RESEARCH ARTICLE

Detection of Alzheimer Disease on Online Handwriting Using 1D Convolutional Neural Network

QUANG DAO[ID][1]**, MOUNÎM A. EL-YACOUBI**[ID][1]**, (Member, IEEE), AND ANNE-SOPHIE RIGAUD**[2,3]
[1]Samovar/Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France
[2]AP-HP, Groupe Hospitalier Cochin Paris Centre, Hôpital Broca, Pôle Gérontologie, 75005 Paris, France
[3]Université Paris Descartes, 75006 Paris, France

Corresponding author: Quang Dao (quangdaovu206@gmail.com)

ABSTRACT Building upon the recent advances and successes in the application of deep learning to the medical field, we propose in this work a new approach to detect and classify early-stage Alzheimer patients using online handwriting (HW) loop patterns. To cope with the lack of training data prevalent in the tasks of classification of neuro-degenerative diseases from behavioral data, we investigate several data augmentation techniques. In this respect, compared to the traditional data augmentation techniques proposed for HW-based Parkinson detection, we investigate a variant of Generative Adversarial Networks (GANs), DoppelGANger, especially tailored for times series and hence suitable for synthesizing realistic online handwriting sequences. Based on a 1D-Convolutional Neural Network (1D-CNN) to perform Alzheimer classification, we show, on a real dataset related to HW and Alzheimer, that our DoppelGANger-based augmentation model allow the CNN to significantly outperform both the current state of the art and the other data augmentation techniques.

INDEX TERMS Alzheimer disease, DoppelGANger, online handwriting, 1D-convolution neural networks.

I. INTRODUCTION

Alzheimer's Disease (AD), a progressive neuro-degenerative disorder, is the cause of memory loss as well as a decline in cognitive functioning [1]. It is the most common type of dementia, affecting typically people of advanced ages. Currently there is no cure to reverse its symptoms, and only medications to delay the progression are available. As a matter of fact, the earlier the patient is diagnosed, the more likely the treatment will be effective.

Because people with AD are significantly impacted by episodic memory impairment, it comes as no surprise that a large number of studies have focused on language disorders involving spelling, grammatical, syntactic or semantic errors, etc. [2], [3]. It has been shown, however, that AD can be predicted by non-cognitive symptoms, in particular by

motor impairment [4]. In this respect, several studies have assessed gait impairment, mild parkinsonian signs, fatigue and frailty [5], [6], [7], [8]. Other studies have investigated fine motor impairment, especially handwriting (HW) changes due to AD [9], [10], [11]. Such studies make sense given that AD induces cognitive and visuospatial impairment that makes the physical act of writing difficult, thereby triggering HW impairment [12].

As a result, there have been numerous studies and researches into applying machine learning methods to build a system that can reliably detect Alzheimer at an early stage, so that the patient can receive a timely and successful treatment. These researches work with a wide range of input data recorded from medical experiments and examinations, with handwriting and graphical gestures being ones of the most prominent sources. Traditionally, research on handwriting analysis applies either statistical tests or traditional classification models. Only recently have deep learning approaches

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy[ID].

been explored, albeit these were mostly for the detection of Parkinson and not Alzheimer. The goal of our work is to fill this gap and to explore this approach by building a deep learning model that can classify whether a subject is an Alzheimer patient, using online handwriting data. We also look to address one of the most prominent problems in deep learning medical applications, the lack of training data, by investigating the latest data augmentation techniques. Concretely, we introduce, in this work, the following key contributions:

- Application of deep learning architecture for Alzheimer detection from online handwriting. We show how our models achieves a new state-of-the-art result on the proposed real dataset. This is the first deep learning based model applied for Alzheimer's disease classification from handwriting in the context of a very limited training dataset.

- Implementation of both traditional and recent data augmentation methods for time series. Especially, we propose a variant of Generative Adversarial Networks (GANs), DoppelGANger, specifically tailored for times series and hence suitable for generating synthetic online handwriting data. Our DoppelGANger-based data augmentation scheme outperforms classical data augmentation methods proposed for Parkinson's disease classification, and lead to new state of the art results for Alzheimer's classification, that are dramatically above previous state of the art.

II. RELATED WORK

Traditionally, online handwritings are analyzed using statistical tests such as ANOVA [9], [10], [11], [13], [14], [15], [16]. Some works also implement classical classification methods based on Linear Discriminant Analysis (LDA) [9], [10], [13]. More recently, [17] propose a semi-supervised learning approach to discover homogenuous cognitive profiles. This work addresses the problem of encoding spatiotemporal dynamics with the full online handwriting trajectory, by applying representation learning for the treatment of sequential data, and shows how such encoding outperforms global or semi-global parameters. [18] follow a similar approach, extracting loop-like patterns from handwriting and modeling velocity trajectory through unsupervised learning.

