information

MDPI

Article

Cognitive Handwriting Insights for Alzheimer's Diagnosis: A Hybrid Framework

Shafiq Ul Rehman [1,*,†] and Uddalak Mitra [2,†]

[1] College of Information Technology, Kingdom University, Riffa 3903, Bahrain
[2] Department of Computer Science and Engineering, JIS College of Engineering, Kalyani 741235, West Bengal, India; uddalak.mitra@jiscollege.ac.in
* Correspondence: s.rehman@ku.edu.bh
† These authors contributed equally to this work.

Abstract: Alzheimer's disease (AD) is a persistent neurologic disorder that has no cure. For a successful treatment to be implemented, it is essential to diagnose AD at an early stage, which may occur up to eight years before dementia manifests. In this regard, a new predictive machine learning model is proposed that works in two stages and takes advantage of both unsupervised and supervised learning approaches to provide a fast, affordable, yet accurate solution. The first stage involved fuzzy partitioning of a gold-standard dataset, DARWIN (Diagnosis AlzheimeR WIth haNdwriting). This dataset consists of clinical features and is designed to detect Alzheimer's disease through handwriting analysis. To determine the optimal number of clusters, four Clustering Validity Indices (CVIs) were averaged, which we refer to as cognitive features. During the second stage, a predictive model was constructed exclusively from these cognitive features. In comparison to models relying on datasets featuring clinical attributes, models incorporating cognitive features showed substantial performance enhancements, ranging from 12% to 26%. Our proposed model surpassed all current state-of-the-art models, achieving a mean accuracy of 99%, mean sensitivity of 98%, mean specificity of 100%, mean precision of 100%, and mean MCC and Cohen's Kappa of 98%, along with a mean AUC-ROC score of 99%. Hence, integrating the output of unsupervised learning into supervised machine learning models significantly improved their performance. In the process of crafting early interventions for individuals with a heightened risk of disease onset, our prognostic framework can aid in both the recruitment and advancement of clinical trials.

Keywords: Alzheimer's disease prediction; hybrid machine learning in disease diagnosis; predicting Alzheimer's disease through handwriting analysis; accurate prediction of Alzheimer's disease using machine learning

check for updates

Academic Editors: Sami Azam, Zubaer Ibna Mannan and Kawsar Ahmad

Received: 11 February 2025
Revised: 6 March 2025
Accepted: 16 March 2025
Published: 20 March 2025

Citation: Ul Rehman, S.; Mitra, U. Cognitive Handwriting Insights for Alzheimer's Diagnosis: A Hybrid Framework. *Information* **2025**, *16*, 249. https://doi.org/10.3390/info16030249

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Dementia poses a formidable global health challenge, impacting over 55 million individuals worldwide, with pronounced effects on low- and middle-income nations [1]. Annually, close to 10 million new cases emerge, positioning it as the seventh leading cause of mortality and placing a significant strain on global healthcare systems [1]. Among the elderly, dementia is a primary contributor to disability and dependence, with informal caregivers devoting an average of 5 h per day to care-giving duties [1,2]. Notably, women shoulder a disproportionate share, accounting for 70% of care-giving hours. Alzheimer's disease (AD) represents the majority of dementia cases, ranging from 60% to 70% [1], and is characterized by progressive cognitive decline and extensive neural damage and

atrophy in muscle tissue [3,4]. With the absence of curative interventions, early diagnosis remains pivotal before dementia symptom onset [5,6]. With global life expectancy on the rise, the incidence of neurodegenerative disorders is projected to surge in the forthcoming decades [7]. Estimates suggest that by 2050, the number of individuals affected by AD and other dementia will soar to 152 million [8] or even more [9,10], underscoring the urgent need for advanced clinical approaches to bolster early detection.

