Predicting Alzheimer's Disease and Mild Cognitive Impairment with Off-line and On-line House Drawing Tests

Nina Hosseini-Kivanani
University of Luxembourg, Luxembourg
nina.hosseinikivanani@uni.lu

Elena Salobrar-García, Lorena Elvira-Hurtado,
Ramon Castroviejo Institute of Ophthalmologic Research, Universidad Complutense of Madrid, Spain
elenasalobrar@med.ucm.es, marelvir@ucm.es,

Mario Salas
Memory Unit, Geriatrics Service, Hospital Clínico San Carlos, Madrid, Spain
mario.salas@salud.madrid.org

Christoph Schommer, Luis A. Leiva
University of Luxembourg, Luxembourg
christoph.schommer@uni.lu, luis.leiva@uni.lu

***Abstract*—There is growing interest in developing reliable, non-invasive, and cost-effective methods for early diagnosis of neurodegenerative diseases such as Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). In this regard, handwriting-based tasks have shown potential in differentiating MCI and AD patients from healthy controls (HCs). However, previous work has reported mixed results when using different symbols and data representations. We address this research gap by developing computational models (convolutional and recurrent neural networks) to differentiate MCI and AD from HCs with off-line (scanned images) and on-line (discrete time series) house drawings. Notably, we observed that augmenting on-line data and then converting it to off-line format, a method we refer to as "OnOff-line", yielded the best performance results in binary classification tasks. These findings highlight the effectiveness of on-line representations in capturing handwriting dynamics more accurately. Ultimately, our work opens new avenues for future research to enhance automated diagnostic of MCI and AD from handwriting analysis.**

***Index Terms*—Mild Cognitive Impairment, Alzheimer's Disease, Handwriting, Drawing, Deep Learning, Classification.**

I. Introduction

Alzheimer's Disease (AD) is the primary cause of dementia, leading to substantial cognitive and behavioral decline [1], and is expected to impact up to 152 million individuals worldwide by 2050 [2]. The early detection and understanding of Mild Cognitive Impairment (MCI), which often precedes dementia, has garnered considerable attention from both researchers and healthcare professionals, as it represents a pivotal phase between healthy aging and AD. MCI is characterized by mild cognitive symptoms (e.g., memory and thinking skills) that do not significantly impede daily life activities. Although not all MCI cases progress to dementia, there is an elevated risk of evolving into AD [3], particularly for people who experience memory deficits [4], [5]. Furthermore, the availability of specific diagnostic tests can be constrained by factors such as economic limitations, healthcare infrastructure, or geographical location [6]. These restrictions can adversely affect the prompt diagnosis and intervention for individuals with cognitive impairments. This limitation underscores the need for alternative, patient-friendly diagnostic approaches.

Current research and technological advancements are focused on improving the accessibility and accuracy of diagnostic tools. Given the challenges and inaccuracies associated with traditional (manual) diagnosis methods for MCI and AD, there is growing interest in using Machine Learning (ML) to enhance these processes. For example, ML can identify biomarkers such as neurofibrillary tangles and senile plaques, which correlate highly with specific structural changes in the brain caused by AD [7]. This shift towards technology-driven diagnostics aims to streamline the early detection of AD and MCI, facilitating timely interventions that could potentially slow the progression of these neurodegenerative diseases.

Handwriting analysis has seen substantial growth recently, following recent advancements in neuroscience, and in particular, it has proven useful in diagnosing AD due to its ability to detect changes in cognitive and motor skills [8]–[10]. The community has acknowledged handwriting-based tests[1] as a viable, non-invasive method for early detection of

[1]In this paper, we consider *handwriting* and *hand-drawing* synonymous because both tasks involve the same neurophysiological and peripheral processes involved in motor control.

cognitive decline [11]–[13], a necessity in light of increasing AD prevalence and an aging population.

In the medical field, it is widely recognized that handwriting deterioration (e.g., irregular size, spacing, and letter formation) is one of the earlier indicators of cognitive disorders. This deterioration stems from the impact of cognitive diseases on motor skills, which involve cognitive, kinesthetic, and perceptual-motor abilities. Therefore, observing changes in handwriting can serve as a critical early sign of cognitive decline, highlighting the need for early diagnosis and interventions to mitigate the severity of these conditions [14].

The choice of symbols for handwriting/drawing tests significantly impacts the accuracy and utility of the results. The Clock Drawing Test (CDT), widely used for cognitive evaluation, has been shown to require large sample sizes to yield reliable results [15]. The Pentagon drawing test (PDT), which measures visuospatial abilities through a copy task, is not well-suited for detecting broader cognitive impairments beyond spatial skills [16]. Sentence-writing tasks [17] are language-dependent, which poses a challenge in multilingual settings or among individuals with language deficits, and handwritten signatures do not perform well for AD screening [18], [19].

Both the Tree Drawing Test (TDT) [20], [21] and the House Drawing Test (HDT) [22] are increasingly being used by clinicians nowadays. These tests evaluate a range of cognitive functions, including planning, organization, spatial awareness, and motor control [22]. Despite their potential, computational modeling using these symbols remains unexplored. The TDT, though useful, presents certain challenges. Trees can vary greatly in their structure and complexity, which may introduce variability in the interpretation of results. In contrast, the HDT offers a more standardized and relatable symbol. Drawing a house typically involves a combination of straight lines, angles, and geometric shapes, which can effectively assess visuoconstructional abilities and cognitive function without the variability introduced by more complex symbols. It is simple to administer, language independent and captures a wide range of cognitive abilities. Given these advantages, we develop computational models using HDT data in this paper.

Another area that remains underexplored is the use of different input representations that can improve model performance. Namely, the vast majority of previous work uses scanned images (off-line representations) for AD classification [15], [18], [23]–[25]. Only a handful of papers have explored the use of time series data (on-line representations) in this regard [11], [26]. We argue that on-line data can capture better handwriting dynamics; however, in the medical domain, these kinds of datasets are really small, so ML models tend to overfit [27]. This is arguably the main reason why previous work has only focused on off-line data. In this paper, we explore a novel approach: convert on-line data to off-line data. This enables the combination of fine-grained handwriting dynamics with static image data, for which computer vision DL models have proved useful for AD classification.

In sum, this paper makes the following contributions:

- Comparison of Deep Learning (DL) models for classification of AD, MCI, and HCs using both On-line and Off-line HDTs.
- On-line to off-line conversion: a novel approach where time series data are encoded into (pixel-based) intensity values as images.

II. Related Work

ML in digital medicine has recently shown considerable promise in enhancing healthcare outcomes and improving the efficiency of the diagnostic process [26], [28]–[30]. ML algorithms have been used to classify MCI, AD, and cognitive normal groups using neurocognitive tests, with high area under the ROC curve (AUC) values, indicating strong predictive performance [31]. Handwriting analysis, particularly in the context of AD and MCI, offers valuable insights for screening and diagnosis (e.g., [32]). Recent studies have identified distinct patterns in the early stages of AD through handwriting, demonstrating the efficacy of drawing skills as indicators of cognitive decline [10], [33]–[35].

