2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A new scheme for the automatic assessment of Alzheimer's disease on a fine motor task with Transfer Learning*

M. Kachouri, N. Houmani, S. Garcia-Salicetti, and A.-S. Rigaud

Abstract— **We present a new scheme for Alzheimer's Disease (AD) automatic assessment, based on Archimedes spiral, drawn on a digitizing tablet. We propose to enrich spiral images generated from the raw sequence of pen coordinates with dynamic information (pressure, altitude, velocity) represented with a semi-global encoding in RGB images. By exploiting Transfer Learning, such hybrid images are given as input to a deep network for an automatic high-level feature extraction. Experiments on 30 AD patients and 45 Healthy Controls (HC) showed that the hybrid representations allow a considerable improvement of classification performance, compared to those obtained on raw spiral images. We reach, with SVM classifiers, an accuracy of 79% with pressure, 76% with velocity, and 70.5% with altitude. The analysis with PCA of internal features of the deep network, showed that dynamic information included in images explain a much higher amount of variance compared to raw images. Moreover, our study demonstrates the need for a semi-global description of dynamic parameters, for a better discrimination of AD and HC classes. This description allows uncovering specific trends on the dynamics for both classes. Finally, combining the decisions of the three SVMs leads to 81.5% of accuracy.**

Clinical Relevance— **This work proposes a decision-aid tool for detecting AD at an early stage, based on a non-invasive simple graphic task, executed on a Wacom digitizer. This task can be considered in the battery of usual clinical tests.**

I. Introduction

Characterizing Alzheimer's disease (AD) at an early stage is a challenge: on one hand, the onset of the disease is insidious; on the other hand, the heterogeneity of AD profiles is important.

AD starts by memory impairment and evolves progressively into motor deficits. The loss of fine motor skills has been investigated in the literature to identify behavioral markers of neurodegenerative diseases [1-7]. Digital tablets allow collecting handwritten productions as spatiotemporal signals ("online" signals), conveying precious dynamic information about the writing gesture. The online signal has been intensively studied since the 1980s through the analysis of fine movements on a digitizer, mostly for Parkinson disease (PD), on different tasks with no semantic content, such as drawing Archimedes spiral, concentric circles, or moving the stylus as fast as possible between two targets [1-3]. Later, the same tasks were used for automatic AD assessment [4-7].

All such works extracted global spatiotemporal markers (as peak velocity) on segments (strokes), and assessed whether there is a significant difference between healthy controls (HC) and pathological populations. This assessment was based on a statistic on the whole gesture of stroke-based features (e.g. its Signal-to-Noise Ratio, the ratio of the mean value of each feature and its standard deviation [6]). The objective of such pioneering works was to uncover specific markers of pathology among handcrafted spatiotemporal features, not performing automatic assessment of the disease. But some of these works were contradictory on which markers were significantly different between healthy and pathological populations. Slavin *et al.* [6] (on loops) and Schröter *et al.* [5] (on superimposed circles) showed that AD patients present more variable peak velocity across strokes compared to HC. Such strokes are defined as half loops in [6] and half circles in [5]. However, Yu *et al.* [7] did not find significant differences in velocity on loops and circles. These contradictory results were due to the small size of databases, and to the fact that dynamic features were extracted globally, first on segments and then on the whole gesture.

More recently, with the emergence of Deep Learning, studies on pathological aging have exploited images of graphical handwritten productions with deep networks. Such works, devoted to PD [8-12], combined the decisions of different classifiers (deep networks or Support Vector Machines) to take a decision on the health status of a subject (classification). The main contribution of this new trend of research is the automatic extraction of features by Deep Neural Networks. This contribution is significant: it is indeed very difficult to find out the proper set of features through classical methods; it requires strong a priori knowledge on the problem to design the feature extraction process. It also requires performing feature selection, sometimes too costly. The work in [13] on PD assessment illustrates well the problem of a handcrafted feature extraction: still after a high cost sequential feature selection process, several hundreds of features are retained.

*Research supported by Institut Mines-Telecom.

M. Kachouri, Télécom SudParis, Institut Polytechnique de Paris, France (marouane.kachouri@telecom-sudparis.eu).
N. Houmani, Associate professor, SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France (nesma.houmani@telecom-sudparis.eu).
S. Garcia-Salicetti, Professor, SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France (sonia.garcia@telecom-sudparis.eu).
A.-S. Rigaud, AP-HP, Groupe Hospitalier Cochin Paris Centre, Pôle Gérontologie, Hôpital Broca, Paris 75013, France.

