	79.7	82.0
28	86.9	86.9	82.1	84.7	85.5	82.3	80.7	**87.9**
29	83.8	80.5	81.5	85.2	80.0	81.3	80.0	**86.7**
30	81.9	83.0	77.9	80.2	**83.8**	80.9	80.2	82.4
31	83.1	84.0	85.2	78.5	82.2	83.0	83.0	**84.7**
32	83.3	85.7	81.6	**86.8**	84.6	**86.8**	83.8	84.1
33	85.8	**88.2**	80.2	87.5	80.9	85.3	82.4	84.6
34	81.2	82.8	83.8	85.4	86.0	86.1	82.4	**87.5**
Avg	83.6	84.0	82.2	83.9	82.8	83.2	81.2	83.7
Std	2.7	3.1	2.2	2.9	2.5	2.2	2.6	2.9

reports the classifier, the feature category, the accuracy and whether this result was achieved using feature selection. From this table we can see that the RFE algorithm used for feature selection achieved better results on twenty-six out of the thirty-four tasks considered, i.e. more than 75% of the cases, confirming the effectiveness of feature selection. As far as the remaining tasks are concerned, namely signature (#1), copying the word "foglio" above a line (#11), copying in reverse the word "casa" (#16), clock drawing test (#24), copying a paragraph (#25), copying an irregular word in the appropriate box (#30 and #31) and memorizing and rewriting a word (#32), most of them require more complex cognitive functions. Most probably, this makes it necessary to use all features to model the way CIs affect these functions. From Table 10 we can also see that XGB is the most performing classifier (twenty times out of thirty-four). For feature types, A and P are the most performing ones, both achieving the best results on eleven tasks, whereas AP and AL both outperform the others on six tasks.

In order to investigate whether a given feature type is best suited for a given category of task, we calculated how many times a given feature type achieved the best results for each task category. These results are shown in Table 11. Note that we considered tasks #26 to #31 as belonging to the copy and reverse copy category because they were split from task #17, and tasks #32 to #34 as belonging to the memory and dictation category because they were split from task #14. To check if for a feature category (A,P, AP or AL) a type of features (dynamic or static, see Table 4) is predominant, Table 11 also shows the percentage of dynamic features selected with respect to the total number of features selected (the percentage of static features is obviously the complement). The comments for each category are reported in the following.

1) GRAPHIC TASKS

From Table 11 we can observe that the in-air features (A) were the best-performing for graphic tasks (tasks: #2, #5, #21), whereas on-paper (P) features never outperformed the others. Although this result seems counterintuitive, because the execution of this kind of task requires the pen tip to be on the paper most of the time, it indicates that in-air traits, even if they are short, contain most of the anomalies of the handwriting of CI patients. On the other hand, the AP (in-air and on-paper) features outperformed the others in tasks #3 and #4 (circle drawings), indicating that in these cases, anomalies are

TABLE 9. Classification results achieved by the DT classifier.

	A		P		AP		AL	
Task	NO	RFE	NO	RFE	NO	RFE	NO	RFE
1	84.4	**87.2**	80.6	78.4	80.3	76.8	82.4	77.0
2	79.1	84.3	**84.6**	79.4	80.9	83.1	76.5	80.7
3	80.3	74.6	82.4	80.2	78.8	77.4	**82.5**	78.7
4	76.7	79.6	**85.0**	80.0	79.7	82.0	78.0	70.2
5	79.5	**85.4**	80.9	80.2	80.3	80.3	75.2	73.5
6	**83.6**	80.7	82.4	79.4	77.4	77.4	77.4	81.6
7	**87.0**	83.3	78.8	82.5	81.9	84.1	81.2	80.4
8	**80.2**	79.9	76.5	76.5	73.5	78.7	79.4	**80.2**
9	73.2	79.7	81.5	**89.6**	83.8	85.3	78.7	80.2
10	**85.7**	**85.7**	81.6	79.4	83.1	84.6	80.2	80.9
11	79.8	**84.7**	80.9	79.4	78.7	80.9	77.9	80.2
12	77.8	**88.2**	80.7	80.7	77.0	80.6	83.7	74.1
13	79.4	**83.6**	82.8	79.9	82.1	80.6	76.1	79.9
14	80.8	**82.3**	82.3	80.0	79.6	81.8	75.0	79.2
15	79.0	81.2	79.4	80.2	80.7	**84.4**	74.8	79.2
16	**84.5**	79.1	77.2	80.2	81.6	75.7	81.6	78.3
17	76.3	81.3	80.7	79.3	82.2	82.2	**85.2**	83.7
18	**84.8**	80.5	77.2	81.6	79.4	83.1	75.0	83.8
19	87.9	88.8	86.1	82.4	87.9	86.2	**90.5**	89.6
20	83.0	76.9	80.0	**87.7**	81.7	77.1	79.4	80.0
21	75.0	**86.8**	82.0	84.0	78.1	77.3	86.7	78.9
22	83.2	**87.5**	79.4	84.0	80.2	84.0	83.2	76.3
23	80.0	79.2	73.3	82.4	78.6	**83.2**	80.9	82.4
24	83.1	83.1	79.2	78.5	84.6	76.9	**86.2**	84.6
25	80.5	83.3	**85.9**	80.5	81.3	82.6	77.5	77.3
26	80.2	**81.9**	74.6	80.2	80.2	81.8	81.0	72.2
27	81.9	78.5	**82.8**	76.6	85.2	75.8	78.1	75.8
28	81.3	77.6	84.6	**88.9**	78.2	82.3	82.3	85.5
29	82.1	78.8	80.0	80.0	81.5	**86.1**	75.6	84.2
30	83.0	80.9	77.2	79.4	80.9	79.4	**83.8**	78.7
31	78.2	**86.8**	80.7	80.0	83.7	83.7	76.3	82.3
32	83.3	85.7	82.4	**89.0**	77.9	86.8	75.7	81.0
33	78.8	**81.9**	80.2	77.9	77.2	77.2	77.9	75.0
34	75.2	77.6	84.6	**86.8**	79.4	84.0	83.8	84.6
Avg	80.8	82.3	80.8	81.3	80.5	81.3	80.0	79.7
Std	3.4	3.6	3.0	3.4	2.7	3.3	3.9	4.0

