IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 15, 2019, accepted April 24, 2019, date of publication April 29, 2019, date of current version May 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2913685

Attentional Pattern Classification for Automatic Dementia Detection

MARIA TERESA ANGELILLO, FABRIZIO BALDUCCI, DONATO IMPEDOVO, (Senior Member, IEEE), GIUSEPPE PIRLO, (Senior Member, IEEE), AND GENNARO VESSIO

Computer Science Department, University of Bari, 70125 Bari, Italy

Corresponding author: Gennaro Vessio (gennaro.vessio@uniba.it)

This work was supported by the Italian Ministry of Education, University and Research through the PRIN2015-HAND Project under Grant H96J16000820001.

ABSTRACT This paper proposes a novel technique for the automatic detection of dementia based on the attentional matrices test (AMT) for selective attention assessment. The original test provides three matrices, of increasing difficulty, and the test taker is asked to mark target digits assigned. In our proposal, AMT was developed on a digitizing tablet, equipped with an electronic pen. Tablet technology enables the acquisition of additional measures to those that can be obtained by observing the execution of the traditional paper-based test. These measures reflect the dynamics of the handwriting process, particularly the pauses and hesitations while the pen is not in contact with the pad surface. Handwriting measures can then serve as an input to machine learning algorithms to automatize disease detection. In contrast to the traditional approach, dynamic handwriting analysis can provide a means to better evaluate the visual search of the patient, as well as motor planning. To evaluate the effectiveness of the proposal, a classification study was carried out involving 29 healthy control subjects and 36 demented patients. We employed different machine learning algorithms and an ensemble scheme. We observed the first matrix to be the most discriminating, while the ensemble of the best classification models over the three matrices provided the best classification performance [i.e., an area under the ROC curve (AUC) of 87.30% and a sensitivity of 86.11%]. Our proposal has the potential to provide a cost-effective and easy-to-use diagnostic tool, which may also support mass screening of the population.

INDEX TERMS Machine learning, decision support systems, medical diagnosis.

I. INTRODUCTION

Dementia is a general term used to refer to a set of symptoms, associated with intellectual and social skills, severe enough to interfere with everyday life activities [1]. One of the most common types of dementia is Alzheimer's disease, which accounts for 60 to 80 percent of cases [2]. Also other degenerative brain disorders can develop in dementia, for example Parkinson's disease [3] and Huntington's disease [4]. The number of people with cognitive impairment will increase dramatically as the elderly population increases. Therefore, there is a pressing need for diagnostic measures at the earlier course of cognitive decline to evaluate the effectiveness of novel drug treatments and for improving the quality of life of the patient.

Despite the recent advancements in *biomarker* research [5], the current clinical diagnostic criteria still rely on paper-based neuropsychological tests. Mini-Mental State Examination [6], for example, is used extensively today to assess cognitive impairment. This traditional approach has several limitations [7], [8]. Although they have been standardized to maximize administration consistency, paper-based methods can suffer from human error and bias (e.g., [9]). Most of traditional techniques are not conceived for continuous repeated measurements; moreover, they are not practically accessible for those in remote areas [10]. More importantly, such an approach may fail to capture subtle, but meaningful patterns, which are likely to improve prediction accuracy [11], [12].

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

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

In order to address these issues, the present paper proposes a digitized, mobile variant of the well-known paper-based Attentional Matrices test (AMT). The original test was conceived to measure selective attention during a visual search [13], which was shown to be compromised in cognitively impaired individuals [14]. To establish a consistency with the traditional paper-based assessment, the digital variant we propose was designed to be as much similar as possible to the original test. The main difference concerns the acquisition tool, which is a professional digitizing tablet equipped with an electronic pen. Such a device is able to capture not only the geometric position of the pen at certain time stamps, but also the pressure exerted over the writing surface as well as measures of pen inclination. In particular, the tablet is able to track the hand movement not only when the pen is on the pad surface, but also when the pen is in proximity of the surface, i.e. ''in-air''.

From the time series raw data sampled by the acquisition device, several measures reflecting the *dynamics* of the handwriting process were computed. These features were then fed into machine learning algorithms to automatize the disease detection. There is a growing body evidence on computerized handwriting analysis, e.g. [15], [16], supporting the hypothesis that physical, cognitive and psychological characteristics of an individual can be captured by these dynamic measures. Handwriting features, and in particular those related to in-air trajectories, in fact, can extend the performance measures that one can obtain by observing the traditional execution of a test. These measures can better reflect the motor planning and its deficiencies.

Dynamic handwriting analysis was employed with successful results to a digital variant of other well-known tests, such as the Trail Making Test [17] and the Clock Drawing Test [18]. Nevertheless, despite the number of promising results, research is still ongoing and dynamic handwriting measures are far from a generally adopted solution. Our goal is to stress the hypothesis about the effectiveness of these measures in supporting cognitive impairment evaluations: this is done by considering a cancellation test originally conceived to evaluate selective attention faculties. To our best knowledge, this is the first time a digital variant of AMT has been investigated.

Hence, the contribution of this paper is two-fold. On one hand, we propose to enrich the batteries of tests at disposal of neuropsychologists with easy-to-use, technological tools. On the other hand, we propose a methodological procedure, based on dynamic handwriting analysis, to evaluate the discriminating power of attentional patterns to instruct learning algorithms on how discriminate between healthy people and individuals with cognitive impairment automatically.

It is worth remarking that the proposed approach is not intended to replace standard techniques or even doctors, but to provide additional evidence to further support the clinical assessment. Currently available advanced diagnostic methods are invasive or expensive, such as cerebrospinal fluid examination (e.g., [19]) and neuroimaging (e.g., [20]). Handwriting analysis, instead, would provide a non-invasive, low-cost decision support tool.

The rest of this paper is organized as follows. Section II discusses the related work. Section III describes the proposed method. Section IV deals with an experimental study aimed at assessing the effectiveness of the proposal. Section V discusses the results obtained and provides concluding remarks and future work.

II. RELATED WORK

The evaluation of the patients' clinical status and their responsiveness to medication is typically achieved through a clinical workup which includes a thorough medical history, a neuropsychological test battery and rating scales. Unfortunately, there is still no one certain test to determine if someone has dementia and the diagnosis can be confirmed only *post-mortem*. Getting a reliable diagnosis can require months and symptoms need to be constantly monitored. Furthermore, determining the exact type of dementia, as well as its degree of severity, is difficult. To this end, identifying accurate biomarkers for the early and differential diagnosis, prognosis and response to therapy is a primary goal of the current research activity on neurodegenerative diseases.