The present work is focused on AD assessment, never explored to our knowledge in the recent literature on deep learning. Our data, collected at Broca Hospital in Paris with a specific protocol involving several graphomotor tasks, consists in spirals of 30 AD patients at an early stage, and 45 HC. As for other works in the literature, the amount of data is rather small, and the heterogeneity of profiles in both AD and HC classes is high.

In this study, we propose an original scheme based on a hybrid representation of spiral images, combining visual and sequential dynamic information of the drawn spirals. This approach allows exploiting Transfer Learning, useful in case of sparse data, for the automatic extraction of features. Indeed, Transfer Learning can cope with the small size of the data and extract automatically pertinent high-level features [8], based on previous knowledge extracted on thousands of images of the ImageNet database [14]. We propose to feed the deep network with spiral images generated from raw sequences of pen coordinates, and later enriched with dynamic information inherited from the online handwritten signal. Different types of dynamic information were included generating RGB color images: pen pressure, pen inclination and pen velocity. Then, one expert, a Support Vector Machine (SVM) classifier, is trained on the internal features extracted by the deep network on each type of enriched image given as input, in order to take a decision on the health status of the subject. Our results show the pertinence of exploiting Transfer Learning with hybrid representations of spiral images: a significant improvement of performance is obtained on the enriched spiral images when compared to raw images. Furthermore, the fusion of the three experts' decisions improves performance, showing that our strategy allows coping with the strong intra-class variability of motor behavior observed in both AD and HC populations.

In our paper, Section 2 presents related works on Transfer Learning and deep learning for pathology assessment with graphomotor tasks. Section 3 presents the database and explains the methods and the experimental setup. Section 4 states our results and analysis, and finally conclusions and perspectives for future work are given in Section 5.

II. Related works

To our knowledge, up to now, there is no work on AD assessment exploiting deep learning techniques on handwritten graphical tasks. The majority of works in the literature on pathology assessment with deep learning are focused on PD. Their aim is to detect essential tremor by means of graphomotor tasks.

Two strategies have emerged in the literature on graphomotor tasks: on one hand, Transfer Learning is exploited only for high-level feature extraction, and the health status of the subject is assessed with another classifier; on the other hand, deep networks are trained and used as classifiers.

A. Transfer Learning for PD assessment

Transfer Learning allows exploiting deep neural networks that were pre-trained for classification on a given problem, as feature extractors or classifiers for another problem [16]. In [8], a pre-trained CNN, AlexNet, originally trained on ImageNet dataset for a classification task of 1000 classes [14], was used as feature extractor for PD assessment. To our knowledge, this is the only work in the literature on pathology assessment by graphomotor tasks, exploiting Transfer Learning. On spiral images and other eight graphomotor tasks, [8] showed the potential of Transfer Learning for extracting pertinent features from raw images of handwritten inputs. For each input image to the network, corresponding to one task, features are extracted at *fc7* layer (4096 features) and later used to train a SVM expert. The architecture combines different experts, and three representations of each image are considered (raw, median residual, edge) for early fusion (low-level fusion into a single feature vector). Majority voting is used to take the final decision on the eight tasks including the spiral. A good performance is reached on the PaHaW database [8], when combining the three representations of the input data and considering different tasks (letter, bigrams, four words and a sentence, plus the spiral). Indeed, with only one image representation, combining the predictions of all tasks showed rather poor results (at most 68% of accuracy). We noted that the spiral task gave the best results among all, reaching 76% of accuracy with the three representations combined. Unfortunately, only the accuracy value was given in [8], making impossible to assess in a refined way the obtained performance, in terms of sensitivity and specificity measures.

B. Deep Learning for PD assessment

Other works have trained and exploited deep networks as classifiers. In [10], PD detection is studied on graphomotor tasks, among which spirals, acquired with a Biometric Smart Pen [9]. The pen captures six time series: fingergrip, axial pressure, tilt and acceleration in the three directions. For each task, the six time series are converted into a grey-level image, where rows are time in milliseconds and columns represent the six signal channels. As the length of time sequences may differ from one person to another, a normalization of sequences was necessary to map time series into an image of fixed size. CNNs are used to classify these specific images of the time series, using two architectures: ImageNet (5 convolution layers, 5 pooling layers, 2 normalization layers) and a more shallow architecture, CIFAR-10 (3 convolution layers and 3 pooling layers). Several baseline classifiers are used, showing again that the spiral task alone is among the most discriminative, and that the best performance is obtained with the CNN-ImageNet architecture (78.26% of accuracy). Unfortunately, sensitivity and specificity measures are not available for further analysis, maybe because the database is very unbalanced (74 PD patients and 18 HC). In spite of this, such results point out the capacity of CNNs to extract pertinent information for pathology assessment compared to other classical methods.