present in both traits. Finally, in the AL category, i.e. where both in-air and on-paper features were used to represent each sample, the best result was achieved in the clock drawing test, confirming that in this case, the interactions between in-air and on-paper features allow an effective modeling of the anomalies of CI patients.

From Table 11, we can also observe that for A and AL features the percentage of dynamic features (values in parentheses) is about 50%, whereas for AP features there is a significant higher percentage for the dynamic features (66%). This result suggests that in this last case anomalies are more present in the dynamics of the handwriting than in its shape.

2) COPY AND REVERSE COPY TASKS

In the copy and reverse copy tasks, from Table 11 we can see that on-paper (P) and in-air (A) features achieved the best result in nine and six (out of twenty) tasks, respectively. In particular, on-paper features achieved the best result in the tasks requiring the copying of: single letters (#6 and #7), joined bigrams (#8 and #9), words above a line (#11 and #13), a paragraph (#25), and regular (#26) and irregular (#28) words in boxes. This confirms that on-paper movements are altered in CI patients when they perform simple movements, not involving a high cognitive load, or movements requiring spatial organization, as is the case of handwriting following a cue (horizontal lines or boxes). In-air movements achieved the best results in tasks requiring the copying of simple words without any cue (#10 and #12), or in reverse order (#15 and #16), a telephone number (#22), and a regular word in a box (#27). This seems to suggest that the execution of simple movements without a cue, as well as that of movements not previously learned (as is the case of the copy in reverse order), alters the in-air movements of CI patients. It is also worth noting that in these cases the percentage of dynamic features is about 50%. This result suggests that anomalies are present both in the dynamics and shape of the handwriting of cognitively impaired people.

AP features (those computed without distinguishing between in-air and on-paper traits) achieved the best performance in task #17, where the features were extracted

TABLE 10. Summary of the classifcation results. The accuracies achieved using the RFE algorithm are in bold.

Task	Classifier	Features	Acc.	RFE
1	XGB	A	88.6	no
2	XGB	A	**90.3**	yes
3	XGB	AP	**87.6**	yes
4	XGB	AP	**86.8**	yes
5	XGB	A	**87.5**	yes
6	RF	P	**87.5**	yes
7	XGB	P	**88.3**	yes
8	RF	P	**87.5**	yes
9	DT	P	**89.6**	yes
10	XGB	A	**90.1**	yes
11	XGB	P	86.7	no
12	DT	A	**88.1**	yes
13	RF	P	**87.3**	yes
14	RF	AL	**87.7**	yes
15	RF	A	**88.7**	yes
16	RF	A	89.1	no
17	XGB	AP	**89.6**	yes
18	XGB	AL	**88.9**	yes
19	XGB	AL	**91.6**	yes
20	DT	P	**87.6**	yes
21	XGB	A	**88.1**	yes
22	XGB	A	**89.6**	yes
23	XGB	P	**88.5**	yes
24	DT	AL	86.1	no
25	DT	P	85.9	no
26	XGB	P	**85.7**	yes
27	XGB	A	**87.9**	yes
28	DT	P	**88.8**	yes
29	XGB	AP	**90.9**	yes
30	XGB	AP	85.3	no
31	XGB	AP	88.1	no
32	XGB	AL	91.1	no
33	RF	A	**88.1**	yes
34	RF	AL	**87.5**	yes

TABLE 11. Best performing feature categories. Values in parentheses represent the of dynamic features selected.