The progression of Alzheimer's disease (AD) can be more accurately described using the 7-stage model [11], as outlined in Table 1. Early diagnosis of AD holds significant potential [6], particularly when patients remain functionally independent (Stage 2 or 3 of the 7-stage model) and are dementia-free [6]. This early detection window may extend up to 8 years before the onset of dementia symptoms [5]. In the pursuit of early Alzheimer's detection, we introduce a novel hybrid machine learning (ML) model that blends unsupervised cognition and supervised learning. This model utilizes a gold-standard dataset called DARWIN (Diagnosis AlzheimeR WIth haNdwriting), comprising 1D features extracted from handwriting analysis following the protocol outlined in Cilia et al.'s work [7]. Leveraging handwriting analysis for detecting Alzheimer's disease has its own advantages, as it is a non-invasive and cost-effective approach. Further, it can function as a pre-screening procedure, before resorting to more costly medical diagnoses such as MRI or PET scans.

Table 1. Stages and levels of cognitive impairment. This table outlines the various stages of cognitive impairment, along with their corresponding levels of decline.

Stage	Level of Impairments	Details
1	No impairment	Healthy
2	Very mild cognitive decline	Healthy
3	Mild cognitive decline	Healthy
4	Moderate cognitive decline	Early-stage dementia
5	Moderately severe cognitive decline	Early mid-stage dementia
6	Severe cognitive decline	Late mid-stage dementia
7	Very severe cognitive decline	Late-stage dementia

Employing two distinct methodologies, unsupervised and supervised learning within a unified model, we characterize our model as a hybrid model operating in two stages. The model first employs a fuzzy-logic-based unsupervised learning algorithm on the DARWIN dataset to generate a fuzzy membership matrix corresponding to the optimal number of cluster centers, crucial for understanding patient condition patterns. Fuzzy logic is predominantly utilized in this study due to its ability to effectively address uncertainties inherent in data, particularly in fields such as medical diagnosis, where precision in diagnosis may be crucially enhanced via fuzzification [12].

We conceptualize the fuzzy membership matrix as a dataset imbued with cognitive features. In this analogy, each cluster center represents a distinct cognitive feature, while the membership degree of a handwriting sample corresponds to the value of that feature. The term "cognitive feature" is employed to signify the output of the Fuzzy C-Means algorithm, as it encapsulates patterns extracted from handwriting data believed to correlate with cognitive function, particularly in the context of Alzheimer's disease prediction. In the second stage, these cognitive features serve as inputs for constructing predictor models using supervised learning techniques. We anticipate that employing unsupervised cognitive values will enhance the performance of supervised ML models, promising more accurate Alzheimer's risk identification and dementia prevention. The main contributions of this study can be summarized as follows:

- Introduction of a Novel Hybrid Model: The research introduces a novel predictive model designed to identify Alzheimer's disease (AD) at an early stage (in stage 2 or 3, as mentioned in Table 1). This hybrid model seamlessly integrates both unsupervised and supervised approaches, delivering a harmonious blend of rapidity, streamlined design, and enhanced accuracy that surpasses current state-of-the-art methodologies.
- Integration of Fuzzy Logic with Machine Learning: A creative strategy merges "Fuzzy Logic" with "Machine Learning", enhancing the efficacy of conventional classifiers and introducing a novel dimension into medical diagnosis.
- Lightweight Model for Medium Datasets: Introducing a lightweight model tailored for training on medium-sized datasets, ensuring practicality and efficiency, especially in settings with limited data accessibility.
- Comprehensive Set of Baselines: Incorporating a diverse set of baseline classifiers for benchmarking, enabling a thorough assessment of the proposed model against existing methods of AD detection.
- Use of Clustering Validity Index: Multiple clustering validity indices are employed to determine the optimal number of clusters, enhancing data pattern comprehension by considering membership values.
- Extensive Hyperparameter Optimization: Automated hyperparameter optimization ensures the identification of the most optimal model, resulting in enhanced accuracy in AD detection.