In recent years, promising biomarkers to be used at a preclinical stage include neuroimaging techniques, such as PET, SPECT and MRI. For example, functional MRI allows one to measure the brain activation during a cognitive task or resting state. Studies examining brain activation changes for the development of a marker for early Alzheimer's disease have been recently reported in the literature, e.g. [21]. A very recent and promising research direction, in particular, consists in reconstructing, from the brain imaging data, the human brain network. Encouraging results have been recently reported both for the case of fMRI [22] and diffusion tensor imaging data [20].

Multimodal approaches exploiting multiple imaging data, such as MRI and PET, have been also investigated, e.g. [23]. When these data are used along with genetic data, such as single nucleotide polymorphism, valuable insights into the brain abnormalities can be provided and the accuracy of dementia diagnosis may improve (see, for example, [24]). Multiview approaches can link neuroimaging data also to EEG signals [25]. Techniques aimed at treating the multi-way nature of these data are a topic of intense research; in this context, matrix/tensor decomposition methods have been successfully used for biomedical data analysis [26].

However, acquiring brain images is still a time consuming and expensive process (the image processing of a single subject can require hours); moreover, it cannot be used for monitoring at the patient's home. Conversely, a growing interest has arisen, in the last years, towards the application of biometric techniques to health [27]. From this perspective, as already part of neuropsychological test batteries, a special role can be covered by handwriting. Handwriting, in fact, is a complex activity entailing cognitive as well as motor components, whose changes seem to represent a prominent biomarker for the evaluation of neurodegenerative diseases [15], [16]. Some recent works, in particular, provided evidence that the automatic discrimination between healthy and unhealthy people can be accomplished on the basis of simple and easy-to-perform handwriting tasks. Within this direction, static features, based on images of the patterns acquired, and dynamic features, acquired through the use of a digitizing tablet, can be used. In [28], simple handwriting exercises, from drawing an Archimedes spiral to writing simple words or short sentences, have been employed for the characterization of Parkinson's disease ''dysgraphia''. In the work, features tailored to capture the dynamics of the handwriting process were used. In other recent works [29], [30], instead, the same classification problem has been addressed through the use of features automatically learned by convolutional neural networks trained on images of the performed tasks. For what specifically concerns cognitive impairment evaluation, particularly of Alzheimer's type, the effectiveness of handwriting features in supporting the disease detection has been recently shown, e.g. [31]–[33].

Generally speaking, the automatic analysis of handwritten exercises provides a robust and complementary alternative to other more expensive approaches based for example on neuroimaging. The data acquisition, in fact, can be carried out even at the patient's home; the task performance is quite simple and natural, and does not require any timing or exhaustive repetition; finally, the computational cost of such a method is very low, due to the small size of data samples.

III. PROPOSED METHOD

A. ORIGINAL TEST

Spinnler and Tognoni proposed the original version of AMT in [13], as a digit cancellation test suitable for assessing selective attention deficits in neurological patients. Three matrices of 13 lines, each including 10 numbers from 0 to 9 in random sequence, are shown on a sheet to the test taker. Then, she is requested to cross out with a pencil as fast as possible target digits assigned (1, 2 and 3 for the three matrices, respectively). The total number of targets correctly marked during the three tasks, within 45 s, is considered: it ranges from 0 to 60 (10, 20 and 30, respectively). Before the task is carried out, the examiner explains that each matrix has to be scanned line by line, from left to right and from top to bottom. This is done with the help of the first line, which is used as example, and the second line, which serves to run-in. The targets to be marked are printed at the top of each matrix. Failures to obey these instructions are not corrected during the test.

Three types of errors may occur: omissions, that are target digits skipped; false alarms, i.e. non-target digits crossed out; perseverative errors, which consist in the cancellation of digits that were targets in the previous matrix. Omissions were found to be the most predominant error [14].

Such a cancellation test calls for the subject's ability to react to a predetermined stimulus (the target), actively ignoring all other stimuli, which serve as distractors. Della Sala et al. [14] claimed that AMT involves three sequentially arranged set of actions: (*i*) trigger representation, i.e. assigning a special salience to the digits predetermined as targets; (*ii*) perceptual decision while scanning, since the subject must decide what is a target and what is not as quickly as possible; (*iii*) cancellation, which involves a motor action planning. Cognitively impaired subjects are likely to fail in particular in the second step [14]. One problem is connected to the unsystematic within-line scanning, despite the left-to-right scanning procedure suggested by the examiner. Another problem deals with the poor trade-off between the gaze moving program and the perceptual decision between targets and non-targets. In other words, people with cognitive deficits may ''look without seeing''. Finally, slowness in making the discriminating decision may occur: patients may not be able to use the almost automatic cancelling routine that healthy people seem to apply. In this view, Della Sala et al. observed that the second and third matrix are harder than the first.

The digitalization of AMT was meant to support the evaluation of both the perceptual decision while scanning and, to a better extent, the motor planning.

B. DESIGN OF THE DIGITAL VARIANT

The digital AMT was developed on the commercially available MobileStudio Pro 13 (Wacom technology), which is a full-featured tablet with visual feedback and computational capabilities. Therefore, it immediately provides feedback and it does not require to be connected to a desktop computer, making the test administration easy and natural. The test was written in Java, version 8 update 144, within the integrated development environment Eclipse Oxygen. Among the others, the program makes use of the JPen software library, which allows one to capture data samples through the low-level Wintab API. Data acquired by the program are stored in .txt files. The human-computer interface of the test was designed for a *face-to-face* interaction between the examiner and the test taker. Some useful buttons allow the examiner to reset the current task and go forward to the next task. Figure 1 depicts the acquisition of a test execution performed by a healthy adult.

Note that the design of the test was determined by a usability evaluation involving 10 elderly people (age: 70 ± 5). Since elderly individuals are not familiar with technological tools, we opted for a thinking aloud-based evaluation [34]. It simply requires the users to verbalize their thoughts as the task proceeds. Users reported they felt comfortable using the electronic pen for writing; moreover, they did not feel any difference with respect to the traditional pen and paper-based writing. The observations drawn from these tests helped identify areas of improvement.

The raw data captured by the device are the x- and y-coordinates of the pen position and their *time stamps*. Moreover, the tablet captures more information than the pen

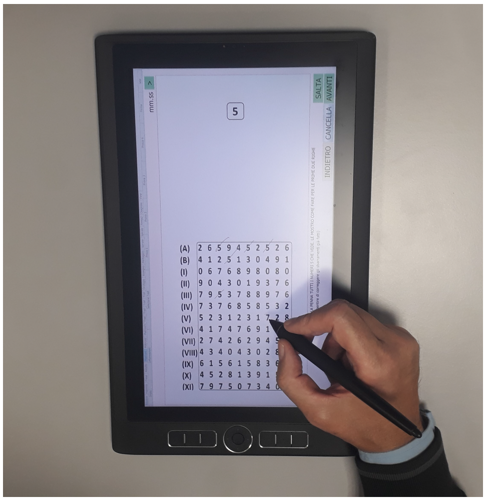


FIGURE 1. Acquisition of the first matrix as performed by a healthy adult.

trajectory, namely pen inclination (*azimuth* and *altitude*) and pen *pressure*. Finally, the tablet also detects the pen trajectory while the tip is not in contact with the surface (no more than 1 cm above), allowing trajectory acquisition pen-ups. One measure, in fact, is the so called *button status*, which is a binary variable evaluating 0 for pen-up state (in-air movement) and 1 for pen-down state (on-surface movement).