With the rapid development of deep learning recently, there have been a lot of research and experiments focused on applying deep learning architectures, as they provide the possibility to learn meaningful features and extract intricacies by themselves just from the raw data, thereby eliminating the need for extensive hand-crafting of features. For problems similar to ours, that is, to deal with medical data that is limited in available samples, and in the form of different channels of long time series, specifically online handwriting, there have been a lot of works on the popular dataset PaHaW [19] for Parkinson detection, notably [20], [21], [22]. Another notable work on Parkinson online handwriting is [23], in which the authors experiment with different data augmentation methods

and achieve significant improvement compared to using only original data. For Alzheimer's, however, there is no equivalent publicly available dataset. For our work, we use the same dataset presented in the study [18] based on K-medoids clustering and Bayesian classifier. This work will serve at the benchmark for our results.

III. METHODOLOGY
A. DATASET
The dataset, the same that was used in [18], was collected at the Broca Hospital in Paris. It consists of 54 subjects, from which 27 are patients who have early stage Alzheimer disease (AD) and 27 healthy control (HC). The Alzheimer patients are diagnosed based on DSM-5 criteria. The selected subjects were all able to understand French fluently and agreed to sign a consent form. The subjects perform multiple handwriting tasks that are recorded on a WACOM Intuos Pro Large tablet with an inking pen. The sampling rate that the tablet records pen position is 125 Hz. The data are recorded in the form of channels of time series, containing static and dynamic information of the pen movement during which the subjects perform the tasks. There are six channels in total: timestamps, x coordinates, y coordinates, pressure, azimuth and altitude.

To compare with the results in [20] and [21], we focus in this work on the task of drawing four series of repetitive *llll* letters, as shown on the left of Figure 1.

B. LOOP SEGMENTATION
Because our dataset consists of only 54 subjects, it is necessary to come up with a method that increases the number of samples, especially for the training of deep learning models. Inspired by [18], we notice the repetitive pattern of the l letter loops that can thus be extracted into individual training samples. As we usually have 16 l letters for each subject, we first split the sample into strokes and then keep only the 16 loops, by discarding the ligatures between them, as illustrated on the right of Figure 1. This gives us about 16 times more data to train with as we now treat the loops and not the subjects as individual training samples.

C. DATA PREPROCESSING
In order to achieve good performance on deep learning models, researches have shown that it is necessary to standardize the input data. For this project we decide to standardize each feature of each subject individually, so that they all have a mean of 0 and standard deviation of 1.

As we choose to split the training samples into loops and consider training based on individual loops, we also subtract the timestamps of each loop by its first timestamps, so that every loop's first timestamp is 0.

To design an adequate CNN architecture that is well-dimensioned for effective model training, it is necessary that the loops have lengths that are not too scattered. We notice in Figure 2 that the maximum length of the loops, i.e. 281 time-steps, is actually due to three loop outliers, while the length of the remaining loops follows roughly a Gaussian
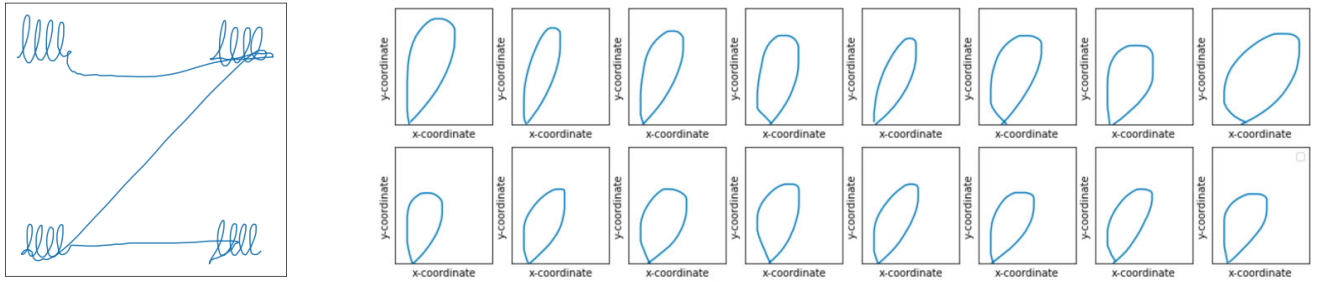
FIGURE 1. Example of the task performed (left) and the extracted loops (right).
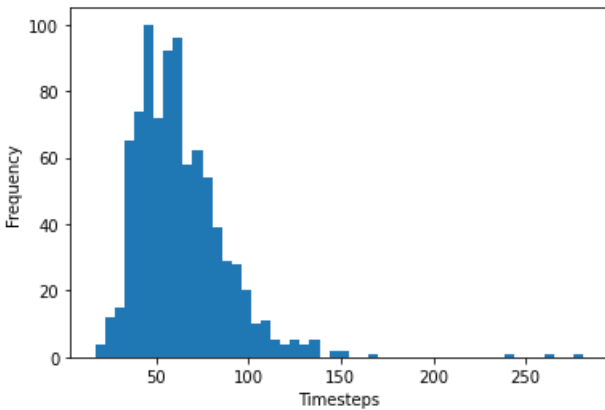

FIGURE 2. Length distribution of extracted loops.

distribution with a mean around 60 and a maximum value of 176. We set, therefore, the maximum loop length to 176, trim the three outlier loops to this length and add zero-padding to all the loops that are shorter than the defined maximum length. This considerably reduces the amount of zero-padding that would be created if the length was kept to the original maximum (i.e. 281). This means the length of the feature vector taken as input by the CNN is significantly optimized, with minimal information loss.