Task category	A	P	AL	AP
Graphic	3 (51)	0	1(50)	2(66)
Copy and reverse copy	6(49)	9(53)	1(0)	4(58)
Memory and dictation	2(45)	2(57)	4(52)	0

from the six words (regular, non-regular, non-words) without splitting them, and tasks #29 to #31. Note that these last tasks, as mentioned above, were obtained splitting task #17, and in particular relate to the copying of a non-regular word and the two non-words. This seems to confirm that when a medium or a high cognitive load is required, CI alters both in-air and on-paper movements. In this case the percentage of dynamic features is 58%. This result suggests that in this case CIs affect more the movements performed to write non-regular or non-words than their shape.

Finally, AL features achieved the best result in task #19, copying the fields of a postal order. This task required precise movements and good spatial organization skills. Therefore, this result seems to confirm that these movements and skills are altered in CI patients and that interaction between in-air and on-paper features allows better modelling of these alterations. Note that in this case no dynamic feature is selected. This result confirms that in the handwriting requiring precise movements and a good spatial organisation, CIs affect the shape of both on-paper and in-air traits.

3) MEMORY AND DICTATION TASKS

For the memory and dictation tasks, from Table 11 we can observe that AL features achieved the best result in four tasks, whereas A and P features were the best in two tasks. More precisely, AL achieved the best result in task #14 (features extracted from the three dictated words without splitting them), task #18 (writing the name of an object shown in a picture), and tasks #32 and #34 (these tasks were also obtained splitting task #14). This result suggests that for this kind of tasks, the interaction between in-air and on-paper features allows an effective distinction between the handwriting movements of CI patients and those of the control group. In this case the percentage of dynamic features is 52%. This result suggests that in these tasks CIs affect both the dynamics and shape of the handwriting of cognitively impaired people.

The A features outperformed the others in the task #1 (signature) and task #33, (split from task #14), whereas the P features achieved the best performance in the tasks requiring writing a simple sentence (#20) and a telephone number (#23) from dictation. The first result suggests that in air features are more discriminating in tasks involving well-known movements, like those of the signature or writing a simple word previously memorized. On the other hand, the results of the on-paper features suggest that CIs alter on-paper traits when cognitive functions using different senses (sight and hearing in this case) are involved. Note that for the A and P features the percentage of dynamic features is 45% and 57%, respectively. This result suggests that CI patients still had the ability to correctly write the on-paper traits of well-known words, whereas they tend to make in-air trajectories different from those of the control group. On the other hand, the value of 57% for the P features suggests that when different senses are used CIs affect more the dynamics of handwriting movements than their shape.

TABLE 12. Average cross-correlation computed on the features selected for the tasks analyzed in Subsection V-C. The values in parentheses (second column) represent the number of features selected.

Task	Features (#selected)	cross-correlation
#2	A (4)	0.47
#4	AP (5)	0.31
#18	AL (15)	0.39
#22	A (2)	0.77

C. FEATURES AND COGNITIVE FUNCTIONS

In the third set of experiments, we investigated the relationships between the feature subset selected for each task and the cognitive functions and skills involved in that task. To this end, and for the sake of brevity, we have chosen a case study for each task category. Note that for each case we analyzed the group of features that achieved the best result, according to data shown in Table 10.

To test the effectiveness of RFE in selecting few correlated features, we computed the cross-correlation between the features selected. Table 12 shows the average cross-correlation between the features selected for each of the tasks analyzed in the following. From Table 12 we can observe that cross-correlation is always below 0.5, except for task #22. This result confirms that RFE typically provides subsets containing few correlated features.

1) GRAPHIC TASKS

As case studies for the graphic tasks we chose tasks #2 (joining two points) and #4 (circle retracing). These tasks require the use of previously learned motor skills. From the literature, we know that these skills may be affected by brain damages due to ND dementia (e.g., brain atrophy, neuronal loss, cellular or synaptic dysfunction). The analysis of the changes in learned movements caused by MCI and AD facilitates the understanding of brain-body functional relationships and allows the identification of patterns of sensory-motor dysfunctions associated with MCI and AD [39], [40].

For task #2 we analyzed the A features. In particular, the RFE algorithm selected the following features: peak vertical velocity, horizontal size, normalized y jerk, #strokes. These features confirm that the learned handwriting movements of cognitively impaired people are more unsteady than the control group. Most probably, as just mentioned, this is due to the brain damage causing cognitive impairments. This result confirms that reported in [2]. In that study, the authors investigated the handwriting movements of twenty eight participants (nine with a diagnosis of probable AD, nine amnestic MCI subjects and ten cognitively normal) in performing tasks very similar to our task #2. They found that several movement characteristics, e.g. smoothness, timing, trajectories, velocity, or acceleration, are correlated to cognitive aging. They also found that the control group and MCI patients were significantly faster and smoother than AD patients.