Related Work

Machine learning (ML) has become a powerful tool for predicting neurodegenerative diseases [13–19], including Alzheimer's disease (AD), through the analysis of various data types, including medical images [20–25], genetic information [26–28], cognitive assessments [29–31], and multi modal approaches [32]. However, the utilization of handwriting-based analysis holds its own significance. This form of cognitive assessment offers a direct measure of brain function, capturing individual motor patterns [33,34] and is a well-established means for detection of neurodegenerative diseases [34–39]. Furthermore, due to its non-invasive nature and cost-effectiveness, such a technique could be considered a pre-screening routine by practitioners, before resorting to expensive diagnostic methods such as MRI and PET scans. This makes it particularly appealing in low- and middle-income countries, which account for 60% of global Alzheimer's disease incidents [1]. Despite being a significant cause of dementia, attempts to predict AD through handwriting analysis remain limited, due to the lack of benchmark datasets.

Crafting a benchmark dataset hinges on establishing a meticulously defined protocol for both data collection and subsequent analysis. Recognizing the significance of a standardized protocol in constructing such a benchmark dataset, Cilia et al. [7] introduced a comprehensive protocol for collecting and analyzing handwriting data to detect Alzheimer's disease. Their protocol recommends the use of 25 specific tasks (detailed descriptions of the tasks are given in Supplementary Table S1), designed based on insights from motor control and neuroscience. Each task aims to delve into the intricate dynamics of how specific brain regions contribute to both handwriting and drawing movements performed by the participants. By elucidating how deviations in these brain regions manifest in the overall progression of Alzheimer's disease, the protocol further describes each task using 18 unique measurable features (detailed descriptions of the features are given in Supplementary Table S2). This collaborative effort culminated in the creation of the DARWIN dataset (Diagnosis AlzheimeR WIth haNdwriting) with 450 features (25 tasks × 18 features per task), representing the largest publicly available, as well as comprehensive and invalu-

able, resource for applying machine learning techniques to analyze handwriting data in the context of Alzheimer's disease diagnosis.

Building upon previous research conducted in [40,41], the present investigation acknowledges an opportunity to refine the existing predictive model for Alzheimer's disease (AD) by leveraging handwriting analysis with the DARWIN dataset. In the study by De Gregorio [40], a set of classifiers, particularly Random Forest (RF), K-Nearest-Neighbor (KNN), Linear Discriminant Analysis (LDA), Gaussian Naive Base (GNB), and Support Vector Machine (SVM) models, were trained using task-specific features, resulting in multiple predictors of AD based on particular tasks. Subsequently, these predictors were amalgamated to form the final prediction model. The development of task-specific predictors entailed the application of a single type of machine learning algorithm. Their findings suggest that, when employing Random Forest (RF) to construct task-specific classifiers and then ensemble them, the achieved performance metrics were notable: an overall accuracy of 91%, with a sensitivity of 83% and a specificity of 100%.

In contrast, Cilia et al. [41] employed diverse machine learning algorithms (Random Forest, Logistic Regression, K-Nearest Neighbor, Linear Discriminant Analysis, Gaussian Naive Bayes, Support Vector Machine, Decision Tree, Multilayer Perceptron, and Learning Vector Quantization) to construct task-specific models, combining the best-performing models for the ultimate prediction by ensembling them. They achieved a highest accuracy of 94.28%, specificity of 88.24%, and sensitivity of 100%. However, both approaches treated the set of task-specific features as a unified entity, overlooking potential correlations among the features.

In a notable study by Parziale et al. [42], an intriguing approach was taken to evaluate the performance of three one-class classifier models—namely, the Negative Selection Algorithm, the Isolation Forest, and the One-Class Support Vector Machine—on the DARWIN dataset. Remarkably, they achieved an accuracy of 97.12%, a sensitivity of 94.23%, and a specificity of 100.00% with the random Negative Selection Algorithm.