D. VELOCITY FEATURES
Beside the available raw features directly provided by the recording tablet, another approach is to create new features that represents better the difference between how a healthy subject and an Alzheimer patient performs the task. The result from [18] shows that velocity features calculated from the coordinates and timestamps are good indicators and can help improve classification performance. Therefore, we create two new channels: x-velocity ($Vx(n)$) and y-velocity ($Vy(n)$), with $Vx(n) = \triangle x(n) / \triangle t(n)$; $Vy(n) = \triangle y(n) / \triangle t(n)$, where $\triangle x(n) = x(n+1) - x(n-1)$, and $\triangle t(n) = t(n+1) - t(n-1)$. Combined with the six channels already available in the dataset, we now have eight time series channels to represent every loop as shown in Figure 3.

E. DATA AUGMENTATION FOR TIME SERIES
In addition to the segmentation of loops, it is also possible to increase the amount of training data by generating new

synthetic samples. There are two principal ways with which a new time series sample can be generated: slight random transformation of real data (jittering, scaling, warping), and data synthesis from the distribution of the real data (pattern mixing, generative models, pattern decomposition methods) [24].

In this work, we experiment with both traditional augmentation techniques as well as implement the latest innovations in synthetic time series generation using Generative Adversarial Networks.

1) TRADITIONAL AUGMENTATION METHODS
As our baseline, we consider different time series augmentation techniques as follows. For the cases of Jittering, Scaling and Time Warping, the parameters are defined as suggested in [25].

- **Jittering**: Addition of Gaussian noise to each time step, with mean $\mu = 0$ and standard deviation $\sigma = 0.03$.
- **Scaling**: Pattern-wise magnitude change by a scalar value determined by a Gaussian distribution ($\mu = 1$, $\sigma = 0.1$)
- **Time Warping**: Time warping uses a smooth warping path to deform the temporal location.
- **Window Warping**: Window warping [26] is a version of time warping. It takes random windows of the time series, then expands by 2 or contracts by 0.5.
- **Suboptimal Warped Time Series Generator (SPAWNER)**: SPAWNER [27], a pattern mixing method, generates synthetic data through suboptimal time warping of intra-class patterns.

2) TIME SERIES DATA AUGMENTATION WITH DoppelGANger
Generative Adversarial Networks (GANs) [28] are a type of generative model that aim to train two neural networks — a generator G and a discriminator D — using an adversarial training workflow. The generator tries to create synthetic samples that the discriminator is unable to distinguish from real data. The goal of GANs is to create models that can draw samples from the same distribution of the training data.

While data augmentation techniques using GANs have been actively researched for the application on images, they are not as often considered for time series. Some notable GANs architectures that have been developed for time series
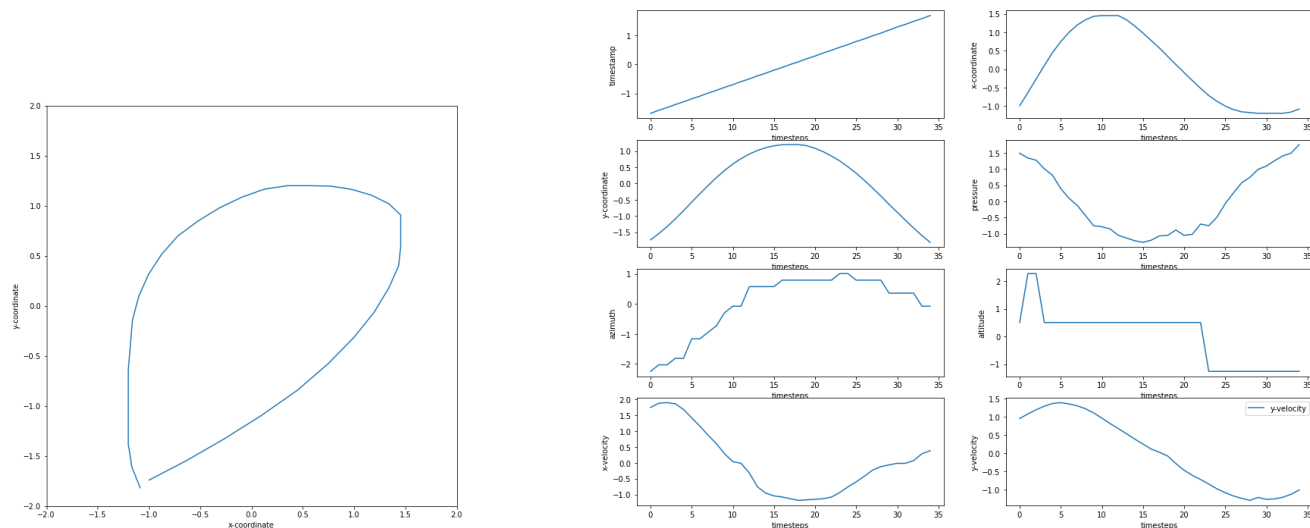
FIGURE 3. Example of an extracted and normalized loop (left) and the eight time series channels that represent it (right).