In task #4 we analyzed the AP features. In this case, the RFE algorithm selected the following features: peak vertical velocity, peak vertical acceleration, slant, relative time to peak vertical velocity, absolute jerk. This result confirmed that of task #2: the unsteady handwriting movements of cognitively impaired people can be effectively identified using the kinematic features. Our results confirm that reported in [7] in which the authors investigated the handwriting movements required to perform tasks very similar to our task #4. They found that the handwriting movements of AD patients were significantly less regular than those of the control group.

2) COPY AND REVERSE COPY TASKS

As a case study for the copy and reverse copy category we chose task #22, i.e. the copying of a telephone number. For this task we analyzed the A features. In this case the RFE algorithm selected two features, namely, normalized jerk and stroke duration. These in-air features allowed us to achieve an accuracy of 89.6% and outperformed the other feature groups both in terms of selected features and accuracy. Indeed the AL features achieved an accuracy of 86.3% (selecting twenty-six features), whereas the P features achieved an accuracy 87.8% (thirteen features); the AP group attained an accuracy of 86.3, but selecting far more features (twenty). Therefore, these results suggest that the handwriting movements of CI patients are more unsteady than those of the control group in the case of copy and reverse copy tasks as well. Similar results were reported in [11]. The authors analyzed the kinematic characteristics of the handwriting process of ninety-four elderly people, while they were performing five copy tasks, using temporal as well as spatial features. They found that in-air time consistently differentiated between the groups (mild AD, MCI and healthy) in four out of the five tasks. In particular, they found that MCI and mild AD patients spent significantly longer with the pen in the air than the control group. Note that this is the only case in which RFE selected correlated features (0.77, see Table 12). This means that in this case, although correlated, both features are needed to achieve the best performance. Indeed, as explained in Section IV, RFE is a wrapper feature selection algorithm. As a consequence it tries to maximizes the classification performance, even if the features selected are correlated.

3) MEMORY AND DICTATION TASKS

We took task #18 as a case study for the memory and dictation tasks (writing the name of the object shown in a picture). This task allowed us to check anomia, i.e. word-finding difficulties. From the literature, we know that anomia may be an early symptom of AD. It may be due to several causes: diminished integrity of semantic representations in memory [41], [42], difficulty in accessing those representations due to the weakening of neural network connectivity and/or cognitive processing limitations, or to both ''storage'' and ''access'' deficits [43], [44].

In this case, we achieved the best accuracy (88.9%) using the AL features (the feature vector contains both in-air and on-paper features). This task required the participant to recognize the object shown and write its name, after having recalled the word s/he associated with it. The RFE algorithm selected the following features:

in-air: peak vertical acceleration, relative time to peak vertical velocity, normalized jerk, number of peak acceleration points, #strokes;

on-paper: start vertical position, vertical size, start horizontal position, slant, loop surface, road length, absolute y jerk, normalized y jerk, pen pressure, age;

It is worth noting that the in-air features selected are related to the kinematics of the movements, whereas most of the on-paper features selected are linked to the shape of the handwriting. This result confirms that CIs alter in-air and on-paper movements differently, with the former being more free than the latter. This freedom is due to two reasons. Firstly, on-paper movements provide the visual feedback given by the ink traits written. Secondly, on-air traits are not constrained by the shapes of the characters to be written. This result also confirms that on-paper and in-air features investigate different cognitive aspects. In summary, we can state that in this memory task: (i) in-air movements of cognitively impaired people are more unsteady because they are less constrained and do not provide any visual feedback; (ii) on-paper movements produce altered shapes even though these are not significantly unsteady in comparison with the control group. Note that these alterations are detectable even using the simple features we used.

VI. CONCLUSION

MCI is a condition in which a person experiences a slight, but still noticeable, decline in mental abilities (memory and thinking skills) compared with others of the same age, and it is considered a prodromal syndrome of AD. Diagnostic signs of MCI and AD also include alterations of spatial organization and poor control of movements, which may affect the handwriting of the people affected. Currently, in the Artificial Intelligence field there is an ever-increasing interest in the development of systems that, through the analysis of handwriting, are able to provide doctors with further evidence of the onset of these diseases. In order to investigate whether and how the diagnostic signs related to handwriting can be used to implement such a system, in a previous study we presented an experimental protocol, made of twenty-five handwriting tasks.

The choice to consider different scenarios is because the assessment of onset and progression of neurocognitive disorders requires the joint analysis of the different cognitive abilities that could be compromised. This is the reason why we defined a protocol in which the different tasks were organized into groups, specifically oriented to the evaluation of a particular ability, namely fine motor control, memory ability and cognitive ability.