In order to obtain the final prediction, the following steps concern with feature extraction and classification: they are described in the following. An overall scheme of the proposed method is depicted in Fig. 2.

C. FEATURE EXTRACTION

The system segments the horizontal and vertical components of the pen position, as sampled during data acquisition, into on-surface and in-air strokes, in accordance with the button status. A *stroke* is a single connected and continuous trait of the handwritten pattern: on-surface strokes correspond to the trace left on the pad surface; in-air strokes are imaginary traces expressing the pauses and hesitations exhibited during writing.

Based on this segmentation, the following features are extracted:

- Number of on-surface strokes;
- Number of in-air strokes.

They are suited to our classification problem, as they may vary between healthy and demented individuals reflecting the complexity of the task as perceived by the test taker. In addition, the following temporal features are computed:

- On-surface time;
- In-air time;
- Total time.

These measures are intended to provide a quantification of the slowness and hesitation of the handwriting movement. Finally, in order to uncover the hidden randomness of handwriting, the following information theory-based measures are calculated for both the horizontal and vertical components of handwriting:

- Horizontal Shannon entropy;
- Vertical Shannon entropy.

The classic formula for Shannon entropy is: $H_S(X) = -\sum_{x \in X} p(x) \log_2\ p(x)$, where $p(X)$ is the probability density function estimated with a Gaussian kernel. These features are likely to provide a way to quantify the systematic/unsystematic line scanning and the automatic/semi-automatic visual search of the subject.

The feature extraction step thus results in 7 features for each matrix. All features are normalized before classification so as to have zero mean and unit variance. Analogous features were successfully used in similar studies, for example [28], [35] and [36].

D. MODEL FITTING

There is no universally best model: a set of assumptions that works well in a feature space may work worst in another. As a consequence, in order to evaluate the performance of the system, we employed the following state-of-the-art classification algorithms:

- K-Nearest Neighbors (KNN) [37]. In particular, we used the usual Euclidean distance as distance metric and we set $K = 5$;
- Logistic regression (LR) [38]. In particular, we used the dual formulation with L2 regularization, which helps avoid overfitting;
- Support Vector Machines (SVMs) [39]. We used both the linear kernel and the radial basis function (RBF) kernel. Note that the bias-variance trade-off of the algorithm is governed by the fine tuning of the penalty parameter C and the kernel coefficient γ in the case of RBF kernel [37]. We set $C = 1$ and $\gamma = \frac{1}{n}$, where n is the number of features. These values represent a typical setting.
- Random Forest (RF) [40]. In the present work, we employed 500 trees for growing the forest, which represents a typical choice.

It is worth remarking that, for each algorithm, the scikit-learn implementation was used [41]. We did not consider other advanced techniques, such as Neural Networks and Deep Learning, as they typically require large sets of data for training, which are difficult to collect in a clinical setting.

Additionally, we used an ensemble of classifiers obtained by combining the best models, i.e. the ones achieving the highest prediction accuracy, over all the three matrices. Combining the classes predicted by different classifiers on the three different tasks, in fact, is likely to provide better predictions, due to diversification. In the proposed method, a majority voting scheme is adopted: the final class label is the most-occurring class label predicted by each individual classifier in the ensemble. With this scheme, every classification model is trained on the features coming from each matrix

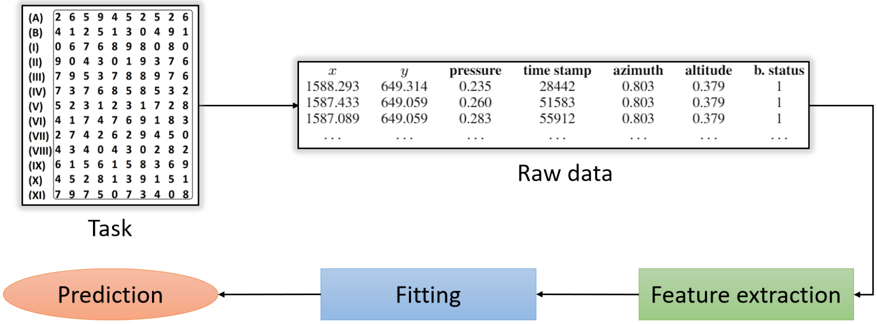


FIGURE 2. **Workflow of the proposed method.**

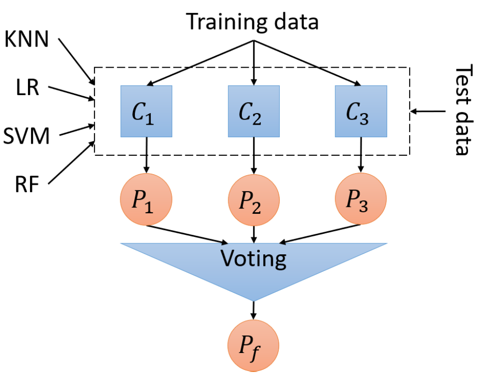


FIGURE 3. **Workflow of the ensemble approach. Each classification model is trained and evaluated on each matrix task. The best three models are then pooled together in a majority voting scheme. Abbreviations: C = classifier; P = prediction; f = final.**

and their performance are evaluated. Then, since different classifiers are likely to provide different results on the same set of data, the best models, i.e. those showing the best result per task, are pooled in the ensemble scheme to achieve the final classification (Fig. 3). Ensemble techniques have shown their effectiveness in the classification problem at hand, for example in the context of neuroimaging [42].

E. FEATURE SELECTION

Note that, although relevant, some features may be redundant with other ones. In particular, the number of in-air strokes is strongly correlated to the number of on-surface strokes, as an in-air stroke consequently follows an on-surface stroke during writing. Similarly, total time is strongly correlated to on-surface and in-air time. To mitigate the effect of redundancy and thus to enhance the generalization power, a feature selection algorithm is applied before classification. The discriminating power of subsets of features is evaluated by a recursive feature elimination strategy based on a linear SVM classifier [43]. Briefly speaking, the model is trained on the initial set of features and a ranking criterion in computed for all of them: the features having small criterion are removed from the feature set. This process is iteratively computed until all the features are removed. The final outcome of the algorithm is thus a ranked feature list: feature selection is achieved by choosing a group of top-ranked features. To automatically tune the best number of features to be retained, an internal 3-fold cross-validation is employed.