In [9], the same architectures are compared to other more shallow, as LeNet (2 convolution layers and 2 pooling layers), on a smaller database of 14 PD patients and 21 HC. With two evaluation protocols splitting training and test data (50% for training and 50% for testing, versus 75% for training and 25% for testing), results show that with the latter protocol, on spirals, the CNN-ImageNet gives 80.19% of accuracy, CIFAR-10 still remains at a high level of performance (78.31%) but LeNet drops at 43.64%. Thus, the deep architecture cannot be too shallow to extract pertinent information from the mapped images.

Finally, Taleb *et al.* [12] followed the same idea of generating images from online handwriting time series, but proposed an alternative for this normalization applied on the length of time series. On PDmultiMC database of 21 PD patients and 21 HC, they used 2D spectrogram images of each feature time series. They first studied a CNN with images of a fixed size, as in [10], and fused the decisions of different experts by a voting approach on 7 tasks (cursive loops, 2 sorts of waves, 2 single tasks of one word, name and family name). Note that there is no spiral task. They also proposed a CNN for feature extraction directly on time series (without converting the sequences into images), using one-dimensional kernels that move through the raw sequence. The output of the CNN, seen as a sequence of vectors, is given as input to a dynamic BLSTM for sequence prediction. The same accuracy is obtained with both proposals (accuracy of 83.33%) after late fusion of all tasks. The CNN approach on spectrogram images seems more reliable (sensitivity of 85.71% and specificity of 80.95%) than the hybrid CNNs-BLSTM model, which gives very unbalanced performance measures: a much lower sensitivity of 71.43% while specificity reaches 95.24%, showing the difficulty in detecting PD patients.

It is difficult to compare all the above-mentioned approaches on PD assessment, since the datasets used are not the same. But, it is clear that pathology assessment with such models is feasible. A novel trend in research emerges among these studies, exploiting deep learning techniques and enriching the input images with raw time series information. The contribution of our work is two-fold. First, this is the only study addressing AD detection based on deep learning, with a single graphomotor task, Archimedes spiral, whose effectiveness is proven for pathology assessment [8,10]. Second, we adopt a novel strategy based on a hybrid modality that consists in embedding directly in spiral images dynamic information of the drawing gesture. In this way, both the image of the produced spiral and the dynamic process associated to its production are simultaneously given as input to the deep network for high-level feature extraction.

III. METHODOLOGY AND EXPERIMENTAL SETUP

A. Data and acquisition protocol

Our private dataset was acquired at Broca Hospital in Paris, in the framework of ALWRITE project [17-19], a French research project on handwriting analysis for AD assessment. All participants freely signed a consent form after receiving information on the study. The participants were asked to draw one Archimedes spiral, among other graphical tasks, on a sheet of a paper fixed on a Wacom Intuos Pro Large tablet, using an inking pen. Examples of spirals from HC and AD patients are displayed in Figures 1 and 2 respectively.

Figure 1. HC raw spiral images generated from coordinate sequences.

Figure 2. AD raw spiral images generated from coordinate sequences.

The tablet captures, with a sampling rate of 125 Hz, five temporal functions: pen coordinates (x,y), pen pressure (P), Azimuth (Az) and Altitude (Alt) pen inclination angles. It also captures the in-air trajectory of the pen (pen-ups) up to 1 cm off the tablet surface.

Our database includes 30 early-stage AD patients and 45 HC subjects, with a mean age of 80.2 ± 8.8 and 73.5 ± 6.1 respectively. Each of the 75 participants performed one spiral. AD patients were diagnosed based on DSM-5 criteria [15], and considered as having an early-stage AD if their MMSE was over 20. HC performed neuropsychological tests to ensure their cognitive profile is normal. Participants with medical problems such as stroke and other neurodegenerative diseases were not included. The average MMSE is 22.1 ± 4 for AD patients and 29 ± 0.98 for HC.

B. Generating hybrid images: an alternative approach

We use hybrid images generated from the online sequence of pen coordinates produced by each participant. First, we generate the image of a spiral as in [8], by exploiting pen coordinates information through time. This results in the images shown in Figures 1 and 2 when displaying only pen-down trajectories (when the pen touches the tablet surface).

Then, we introduce dynamic information directly on the spiral, in two different ways: locally (pointwise in the pen trajectory), and more globally considering meaningful resolution levels after a quantization process. We exploit the pen pressure (P) and altitude angle (Alt) captured by the digitizer. Also, we extract pen local velocity values since it is a good marker for pathology detection [5,17,18].

We use K-means clustering to generate K meaningful resolution levels (K clusters) of temporal functions, considering all the values of the given features for the whole population. Each cluster is represented by its mean value, following a heat map. Examples of hybrid images with K=3 pressure levels are given in Figures 3 and 4. In-air trajectories are displayed in dotted red points.