Following this line of research, in this paper we presented the results of a study in which we used a well-known and widely-used feature selection approach to determine, among the extracted features, which, if any, allow an effective prediction of the symptoms related to CI and AD. We performed this analysis task by task, with the purpose of investigating which cognitive skills are affected by these diseases, especially during the early stages. We tested different types of features and performed three sets of experiments. In the first set we tested the importance of each feature across the tasks of the protocol, and in the second one we assessed the effectiveness of the feature selection algorithms used. Finally, in the third set, we investigated the relationship between the selected features and the cognitive functions and skills involved in the handwriting process. The results proved that except for a few features, common to most tasks, (such as, for example, execution time and speed), each task of the protocol was characterized by a different set of relevant features. This diversity confirms our hypothesis that each of the cognitive abilities tested by the tasks of the protocol is affected by the damage to the different areas of the brain caused by neurodegenerative diseases. The results also confirmed that handwriting analysis can be used to develop inexpensive and non-invasive systems for the assessment of the mental status of the people involved.

Furthermore, we have considered only one feature selection algorithm for two reasons: firstly because the RFE algorithm is among the most effective and widely used. RFE belongs to the category of wrapper approaches, which measure the effectiveness of features based on classification performance: in our case, we used as evaluation function the accuracy achieved by using the xgboost classifier with the K-fold cross-validation strategy. Secondly, because the goal of our study is not to identify the best performing feature set, but to demonstrate that performance significantly improves by carrying out a specific feature selection phase for each task. The results obtained confirmed that the different aspects of cognitive impairment can be better highlighted by selecting a specific set of features for each task.

Finally, as previously mentioned, we preferred to present the results of all tasks together because it is precisely the joint evaluation of the different tasks that allowed us to characterize the different aspects of cognitive impairment due to the onset and the progression of neurocognitive disorders. Splitting our paper into different studies could simplify the presentation of the results, but it would not allow for the overall view of the results that we consider very useful for this kind of application.

The limitations of our study are essentially related to the number of subjects involved in the experiments and to the type of cognitive disorder considered. As regards the first point, the number of people who participated in the experiments is quite high for this type of study even if it is limited from the point of view of machine learning techniques. However, it is necessary to consider that it is not easy to involve a very large group of participants, because they must be carefully selected in collaboration with the hospitals. As for the second point, it would be very useful to have patients with different levels of cognitive impairment, ranging from an initial state where there are no signs of alteration in cognitive abilities, up to a more advanced level of disease, where the patients however are still able to carry out an autonomous life and are able to perform the writing tasks of our protocol. These aspects will be the subject of our future research activity.

Future work will also include: (i) a further analysis of these results, which will involve doctors specialized in brain disease and dementia; (ii) more feature selection techniques [45], [46]; (iii) the development of classification systems based on the combination of the predictions provided by the classifiers trained on the data from the single tasks [47]–[50].

COMPLIANCE WITH ETHICAL STANDARDS

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Conflict of interest: The authors declare no conflict of interest.

REFERENCES

[1] M. Grundman *et al.*, "Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials," *Arch. Neurol.*, vol. 61, no. 1, pp. 59–66, 2004.

[2] J. H. Yan, S. Rountree, P. Massman, R. S. Doody, and H. Li, "Alzheimer's disease and mild cognitive impairment deteriorate fine movement control," *J. Psychiatric Res.*, vol. 42, no. 14, pp. 1203–1212, Oct. 2008.

[3] G. Di Mauro, V. Bevilacqua, L. Colizzi, and D. Di Pierro, "Testgraphia, a software system for the early diagnosis of dysgraphia," *IEEE Access*, vol. 8, pp. 19564–19575, 2020.

[4] E. Onofri, M. Mercuri, M. Salesi, M. Ricciardi, and T. Archer, "Dysgraphia in relation to cognitive performance in patients with Alzheimer's disease," *J. Intellectual Disab.-Diagnosis Treatment*, vol. 1, no. 2, pp. 113–124, Dec. 2013.

[5] S. Müller, O. Preische, P. Heymann, U. Elbing, and C. Laske, "Diagnostic value of a tablet-based drawing task for discrimination of patients in the early course of Alzheimer's disease from healthy individuals," *J. Alzheimer's Disease*, vol. 55, no. 4, pp. 1463–1469, Dec. 2016.

[6] J. A. Small and N. Sandhu, "Episodic and semantic memory influences on picture naming in Alzheimer's disease," *Brain Lang.*, vol. 104, no. 1, pp. 1–9, Jan. 2008.