data augmentation are RCGAN [29], TimeGAN [30] and DoppelGANger [31]. After experimentation, we decide to implement DoppelGANger as our GAN model to generate online handwriting time series, as it addresses the weaknesses of the previous works [31].

DoppelGANger attempts to capture the relationship between attributes A^i (metadata) and time series values T^i by using two separated generators, one to generate metadata and one to generate time series conditioned on metadata: $P(A^i, T^i) = P(A^i) \cdot P(T^i | A^i)$. For the metadata generator, a multi-layer perceptron network is used. The time series generator is a recurrent neural network (RNN). At every time step the generated metadata A^i is added to the RNN as input. In order to retain the long-term correlations within time series, the RNN is combined with a novel idea called *batch generation*. The goal of batch generation is to reduce the number of RNN passes by generating S records for each pass instead of one.

Furthermore, an auxiliary discriminator is also introduced to discriminate only atributes. The final loss function is a combination of the losses from the two discriminators, with a weighting parameter α: $\min_G \max_{D_1,D_2} \mathcal{L}_1(G, D_1) + \alpha \mathcal{L}_2(G, D_2)$ with $\mathcal{L}_1$ as the Wasserstein loss of the original discriminator and $\mathcal{L}_2$ of the auxiliary.

As GANs are notoriously difficult to train, and according to the loop distribution in Figure 2 the number of loops with length greater than 100 is very small. As the more time steps we have the more parameters there are to optimize, we decide to train the model only on loops with maximum length 96. The batch generation parameter S is set as 8. The training labels of our dataset (AD or HC) are considered as attributes and are trained together with the time series. This allows us to later conditionally generate synthetic samples of each class by specifying the attributes. The GAN model is trained for 1000 epochs with learning rate 0.001.

F. MODEL SELECTION
As the data we work with are time series channels, it is necessary to select a model architecture that is suitable for the task. Using recurrent neural networks (RNNs) is a possible solution as it is designed to deal with sequential data. However it has been shown that for very long sequences RNN struggles to capture the temporal correlation [31]. Considering the promising results that has been obtained in other works using 1D Convolutional Neural Network (1DCNN) on medical time series, especially EEG signals [32], [33], we decide to select it as the architecture to our model.

To figure out the best 1DCNN architecture for our task, we perform random search on a range of different hyperparameters such as: number of convolutional layers, number of convolutional filters, convolutional filter size, pool filter size, etc. In order to reduce overfitting, dropout [34] is also applied.

IV. EXPERIMENTS AND RESULTS
A. TOOLS
The experiments are coded in Python, and the process of building and training the model performed with the help of the libraries Pytorch and sklearn. In order to have access to GPUs and improve the traning speed, we make use of the resources available on Google Colab.

B. EVALUATION METRICS
In order to maintain consistency and make comparisons with previous works, we will evaluate our results by three main metrics: accuracy, sensitivity and specificity, which are calculated from the percentage of true positives (*tp*), true negatives (*tn*), false positives (*fp*) and false negatives (*fn*):

a **Accuracy**: the overall ability of the model to make correct classification $((tp + tn)/(tp + tn + fp + fn))$.

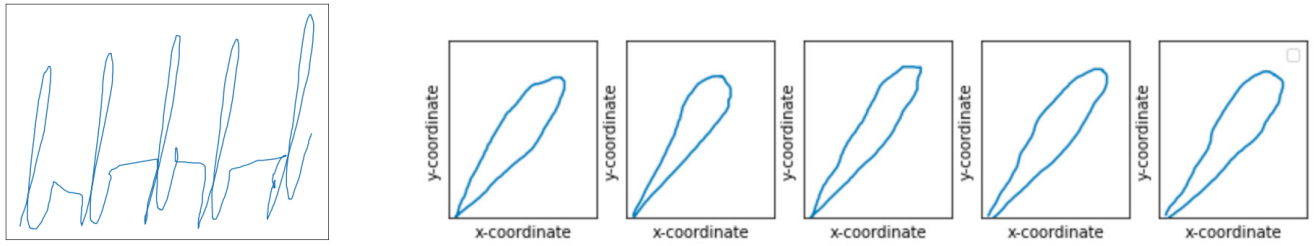b **Sensitivity**: the ability to correctly classify Alzheimer patients $(tp/(tp + fn))$.