Handwriting data can be captured either as scanned images, also known as *off-line* data, or as discrete sequences of $\{x, y\}$ (sometimes $\{x, y, t\}$) points, also known as *on-line* data. The former has been the focus of most of the previous work in AD screening [23], [29]. The latter has gained attention only recently [11], [26], which is surprising because, in other domains, it has been shown that on-line data representations provide richer movement dynamics, including, for example, detailed timing and (sometimes) pressure information [14], [36]–[38], which are features not available in off-line data representations.

When combined with cognitive functioning assessments, handwriting kinematic measures can differentiate between MCI, AD, and healthy controls (HCs). The reported classification accuracy ranges from 69% to 72% in differentiating participants, although the classification accuracy for the MCI group alone is relatively poor [39]. Müller et al. [40] reported that the digital Clock Drawing Test (dCDT) has a higher diagnostic accuracy for discriminating MCI patients from HCs compared to the conventional CDT (cCDT) (81.3% vs. 62.5%). Robens et al. [41] used the digital Tree Drawing Test (dTDT), achieving 77% AUC when discriminating MCI from HCs and 90% AUC when discriminating AD from HCs. Faundex et al. [22] used the House Drawing Test to analyze (not classify) handwriting movements in AD patients and HCs. They noted that on-line data revealed subtler motor impairments that traditional off-line methods would miss.

A study by Garre et al. [27] investigated handwriting and drawing copy-tasks involving several symbols (e.g., spiral, house, or crossed pentagons) to differentiate between AD, MCI, and HC. They found that kinematic features such as pen tip velocity and pressure could classify participants based on their cognitive status with varying accuracy, from 63.5% (drawing of a 3D house and CDTs) to 100% (drawing of a spiral). Unfortunately, the spiral symbol does not include considerations of fine-grained details and facets of spatial

awareness, planning, and memory, all of which are particularly affected in MCI and AD patients.

El-Yacoubi et al. [38] tried to identify cognitive states based on handwriting characteristics (e.g., velocity, acceleration, or stroke length) when copying predefined sentences. Their study revealed distinct clusters that corresponded to different cognitive profiles. For example, one cluster was dominated by HC and MCI patients, while another was dominated by MCI and early-stage AD patients. This clustering highlighted that MCI patients exhibited handwriting behaviors that were intermediate between HCs and early-stage AD patients. Another study by Raksasat et al. [42] developed an Attentive Pairwise Interaction Network (API-Net) aimed at enhancing the automatic scoring of the CDT. They achieved 79% F1-score, slightly outperforming a convolutional neural network (ResNet-152 by 3%) for multi-class classification.

As shown before, there is no consensus about which performance metric should be reported and which handwriting task should be administered. Further, very few datasets are available for AD and MCI screening, mostly in off-line form. To address these issues, we collected a comprehensive dataset [2] featuring HDTs, drawn by MCI and AD patients as well as HC, using *both* on-line and off-line data. Our dataset is, therefore, unique in the sense that it encompassed off-line and on-line representations of the same drawing from the same subject. Accordingly, we compare and contrast the classification performance of various computational models (neural networks) using either off-line or on-line handwriting to detect MCI and AD. We also investigate a novel approach, which consists of converting on-line data to off-line, which turned out to outperform models trained from scratch on either on-line or off-line data. Taken together, our results provide new insights into the performance of ML models when classifying MCI and AD patients using different handwriting data representations.

III. Experimental Setup

Digital pens bring new challenges, including increased costs and a potential compromise in handwriting input due to variations in pen grip and user familiarity with digital interfaces [43]. To address these concerns, we used a Repaper tablet,[3] featuring conventional pencils equipped with a small accelerometer, which allowed us to replicate the natural paper-based handwriting experience while at the same time capturing discrete time series data of $\{x, y\}$ points. The on-line data capture of these $\{x, y\}$ points is crucial as it provides a continuous, real-time digital record of the writing process, enabling detailed analysis of handwriting trajectories. These trajectories are particularly meaningful in our study as they offer insights into neuromotor control in elderly populations, helping to differentiate between normal aging processes and specific impairments associated with MCI and AD.

In addition to on-line data, off-line data captures static end results of the handwriting, such as the final drawing. The key difference between off-line and on-line data is that while off-line data provides a snapshot of the final product, on-line data captures the process. By comparing both off-line and on-line versions from the same patient, we can gain a comprehensive understanding of both the result and the process of handwriting, which is essential for diagnosing and differentiating between MCI and AD.

A. Participants

TABLE I
Demographic information of our dataset.

	HC	MCI	AD
	(n = 11)	(n = 25)	(n = 22)
Gender (male + female)	3 + 8	15 + 10	5 + 17
Age (M ± SD)	82.63 ± 2.46	81.44 ± 5.89	79.36 ± 4.09
MMSE (M ± SD)	29.91 ± 0.83	27.60 ± 2.18	23.45 ± 3.57

We recruited 58 individuals aged between 70 and 89 years at the Memory Unit of the Hospital Clinico San Carlos (HCSC) in Madrid, including 25 patients with MCI, 22 patients with AD, and 11 HC; see Table I and Figure 1. All individuals had normal vision and no hearing problems. Cognitive capabilities were assessed in a clinical setting optimized to minimize distractions and background noise. This assessment used the Mini-Mental State Examination (MMSE) [44] to evaluate cognitive status. Patients were classified into MCI or AD based on criteria from the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS), Alzheimer's Disease and Related Disorders Association (ADRDA) [45], and the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [46]. Statistical analysis with one-way ANOVA tests revealed no significant age differences between the three groups ($F(2, 55) = 2.04, p > .139$), indicating a balanced distribution across participants. Conversely, significant differences in MMSE scores were found among the groups ($F(2, 55) = 25.64, p < .0001$), with the HC group scoring higher than the MCI and AD groups ($p < .0001$), as expected.

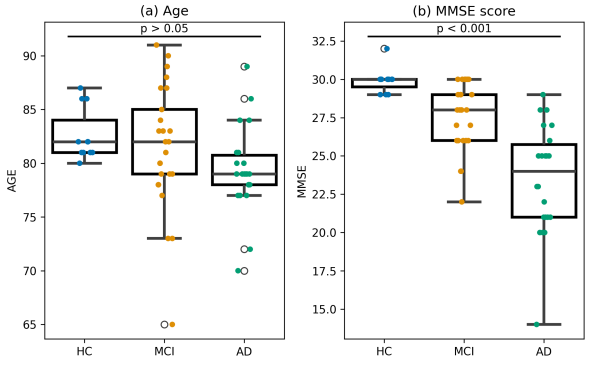

Fig. 1. Data distribution of (a) age and (b) MMSE score of HC, MCI, and AD groups.