[7] A. Schröter, R. Mergl, K. Bürger, H. Hampel, H.-J. Möller, and U. Hegerl, "Kinematic analysis of handwriting movements in patients with alzheimer's disease, mild cognitive impairment, depression and healthy subjects," *Dementia Geriatric Cogn. Disorders*, vol. 15, no. 3, pp. 132–142, 2003.

[8] R. Senatore, A. D. Cioppa, and A. Marcelli, "Automatic diagnosis of neurodegenerative diseases: An evolutionary approach for facing the interpretability problem," *Information*, vol. 10, no. 1, p. 30, Jan. 2019. [Online]. Available: https://www.mdpi.com/2078-2489/10/1/30

[9] D. Impedovo, G. Pirlo, G. Vessio, and M. T. Angelillo, "A handwriting-based protocol for assessing neurodegenerative dementia," *Cogn. Comput.*, vol. 11, no. 4, pp. 576–586, Aug. 2019.

[10] N. D. Cilia, C. De Stefano, F. Fontanella, and A. S. D. Freca, "Using genetic algorithms for the prediction of cognitive impairments," in *Proc. Int. Conf. Appl. Evol. Comput.*, in Lecture Notes in Computer Science, vol. 12104, 2020, pp. 479–493.

[11] P. Werner, S. Rosenblum, G. Bar-On, J. Heinik, and A. Korczyn, "Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment," *J. Gerontol., Psychol. Sci.*, vol. 61, no. 4, pp. 228–236, 2006.

[12] G. Pirlo, M. D. Cabrera, M. A. Ferrer-Ballester, D. Impedovo, F. Occhionero, and U. Zurlo, "Early diagnosis of neurodegenerative diseases by handwritten signature analysis," in *Proc. ICIAP Workshops*, 2015, pp. 290–297.

[13] D. Impedovo, G. Pirlo, F. M. Mangini, D. Barbuzzi, A. Rollo, A. Balestrucci, S. Impedovo, L. Sarcinella, C. O'Reilly, and R. Plamondon, "Writing generation model for health care neuromuscular system investigation," in *Proc. CIBB*. Cham, Switzerland: Springer, 2014, pp. 137–148.

[14] J. Garre-Olmo, M. Faúndez-Zanuy, K. López-de-Ipiña, L. Calvó-Perxas, and 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.*, vol. 14, no. 9, pp. 1–9, Aug. 2017.

[15] N. Cilia, C. De Stefano, F. Fontanella, M. Molinara, and A. S. D. Freca, "Using handwriting features to characterize cognitive impairment," in *Proc. Int. Conf. Image Anal. Process.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11752, 2019, pp. 683–693.

[16] G. Dimauro, D. Caivano, V. Bevilacqua, F. Girardi, and V. Napoletano, "VoxTester, software for digital evaluation of speech changes in Parkinson disease," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, May 2016, pp. 1–6.

[17] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi, "Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system," *IEEE Access*, vol. 5, pp. 22199–22208, 2017.

[18] N. D. Cilia, C. D. Stefano, F. Fontanella, and A. S. D. Freca, "An experimental protocol to support cognitive impairment diagnosis by using handwriting analysis," *Procedia Comput. Sci.*, vol. 141, pp. 466–471, Jan. 2018.

[19] N. D. Cilia, C. D. Stefano, F. Fontanella, M. Molinara, and A. S. D. Freca, "Handwriting analysis to support Alzheimer's disease diagnosis: A preliminary study, notes in computer science," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, in Lecture Notes in Computer Science, vol. 11679, 2019, pp. 143–151.

[20] D. Impedovo and G. Pirlo, "Dynamic handwriting analysis for the assessment of neurodegenerative diseases: A pattern recognition perspective," *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 209–220, 2019.

[21] S. Rosenblum, B. Engel-Yeger, and Y. Fogel, "Age-related changes in executive control and their relationships with activity performance in handwriting," *Hum. Movement Sci.*, vol. 32, no. 2, pp. 363–376, Apr. 2013.

[22] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faúndez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," *Artif. Intell. Med.*, vol. 67, pp. 39–46, Feb. 2016.

[23] G. Vessio, "Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review," *Appl. Sci.*, vol. 9, no. 21, p. 4666, Nov. 2019.

[24] J. Kawa, A. Bednorz, P. Stępień, J. Derejczyk, and M. Bugdol, "Spatial and dynamical handwriting analysis in mild cognitive impairment," *Comput. Biol. Med.*, vol. 82, pp. 21–28, Mar. 2017.

[25] K. L. de Ipina, M. Iturrate, P. M. Calvo, B. Beitia, J. Garcia-Melero, A. Bergareche, P. De la Riva, J. F. Marti-Masso, M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure, and J. Sole-Casals, "Selection of entropy based features for the analysis of the Archimedes' spiral applied to essential tremor," in *Proc. 4th Int. Work Conf. Bioinspired Intell. (IWOBI)*, Jun. 2015, pp. 157–162.