In the study of Subha R. et al. [43], a hybrid ML model with Swarm Intelligence (SI)-based feature selection was developed for detection of Alzheimer's disease. They employed several machine learning models, specifically LR, KNN, SVM, DT, RF, and AdaBoost, achieving a highest accuracy of 90%, precision of 88%, recall of 92%, F1-score 90%, and AUC-ROC score 90% with the RF and AdaBoost classifiers.

In Gattulli V. et al. [44], a wide set of classification models, including Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Bayesian Networks (BN), Gaussian Naïve Bayes (GNB), Multilayer Perceptron (MP), and Learning Vector Quantization (LVQ), were used for a group of selected tasks, encompassing different types of writing and drawing tasks, and varying levels of complexity, from simple to more complex tasks. Their reported highest accuracy was 88%, with a specificity of 86% and sensitivity of 90% with RF classifiers.

In the work of Onder M. et al. [45], the authors conducted a comprehensive study to diagnose AD using four different categorization methods. These methods included XGBoost, GradientBoost, AdaBoost, and Voting Classification algorithms. The highest accuracy obtained was 85% by the XGBoost classifier.

In Hakan Ö et al. [46], a tipple ensemble machine learning model was developed and applied for AD detection, using Light Gradient Boosting Machine, Categorical Boosting, and Adaptive Boosting machine learning classification algorithms combined with a Hard Voting Classifier. The reported results of the experimental studies using the proposed ensemble methodology were 97.14% Acc, 95% Prec, 100% Recall, 90.25% Spec, and 97.44% F1-score.

In Mitra and Rehman [47], a stacking-based ensemble integrating multiple classifiers achieved superior performance metrics, including 97.14% accuracy, 95% sensitivity, and 100% specificity. Feature selection methods like ANOVA and RFE further enhanced the model performance.

Interestingly, along with machine-learning-based approaches, deep-learning-based models have also been applied on the DARWIN dataset for detection of AD. On the other hand, in Erdogmus P. et al. [48], a Convolutional Neural Network (CNN) was employed on the DARWIN dataset designed for AD detection through handwriting. They extracted 2D features from the original 1D dataset and applied these to their proposed model. Training and evaluation on this 2D dataset resulted in a 90.4% accuracy for their proposed model. This suggests the potential of deep learning approaches for early and effective AD diagnosis.

In conclusion, while current methods for Alzheimer's Disease (AD) detection show promise, there is a clear need for methodological advancement. Despite achieving commendable accuracies, future research should focus on refining feature selection, exploring new algorithmic approaches, and incorporating additional data sources. By continually improving upon existing methodologies, the field of AD detection can advance towards earlier and more effective diagnosis. Table 2 presents a summary of the recent methods targeting the DARWIN dataset, while in Figure 1, we present a workflow of the proposed approach and Table 2 presents it in detail.