[2]Available upon reasonable request.
[3]https://www.iskn.co/

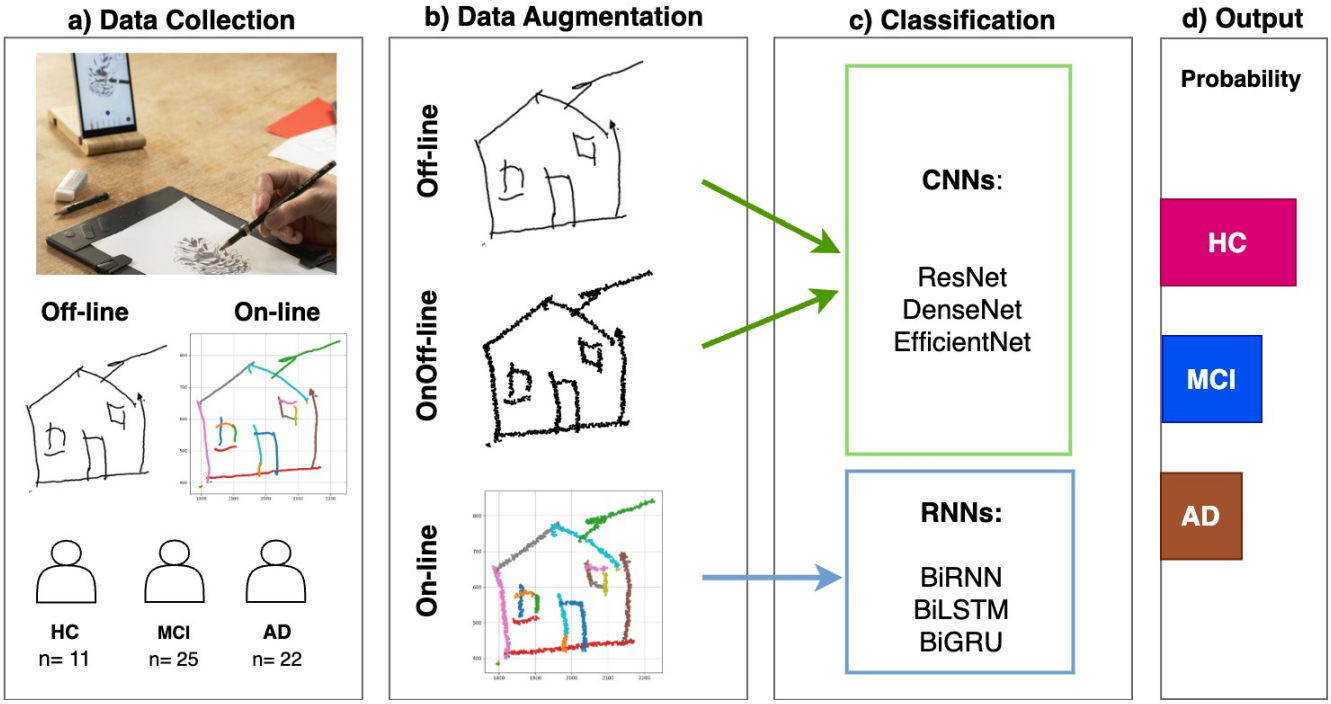

Fig. 2. Full system pipeline, from data collection and preprocessing to model training and classification.

Figure 1 presents the distribution of (a) age and (b) MMSE scores for the three groups. The HC group displays a broader age range with a slightly higher median age compared to the MCI and AD groups. Notably, the statistical test indicates no significant difference ($p > .05$) in age distribution among the groups, suggesting that age alone does not allow to differentiate between these groups. The boxplots reveal a clear descending trend in median MMSE scores from HC to AD, indicating a progressive cognitive decline. The MCI scores lie between those of the HC and AD groups, consistent with their intermediary diagnostic status. The difference across groups was statistically significant ($p < .001$).

B. Procedure

Participants were tasked to draw a house with the instrumented pencil (1 drawing per participant, totaling 58 drawings). The Repaper tablet simultaneously recorded off-line and on-line data. The on-line data were stored as SVG files (default format in Repaper) and then converted to JSON files for later post-processing. The off-line data were digitized with an HP Color LaserJet Pro scanner, stored as PDF files (default format in LaserJet), and then converted into PNG files for later post-processing. The PNG files were enhanced with the Canny edge detector [47] and resized to a standard resolution of 224×224 px.

IV. Modeling Methodology

Figure 2 shows the full system pipeline. As explained later, we use Convolutional Neural Networks (CNNs) to classify off-line data and Recurrent Neural Networks (RNNs) to classify on-line data. We chose CNNs and RNNs for our tasks because of their effectiveness in analyzing complex drawing tasks such as the CDT [15], [48], HDT [22] and PDT tasks [16], which are commonly used to assess cognitive impairment [49] for both off-line and on-line data. CNNs are particularly effective for processing off-line data, such as pixel-based images, where they can detect spatial patterns and intricate details within the static drawings, essential for identifying signs of cognitive decline. On the other hand, RNNs, especially Long Short-Term Memory networks (LSTMs) and Gated Recurrent Units (GRUs), excel in analyzing on-line data by capturing the sequential nature of stroke movements, which is critical for understanding the temporal progression in tasks like drawing tests. This approach allows us to capitalize on the specific strengths of CNNs for off-line image analysis and RNNs for on-line sequential data, providing a comprehensive assessment of cognitive function.

A. Data Augmentation

Data augmentation is essential to increase the robustness and generalizability of ML models, especially in digital medicine, where sample sizes are quite often too small for today's ML standards. We considered three versions of data augmentation (Figure 2b), as follows.

Off-line version: We applied a series of standalone geometric transformations to the images, such as scaling and small rotations, in a selective manner to preserve the image semantics (e.g., no vertical flipping). For this, we used the

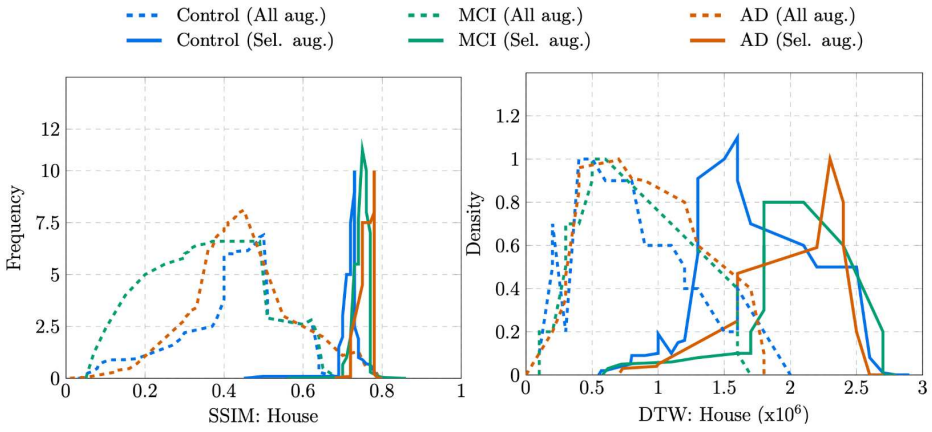

Fig. 3. Distributions of SSIM and DTW distributions across HC, MCI, and AD. Dashed plots correspond to the results considering all the augmentation techniques collectively (labeled as "All aug."). The solid plots show the results from the selected augmentation (labeled as "Sel. aug.") techniques.

Albumentations library.[4]

On-line version: The discrete point sequences were modified according to commonly used transformation methods for time series, such as jittering and scaling [50].