FIGURE 4. Example of the task performed (left) and the extracted loops (right) for PaHaW dataset.
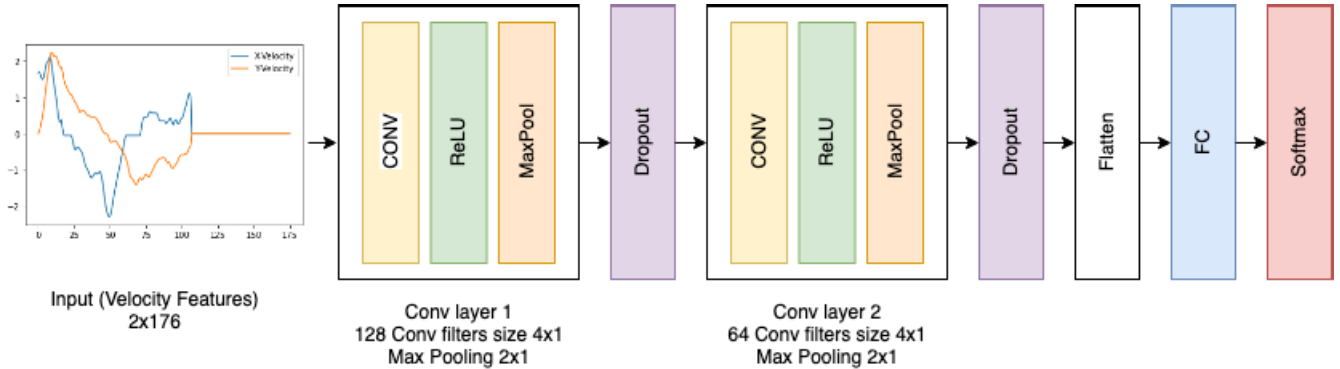

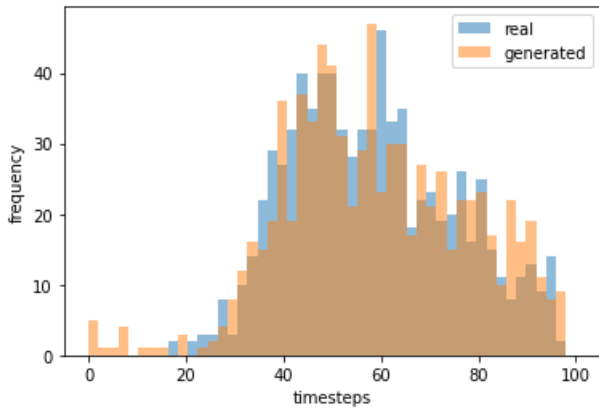**FIGURE 5.** The proposed 1DCNN architecture.

FIGURE 6. Loop length distribution of real data vs GAN-generated data.

c **Specificity**: the ability to correctly classify healthy subjects ($tn/(tn + fp)$).

In all our experiments except for the final, the model is trained using 10-fold stratified cross validation. For each fold, the model will be trained for 40 epochs, with a learning rate of 0.001. For training, each of the loops is considered as an individual training sample, and for validation we average the output score of all the loops that belonged to one subject to get the final classification for that subject.

C. HYPERPARAMETER OPTIMIZATION
As indicated in Section III-F, we perform random search in order to find the best hyperparameter combination. To avoid experimental bias, we decide to use, as training/validation data for the process of hyperparameter optimization, the

similarly recorded online handwriting dataset PaHaW [19] associated with Parkinson disease. The second task of PaHaW, as shown in Figure 4 is also to write repetitive *l* letters, which then can be splitted into loops using our segmentation process.

We show in Figure 5 the 1D CNN architecture that is obtained as the result. The model consists of two 1D convolutional layers, each followed by ReLU activation and Max Pooling. There are 128 1D filters in the first convolutional layer and 64 in the second, all of them have size 4×1. The Max Pooling filters for both layers have size 2×1. Dropout is applied after both poolings with a dropout rate set as 0.2. The output is then flattened and put into a fully connected layer (FC) and finally softmax for classification. This model achieves 78% accuracy, 85% sensitivity and 73% specificity when training with our dataset using all eight available channels as the input. As we work with limited data, such a simple architecture can be enough to learn well without overfitting, in contrast to deeper models.