IV. EXPERIMENT

The final version of the digital test was administered to a sample of healthy and demented subjects to evaluate its effectiveness in supporting dementia assessment.

A. PARTICIPANTS

We acquired data from 65 subjects, both male and female. All subjects were right-handed and had completed at least five years of education. In accordance with their medical history, we excluded, from our study, all subjects showing psychiatric disorders or any injury that could have significantly affected handwriting. Moreover, we excluded those with brain damages, such as vascular or traumatic damages. In accordance with a thorough clinical diagnosis, participants were grouped into 29 healthy control (H) subjects (age: 65 $\pm$ 13) and 36 demented (D) patients (age: 75 $\pm$ 9).

In particular, all participants underwent a battery of neuropsychological tests, including Mini-Mental State Examination, Trail Making Test [44], Mini-Cog [45] and a number of other standard assessments. A thorough medical history and the assessment of independent functions and daily activities were also included. The digital AMT, instead, was not used in the diagnostic process.

All participants first signed an informed consent. Then, the tablet was placed flat in front of them and each participant was read the same set of instructions provided by the paper version of the test. The subjects performed the test without previous knowledge. However, a training process was scheduled before the test to allow the subjects to familiarize with the equipment. These preliminary trials consisted in writing in cursive the word "ciao" (*hello* in Italian), connecting two horizontal and vertical points with a straight line repeatedly and copying a square.

It is worth remarking that, although the examiner asked the subject to perform the tasks as quickly as possible, each task was not interrupted after 45 s, but we let participants to continue until task completion (i.e., until they believed the task to be completed). This practice is usually carried out in the daily clinical trials to evaluate whether the test taker is able to complete the test provided enough time. In other words, this makes easier to distinguish between the slowness of movement and the inability to perform the task at all. Allowing participants to continue may allowed us to capture, to a some extent, this difference. In addition, it must be underlined that clinicians are only interested in evaluating the errors occurred within 45 s; since our analysis concerned with different measures, our goal was to acquire as much data as possible to uncover hidden, possibly discriminating attentional patterns.

We also remark that any point sampled within the first two lines of each matrix was not acquired, since, as previously mentioned, they serve as example and to run-in.

B. VALIDATION

Since the set of data is small, the classification performance was validated through a 5-fold cross-validation. With this scheme, the set of examples is divided into five subsets/folds: one fold is treated as test set; the others form the training set. The overall procedure is repeated five times, until each fold is used as test set once. Note that the splitting was *stratified* by diagnosis, so that each fold contained roughly the same number of subjects from each diagnostic group (healthy vs. demented).

We remark that feature selection was *nested* within cross-validation, so that the most discriminating features were chosen based only on the training set, blindly to the test set. Applying an a priori selection of features on the entire dataset, in fact, inadvertently introduces a bias in the classification model which may lead to overoptimistic results [37].

C. RESULTS

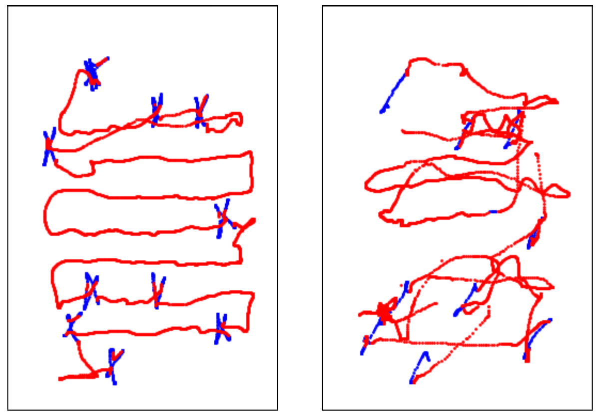
1) EXPLORATORY ANALYSIS

It is worth remarking that the two groups under investigation (healthy vs. demented) have non-comparable age; in addition, the healthy sample age is characterized by a higher standard deviation. To evaluate if the younger healthy subjects affected significantly the between-group differences, we performed a within-group clustering analysis on only the healthy sample. First, we run the k-means algorithm [38] over this sample, by varying the parameter k indicating the number of clusters in which to separate the data. Then, for each data separation, we computed the traditional silhouette coefficient to evaluate if the algorithm found well-defined clusters. This metric ranges from -1 to 1, with high values indicating a good separation, while values near zero indicate overlapping clusters. The silhouette scores obtained range between ~ 0.03 and ~ 0.05, for values of k ranging from 2 to 4. This indicates that there is no well-defined separation between the clusters into which data were partitioned. In other words, there is not a strong tendency to form clusters. This seems to suggest that the healthy sample we considered forms a quite homogeneous group with no significant within-group difference in terms of handwriting performance: this despite the presence of participants with different ages.

Furthermore, to obtain some preliminary insights on which features are relevant for the classification problem at hand, we computed the Spearman's correlation coefficient ρ between the overall feature vectors and the corresponding class labels. Table 1 summarizes the 9 handwriting measures with largest absolute correlation coefficient. It can be noted that 4 of 9 features, included the most correlated ones, come from the first matrix, suggesting that this task may be more discriminating than the other two. Moreover, it can be observed that most of the features concerns the time of completion, in particular the time spent in-air during line scanning. Three features, instead, are related to the randomness of the visual search (horizontal and vertical Shannon entropy).

TABLE 1. Features with largest relevance to class label.

Feature	Matrix	$\lvert\rho\rvert$
In-air time	1	0.471
In-air time	3	0.450
Total time	3	0.450
Total time	1	0.447
In-air time	2	0.447
Total time	2	0.428
Horizontal Shannon entropy	1	0.410
Vertical Shannon entropy	1	0.399
Horizontal Shannon entropy	2	0.377



(a) (b)

FIGURE 4. First matrix performed by a healthy control subject (a) and a demented patient (b). The on-surface movement is in blue color; the in-air movement in red.

Illustrative outcomes of the first matrix as performed by a healthy subject and a demented patient are shown in Fig. 4a and 4b. The visual inspection of only the on-surface traits, although the patient committed an omission error, seems to suggest an unimpaired performance. However, the visual information provided by the in-air movements highlights a less systematic visual search in the demented patient.