OnOff-line version: The augmented on-line data were converted to off-line data (PNG format), following the same steps indicated in Section III-B. This variant is a compromise solution between on-line and off-line approaches; the idea is to leverage the potentially larger variability produced by time series transformations and combine it with the high performance of pre-trained CNNs.

Thanks to data augmentation, we expanded our dataset to include 300 images (off-line representations) and 300 sequences (on-line representations), ensuring a balanced representation across categories. As an example, for the MCI category, we generated 75 unique variations from the initial set of 25 house drawings.

While augmented data should introduce variability, it should still retain the essential features of the original data to be useful for training. To ensure this, the quality and variability of the augmented dataset were rigorously assessed using the Structural Similarity Index Measure (SSIM) [51] for off-line data and the Dynamic Time Warping (DTW) [52] metric for on-line data. SSIM is a similarity metric where 0 indicates no similarity, and 1 means full similarity. SSIM is commonly used to measure the visual similarity between images by considering changes in structural information, luminance, and contrast. DTW is a distance metric where values greater than 0 indicate deviation from full similarity. DTW, on the other hand, compares time-series data by aligning sequences in a way that minimizes the cumulative distance.

Note that, in this context, "similarity" refers to maintaining the structural integrity of the original data while introducing the desired variations. The goal of using similarity measures in the context of data augmentation is not to create augmented data identical to the originals but rather to ensure that the transformations do not lead to the loss of important features or introduce unrealistic characteristics. Too high SSIM or DTW values might indicate insufficient augmentation, whereas too high SSIM values might indicate that the augmented data is too similar to the original, suggesting insufficient variability and, therefore, potentially less effective augmentation. Conversely, SSIM values that are too low might indicate excessive distortion, where the structural integrity of the data is compromised. Similarly, higher values suggest greater deviation from the original time-series patterns for DTW, while very low values could imply that the augmentation did not introduce meaningful variations. Thus, both SSIM and DTW provide critical insights into the balance between maintaining essential features and introducing sufficient variability, ensuring that the augmented data remains valid for model training.

Figure 3 represents the plots for SSIM and DTW. With SSIM scores ranging between 0.7 and 0.75 (see Figure 3 b) and DTW ranging from 180 to 4987 (see Figure 3b), we confirm that the augmented data comprise novel variations rather than mere replicas of the originals.

B. Convolutional Neural Networks

CNNs are inspired by the hierarchical structure of the human visual cortex [53] and are widely used for image classification tasks in digital medicine (e.g., [54]). We selected three state-of-the-art CNNs pre-trained on the popular ImageNet dataset, which offers a vast range of images across multiple categories:[5] ResNet50 [55], DenseNet121 [56], and EfficientNet [57]. ResNet uses skip connections to allow gradients to flow through the network directly, preventing the vanishing gradient problem and enabling the training of very deep networks. DenseNet features a unique architecture where each layer is connected to every other layer in a feed-forward

[4]https://albumentations.ai/

[5]https://www.image-net.org/

fashion, significantly reducing the number of parameters and enhancing feature propagation. EfficientNet scales up CNNs in a more structured manner using a compound coefficient to ensure that depth, width, and resolution grow uniformly. These models were fine-tuned to our dataset using transfer learning [58].

C. Recurrent Neural Networks

RNNs are preferred to handle sequential data, where understanding spatiotemporal dynamics is important [59]. Since no pre-trained RNN models for sequence classification are currently available as open source, we designed three models from scratch, each based on one type of RNN memory cell.

Bidirectional vanilla RNN (BiRNN, with no memory), which operates without the use of gating mechanisms, makes it simpler and faster for tasks where long-term dependencies are less critical. This model analyzes the sequence dynamics in a straightforward manner, though it may struggle with longer sequences due to the vanishing gradient problem.

Bidirectional Long Short-Term Memory (BiLSTM) [60] uses LSTM units to capture long-range dependencies within the data effectively. This model is particularly adept at handling the challenges of sequence classification, where understanding across large time lags is crucial. The bidirectional architecture enhances its capability to integrate context from both past and future inputs, providing a robust analysis of the sequence's temporal features.

Bidirectional Gated Recurrent Unit (BiGRU) [61] uses GRU cells that streamline the architecture of LSTMs while retaining their ability to manage long-term dependencies. GRUs simplify the gating mechanism found in LSTMs, leading to faster training times without a significant trade-off in performance. Like BiLSTM, the bidirectional approach allows the BiGRU to glean comprehensive insights from both directions of the sequence, enhancing its predictive accuracy in complex scenarios.

Generally, bi-directionality allows the models to analyze an input sequence in the forward and backward directions, offering a more comprehensive understanding of the sequence's temporal features. Our three RNNs have one hidden layer comprising 128 neurons selected through Bayesian optimization with the *Keras Tuner* library [62], hyperbolic tangent as an activation function, and a dropout rate of 0.1 to prevent overfitting. This setup precedes a softmax output layer, ensuring a probabilistic distribution over the classification labels as model output (Figure 2d).

D. Training Procedure

We split our augmented dataset as disjoint partitions of: 80% training, and 20% testing. The test partition simulates unseen data, as it is only used for final model evaluation. We train our CNN and RNN models using the Adam optimizer with a learning rate of $\eta = 0.001$ and momentums $\beta_1 = \beta_2 = 0.99$. The loss function is binary cross-entropy for two-class classification tasks (in our case, two categories: HC and patient) and categorical cross-entropy for multi-class classification tasks (in our case, three categories: HC, MCI, and AD). We use a batch size of 32 and train each model for up to 100 epochs with early stopping (patience of 40 epochs, meaning that if validation loss does not improve over 40 consecutive epochs, training stops, preserving the optimal model weights).

To ensure a consistent proportion of samples across different classes, we used a stratified 5-Fold Cross-Validation, a type of K-Fold Cross-Validation that splits the entire set into k number of folds. The number of folds, k, was set to 5. Consequently, the training set–which represents 80% of the whole dataset–was split into 5 folds. The first fold was used as the validation set, while the remaining 4 folds served as the training set. This process was repeated 5 times to guarantee that the entire set was used for both training and validation purposes.

E. Evaluation Metrics

To assess the performance of our models, we compute classification accuracy (Acc) and Area Under the Receiver Operating Characteristic curve (AUC). Together, these two metrics serve as fundamental tools in the assessment of predictive models, particularly in applications where the balance between sensitivity (true positive rate) and specificity (true negative rate) is crucial.

Accuracy is defined as the percentage of true cases (true positives and true negatives) that are correctly identified relative to the total number of cases. This metric offers a straightforward measure of a model's overall performance in correctly predicting outcomes. The AUC, on the other hand, provides insight into the discriminative power of any classifier. It is calculated by plotting the true positive rate against the false positive rate and measuring the area under the resulting curve. The true positive rate quantifies the model's ability to correctly identify actual positives, while the false positive rate measures how often the model incorrectly classifies negatives as positives. If AUC = 50%, the classifier is no better than random guessing, highlighting its ineffectiveness.

V. Results and Discussion

We first evaluated the performance of the proposed CNN and RNN models, with and without data augmentation, for the three data representations considered (Off-line, OnOff-line, and On-line). Figure 4 shows that data augmentation yields notable improvements in all cases. Without it, most models behave like a random classifier, especially when considering binary classification tasks; see the leftmost plots in Figure 4. This emphasizes that data augmentation is essential to train competent computational models.