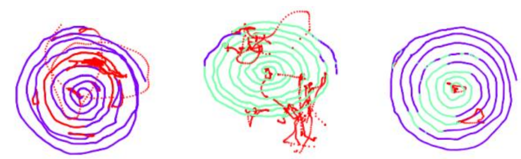

Figure 3. AD spiral images. Red points stand for zero pressure values, green and purple for medium- and high- pressure levels, respectively.

Figure 4. HC spiral images. Red Points stand for zero pressure values, green and purple for medium- and high- pressure levels, respectively.

Spiral color images are then fed to AlexNet for automatic feature extraction. AlexNet is indeed able to process RGB images, and we take advantage of this in a Transfer Learning framework, for high-level feature extraction.

C. *Automatic feature extraction and evaluation set-up*

As shown in Figure 5, we consider 4096 features extracted at *fc7* layer from AlexNet, to perform classification with a SVM. Such features were extracted for each type of hybrid image given as input to the deep network. According to the type of hybrid image, the output feature vector is used to train its corresponding SVM expert. We use a RBF kernel to manage the high dimension of the feature vector. Since we investigate three hybrid representations of spiral images, including pressure, altitude and velocity information, we build three experts to analyze the effectiveness of each representation for AD assessment.

As we have at disposal one spiral per person, for 45 HC and 30 AD patients, we perform 10 random samplings of 30 HC among the 45 HC, in order to have the same number of samples per class when training the SVM classifier. In each sampling, the SVM is trained on 20 HC and 20 AD and tested on the remaining 10 HC and 10 AD. Average performance is given in terms of accuracy, specificity (percentage of HC well classified) and sensitivity (percentage of AD patients well classified), considering the 10 samplings, along with their standard deviations.

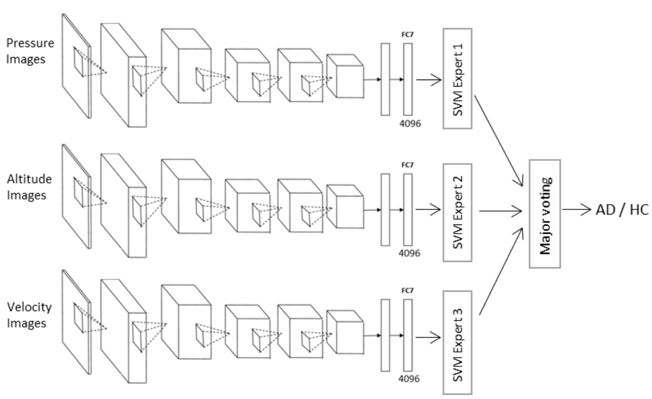

Figure 5. System overview of the three AlexNet networks and fusion of experts' decisions.

IV. EXPERIMENTS AND RESULTS

A. *Raw pen-down spiral images vs. in-air hybrid images*

Table I shows the performance of the SVM classifier based on the features extracted by AlexNet, when considering raw spiral images generated from the online sequence considering only pen-down trajectories, and also spiral images with in-air trajectories. Note that in the literature, only raw spiral images are usually fed to AlexNet for automatic feature extraction [8].

With raw spiral images, sensitivity and specificity values are very unbalanced, revealing numerous false negatives. When considering pen-ups on such spiral images, the sensitivity shows an absolute improvement of 11% at the price of a lower specificity. The accuracy is thus slightly improved compared to the raw images, reaching 72%, but the standard deviation is decreased.

TABLE I. PERFORMANCE ON RAW SPIRAL IMAGES AND SPIRAL IMAGES WITH IN-AIR TRAJECTORY INFORMATION

Performance	**Sensitivity**	**Specificity**	**Accuracy**
Raw spiral (Baseline)	50 ± 19.5	88 ± 6	69 ± 8.8
Raw spiral with pen-ups	61 ± 11.3	83 ± 12.6	72 ± 6.4

In the following, we propose to introduce, additionally to pen-ups, pressure information on pen-down trajectories, pointwise and later at different resolution levels (after a quantization process, as explained in Section III.B). In the same way, we also exploit altitude and velocity information.

B. *Pressure-based hybrid images*

For pressure-based hybrid images (Figures 3 and 4), Table II shows the considerable improvement of classifier performance, obtained on images containing pressure values taken pointwise, comparatively to raw images. The sensitivity is improved in average from 50% (raw image) to 68%, and the standard deviation obtained on the 10 samplings is decreased from 19.5 to 12.4. Compared to raw spirals with pen-ups (Table I), the accuracy values are in the same range; however, the sensitivity and specificity are more balanced.