[26] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, and M. Arnaoutoglou, "Machine learning-based classification of simple drawing movements in Parkinson's disease," *Biomed. Signal Process. Control*, vol. 31, pp. 174–180, Jan. 2017.

[27] R. Plamondon, "A kinematic theory of rapid human movements: Part I. Movement representation and generation," *Biol. Cybern.*, vol. 72, no. 4, pp. 295–307, Mar. 1995.

[28] R. Plamondon, "A kinematic theory of rapid human movements: Part II. Movement time and control," *Biol. Cybern.*, vol. 72, no. 4, pp. 309–320, 1995.

[29] D. Impedovo, G. Pirlo, D. Barbuzzi, A. Balestrucci, and S. Impedovo, "Handwritten processing for pre diagnosis of Alzheimer disease," in *Proc. BIOSTEC*. Setúbal, Portugal: SCITEPRESS, 2014, pp. 193–199.

[30] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. S. di Freca, "Handwriting analysis to support neurodegenerative diseases diagnosis: A review," *Pattern Recognit. Lett.*, vol. 121, pp. 37–45, Apr. 2019.

[31] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "'Mini-mental state': A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatric Res.*, vol. 12, no. 3, pp. 189–198, 1975.

[32] A. Iavarone, B. Ronga, L. Pellegrino, E. Loré, S. Vitaliano, F. Galeone, and S. Carlomagno, "The frontal assessment battery (FAB): Normative data from an Italian sample and performances of patients with Alzheimer's disease and frontotemporal dementia," *Funct Neurol.*, vol. 19, pp. 191–195, Jul. 2004.

[33] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment," *J. Amer. Geriatrics Soc.*, vol. 53, no. 4, pp. 695–699, Apr. 2005.

[34] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 389–422, Mar. 2002.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[36] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 785–794.

[37] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[38] J. R. Quinlan, *C4.5: Programs for Machine Learning* (Morgan Kaufmann Series in Machine Learning). San Francisco, CA, USA: Morgan Kaufmann, 1993.

[39] C. Pennanen, M. Kivipelto, S. Tuomainen, P. Härtikainen, T. Hanninen, M. P. Laakso, M. Hallikainen, M. Vanhanen, A. Nissinen, E. Helkala, P. Vainio, R. Vanninen, K. Partanen, and H. Soininen, "Hippocampus and entorhinal cortex in mild cognitive impairment and early AD," *Neurobiol. Aging Physiol. Rev.*, vol. 25, no. 3, pp. 303–310, Mar. 2004.

[40] D.-W. Yang, B.-S. Kim, J.-K. Park, S.-Y. Kim, E.-N. Kim, and H.-S. Sohn, "Analysis of cerebral blood flow of subcortical vascular dementia with single photon emission computed tomography: Adaptation of statistical parametric mapping," *J. Neurol. Sci.*, vols. 203–204, pp. 199–205, Nov. 2002.

[41] P. Garrard, M. A. L. Ralph, K. Patterson, K. H. Pratt, and J. R. Hodges, "Semantic feature knowledge and picture naming in dementia of Alzheimer's type: A new approach," *Brain Lang.*, vol. 93, no. 1, pp. 79–94, 2005.

[42] J. R. Hodges, D. P. Salmon, and N. Butters, "Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge?" *Neuropsychologia*, vol. 30, no. 4, pp. 301–314, Apr. 1992.

[43] J. L. Bonilla and M. K. Johnson, "Semantic space in Alzheimer's disease patients," *Neuropsychology*, vol. 9, no. 3, pp. 345–353, 1995.

[44] H. J. Chenery, B. E. Murdoch, and J. C. L. Ingram, "An investigation of confrontation naming performance in Alzheimer's dementia as a function of disease severity," *Aphasiology*, vol. 10, no. 5, pp. 423–441, Jul. 1996.

[45] C. D. Stefano, F. Fontanella, and A. D. Freca, "Feature selection in high dimensional data by a filter-based genetic algorithm," in *Proc. Eur. Conf. Appl. Evol. Comput.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 10199, 2017, pp. 506–521.

[46] N. Cilia, C. De Stefano, F. Fontanella, and A. S. di Freca, "Variable-length representation for EC-based feature selection in high-dimensional data," in *Proc. Int. Conf. Appl. Evol. Comput.*, in Lecture Notes in Computer Science: Including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11454, 2019, pp. 325–340.

[47] C. D'Elia, C. Marrocco, M. Molinara, and F. Tortorella, "Detection of clusters of microcalcifications in mammograms: A multi classifier approach," in *Proc. 21st IEEE Int. Symp. Comp.-Based Med. Syst.*, Jun. 2008, pp. 572–577.