Subsequently, we found that both jittering and scaling were particularly effective in improving classification accuracy for models that rely on on-line data representations. This contradicts previous findings in Parkinson's disease (PD) screening, where jittering failed to improve performance due to the introduction of noise that mimicked dyskinesia charateristics [50]. It is important to note that time series data should not be excessively modified to avoid significant distortions after data augmentation. Contrary to studies in PD (e.g., [50]),

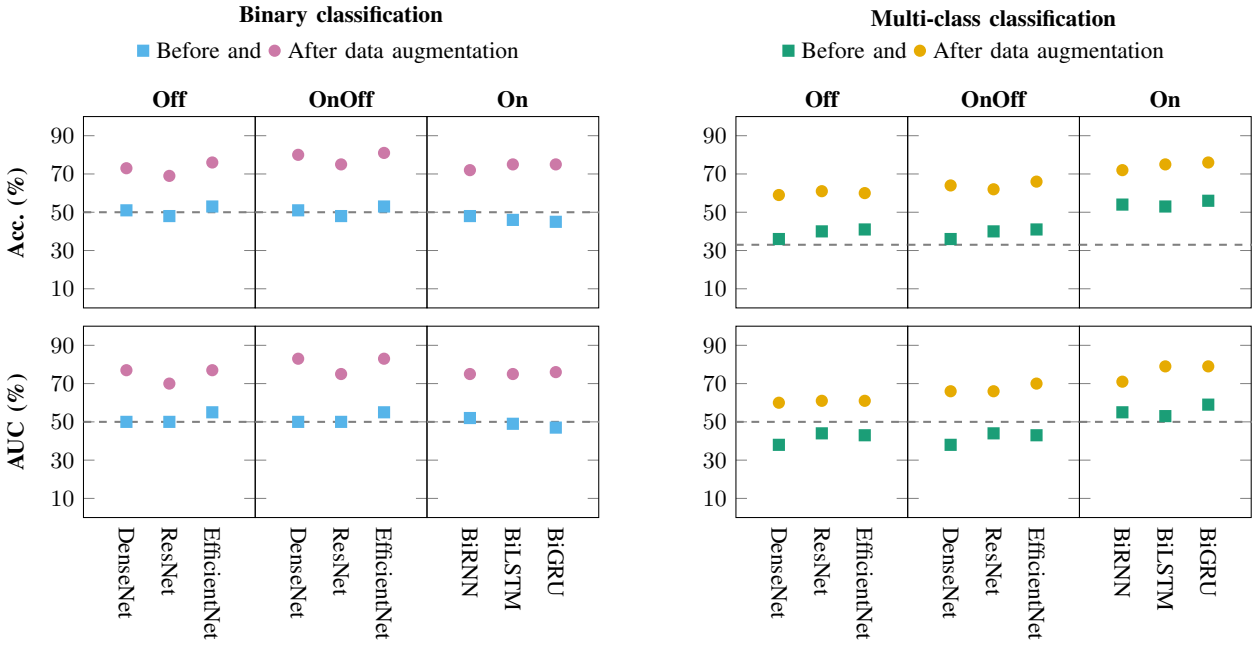

Fig. 4. Classification accuracy (top row) and AUC scores (bottom row) using different models. Dashed lines indicate the performance of a random classifier, serving as an empirical lower bound for comparison.

our methods proved effective, suggesting that the type of data augmentation and the nature of the disease significantly influence model outcomes.

Our study aims to go beyond merely achieving accurate and high-performance classification results; it seeks to understand the effectiveness of different input data formats as novel diagnostic methods. Overall, our CNN and RNN models succeeded in distinguishing AD patients from HC or MCI, outperforming random guessing by a large margin. Interestingly, our experiments achieved the best results in binary classification tasks when augmenting on-line (time series) data and then converting them to off-line (images) data, referred to as OnOff-line in our experiments. To the best of our knowledge, this is the first work to report this finding.

Conversely, RNNs, specifically BiGRU, achieved the best performance for multi-class classification tasks. In binary classification tasks, MCI and AD patients are considered in the same group, which may introduce some noise and ambiguities in the multi-class case, affecting the BiGRU model's performance. However, for multi-class classification tasks, where HC, MCI, and AD are treated as separate classes, BiGRU models can effectively use their architecture to distinguish between these groups more accurately. BiGRU worked best for multi-class classifications and performed as well as EfficientNet for binary classifications, making it the most versatile model for our dataset.

CNNs have previously been recognized for their high accuracy in predicting the conversion from MCI to AD [33], [63]. ML models have also been used to detect the progression of AD stages. For example, Bucholc et al. [64] introduced a hybrid Random Forest model that achieved 87.5% accuracy; however, the reported performance varied across different measures (e.g., MRI, age, and cognitive measures). In contrast, our models deliver higher accuracy across classification tasks and are robust across different data formats. Another study by Piers et al. [65] used on-line data to examine neurocognitive behavior over time, demonstrated the utility of digital pen technology in cognitive evaluation. In line with these observations, our study shows that on-line data representations are preferred, even when converted to off-line data.

Table II presents the classification results from various studies involving handwriting analysis to diagnose AD, using both off-line and on-line representations and across different datasets. Where applicable, each model's performance is evaluated under binary and multi-class classification scenarios. When using off-line data representations, EfficientNet models on pentagon drawings achieved the highest binary classification accuracy. This is followed closely by the InceptionResNetV2 models on letter drawings, which also show robust performance. In multi-class classification, where the complexity of class differentiation increases, all models generally show reduced performance, exemplifying the challenges posed by more complex classification tasks. On-line data representations performed best in this case.

Classification accuracy varies notably across the different drawings and formats, underscoring the influence of drawing complexity and data format on model performance. Our findings suggest that simpler shapes may facilitate higher accuracy in binary classification due to fewer complexities distinguishing ADs from HCs. However, as the task complex-

ity increases in multi-class scenarios, where the model must differentiate between multiple stages or types of cognitive impairment, model performance generally declines. This is evident in the more complex "Clock" and "Letter" drawings, which involve more intricate details and potentially more variation in individual execution.

TABLE II
CLASSIFICATION RESULTS AND COMPARISON TO THE STATE OF THE ART.

	Drawing	Models	Binary	Multi-class
Off-line data	Clock [15]	DenseNet-121	74%	61%
	Letter [66]	InceptionResNetV2	74.6%	N/A
	Pentagon [16]	EfficientNet	87%	76%
	House (Off), ours	EfficientNet	76%	60%
	House (OnOff), ours	EfficientNet	82%	66%
On-line data	Clock [67]	DL	83.44%	N/A
	Signature [68]	SVM	75.71%	N/A
	Letter [11]	1D-CNN	89%	N/A
	House (On), ours	BiGRU	72%	61%
	House (OnOff), ours	EfficientNet	82%	66%

Note: All these studies used data augmentation to improve their model performance.