TABLE II. PERFORMANCE ON PRESSURE-BASED HYBRID IMAGES.

Performance	**Sensitivity**	**Specificity**	**Accuracy**
Baseline: Raw spiral	50 ± 19.5	88 ± 6	69 ± 8.8
Pointwise pressure values	68 ± 12.4	74 ± 11.2	71 ± 7.2
Pressure P=0 & *K*=2 levels	75 ± 10.2	83 ± 11	79 ± 8.8
K=3 levels	62 ±10.7	82 ± 8.7	72 ± 7.1
Pressure P=0 & *K*=4 levels	70 ± 12.6	81 ± 11.3	75.5 ± 6.1
K=5 levels	62 ± 14.7	72 ± 18.8	67 ± 12.1

Then, we studied different configurations for quantizing pressure values, from 3 to 5 pressure levels. As shown in Table II, the best result was obtained when considering zero pressure values apart and performing quantization on two levels on pen-downs (case of Pressure P=0 & *K*=2 levels). We obtain in this case a significant improvement of all performance measures, reaching 79% of accuracy with a good balance between specificity (83%) and sensitivity (75%).

To better assess the impact of quantization on the classification process, in Figure 6, we display for each class the percentage of points in spiral drawings that are assigned to each cluster.

When considering the best resolution for discriminating between the two populations (P=0 & *K*=2), we notice that more than half of the points in spiral drawings of HC show high pressure values, while most of the points in spiral drawings of AD patients (65.7%) show low pressure values.

Higher quantization resolutions, as reported in Figures 6.c and 6.d, confirm the lower discrimination capacity of the system between HC and AD.

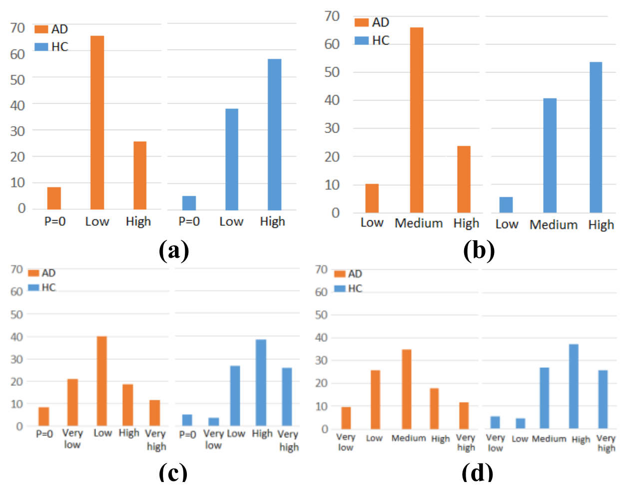

Figure 6. Pressure distribution of AD and HC across clusters for: (a) P=0 & *K*=2, (b) *K*=3, (c) P=0 & *K*=4, and (d) *K*=5.

C. *Altitude-based images*

Altitude hybrid images lead to similar accuracy values with those of raw spiral images (see Table III), but still improve sensitivity considerably. When comparing different quantization resolutions of altitude values, we notice that the case of three resolution levels of altitude (*K*=3 levels) leads to a good balance between specificity and sensitivity, with even lower standard deviation on the 10 samplings of the HC.

TABLE III. PERFORMANCE ON ALTITUDE-BASED HYBRID IMAGES

Performance	Sensitivity	Specificity	Accuracy
Baseline: Raw spiral	50 ± 19.5	88 ± 6	69 ± 8.8
Pointwise altitude values	83 ± 10	62 ± 21.8	72.5 ± 12.9
K=3 levels	65 ± 12.8	76 ± 12.8	70.5 ± 8.2
K=5 levels	65 ± 12	71 ± 19.2	68 ± 12.1

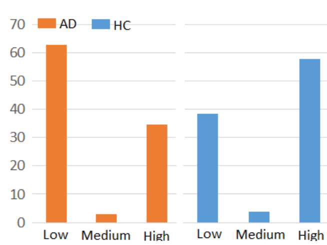

Figure 7. Altitude distribution of AD and HC across *K*= 3 clusters.

This result is very interesting since the altitude parameter is not widely exploited in the literature for pathology detection. This finding is in accordance with our previous work on AD detection based on online signatures [19]. Indeed, the case in Figure 7 (*K*=3 levels) shows that AD patients have a tendency to produce spirals with the lowest altitude values, while HC show the opposite trend. This result may reveal that AD patients tend to have less tonus on their way of holding the pen, compared to control elderly subjects, as found on online signatures in [19].