D. FEATURE SELECTION
By adding the two extra velocity features, we get eight time series channels in total: x-coordinate, y-coordinate, timestamps, pressure, azimuth, altitude, x-velocity, y-velocity. However, not all of these channels contain equal information and some of them may even introduce noise. Therefore, it is necessary to train the model with different feature combinations in order to figure out the one that works best. We start out with the two basic combinations (x-coordinate, y-coordinate) and (x-velocity, y-velocity) then try to improve the result by combining them with the rest.
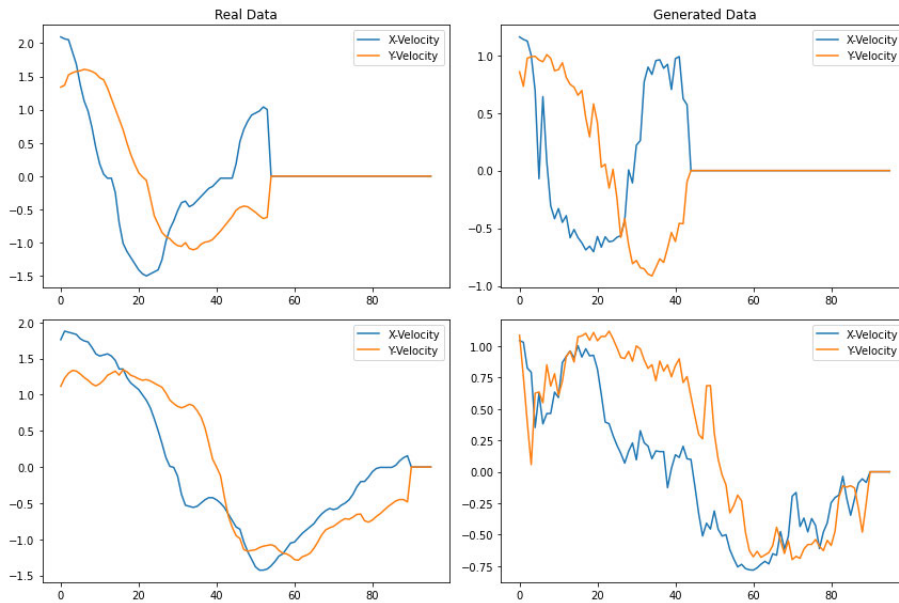
FIGURE 7. Examples of real data (left) and GAN-generated data (right).

From the results obtained in Table 1, we observe that the best performance is achieved by using only the two velocity features. Adding additional features to them does not further improve accuracy. Another notable observation is that the pressure feature is very noisy and reduces significantly the performance of the combinations that include it.

E. DATA AUGMENTATION

We apply data augmentation on input data that are time series of two channels: x-velocity and y-velocity, according to the results of the experiment in Section IV-D. Table 2 gives us the classification results comparing different data augmentation techniques. For each test fold of cross validation, the training data are augmented once while the validation data are kept intact. This means we perform training with twice the original amount of data. We observe that using DoppelGANger [31], we obtain an accuracy of 89%, the highest among all the augmentation schemes. With Jittering and SPAWNER, the accuracy also improves but only slightly (2%). Scaling does not affect the performance in terms of accuracy, while the data generated with warping methods are noisy, which is detriment to classification. We also attempt to combine different augmentation methods with DoppelGANger, but this does not further improve the results.

F. ANALYSIS OF GAN-GENERATED SYNTHETIC DATA

In order to evaluate the fidelity of the time series data generated by DoppelGANger compared to the original training data, we compare the length distribution of real and synthetic data (Figure 6). Note that as we mentioned in Section III-E2, the maximum length of the time series that are used for GAN training is 96. As observed in Figure 6, apart from a few outliers that have too short lengths, the synthetic data that

TABLE 1. 1DCNN performance with different feature combinations.

Feature Combination	Accuracy	Sensitivity	Specificity
x,y	79	88	73
x,y,timestamps	79	82	78
x,y,pressure	74	82	68
x,y,azimuth	78	87	73
x,y,altitude	78	87	72
x-velocity,y-velocity	**85**	83	88
x-velocity,y-velocity,altitude	84	83	87
x-velocity,y-velocity,azimuth	82	93	73
x-velocity,y-velocity,pressure	79	82	80
all features	78	85	73

TABLE 2. 1DCNN performance with different data augmentation techniques.

Augmentation Method	Accuracy	Sensitivity	Specificity
No Augmentation	85	83	88
Jittering	87	90	87
Time Warping	78	83	75
Window Warping	80	80	82
Scaling	85	90	83
SPAWNER	87	90	85
DoppelGANger	**89**	90	88
DoppelGANger+Jiterring	84	83	87
DoppelGANger+SPAWNER	**89**	93	85

we generate match closely the length distribution of the real training data.

We also perform a qualitative assessment, by comparing the generated series by our DoppelGANger model with the original ones (Figure 7). The comparison shows basically that even though the generated data are not as smooth, they do approximate the velocity curves of the original data at different lengths. This shows that DoppelGANger are suitable GAN models for generating good quality online handwriting time series, which ultimately significantly improves the
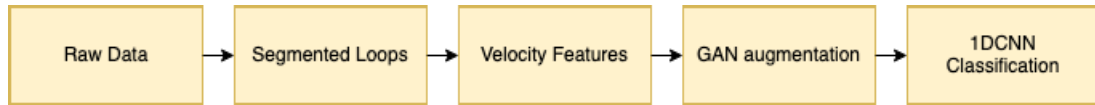
FIGURE 8. The finalized pipeline of data processing and classification.