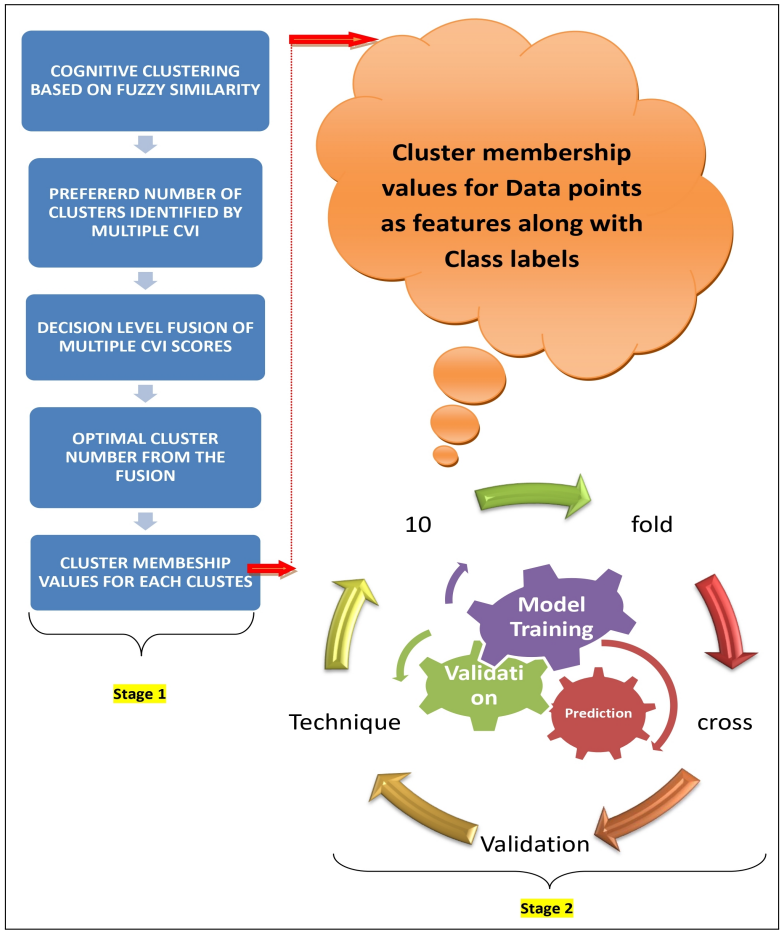


Figure 1. The workflow of the proposed approach.

Table 2. Workflow stages and their purpose.

Stage	Steps	Purpose
Stage 1:	Cognitive clustering based on fuzzy similarity	Group data points based on fuzzy similarity to handle uncertainty in handwriting patterns.
	Preferred number of clusters identified by multiple CVI	Use multiple cluster validity indices (CVI) to determine the optimal number of clusters.
	Decision-level fusion of multiple CVI scores	Combine multiple CVI scores to make a robust decision on the optimal cluster count.
	Optimal cluster number from the fusion	Select the best cluster number that represents the data structure effectively.
	Cluster membership values for each cluster extracted as features	Use cluster membership values as features for supervised classification.
Stage 2:	Cluster membership values used as features, along with class labels	Integrate clustering results into the machine learning model as additional features.
	Ten-fold cross-validation technique applied	Ensure model generalization and prevent overfitting by using cross-validation.
	Model training performed	Train a supervised learning model using the extracted features.
	Prediction and validation processes executed	Evaluate model performance and validate predictions for classification accuracy.

The rest of the paper is structured as follows: Section 2 describes the materials and methods used for this study. Section 3 presents the experimental results and discussion. Finally, Section 4 concludes the paper by summarizing the findings, with directions for future work.

2. Materials and Methods

In this section, we present a description of the (i) dataset construction, (ii) software stack used for end-to-end implementation of the model, (iii) development of the proposed novel model, (iv) performance evaluation metrics, and (v) training of the proposed model.

2.1. Dataset Construction

The DARWIN (Diagnosis AlzheimeR WIth haNdwriting) dataset has been meticulously curated to facilitate the early detection of Alzheimer's disease by analyzing handwriting features based on clinical processes. This valuable dataset is openly accessible at https://archive.ics.uci.edu/dataset/732/darwin (accessed on 15 March 2025), containing data from 174 individuals. Among them, 89 persons were diagnosed with Alzheimer's disease, while 85 were healthy participants. The demographic information about the dataset is presented in Table 3. The dataset was created by combining data from 25 unique neuroscience guide tasks, categorized as Memory and dictation (M), Graphic (G), and Copy (C), as described in Supplementary Table S1. Each task contributes 18 distinct handwriting features, detailed in Supplementary Table S2. In total, this compilation comprises 450 features (25 tasks $\times$ 18 features per task), providing a comprehensive resource for researchers and professionals involved in the early diagnosis of Alzheimer's disease. The term clinical feature is henceforth used to denote these 450 features.

Table 3. Summary of studies on Alzheimer's disease prediction using the DARWIN dataset.