D. *Velocity-based images*

Table IV shows that velocity hybrid images provide significant improvement of performance compared to raw spiral images, especially in terms of sensitivity, whatever the quantization resolution. However, we notice that a pointwise description of velocity leads to poor sensitivity compared to the cases using a semi-global description of such parameter thanks to quantization (*K*=3 and *K*=5).

Moreover, we notice a considerable decrease of the standard deviation in sensitivity for *K*=5 compared to raw spiral images (from 19.5 to 8.7). In this case, an absolute improvement of 22% in sensitivity and of 7% in accuracy are observed, compared to raw spiral images.

TABLE IV. PERFORMANCE ON VELOCITY-BASED HYBRID IMAGES

Performance	Sensitivity	Specificity	Accuracy
Baseline: Raw spiral	50 ± 19.5	88 ± 6	69 ± 8.8
Pointwise velocity values	65 ± 10.2	85 ± 10.2	75 ± 7.1
K=3 levels	74 ± 10.2	78 ± 12.5	76 ± 7
K=5 levels	72 ± 8.7	80 ± 10	76 ± 5.8

Figure 8 shows that AD patients exhibit very low velocity values. Indeed, 48% of the points in spiral drawings of AD patients are produced with very low velocity values, while only 30% of the points in spirals drawn by HC have the lowest velocity. Besides, Figure 8 also shows that HC tend to exhibit higher velocity values. This result confirms a slower motion trend in AD since the early stage of the disease, when drawing the spiral.

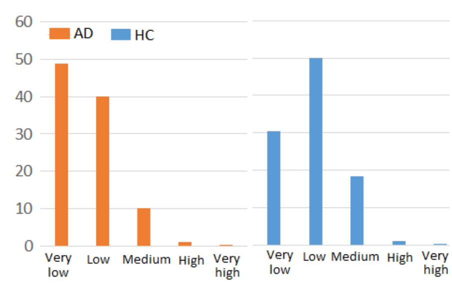

Figure 8. Velocity distribution of AD and HC across *K*=5 clusters.

E. *Fusion of expert's decisions*

As shown in Figure 5, we combine the decisions of the three best experts, assessed in sections IV.B, IV.C and IV.D, in order to take a final decision with a major voting scheme. These three experts exploit hybrid images containing different information, which might be complementary, leading to a more reliable AD detection.

Table V reports the results of the best configuration for each type of hybrid image and the late fusion of the three experts. Compared to raw spiral images (Table I), we notice an absolute improvement of 12.5% in accuracy and of 29% in sensitivity, while specificity is maintained. Besides, the standard deviation in sensitivity decreases from 19.5 to 9.4, when fusing the three experts' decisions. The standard deviation of the accuracy also decreases from 8.8 to 5.5.

TABLE V. PERFORMANCE OF THE THREE BEST EXPERTS AND THEIR FUSION

Performance	Sensitivity	Specificity	Accuracy
Expert 1 (Pressure) Pressure P=0 & K=2 levels	75 ± 10.2	83 ± 11	79 ± 8.8
Expert 2 (Altitude) K=3 levels	65 ± 12.8	76 ± 12.8	70.5 ± 8.2
Expert 3 (velocity) K=5 levels	72 ± 8.7	80 ± 10	76 ± 5.8
Fusion of the three experts	79 ± 9.4	84 ± 6.6	81.5 ± 5.5

When comparing fusion results to those of each expert separately (Table V), we notice that our fusion scheme is more reliable in terms of average performance on one hand, for both sensitivity and specificity, and on the other hand in terms of standard deviation across the 10 samplings of the HC population.

F. *Analysis of internal high-level features with PCA*

We performed a Principal Component Analysis (PCA) on the 4096 features extracted by AlexNet on raw images and each type of hybrid image considered in Table V. This analysis helps understanding which dynamic information better describes a given class, in terms of reduced intra-class heterogeneity.

Figures 9 and 10 show, respectively for HC and AD classes, the percentage of variance explained as a function of the retained principal components (PC).

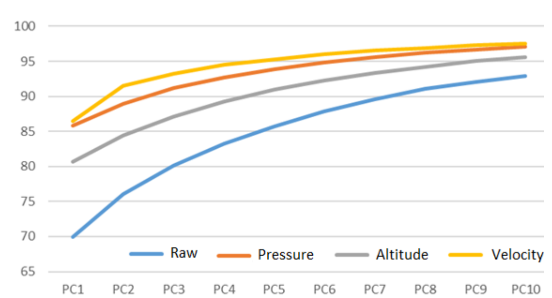

Figure 9. Percentage of variance explained with PCA as a function of the retained principal components (PC) for the HC population.