The MCI group showed characteristics that were intermediate between HCs and ADs, which made it challenging to distinguish them clearly, as it was also stated in Werner et al. work [10]. However, our study demonstrated better performance in distinguishing MCI from HC and AD with higher accuracy in our multi-class classification task, with an accuracy of (See Table II). Compared with previous work, several factors of our study stand out:

1) **Data augmentation**: Our results demonstrate that data augmentation significantly enhances model performance, which aligns with findings by Dao et al. [11] and Bensalah et al. [26] (see Table II). Unlike their studies, which focused on GANs, our work shows that traditional augmentation techniques like jittering and scaling can also yield high accuracy. This highlights the versatility of these simple augmentation techniques across different datasets and model architectures.
2) **Model performance**: Similar to Bucholc et al. [64], who reported high accuracy with a hybrid Random Forest model, our study finds that BiGRU performs exceptionally well in multi-class classification tasks. This suggests that RNNs, particularly BiGRU, are highly effective in capturing the temporal dynamics of handwriting data, which is critical for distinguishing between HC, MCI, and AD.
3) **Data Representations**: Our findings that on-line data representations are preferred, even when converted to off-line data (here OnOff-line format), are consistent with Piers et al. [12], who suggested that using both on-line and off-line handwriting analysis with deep transfer learning and GANs can improve early-stage Alzheimer's disease detection. This underscores the importance of preserving temporal information in handwriting data for accurate classification.
4) **Digital Drawing Tools**: The use of digital drawing tools, as highlighted in previous studies [41], [65], [69]), supports our approach of using on-line pen stroke data to analyze drawing characteristics indicative of cognitive impairment. All those previous studies emphasized the potential of these tools for early dementia detection but only a handful of them actually did some computational modeling tasks.

VI. LIMITATIONS AND FUTURE WORK

We acknowledge that relying solely on a single type of drawing task (in our case, a house) may limit the generalizability of our findings across different clinical settings where other forms of cognitive assessments are used. Although, as discussed in the Introduction section, the HDT evaluates a range of cognitive functions [22], therefore, in line with recent work, it should become the standard task to screen MCI and AD patients. We also should note that previous work has considered pentagons [70], clocks [15], and signatures [18], [71], achieving similar results (sometimes lower) to ours.

On the other hand, our experimental findings are drawn from a relatively small sample size, which might limit their applicability to a broader population of patients. However, the challenge of recruiting a large and diverse cohort is a common and pervasive issue in digital medicine [14], [72]. Despite these limitations, our results hold promise and could pave the way for future clinical applications using a simple handwriting test as a non-invasive, low-cost, and accurate method.

VII. CONCLUSION

Handwriting analysis has been used by neurologists assessing suspected dementia patients in conjunction with a range of other measurements and tests. We have investigated how neural networks can use off-line and on-line house drawings to classify AD and MCI patients. We have observed that on-line data converted to off-line data is the most efficient approach to distinguish patients from HC (binary classification), whereas on-line data representations are preferred over off-line data for distinguishing AD, MCI, and HC (multi-class classification). Taken together, our results show the potential to enhance home-based healthcare services, using non-invasive, low-cost handwriting tests like the one we have investigated in this paper.

ACKNOWLEGMENTS

Work supported by the UCM research group (Grupo de Investigación básica en Ciencias de la Visión del IIORC, UCM-GR17-920105), the Horizon 2020 FET program of the European Union (grant CHIST-ERA-20-BCI-001), and the European Innovation Council Pathfinder program (grant 101071147).

REFERENCES

[1] Alzheimer's Association, "2022 Alzheimer's disease facts and figures," *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, vol. 18, no. 4, pp. 700–789, Apr. 2022.

[2] Z. Breijyeh and R. Karaman, "Comprehensive review on alzheimer's disease: causes and treatment," *Molecules*, vol. 25, no. 24, p. 5789, 2020.

[3] C. DeCarli, "Mild cognitive impairment: prevalence, prognosis, aetiology, and treatment," *The Lancet Neurology*, vol. 2, no. 1, pp. 15–21, Jan. 2003.

[4] K. Kantarci, S. D. Weigand, S. A. Przybelski, M. M. Shiung, J. L. Whitwell, S. Negash, D. S. Knopman, B. F. Boeve, P. C. O'Brien, R. C. Petersen, and C. R. Jack, "Risk of dementia in MCI: combined effect of cerebrovascular disease, volumetric MRI, and 1H MRS," *Neurology*, vol. 72, no. 17, pp. 1519–1525, Apr. 2009.

[5] R. C. Petersen, B. Caracciolo, C. Brayne, S. Gauthier, V. Jelic, and L. Fratiglioni, "Mild cognitive impairment: a concept in evolution," *Journal of Internal Medicine*, vol. 275, no. 3, pp. 214–228, 2014.

[6] Y. H. El-Hayek, R. E. Wiley, C. P. Khoury, R. P. Daya, C. Ballard, A. R. Evans, M. Karran, J. L. Molinuevo, M. Norton, and A. Atri, "Tip of the iceberg: assessing the global socioeconomic costs of alzheimer's disease and related dementias and strategic implications for stakeholders," *Journal of Alzheimer's Disease*, vol. 70, no. 2, pp. 323–341, 2019.

[7] D. G. Munoz and H. Feldman, "Causes of Alzheimer's disease," *Cmaj*, vol. 162, no. 1, pp. 65–72, 2000.

[8] M. Delazer, L. Zamarian, and A. Djamshidian, "Handwriting in Alzheimer's Disease," *Journal of Alzheimer's Disease*, vol. 82, no. 2, pp. 727–735, Jan. 2021.

[9] N. D. Cilia, C. De Stefano, F. Fontanella, M. Molinara, and A. Scotto Di Freca, "Handwriting analysis to support alzheimer's disease diagnosis: a preliminary study," in *Computer Analysis of Images and Patterns: 18th International Conference, CAIP 2019, Salerno, Italy, September 3–5, 2019, Proceedings, Part II 18*. Springer, 2019, pp. 143–151.

[10] P. Werner, S. Rosenblum, G. Bar-On, J. Heinik, and A. Korczyn, "Handwriting Process Variables Discriminating Mild Alzheimer's Disease and Mild Cognitive Impairment," *The Journals of Gerontology: Series B*, vol. 61, no. 4, pp. P228–P236, Jul. 2006.

[11] Q. Dao, M. A. El-Yacoubi, and A.-S. Rigaud, "Detection of Alzheimer Disease on Online Handwriting Using 1D Convolutional Neural Network," *IEEE Access*, vol. 11, pp. 2148–2155, 2023.

[12] N. D. Cilia, T. D'Alessandro, C. De Stefano, F. Fontanella, and M. Molinara, "From online handwriting to synthetic images for Alzheimer's disease detection using a deep transfer learning approach," *IEEE Journal of Biomedical and Health Informatics*, 2021.

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

[14] D. Impedovo and G. Pirlo, "Dynamic Handwriting Analysis for the Assessment of Neurodegenerative Diseases: A Pattern Recognition Perspective," *IEEE reviews in biomedical engineering*, vol. 12, pp. 209–220, 2018.

[15] N. Hosseini-Kivanani, C. Schommer, and L. A. Leiva, "The Magic Number: Impact of Sample Size for Dementia Screening Using Transfer Learning and Data Augmentation of Clock Drawing Test Images," in *Healthcom*, China, 2023, pp. 23–28.