TABLE 3. Final classification result of model with LOOCV and comparison to previous state of the art.

Input Data	Methodology	Accuracy	Sensitivity	Specificity
Segmented loops	Velocity features + K-medoids + Bayes classifier (Kahindo et al 2019)	74	72.2	75.6
Segmented loops	Velocity features + 1D CNN + GAN augmentation	**87.04**	**85.19**	**88.89**

training of our CNN classifier of Alzheimer's patients vs. Healthy Controls.

G. FINAL EVALUATION WITH LEAVE-ONE-OUT CROSS VALIDATION

In order to better evaluate the generalization ability of the model, as well as to make comparison with the previous state of the art on the dataset obtained in [18], we decide to perform leave-on-out cross validation (LOOCV) in addition to stratified 10-fold cross validation that is used in previous experiments. This means that from the dataset of 54 subjects, we have 54 folds, and for each fold we train on 53 subjects and test on the one left out. For each fold, we also randomly initialize the weights of the model. The model architecture as well as the feature combination (x-velocity, y-velocity) and augmentation method (DoppelGANger) are chosen from the previous best results. We illustrate this finalized training procedure in Figure 8.

We are also able to compare our final results in Table 3 with the results of [18], which are also evaluated using LOOCV. We observe that using the same data and features, our approach with deep learning using 1DCNN and data augmentation with DoppelGANger is able to improve the result significantly. This confirms the viability of applying deep learning solutions to this problem, even if the data available is very limited, thanks to our GAN-based augmentation scheme, specifically tailored for time series.

V. CONCLUSION

In this work, we have been able to develop an effective deep learning approach for the problem of classifying Alzheimer patients. Despite the lack of sufficient training data, the result obtained from training 1D Convolutional Neural Network model is the new state of the art on the dataset, at 87.04% accuracy, 85.19% sensitivity and 88.89% specificity. This is very promising and opens up further possibilities for research.

The newly achieved state of the art has been made possible by tackling the problem of limited data through synthetic data generation, based on the application of of Generative Adversarial Network adapted to time series, namely Doppel-GANger. The generated data have been shown to approximate the quality and distribution of the real data, which helped significantly improve the classification performance.

For future improvements, several research directions can be taken. [20] has shown that it is possible to apply transfer learning from unrelated data such as ImageNet [35] and perform classification based on the images created by the handwriting coordinates. Another possibility is to combine online handwriting with other biomarkers such as voice and facial expressions in order to make better detection.

REFERENCES

[1] Alzheimer's Association, "Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 18, no. 4, pp. 327–406, Apr. 2021.

[2] M. L. G. Bencini, L. Pozzan, R. Biundo, J. W. McGeown, V. V. Valian, A. Venneri, and C. Semenza, "Language-specific effects in Alzheimer's disease: Subject omission in Italian and English," *J. Neurolinguistics*, vol. 24, no. 1, pp. 25–40, Jan. 2011.

[3] S. Pekkala, D. Wiener, J. J. Himali, A. S. Beiser, L. K. Obler, Y. Liu, A. McKee, S. Auerbach, S. Seshadri, P. A. Wolf, and R. Au, "Lexical retrieval in discourse: An early indicator of Alzheimer's dementia," *Clin. Linguistics Phonetics*, vol. 27, no. 12, pp. 905–921, Aug. 2013.

[4] A. S. Buchman and D. A. Bennett, "Loss of motor function in preclinical Alzheimer's disease," *Exp. Rev. Neurotherapeutics*, vol. 11, no. 5, pp. 665–676, 2011.

[5] P. A. Boyle, A. S. Buchman, R. S. Wilson, S. E. Leurgans, and D. A. Bennett, "Physical frailty is associated with incident mild cognitive impairment in community-based older persons," *J. Amer. Geriatrics Soc.*, vol. 58, no. 2, pp. 248–255, Jan. 2010.

[6] A. S. Buchman, P. A. Boyle, R. S. Wilson, Y. Tang, and D. A. Bennett, "Frailty is associated with incident Alzheimer's disease and cognitive decline in the elderly," *Psychosomatic Med.*, vol. 69, no. 5, pp. 483–489, Jun. 2007.

[7] L. E. Hebert, J. L. Bienias, J. J. McCann, P. A. Scherr, R. S. Wilson, and D. A. Evans, "Upper and lower extremity motor performance and functional impairment in Alzheimer's disease," *Amer. J. Alzheimer's Disease Other Dementiasr*, vol. 25, no. 5, pp. 425–431, May 2010.

[8] E. D. Louis, N. Schupf, J. Manly, K. Marder, M. X. Tang, and R. Mayeux, "Association between mild parkinsonian signs and mild cognitive impairment in a community," *Neurology*, vol. 64, no. 7, pp. 1157–1161, Apr. 2005.