Figure 9 shows that images including dynamic information explain a much higher amount of variance compared to raw images. Indeed, in HC class, with velocity and pressure, we reach 91,47% and 91.15% of variance explained respectively, with two and three components. At the opposite, with raw images, eight components are required to reach 91.07% of variance explained. We also notice that velocity and pressure better characterize the HC population, compared to altitude, which requires five components to explain 91% of the variance.

Figure 10 shows that pressure information characterizes better the AD population than other dynamic information (altitude and velocity). Indeed, with pressure images, only three components are required to reach 91.68% of variance explained, while four components are needed to reach 90.52% of variance explained with raw images. Also, interestingly, we notice that the AD population shows more heterogeneity when described by altitude and velocity hybrid images.

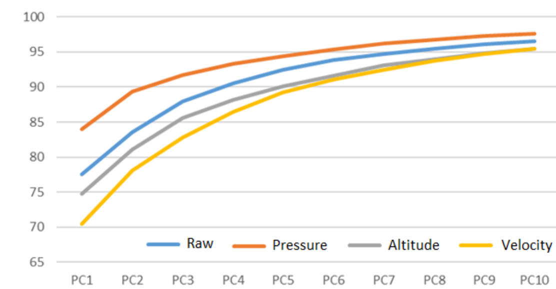

Figure 10. Percentage of variance explained with PCA as a function of the retained principal components (PC) for the AD population.

Finally, pressure information seems to be a good dynamic feature for characterizing both HC and AD classes. Indeed, only one principal component is sufficient to explain 84% of the variance in the AD class and 86% of the variance in the HC class.

V. CONCLUSION AND PERSPECTIVES

The present study proposes a novel scheme for the automatic assessment of early-stage AD, based on Archimedes spiral. It exploits Transfer Learning for feature extraction using hybrid spiral images. Such images are generated from the online sequences captured by a digitizer, and enriched with dynamic information displayed in RGB at different resolution levels. Three types of hybrid images are considered: images including pressure, altitude and velocity.

Experiments have shown that these hybrid representations of spiral images allow a considerable improvement of performance compared to raw spiral images. In the case of pressure-based hybrid images, when considering the pen-ups in the trajectory and two levels of pen pressure values on pen-downs, we reach 79% of accuracy (sensitivity = 75% and specificity= 83%) with a SVM classifier. This result confirms the necessity of a semi-global description of pressure instead of a local pointwise encoding, to discriminate AD and HC classes. Actually, this level of description is well suited to uncover that the AD population shows a trend of rather low-pressure values, while the HC one tends to show the opposite trend. Furthermore, the semi-global description of pressure reduces significantly the important heterogeneity of each class.

When considering hybrid images with three pen altitude levels, the obtained accuracy is comparable to that of raw spiral images, but the sensitivity improves considerably, for all quantization resolutions. We reach with a SVM classifier 70.5% of accuracy, 65% of sensitivity and 76% of specificity. Our analysis also reveals a trend of rather low altitude values in the AD population. This result reflects their lower tonus when holding the pen comparatively to HC, which is in accordance with the above-mentioned finding of lower pressure values in AD patients. The same relative behavior between pressure and altitude was indeed observed in AD patients when signing [19].

Regarding hybrid images with five velocity levels, performance also improves considerably compared to raw spiral images. Indeed, we reach in this case 76% of accuracy, 72% of sensitivity and 80% of specificity. Compared to raw images on which a sensitivity of 50% is obtained with a high standard deviation (19.5) on the 10 samplings of HC, including velocity information leads to better performance with a considerable reduction of the standard deviation (8.7). Again, a semi-global encoding of velocity into five levels uncovers a trend of lower velocity values in AD population compared to HC, which also show low velocity values due to normal aging.

Finally, combining the decisions of the three experts leads to the best accuracy value (81.5%) with a good balance of sensitivity (79%) and specificity (84%). Also, fusion allows decreasing globally the standard deviation on performance measures, due to random samplings of HC.

The analysis of high-level internal features with PCA showed that dynamic information included in images explain a much higher amount of variance compared to raw images, for the healthy population. For the AD population instead, pressure is the only dynamic information that explains a higher amount of variance compared to raw images. This result shows the pertinence of pressure information in characterizing both HC and AD classes through hybrid images.

The majority of studies in the literature on deep learning for pathology detection have focused on PD assessment. They combined different handwritten tasks and different representations of the same original images, on one hand after applying classical image processing transformations, as in [8], on the other hand by generating images with no visual information of the drawing tasks [9-12]. Our work proposes to exploit only one task for AD assessment and to enrich the input images in RGB with dynamic information represented with a semi-global encoding. This new scheme takes advantage simultaneously of both visual and dynamic information.