The probability density functions of the 9 most highly ranked features from Table 1 are shown in Fig. 5. The vertical axes are the probability densities of the normalized measures

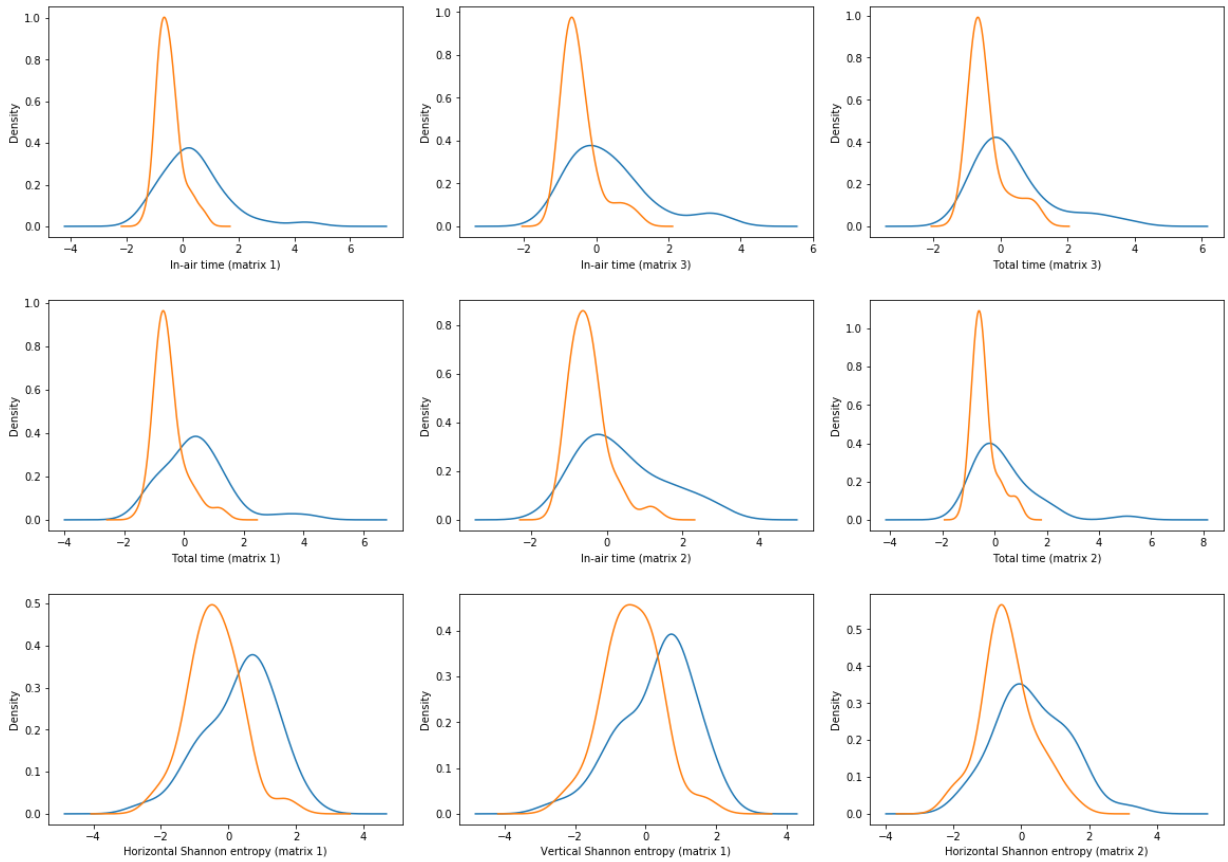


FIGURE 5. Probability densities of the features with largest relevance to class label: orange curves are for healthy subjects; blue curves for demented patients.

TABLE 2. Classification performance on the first matrix.

Classifier	Accuracy	AUC	Sensitivity	Specificity
KNN	73.30%	80.20%	77.78%	68.97%
LR	71.67%	76.29%	72.22%	72.41%
SVM-linear	77.95%	76.67%	72.22%	86.21%
SVM-RBF	77.82%	83.43%	77.78%	79.31%
RF	80.40%	84.26%	83.33%	75.86%

TABLE 3. Classification performance on the second matrix.

Classifier	Accuracy	AUC	Sensitivity	Specificity
KNN	53.00%	63.86%	55.56%	51.72%
LR	76.41%	73.90%	69.44%	86.21%
SVM-linear	71.89%	70.29%	63.89%	82.76%
SVM-RBF	70.22%	71.80%	63.89%	79.31%
RF	70.46%	75.00%	77.78%	62.07%

estimated by using a kernel density estimation with Gaussian kernel. The curves of the measures for the demented patients show a clear difference from the probability densities of the healthy subjects.

2) CLASSIFICATION HEALTHY/DEMENTED

The classification performance are expressed in terms of some traditional performance metrics: accuracy, area under

TABLE 4. Classification performance on the third matrix.

Classifier	Accuracy	AUC	Sensitivity	Specificity
KNN	63.00%	67.35%	69.44%	55.17%
LR	72.01%	75.38%	66.67%	79.31%
SVM-linear	73.55%	77.25%	63.89%	86.21%
SVM-RBF	75.09%	78.49%	72.22%	79.31%
RF	76.52%	69.54%	83.33%	68.97%

TABLE 5. Classification performance of the ensemble.

Classifier	Accuracy	AUC	Sensitivity	Specificity
Ensemble	84.10%	87.30%	86.11%	82.76%

the ROC curve (AUC), sensitivity and specificity. For each metric, the mean value, averaged over all the cross-validation iterations, is reported.

Tables 2–4 reports the results obtained by each classification model on each task. Generally speaking, all classifiers agree that the best performing matrix is the first one. RF achieved the best sensitivity values over all matrices, even if this does not apply to accuracy, as the algorithm was surpassed by LR on the second task. Good performance were generally obtained providing an initial confidence on the effectiveness of their ensemble.

In fact, as expected, the ensemble of the more accurate models (RF for the first and third matrix and LR for the second one), improved the results obtained over the single tasks (Table 5). In particular, an AUC of 87.30% and a sensitivity of 86.11% were achieved. Interestingly, the individual matrices generally showed specificity higher than sensitivity. An inversion, instead, can be observed in the ensemble of classifiers, where sensitivity is higher than specificity.

In addition, in order to mitigate the bias due to the non-comparable age between the groups, we applied the same classification setting to a sub-sample of subjects including 18 H and 18 D with an age of 72 ± 11 and 71 ± 7, for the healthy and pathological group respectively. The ensemble over the three matrices resulted in a prediction accuracy of 82.50%, an AUC of 83.61%, a sensitivity of 72.22% and a specificity of 94.44%. A slight performance deterioration can be observed in the values of accuracy and AUC; specificity, instead, increased, at the expense of a detrimental effect to sensitivity. We acknowledge that this set of data was very small, so we warn the reader to use caution when interpreting these results, since the resulting model may have developed a low generalization power.

V. DISCUSSION AND CONCLUSION

In this paper, a digital, mobile variant of the well-known Attentional Matrices test for cognitive impairment assessment has been proposed. The digitalization of such a test allows one to capture a larger set of performance measures than those that can be obtained by only observing the paper-based test. In particular, these measures are able to capture characteristics of the dynamics of the writing process carried out during the test execution.