[16] N. Hosseini-Kivanani, E. Salobrar-Gracía, L. Elvira-Hurtado, I. López-Cuenca, R. de Hoz, J. M. Ramírez, P. Gil, M. Salas, C. Schommer, and L. A. Leiva, "Ink of insight: Data augmentation for dementia screening through deep learning," in *Proceedings of the 8th International Conference on Medical and Health Informatics (ICMHI)*, Japan, Yokohama, 2024, pp. –.

[17] M. Faundez-Zanuy, J. Mekyska, and D. Impedovo, "Online Handwriting, Signature and Touch Dynamics: Tasks and Potential Applications in the Field of Security and Health," *Cognitive Computation*, vol. 13, no. 5, pp. 1406–1421, Sep. 2021.

[18] N. Hosseini-Kivanani, E. Salobrar-Gracía, L. Elvira-Hurtado, I. López-Cuenca, R. de Hoz, J. M. Ramírez, P. Gil, M. Salas, C. Schommer, and L. A. Leiva, "Better Together: Combining Different Handwriting Input Sources Improves Dementia Screening," in *IEEE e-Science*, Cyprus, 2023, pp. 1–7.

[19] M. Faundez-Zanuy, A. Hussain, J. Mekyska, E. Sesa-Nogueras, E. Monte-Moreno, A. Esposito, M. Chetouani, J. Garre-Olmo, A. Abel, Z. Smekal *et al.*, "Biometric applications related to human beings: there is life beyond security," *Cognitive Computation*, vol. 5, pp. 136–151, 2013.

[20] S. Robens, P. Heymann, R. Gienger, A. Hett, S. Müller, C. Laske, R. Loy, T. Ostermann, and U. Elbing, "The digital tree drawing test for screening of early dementia: an explorative study comparing healthy controls, patients with mild cognitive impairment, and patients with early dementia of the alzheimer type," *Journal of Alzheimer's Disease*, vol. 68, no. 4, pp. 1561–1574, 2019.

[21] M. Stanzani Maserati, C. Matacena, L. Sambati, F. Oppi, R. Poda, M. De Matteis, R. Gallassi *et al.*, "The tree-drawing test (koch's baum test): a useful aid to diagnose cognitive impairment," *Behavioural neurology*, vol. 2015, 2015.

[22] M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcobé, J. Garré-Olmo, K. Lopez-de Ipiña, and J. Solé-Casals, "Online drawings for dementia diagnose: in-air and pressure information analysis," in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013: MEDICON 2013, 25-28 September 2013, Seville, Spain*. Springer, 2014, pp. 567–570.

[23] C. Jiménez-Mesa, J. E. Arco, M. Valentí-Soler, B. Frades-Payo, M. A. Zea-Sevilla, A. Ortiz, M. Ávila Villanueva, D. Castillo-Barnes, J. Ramírez, T. Del Ser-Quijano, C. Carnero-Pardo, and J. M. Górriz, "Using Explainable Artificial Intelligence in the Clock Drawing Test to Reveal the Cognitive Impairment Pattern," *International Journal of Neural Systems*, vol. 33, no. 04, p. 2350015, Apr. 2023.

[24] J. Maruta, K. Uchida, H. Kurozumi, S. Nogi, S. Akada, A. Nakanishi, M. Shinoda, M. Shiba, and K. Inoue, "Deep convolutional neural networks for automated scoring of pentagon copying test results," *Scientific Reports*, vol. 12, no. 1, p. 9881, Dec. 2022.

[25] S. Chen, D. Stromer, H. A. Alabdalrahim, S. Schwab, M. Weih, and A. Maier, "Automatic dementia screening and scoring by applying deep learning on clock-drawing tests," *Scientific Reports 2020 10:1*, vol. 10, no. 1, pp. 1–11, Nov. 2020.

[26] A. Bensalah, A. Parziale, G. De Gregorio, A. Marcelli, A. Fornés, and J. Lladós, "I Can't Believe It's Not Better: In-air Movement for Alzheimer Handwriting Synthetic Generation," in *Graphonomics in Human Body Movement. Bridging Research and Practice from Motor Control to Handwriting Analysis and Recognition*, A. Parziale, M. Diaz, and F. Melo, Eds., Cham, 2023, pp. 136–148.

[27] 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 Research*, vol. 14, no. 9, pp. 960–968, 2017.

[28] J. Y. C. Chan, B. K. K. Bat, A. Wong, T. K. Chan, Z. Huo, B. H. K. Yip, T. C. Y. Kowk, and K. K. F. Tsoi, "Evaluation of Digital Drawing Tests and Paper-and-Pencil Drawing Tests for the Screening of Mild Cognitive Impairment and Dementia: A Systematic Review and Meta-analysis of Diagnostic Studies," *Neuropsychology Review*, vol. 32, no. 3, pp. 566–576, 2022.

[29] N. Cilia, T. D'Alessandro, C. De Stefano, and F. Fontanella, "Deep transfer learning algorithms applied to synthetic drawing images as a tool for supporting Alzheimer's disease prediction," *Machine Vision and Applications*, vol. 33, May 2022.

[30] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017.

[31] S. Nagaraj and T. Q. Duong, "Deep learning and risk score classification of mild cognitive impairment and alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 80, no. 3, pp. 1079–1090, 2021.

[32] J. Kawa, A. Bednorz, P. Stępień, J. Derejczyk, and M. Bugdol, "Spatial and dynamical handwriting analysis in mild cognitive impairment," *Computers in Biology and Medicine*, vol. 82, pp. 21–28, 2017.

[33] G. Poirier, A. Ohayon, A. Juranville, F. Mourey, and J. Gaveau, "Deterioration, Compensation and Motor Control Processes in Healthy Aging, Mild Cognitive Impairment and Alzheimer's Disease," *Geriatrics*, vol. 6, no. 1, p. 33, Mar. 2021.

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

[35] K. E. Forbes, M. F. Shanks, and A. Venneri, "The evolution of dysgraphia in Alzheimer's disease," *Brain Research Bulletin*, vol. 63, no. 1, pp. 19–24, Mar. 2004.

[36] L. A. Leiva, M. Diaz, M. A. Ferrer, and R. Plamondon, "Human or Machine? It Is Not What You Write, But How You Write It," in *ICPR*, Jan. 2021, pp. 2612–2619.

[37] N. Mwamsojo, F. Lehmann, M. A. El-Yacoubi, K. Merghem, Y. Frignac, B.-E. Benkelfat, and A.-S. Rigaud, "Reservoir Computing for Early Stage Alzheimer's Disease Detection," *IEEE*, vol. 10, pp. 59 821–59 831, 2022.

[38] 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 Recognition*, vol. 86, pp. 112–133, 2019.

[39] P. Werner, S. Rosenblum, G. Baron, J. Heinik, and A. Korczyn, "Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment." *The journals of gerontology. Series B, Psychological sciences and social sciences*, vol. 61 4, pp. P228–36, 2006.

[40] 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 in Aging Neuroscience*, vol. 9, 2017.