In the future, we aim at studying other dynamic features such as acceleration and jerk, as well as other graphical tasks available in our database. We will also go further in the analysis of the internal high-level representations represented by the extracted features for a better understanding of the decisions taken by the classifiers. We will also study other architectures than AlexNet for feature extraction.

ACKNOWLEDGMENT

We would like to thank Hospital Broca (Paris, France) for enabling the data acquisition campaign, in the framework of the French ALWRITE research project.

REFERENCES

[1] Teulings, H.L., Stelmach, G.E., 1991. Control of stroke size, peak acceleration, and stroke duration in Parkinsonian handwriting," Hum. Movement Sci., 10(2), 315–334.

[2] Teulings,, H.L., Contreras-Vidal, J.L., Stelmach, G. E., Adler, C.H., 1997. Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control," Exp. Neurol., 146 (1), 159–170.

[3] Gemmert, V., Arend, W., Teulings, H.L., Stelmach, G.E., 1998. The influence of mental and motor load on handwriting movements in Parkinsonian patients, Acta Psychol., 100 (1), 161–175.

[4] Yan, J.H., Rountree, S., Massman, P., Smith Doody, R., Li, H., 2008. Alzheimer's disease and mild cognitive impairment deteriorate fine movement control, J. Psychiatric Res., 42(14), 1203–1212.

[5] Schröter, A., Mergl, R., Bürger, K., Hampel, H., Möller, H.J., Hegerl, U., 2003. Kinematic analysis of handwriting movements in patients with Alzheimer's disease, mild cognitive impairment, depression and healthy subjects. Dement Geriatr Cogn Disord., 15(3), 132-142.

[6] Slavin, M.J., Phillips, J. G., Bradshaw, J. L., Hall, K. A., Presnell, I., 1999. Consistency of handwriting movements in dementia of the Alzheimer's type: A comparison with Huntington's and Parkinson's diseases, J. Int. Neuropsychol. Soc., 5(1), 20–25.

[7] Yu, N.Y., Chang, S.H., 2016. Kinematic analyses of graphomotor functions in individuals with Alzheimer's disease and amnestic mild cognitive impairment, J. Med. Biol. Eng., 36(3), 334–343.

[8] Moetesum, M., Siddiqi, I., Vincent, N., Cloppet, F., 2019. Assessing visual attributes of handwriting for prediction of neurological disorders—A case study on Parkinson's disease, Pattern Recognition Letters, 121, 19-27.

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

[10] Pereira, C.R., Pereira, D.R., Rosa, G.H., Albuquerque, V.H., Weber, S.A., Hook, C., Papa, J.P., 2018. Handwritten Dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification, Artificial Intelligence in Medicine, 87, 67-77.

[11] Taleb, C., Khachab, M., Mokbel, C., Likforman-Sulem, L., 2017. Feature Selection for an improved Parkinson's Disease Identification based on Handwriting, IEEE International Workshop on Arabic Script Analysis and Recognition (ASAR), 52-56.

[12] Taleb, C., Khachab, M., Mokbel, C., Likforman-Sulem, L., 2019. Visual Representation of Online Handwriting Time Series for Deep Learning Parkinson's Disease Detection, International Conference on Document Analysis and Recognition Workshops (ICDARW), 25-30.

[13] Peuker, D., Scharfenberg, G., Hook, C., 2011. Feature Selection for the detection of fine motor movement disorders in Parkinson's patients, in Advanced Research Conference, ser. ARC'11, Shaker Verlag.

[14] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 1097–1105.

[15] Diagnostic and Statistical Manual of Mental Disorders, Fifth ed. (DSM-5). Arlington, VA, USA: Amer. Psychiatric Assoc., 2013.

[16] Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. CNN features off-the-shelf: an astounding baseline for recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognitionworkshops, 806-813.

[17] El-Yacoubi, M., Garcia-Salicetti, S., Kahindo, C., Rigaud, A.S., Cristancho-Lacroix, V., 2019. From Aging to Early-Stage Alzheimer's : Uncovering Handwriting Multimodal Behaviors by Semi-Supervised Learning and Sequential Representation Learning, Pattern Recognition 86, 112-133.

[18] Kahindo, C., El-Yacoubi, M., Garcia-Salicetti, S., Rigaud, A-S., Cristancho- Lacroix, V., 2018. Characterizing Early Stage Alzheimer through Spatiotemporal Dynamics of Handwriting, IEEE Signal Processing Letters.

[19] Wang, Z., Abazid, M., Houmani, N., Garcia-Salicetti, S., Rigaud A-S., 2019. Online signature analysis for characterizing early stage Alzheimer's disease: a feasibility study. Entropy, MDPI, 21 (10), 956-1 - 956-12.