[48] C. D. Stefano, C. D'Elia, A. S. Di Freca, and A. Marcelli, "Classifier combination by Bayesian networks for handwriting recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 05, pp. 887–905, Aug. 2009.

[49] C. D. Stefano, F. Fontanella, C. Marrocco, and A. S. D. Freca, "A hybrid evolutionary algorithm for Bayesian networks learning: An application to classifier combination," in *Proc. Eur. Conf. Appl. Evol. Comput.*, in Lecture Notes in Computer Science, vol. 6024, 2010, pp. 221–230.

[50] C. D. Stefano, F. Fontanella, G. Folino, and A. S. D. Freca, "A Bayesian approach for combining ensembles of GP classifiers," in *Proc. Int. Workshop Multiple Classifier Systems*, in Lecture Notes in Computer Science, vol. 6713, 2011, pp. 26–35.

NICOLE DALIA CILIA is currently a Postdoctoral Researcher with the Department of Electrical and Information Engineering, University of Cassino and Southern Lazio. She participated in the PANOPTESEC EU project on Cyber Security in the Ontology for Reachability Matrix Computation package and in national HAND project on Handwriting Analysis Against Neuromuscular Disease. She has been an invited speaker in international conferences. She has been awarded two starting grants regarding eyetracking and handtracking tools to investigate analogical reasoning. She has published about 40 articles and she edited two volumes in the field of AI and cognitive sciences. In her research activity, she mainly deals with the artificial intelligence, embodied cognition methodological problems, and with the machine learning and deep learning techniques-based on handwriting and eyetracking analysis for support neurodegenerative and neurodevelopmental disorders diagnosis. She is a member of CVPL. She is a Review Editor of *Frontiers in Robotics and AI* and *La mente e i sistemi cognitive*.

CLAUDIO DE STEFANO received the Laurea degree (Hons.) in electronic engineering and the Ph.D. degree in electronic and computer engineering from the University of Naples Federico II, Italy, in 1990 and 1994, respectively. From October 1996 to October 2001, he was an Assistant Professor of computer science with the University of Sannio, Benevento, Italy. He joined the University of Cassino and Southern Lazio, in November 2001, where he is currently a Full Professor of computer science. He has authored over 150 publications on international journals and congresses. He is the co-editor of two books. In particular, he has been active in the fields of document analysis and recognition, data mining and classification systems-based on both neural networks, and statistical learning paradigms. His current research interests include artificial vision, image processing, pattern recognition, automatic learning systems, techniques for automatic segmentation and classification of on-line and off-line cursive handwriting, classifier combination paradigms, and automatic learning methods-based on the use of evolutionary algorithms and Bayesian networks. Recently, writing analysis techniques have been applied to the early detection of cognitive disorders. He is a member of the IAPR. He is a member of the Board of Directors of the National Interuniversity Consortium for Computer Science. He is the President of the International Graphonomics Society (IGS). He is a guest editor of several special issues of international journals. He is an Associate Editor of *Pattern Recognition Letters* journal.

FRANCESCO FONTANELLA received the Laurea degree in physics and the Ph.D. degree in electronic and computer engineering from the University of Naples Federico II, Naples, Italy, in 2001 and 2005, respectively. He is currently an Assistant Professor with the Department of Electrical and Information Engineering, University of Cassino and Southern Lazio. He has authored over 60 scientific articles in journals and international conference proceedings. His research interests include pattern recognition, evolutionary computation, and bio-inspired computing. He has been a member of the Board of the Society for the Promotion of Evolutionary Computation in Europe and its Surroundings (SPECIES). He is a member of the IEEE Computational Intelligence Society Task Force on Evolutionary Computer Vision and Image Processing. He has been the Co-Chair of the International Workshop on Pattern Recognition of Cultural Heritage (PatReCH 2019, held in conjuntion with ICIAP 2019) and the Technical Co-Chair of the 2018 IEEE International Conference on Metrology for Archaeology and Cultural Heritage (MetroArcheo 2018). He has been guest editor of two special issues published on *Pattern Recognition Letters*.

ALESSANDRA SCOTTO DI FRECA received the Laurea degree in telecommunications engineering and the Ph.D. degree in electronic and computer engineering from the University of Cassino and Southern Lazio, Cassino, Italy, in 2002 and 2006, respectively. She has authored over 50 scientific articles in journals and international conference proceedings. Her research interests include pattern recognition, evolutionary computation, and bio-inspired computing. She is also a member of the IEEE Computational Intelligence Society Task Force on Evolutionary Computer Vision and Image Processing. She has been the Co-Chair of the International Workshop on Pattern Recognition of Cultural Heritage (PatReCH 2020, held in conjuntion with ICPR 2020). She has been a guest editor of a special issue published on *Pattern Recognition Letters*.