In order to investigate if and to which extent the digital AMT we propose can support the discrimination between demented patients and controls, a classification study was carried out on a sample of subjects, including healthy elders and individuals suffering from dementia as well. Both an exploratory analysis performed before classification and classification performance indicated that the most discriminating matrix is the first one. This contrasts with [14], in which the authors claimed the second and third matrix to be more effective. This may be explained considering that the simplicity of the first matrix may have emphasized the differences between the groups in terms of the handwriting measures here adopted (in particular, in-air movement and its irregularities). The second and third matrix, instead, may have been perceived much more difficult by both groups, so thinning the differences in these measures. Note that the use of only the first matrix was preferred in the Milan Overall Dementia Assessment (MODA) test [46] for reasons of simplicity. Another possible motivation is that the adopted features may have overlooked patterns of visual search that could have been captured by other measures. Future work should explore more in depth this issue.

In our study, the variable best for use in screening for cognitive impairment is reflected in the prolonged in-air time. In-air trajectories, in fact, are tailored to reflect the impairments of demented patients from the point of view of selective attention. The usefulness of in-air movement analysis to support pathology evaluation has been exploited also in other works, for example [47] and [48].

The highest classification performance was obtained with the combination of all the three tasks, suggesting that the three matrices are all important for classification. In particular, very good AUC (87.30%) and very good sensitivity (86.11%) were achieved. This suggests that a screening routine test based on our proposal may be able to correctly detect dementia signs in the pathological population, and thus may be useful for ruling out disease when resulting in a negative response. Another possible interpretation is that the unbalanced dataset (29 H vs. 36 D) may have led to a predictive preference for the majority class, even if the over-representation of the pathological sample is quite small.

The digital test used in this study has the potential to assist clinicians at the point of care, providing a novel diagnostic tool while reducing the expenditure of public health. It can also be used to quantify aspects of the motor system and its disorders in order to better understand the mechanisms underlying the neurodegeneration, e.g. the difficulties in coordinating the components of a motor sequence movement. Furthermore, it can help study the effects of medication on handwriting (e.g., [49]): these effects can be studied to track the disease progression as well as to monitor the responsiveness of the patient to novel therapies. More in general, tablet technology can provide a cost-effective, user-friendly, easy-to-use and even customizable tool to support the daily clinical trials. Such a technology would also provide a means for more accurate data recordings, paper saving, rapid administration and fast recall. Finally, a purely automatic diagnostic tool paves the way for a quick instrument for a mass screening of the population.

Tablet technology enables the implementation of a multimodal interaction system in which not only the input provided by the electronic pen, but also tactile, speech or visual input can be acquired. The information related to handwriting, in fact, may complement the information coming from speech, as well other biometric traits, providing different findings which may support novel clinical insights and a better understanding of the pathology. For example, research is ongoing (e.g., [50]) to develop a common framework for the evaluation of Parkinson's disease based on both voice and handwriting. Another key feature of the proposed method is its extensibility, as the implemented task is decoupled from the logic of the application. Thanks to this, the acquisition protocol may be extended with different tasks useful to investigate other cognitive domains.

The major limitation of this study is the small size of the population under investigation, which restricts the reliability of our conclusions. Moreover, the results obtained may have been biased by the difference between the groups in terms of age, although no significant difference in handwriting was found between participants of different age within the healthy group. In addition, results comparable to those obtained with all data were achieved with sub-samples of the groups characterized by a comparable age. Future work should address these issues. In addition, future developments of the present research should explore other classification strategies to further improve the prediction performance. Novel insights, for example, could be obtained by considering other kinematic properties of the human motor control as features. For instance, in the present work, we employed measures tailored to analyze the task execution only from a global perspective. Local attentional patterns may be then investigated to provide additional insights. Novel findings and better performance could also be obtained by combining dynamic to static features of handwriting, that are the ones based on static images of the patterns acquired.

REFERENCES

[1] J. Dewing and S. Dijk, "What is the current state of care for older people with dementia in general hospitals? A literature review," *Dementia*, vol. 15, no. 1, pp. 106–124, 2016.

[2] Alzheimer's Association, "2017 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 13, no. 4, pp. 325–373, 2017.

[3] J. Gratwicke, M. Jahanshahi, and T. Foltynie, "Parkinson's disease dementia: A neural networks perspective," *Brain*, vol. 138, no. 6, pp. 1454–1476, 2015.

[4] G. Bates, S. Tabrizi, and L. Jones, *Huntington's Disease*. Oxford, U.K.: Oxford Univ. Press, USA, 2014, p. 64.

[5] R. Sperling and K. Johnson, "Biomarkers of Alzheimer disease: Current and future applications to diagnostic criteria," *CONTINUUM, Lifelong Learn. Neurol.*, vol. 19, no. 2, pp. 325–338, 2013.

[6] M. Folstein, L. Robins, and J. Helzer, "The mini-mental state examination," *Arch. General Psychiatry*, vol. 40, no. 7, p. 812, 1983.

[7] A. W. Wind, F. G. Schellevis, G. Van Staveren, R. J. P. M. Scholten, C. Jonker, and J. T. M. Van Eijk, "Limitations of the mini-mental state examination in diagnosing dementia in general practice," *Int. J. Geriatric Psychiatry*, vol. 12, no. 1, pp. 101–108, 1997.

[8] J. S. Shiroky, H. M. Schipper, H. Bergman, and H. Chertkow, "Can You Have Dementia With an MMSE Score of 30?" *Amer. J. Alzheimer's Disease Dementias*, vol. 22, no. 5, pp. 406–415, 2007.

[9] D. O'Connor, P. Pollitt, J. Hyde, C. Brook, B. Reiss, and M. Roth, "Do general practitioners miss dementia in elderly patients?" *Proc. BMJ*, vol. 297, no. 6656, pp. 1107–1110, 1988.

[10] P. Schatz and J. Browndyke, "Applications of computer-based neuropsychological assessment," *J. Head Trauma Rehabil.*, vol. 17, no. 5, pp. 395–410, 2002.

[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. B, Psychol. Sci. Social Sci.*, vol. 61, no. 4, pp. P228–P236, 2006.

[12] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.

[13] H. Spinnler and G. Tognoni, "Italian group on the neuropsychological study of ageing: Italian standardization and classification of neuropsychological tests," *Ital J. Neurol. Sci.*, vol. 6, no. 8, pp. 1–120, 1987.

[14] S. D. Sala, M. Laiacona, H. Spinnler, and C. Ubezio, "A cancellation test: Its reliability in assessing attentional deficits in Alzheimer's disease," *Psychol. Med.*, vol. 22, no. 4, pp. 885–901, 1992.

[15] 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. 2018.