[9] J. Garre-Olmo, M. Faúndez-Zanuy, K. López-de-Ipiña, L. Calvó-Perxas, and O. Turró-Garriga, "Kinematic and pressure features of handwriting and drawing: Preliminary results between patients with mild cognitive impairment, Alzheimer disease and healthy controls," *Current Alzheimer Res.*, vol. 14, no. 9, Aug. 2017.

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

[11] N.-Y. Yu and S.-H. Chang, "Kinematic analyses of graphomotor functions in individuals with Alzheimer's disease and amnestic mild cognitive impairment," *J. Med. Biol. Eng.*, vol. 36, no. 3, pp. 334–343, Jun. 2016.

[12] J. Firger, "Handwriting changes can indicate Alzheimer's progression," Tech. Rep., 2013.

[13] 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, Jul. 2006.

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

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

[16] M. J. Slavin, J. G. Phillips, J. L. Bradshaw, K. A. Hall, and I. Presnell, "Consistency of handwriting movements in dementia of the Alzheimer's type: A comparison with Huntington's and Parkinson's diseases," *J. Int. Neuropsychological Soc.*, vol. 5, no. 1, pp. 20–25, Jan. 1999.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] M. Diaz, M. Moetesum, I. Siddiqi, and G. Vessio, "Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs," *Exp. Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114405.

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

[23] C. Taleb, L. Likforman-Sulem, C. Mokbel, and M. Khachab, "Detection of Parkinson's disease from handwriting using deep learning: A comparative study," *Evolutionary Intell.*, pp. 1–12, Sep. 2020.

[24] B. Kenji Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," 2020, *arXiv:2007.15951*.

[25] T. T. Um, F. M. J. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, and D. Kulić, "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, Nov. 2017, pp. 216–220.

[26] A. L. Guennec, S. Malinowski, and R. Tavenard, "Data augmentation for time series classification using convolutional neural networks," in *Proc. ECML/PKDD Workshop Adv. Analytics Learn. Temporal Data*, 2016.

[27] K. Kamycki, T. Kapuscinski, and M. Oszust, "Data augmentation with suboptimal warping for time-series classification," *Sensors*, vol. 20, no. 1, p. 98, Dec. 2019.

[28] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*.

[29] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional GANs," 2017, *arXiv:1706.02633*.

[30] J. Yoon, D. Jarrett, and M. V. D. Schaar, "Time-series generative adversarial networks," Tech. Rep., 2019.

[31] Z. Lin, A. Jain, C. Wang, G. Fanti, and V. Sekar, "Using GANs for sharing networked time series data: Challenges, initial promise, and open questions," in *Proc. ACM Internet Meas. Conf.*, Oct. 2020, pp. 464–483.

[32] I. Ullah, M. Hussain, E.-U.-H. Qazi, and H. Aboalsamh, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach," *Exp. Syst. Appl.*, vol. 107, pp. 61–71, Oct. 2018.

[33] Ö. Yıldırım, U. B. Baloglu, and U. R. Acharya, "A deep convolutional neural network model for automated identification of abnormal eeg signals," *Neural Comput. Appl.*, vol. 32, pp. 1–12, Nov. 2018.

[34] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*.

[35] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

QUANG DAO received the master's degree in data analysis and pattern classification from Telecom SudParis, University Paris Saclay. He is currently a Data Scientist and a Computer Vision Engineer. His expertise includes the application of machine learning on medical data, such as signals and images.

MOUNÎM A. EL-YACOUBI (Member, IEEE) received the Ph.D. degree from the University of Rennes, France, in 1996. He was with the Service de Recherche Technique de la Poste (SRTP), France, in 1992 and 1996. He was a Visiting Scientist with the Centre for Pattern Recognition and Machine Intelligence (CENPARMI), Montreal, Canada, and an Associate Professor (1998–2000) with the Catholic University of Parana, Curitiba, Brazil. From 2001 to 2008, he was a Senior Software Engineer at Parascript, Boulder, Colorado. At SRTP and Parascript, he has developed handwriting recognition software for real-life automatic mail sorting, bank check reading, and form processing. Since June 2008, he has been a Professor at Telecom SudParis, Institut Polytechnique de Paris. His research interests include artificial intelligence, data science, machine learning, and modeling human user data, especially behavioral signals like handwriting, voice, gesture and activity recognition, biometrics, e-health, smart agriculture, and smart cities.

ANNE-SOPHIE RIGAUD received the M.D. and Ph.D. degrees in neurosciences. She is currently a Geriatrician and a Psychiatrist. She is also a Professor in geriatrics with University Paris 5. She heads the Geriatric Department, Broca Hospital, and the Research Unit EA 4468, including LUSAGE Laboratory. She has expertise in dementia and gerontechnology.

• • •