Study	ML Model Used	Feature/Data Modality	Techniques Used	Accuracy
De Gregorio et al. [40]	RF, KNN, LDA, GNB, SVM	Task specific	Ensemble multiple task-specific classifiers built upon a single type of ML model	91%
Cilia et al. [41]	RF, LR, KNN, LDA, GNB, SVM, DT, MLP, LVQ	Task specific	Ensemble multiple task-specific classifiers built upon top-performing and different types of ML models	94.28%
Parziale et al. [42]	Negative Selection Algorithm, Isolation Forest, One-Class SVM	All features	One-class classifiers are utilized	97%
Subha et al. [43]	LR, KNN, SVM, DT, RF, AdaBoost with Swarm Intelligence	All features	Swarm intelligence-based feature selection employed with ML models	90%
Gattulli et al. [44]	RF, LR, KNN, LDA, SVM, BN, GNB, MP, LVQ	Task specific	Mixed tasks of different levels of complexity	88%
Onder et al. [45]	XGBoost, GradientBoost, AdaBoost	All features	Categorization methods are employed	85%
Hakan et al. [46]	Ensemble of LGBM, CatBoost, AdaBoost	All features	Hard ensemble of the employed models	97%
Mitra and Rehman [47]	Stacking	Selected top-k features	Classifier-specific top-k features are utilized for training the classifiers with LR as the meta-level	97%
Erdogmus et al. [48]	CNN	1D features converted into 2D images	Pre-trained models were utilized for training the classifiers with the constructed 2D images	90%

2.1.1. Handwriting Data Collection Process

To assess handwriting characteristics in individuals with Alzheimer's Disease (AD), a structured set of 25 handwriting tasks was used. These tasks were carefully designed to evaluate different aspects of motor function (M), cognitive abilities (C), and graphical skills (G), which are often affected in AD.

Participants were given specific instructions to perform tasks such as drawing shapes, copying letters or words, writing dictated sentences, and recalling written information (details of these tasks are described in Supplementary Table S1). These tasks were performed using a digital tablet and stylus, which captured detailed handwriting movements in real time. The device recorded stroke patterns, speed, pressure, and tremor-related variations that are not visible to the naked eye. The collected handwriting data serve as a potential indicator of neurological and cognitive decline in AD patients.

Healthcare professionals supervised the process to ensure participants followed the instructions correctly, minimizing external influences like fatigue or stress that could have affected the results. The data were then stored digitally for further analysis.

2.1.2. Feature Extraction from Handwriting Data

After collecting handwriting samples, various features (details of these features are described in Supplementary Table S2) were extracted to quantify writing performance and detect abnormalities associated with AD. These features fall into different categories:

2.1.3. Time-Based Features (How Long the Writing Takes)

- Total Time (TT): Measures the total duration required to complete a task.
- Air Time (AT) & Paper Time (PT): Differentiates between the time spent writing on the paper and the time when the pen is lifted. A longer air time may indicate hesitation or difficulty in initiating movement.

2.1.4. Speed and Movement Features (How the Hand Moves During Writing)

- Mean Speed on-paper (MSP) & Mean Speed in-air (MSA): Measure the average speed of writing and movement in the air, which may be slower in AD patients due to motor impairments.
- Mean Acceleration (MAP, MAA): Assesses how quickly writing speed changes, indicating the ability to control movement.
- Mean Jerk (MJP, MJA): Measures how smoothly or abruptly the hand moves, providing insights into motor control and stability.

2.1.5. Pressure-Based Features (How Hard the Pen Is Pressed on the Paper)

- Pressure Mean (PM) & Pressure Variance (PV): Assess the force applied while writing. Inconsistencies in pressure may suggest declining motor function or hand tremors.

2.1.6. Tremor and Stability Features (How Steady the Writing Is)

- Generalized Mean Relative Tremor (GMRT, GMRTP, GMRTA): Identifies hand tremors, which are common in neurological disorders.