[16] 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, 2018.

[17] J. Dahmen, D. Cook, R. Fellows, and M. Schmitter-Edgecombe, "An analysis of a digital variant of the trail making test using machine learning techniques," *Technol. Health Care*, vol. 25, no. 2, pp. 251–264, 2017.

[18] S. Müller, O. Preische, P. Heymann, U. Elbing, and C. Laske, "Increased diagnostic accuracy of digital vs. conventional clock drawing test for discrimination of patients in the early course of Alzheimer's disease from cognitively healthy individuals," *Frontiers Aging Neurosci.*, vol. 9, p. 101, Apr. 2017.

[19] A. Lleó *et al.*, "Longitudinal cerebrospinal fluid biomarker trajectories along the Alzheimer's disease continuum: A multicentre European study," *Alzheimer's Dementia, J. Alzheimer's Assoc.*, vol. 13, no. 7, p. P924, 2017.

[20] E. Lella, N. Amoroso, A. Lombardi, T. Maggipinto, S. Tangaro, and R. Bellotti, "Communicability disruption in Alzheimer's disease connectivity networks," *J. Complex Netw.*, vol. 7, no. 1, pp. 83–100, 2018.

[21] Y. Zhang, H. Zhang, X. Chen, S.-W. Lee, and D. Shen, "Hybrid high-order functional connectivity networks using resting-state functional MRI for mild cognitive impairment diagnosis," *Sci. Rep.*, vol. 7, p. 6530, Jul. 2017.

[22] Y. Zhang *et al.*, "Strength and similarity guided group-level brain functional network construction for MCI diagnosis," *Pattern Recognit.*, vol. 88, pp. 421–430, Apr. 2019.

[23] T. Zhou, K.-H. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Hum. Brain Mapping*, vol. 40, no. 3, pp. 1001–1016, 2019.

[24] T. Zhou, K.-H. Thung, M. Liu, and D. Shen, "Brain-wide genome-wide association study for Alzheimer's disease via joint projection learning and sparse regression model," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 1, pp. 165–175, Jan. 2019.

[25] Z. Jin, G. Zhou, D. Gao, and Y. Zhang, "EEG classification using sparse Bayesian extreme learning machine for brain–computer interface," *Neural Comput. Appl.*, pp. 1–9, Oct. 2018.

[26] G. Zhou *et al.*, "Linked component analysis from matrices to high-order tensors: Applications to biomedical data," *Proc. IEEE*, vol. 104, no. 2, pp. 310–331, Feb. 2016.

[27] M. Faundez-Zanuy *et al.*, "Biometric applications related to human beings: There is life beyond security," *Cognit. Comput.*, vol. 5, no. 1, pp. 136–151, 2013.

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

[29] M. Moetesum, I. Siddiqi, N. Vincent, and F. Cloppet, "Assessing visual attributes of handwriting for prediction of neurological disorders—A case study on Parkinson's disease," *Pattern Recognit. Lett.*, vol. 121, pp. 19–27, Apr. 2019.

[30] C. R. Pereira *et al.*, "Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification," *Artif. Intell. Med.*, vol. 87, pp. 67–77, May 2018.

[31] P. Ghaderyan, A. Abbasi, and S. Saber, "A new algorithm for kinematic analysis of handwriting data; towards a reliable handwriting-based tool for early detection of Alzheimer's disease," *Expert Syst. Appl.*, vol. 114, pp. 428–440, 2018.

[32] C. Kahindo, M. A. El-Yacoubi, S. Garcia-Salicetti, A.-S. Rigaud, and V. Cristancho-Lacroix, "Characterizing early-stage Alzheimer through spatiotemporal dynamics of handwriting," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1136–1140, Aug. 2018.

[33] M. A. El-Yacoubi, S. Garcia-Salicetti, C. Kahindo, A.-S. Rigaud, and V. Cristancho-Lacroix, "From aging to early-stage Alzheimer's: Uncovering handwriting multimodal behaviors by semi-supervised learning and sequential representation learning," *Pattern Recognit.*, vol. 86, pp. 112–133, Feb. 2019.

[34] R. Jääskeläinen, "Think-aloud protocol," *Handbook Translation Stud.*, vol. 1, pp. 371–374, 2010.

[35] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, and I. Schlesinger, "Handwriting as an objective tool for Parkinson's disease diagnosis," *J. Neurol.*, vol. 260, no. 9, pp. 2357–2361, 2013.

[36] D. Impedovo, G. Pirlo, and G. Vessio, "Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis," *Information*, vol. 9, no. 10, p. 247, Oct. 2018.

[37] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2009.

[38] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 112. New York, NY, USA: Springer, 2013.
[39] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 2013.
[40] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[41] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[42] R. Raeper, A. Lisowska, and I. Rekik, "Cooperative correlational and discriminative ensemble classifier learning for early dementia diagnosis using morphological brain multiplexes," *IEEE Access*, vol. 6, pp. 43830–43839, 2018.
[43] 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, 2002.
[44] R. M. Reitan, "Validity of the trail making test as an indicator of organic brain damage," *Perceptual Motor Skills*, vol. 8, no. 3, pp. 271–276, 1958.
[45] S. Borson, J. Scanlan, M. Brush, P. Vitaliano, and A. Dokmak, "The mini-cog: A cognitive 'vital signs' measure for dementia screening in multi-lingual elderly," *Int. J. Geriatric Psychiatry*, vol. 15, no. 11, pp. 1021–1027, 2000.
[46] M. Brazzelli, E. Capitani, S. D. Sala, H. Spinnler, and M. Zuffi, "A neuropsychological instrument adding to the description of patients with suspected cortical dementia: The Milan overall dementia assessment," *J. Neurol., Neurosurgery Psychiatry*, vol. 57, no. 12, pp. 1510–1517, 1994.
[47] E. Sesa-Nogueras, M. Faundez-Zanuy, and J. Mekyska, "An information analysis of in-air and on-surface trajectories in online handwriting," *Cognit. Comput.*, vol. 4, no. 2, pp. 195–205, 2012.
[48] C. Alonso-Martinez, M. Faundez-Zanuy, and J. Mekyska, "A comparative study of in-air trajectories at short and long distances in online handwriting," *Cognit. Comput.*, vol. 9, no. 5, pp. 712–720, 2017.
[49] O. Tucha *et al.*, "Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease," *J. Neural Transmiss.*, vol. 113, no. 5, pp. 609–623, 2006.
[50] M. Faundez-Zanuy *et al.*, "A multimodal approach for Parkinson disease analysis," in *Advances in Neural Networks: Computational and Theoretical Issues*. Cham, Switzerland: Springer, 2015, pp. 311–318.