[41] S. Robens, P. Heymann, R. Gienger, A. Hett, S. Müller, C. Laske, R. Loy, T. Ostermann, and U. Elbing, "The digital tree drawing test for screening of early dementia: An explorative study comparing healthy controls, patients with mild cognitive impairment, and patients with early dementia of the alzheimer type." *Journal of Alzheimer's disease : JAD*, vol. 68 4, pp. 1561–1574, 2019.

[42] R. Raksasat, S. Teerapittayanon, S. Itthipuripat, K. Praditpornsilpa, A. Petchlorlian, T. Chotibut, C. Chunharas, and I. Chatnuntawech, "Attentive pairwise interaction network for ai-assisted clock drawing test assessment of early visuospatial deficits," *Scientific Reports*, vol. 13, no. 1, p. 18113, 2023.

[43] S. Marullo, M. Pozzi, M. Malvezzi, and D. Prattichizzo, "Analysis of postures for handwriting on touch screens without using tools," *Scientific Reports*, vol. 12, no. 1, p. 296, Jan. 2022.

[44] T. Tombaugh and N. J. McIntyre, "The Mini-Mental State Examination: A Comprehensive Review," *Journal of the American Geriatrics Society*, vol. 40, no. 9, pp. 922–935, 1992.

[45] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology*, vol. 34, no. 7, pp. 939–939, Jul. 1984.

[46] M. Guha, "Diagnostic and Statistical Manual of Mental Disorders: DSM-5 (5th edition)," *Reference Reviews*, vol. 28, no. 3, pp. 36–37, Jan. 2014.

[47] G. Bradski, "The openCV library," *Miller Freeman Inc*, vol. 25, no. 11, pp. 120–123, 2000.

[48] C. Jiménez-Mesa, J. E. Arco, M. Valentí-Soler, B. Frades-Payo, M. A. Zea-Sevilla, A. Ortiz, M. Ávila Villanueva, D. Castillo-Barnes, J. Ramírez, T. del Ser-Quijano, C. Carnero-Pardo, and J. M. Górriz, "Automatic Classification System for Diagnosis of Cognitive Impairment Based on the Clock-Drawing Test," in *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications*, ser. Lecture Notes in Computer Science, J. M. Ferrández Vicente, J. R. Álvarez Sánchez, F. de la Paz López, and H. Adeli, Eds. Springer International Publishing, 2022, pp. 34–42.

[49] T. Charernboon, "Diagnostic accuracy of the overlapping infinity loops, wire cube, and clock drawing tests for cognitive impairment in mild cognitive impairment and dementia," *International Journal of Alzheimer's Disease*, vol. 2017, 2017.

[50] 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. ICMI '17*, Nov. 2017, pp. 216–220.

[51] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[52] P. Senin, "Dynamic time warping algorithm review," *Information and Computer Science Department University of Hawaii, USA*, vol. 855, no. 1-23, p. 40, 2008.

[53] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[54] B.-L. Chen, K.-T. Hu, K.-S. Cheng, and C.-Y. Chen, "Automatic CDT Scoring Using Machine Learning with Interpretable Feature," in *Proc. ICBBB' 24*, Mar. 2024, pp. 55–59.

[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. the IEEE CVPR*, 2016, pp. 770–778.

[56] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proc. the IEEE CVPR*, 2017, pp. 4700–4708.

[57] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. the 36th International Conference on Machine Learning*. PMLR, May 2019, pp. 6105–6114.

[58] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, May 2016.

[59] R. Ghosh, "A Recurrent Neural Network based deep learning model for offline signature verification and recognition system," *Expert Systems with Applications*, vol. 168, p. 114249, Apr. 2021.

[60] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[61] K. Cho, B. V. Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," in *EMNLP*, 2014.

[62] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi *et al.*, "Kerastuner," https://github.com/keras-team/keras-tuner, 2019.

[63] Y. C. Youn, J.-M. Pyun, N. Ryu, M. J. Baek, J.-W. Jang, Y. H. Park, S.-W. Ahn, H.-W. Shin, K.-Y. Park, and S. Y. Kim, "Use of the Clock Drawing Test and the Rey–Osterrieth Complex Figure Test-copy with convolutional neural networks to predict cognitive impairment," *Alzheimer's Research & Therapy*, vol. 13, no. 1, p. 85, Apr. 2021.

[64] M. Bucholc, S. Titarenko, X. Ding, C. Canavan, and T. Chen, "A hybrid machine learning approach for prediction of conversion from mild cognitive impairment to dementia," *Expert Systems with Applications*, vol. 217, p. 119541, May 2023.

[65] R. J. Piers, K. N. Devlin, B. Ning, Y. Liu, B. Wasserman, J. M. Massaro, M. Lamar, C. C. Price, R. Swenson, R. Davis, D. L. Penney, R. Au, and D. J. Libon, "Age and Graphomotor Decision Making Assessed with the Digital Clock Drawing Test: The Framingham Heart Study," *Journal of Alzheimer's Disease*, vol. 60, no. 4, pp. 1611–1620, Nov. 2017.

[66] N. D. Cilia, T. D'Alessandro, C. De Stefano, and F. Fontanella, "Offline handwriting image analysis to predict alzheimer's disease via deep learning," in *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 2807–2813.

[67] R. Binaco, N. Calzaretto, J. Epifano, S. McGuire, M. Umer, S. Emrani, V. Wasserman, D. J. Libon, and R. Polikar, "Machine Learning Analysis of Digital Clock Drawing Test Performance for Differential Classification of Mild Cognitive Impairment Subtypes Versus Alzheimer's Disease," *Journal of the International Neuropsychological Society*, vol. 26, no. 7, pp. 690–700, Aug. 2020.

[68] Z. Wang, M. Abazid, N. Houmani, S. Garcia-Salicetti, and A.-S. Rigaud, "Online signature analysis for characterizing early stage alzheimer's disease: A feasibility study," *Entropy*, vol. 21, no. 10, p. 956, 2019.

[69] S. Müller, L. Herde, O. Preische, A. Zeller, P. Heymann, S. Robens, U. Elbing, and C. Laske, "Diagnostic value of digital clock drawing test in comparison with CERAD neuropsychological battery total score for discrimination of patients in the early course of Alzheimer's disease from healthy individuals," *Scientific Reports*, vol. 9, no. 1, p. 3543, Mar. 2019.

[70] I. Park, Y. J. Kim, Y. J. Kim, and U. Lee, "Automatic, Qualitative Scoring of the Interlocking Pentagon Drawing Test (PDT) Based on U-Net and Mobile Sensor Data," *Sensors*, vol. 20, no. 5, p. 1283, Jan. 2020.

[71] G. Pirlo, M. Diaz, M. A. Ferrer, D. Impedovo, F. Occhionero, and U. Zurlo, "Early diagnosis of neurodegenerative diseases by handwritten signature analysis," *Lecture Notes in Computer Science*, vol. 9281, pp. 290–297, 2015.

[72] G. Vessio, "Dynamic Handwriting Analysis for Neurodegenerative Disease Assessment: A Literary Review," *Applied Sciences 2019, Vol. 9, Page 4666*, vol. 9, no. 21, pp. 4666–4666, Nov. 2019.