MARIA TERESA ANGELILLO is currently a member of the Research Unit in Affective and Social Neuroscience, Catholic University of Milan, Italy. Her research interests concern clinical neuropshychology, social and well-being neuroscience, and medical informatics. Her studies use different recording and stimulation tools (EEG, biofeedback, ERP, and computer-aided diagnosis) for psychophysiological measures and non-invasive brain stimulation techniques (tDCS), using a multimodal approach. Following her specialization in cognitive behavioral psychology and her studies in neuropsychology, psychophysiology, and medical informatics, she currently works on neuropsychology applied to neurodegenerative diseases and severe traumatic brain injury (TBI) in disorders of consciousness (DOCs). She gained well-established expertise in neuropsychological assessment, biosignal recording and processing (autonomic indices and EEG/ERPs), and quantitative–qualitative analysis. She is a member of the Italian Society of Neuropsychology (SINP) and an Associate Member of the Italian Association of Psychology (AIP). She is involved in national research projects.

FABRIZIO BALDUCCI received the B.Sc. degree in computer science from the Computer Science Department, University of Bari, Italy, the M.Sc. degree (Hons.) from the University of Bari, in 2014, and the Ph.D. degree in computer engineering from the Computer Engineering Department, University of Modena and Reggio Emilia, in 2018. In 2014, he moved to the Computer Engineering Department, University of Modena and Reggio Emilia. He is currently a Postdoctoral Fellow with the University of Bari, Italy. His research interests include pattern recognition, machine learning, affective ludology, gamification, human–computer interaction, and information visualization. He serves as a Reviewer for international journals, such as the IEEE Access and *MDPI Information*, while his academic services include reviewing activity in the program committee of many international conferences, such as the IEEE GEM 2018 and AEIT SCSE 2018.

DONATO IMPEDOVO (M'08–SM'17) received the M.Eng. degree *(cum laude)* in computer engineering and the Ph.D. degree in computer engineering. He is currently an Associate Professor with the Computer Science Department, University of Bari, Italy. His research interests are in the field of signal processing, pattern recognition, machine learning, and biometrics. He has coauthored more than 80 articles in these fields in both international journals and conference proceedings. He is an IAPR Senior Member. He received the "Distinction" Award, in 2009, at the International Conference on Computer Recognition Systems (CORES-endorsed by IAPR), and the First Prize of the first Nereus-Euroavia Academic Competition on GMES, in 2012. He was the General Co-Chair of the International Workshop on Smart Cities and Smart Enterprises (SCSE 2018), the International Workshop on Artificial Intelligence with Application in Health (WAIAH 2017), the International Workshop on Emergent Aspects in Handwritten Signature Processing (EAHSP 2013), and the International Workshop on Image-Based Smart City Application (ISCA 2015). He is also involved in research transfer activities as well as in industrial research. He has managed more than 25 projects funded by public institutions as well as by private SMEs. He serves as a Reviewer and a Rapporteur for the EU and national projects evaluation. He is an Associate Editor of the IEEE Access. He serves as a Reviewer for many international journals, including the IEEE T-SMC, IEEE T-TETC, *Pattern Recognition*, *Information Sciences*, and many others. He was an Editor of the Special Issue on "Artificial Intelligence Applications to Smart City and Smart Enterprise" of the *MDPI Applied Sciences* journal, in 2019, the Special Issue on "eHealth and Artificial Intelligence" of the *MDPI Information* journal, in 2018, the Special Issue on "Handwriting Recognition and Other PR Applications" of the *Pattern Recognition* journal, in 2014, and the Special Issue on "Handwriting Biometrics" of the *IET Biometrics Journal*, in 2014. He was a Reviewer in the Scientific Committee and the Program Committee of many international conferences in the field of computer science, pattern recognition, and signal processing, such as ICPR and ICASSP.

GIUSEPPE PIRLO received the degree *(cum laude)* in computer science from the Computer Science Department, University of Bari, Italy, in 1986. Since 1986, he has been carrying out research in the field of computer science and neuroscience, signal processing, handwriting processing, automatic signature verification, biometrics, pattern recognition, and statistical data processing. Since 1991, he has been an Assistant Professor with the Computer Science Department, University of Bari, where he is currently a Full Professor. He developed several scientific projects and has authored over 250 papers on international journals, scientific books, and proceedings. He is a member of the Governing Board of the Consorzio Interuniversitario Nazionale per l'Informatica (CINI) and the Governing Board of the Società Italiana di e-Learning and the e-Learning Committee of the University of Bari. He is currently the Deputy Representative of the University of Bari in the Governing Board of CINI. He is also the Managing Advisor of the University of Bari for the Digital Agenda and Smart Cities. He is the Chair of the Associazione Italiana Calcolo Automatico-Puglia. He is also a member of the Gruppo Italiano Ricercatori Pattern Recognition, the International Association Pattern Recognition, the Stati Generali dell'Innovazione, and the Gruppo Ingegneria Informatica. He was the General Chair of the International Workshop on Emerging Aspects in Handwriting Signature Processing, Naples, in 2013, and the International Workshop on Image-based Smart City Applications, Genoa, in 2015, and the General Co-Chair of the International Conference on Frontiers in Handwriting Recognition, Bari, in 2012. He was a Reviewer in the Scientific Committee and the Program Committee of many international conferences in the field of computer science, pattern recognition, and signal processing, such as ICPR, ICDAR, ICFHR, IWFHR, ICIAP, VECIMS, and CISMA. He is also the Editor of several books. He was an Editor of the Special Issue ''Handwriting Recognition and Other PR Applications'' of the *Pattern Recognition* journal, in 2014 and the Special Issue ''Handwriting Biometrics'' of the *IET Biometrics Journal*, in 2014. He was the Guest Co-Editor of the Special Issue of the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS on Drawing and Handwriting Processing for User-Centered Systems, in 2017. He was also the Guest Editor of the Special Issue of the Je-LKS journal on Steps toward the Digital Agenda: Open Data to Open Knowledge, in 2014. He is currently an Associate Editor of the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS. He also serves as a Reviewer for many international journals, including the IEEE T-PAMI, IEEE T-FS, IEEE T-SMC, IEEE T-EC, IEEE T-IP, IEEE T-IFS, *Pattern Recognition*, IJDAR, and IPL.

GENNARO VESSIO received the M.Sc. degree (Hons.) in computer science and the Ph.D. degree in computer science and mathematics from the Computer Science Department, University of Bari, Italy, in 2013 and 2017, respectively, where he is currently a Postdoctoral Fellow. His research interests include pattern recognition, machine learning, formal methods, and network protocols. He has coauthored articles in these fields, appeared in both international journals and conferences, such as *MDPI Information* and IJPEDS. He served as a Reviewer for many international journals, including *Information Sciences* and *Information and Software Technology* and has been a part of the Program Committee of international conferences, such as SEKE.