2.1.7. Spatial and Structural Features (How Writing Is Organized on the Page)

- Max X and Y Extensions (XE, YE): Measures the size of handwriting strokes. Changes in stroke size can indicate cognitive and motor decline.
- Dispersion Index (DI): Evaluates whether handwriting is spread evenly or scattered, which may reflect spatial planning difficulties.

2.1.8. Task-Specific Features (Additional Behavioral Indicators)

- Pen-down Number (PWN): Counts how many times the pen touches the paper, reflecting writing fluency and control.

Once these features had been extracted, they were analyzed using statistical and machine learning techniques to identify patterns that differentiated AD patients from healthy individuals. This approach provides objective, quantifiable measures that can complement traditional clinical assessments for early detection of Alzheimer's disease.

2.2. Software Stack

The software stack employed for system development encompassed a collection of Python libraries optimized for various stages of machine learning development, from data preprocessing to model evaluation. These libraries included

1. `pandas` and `numpy`: These libraries are fundamental for data manipulation, providing efficient data structures and array operations necessary for handling and processing datasets.

2. `sklearn.preprocessing.StandardScaler`: Used for standardizing data, ensuring features are on the same scale, which is crucial for many machine learning algorithms to perform optimally.
3. `scikit-learn (sklearn)`: This library forms the backbone of the machine learning pipeline, offering a comprehensive suite of algorithms and tools for modeling, evaluation, and cross-validation. Specific classifiers utilized included
    - `DecisionTreeClassifier` (DT)
    - `RandomForestClassifier` (RF)
    - `ExtraTreesClassifier` (ET)
    - `LogisticRegression` (LR)
    - `LinearDiscriminantAnalysis` (LDA)
    - `GaussianNB` (GNB)
    - `XGBClassifier` (XGB)
    - `DecisionTreeClassifier` (DT)
    - `MLPClassifier` (MLP)
    - `KNeighborsClassifier` (KNN)
4. `matplotlib`: Employed for data visualization, enabling the generation of insightful plots and graphs for data exploration, model evaluation, and comparison of results across different algorithms.

This software stack provides a robust environment for developing and analyzing machine learning models, covering a diverse range of algorithms and techniques suitable for various types of datasets and problem domains.

2.3. Development of the Proposed Model

The hybrid prognostic model presented in this study was formulated through a two-stage approach. A data-centric schematic representation of the model development process is illustrated in Figure 2.

2.3.1. Stage 1: Cognitive Feature Generation Using Unsupervised Learning

In the initial phase, we employed the Fuzzy C-Means (FCM) algorithm to facilitate grouping of the handwriting samples in the DARWIN dataset based on clinical features. Let, $X = \{x_1, x_2, ..., x_n\}$ be the set of handwriting samples, where each $x_i \in \mathbb{R}^d$ and $d = 450$ in our case. The objective function of the Fuzzy C-means algorithm is given by

$$J_m(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^m \|x_i - v_j\|^2 \tag{1}$$

where

- $U = [u_{ij}]$ is the fuzzy membership matrix, with $u_{ij} \in [0, 1]$ representing the membership of data point x_i to cluster j.
- $V = [v_j]$ is the matrix of cluster centers, with $v_j \in \mathbb{R}^d$.
- $m \in (1, \infty)$ is the fuzzifier parameter, controlling the degree of fuzziness.

Being an iterative algorithm, the steps involved in the Fuzzy C-means algorithm can be described as as follows:

1. Initialize U and V.
2. Repeat until convergence:
    (a) Update cluster centers v_j:

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m} \tag{2}$$

(b) Update the fuzzy membership matrix U:

$$u_{ij} = \left(\sum_{k=1}^{c} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (3)$$

We conceptualized the fuzzy membership matrix U as a dataset imbued with cognitive features. In this analogy, each cluster center represents a distinct cognitive feature, while the membership value of a handwriting sample corresponds to the value of that feature. Consequently, this perspective enabled us to interpret the transformation of 450 clinical features into cognitive features, leveraging the proficiency of FCM in handling the uncertainty inherent in medical data.

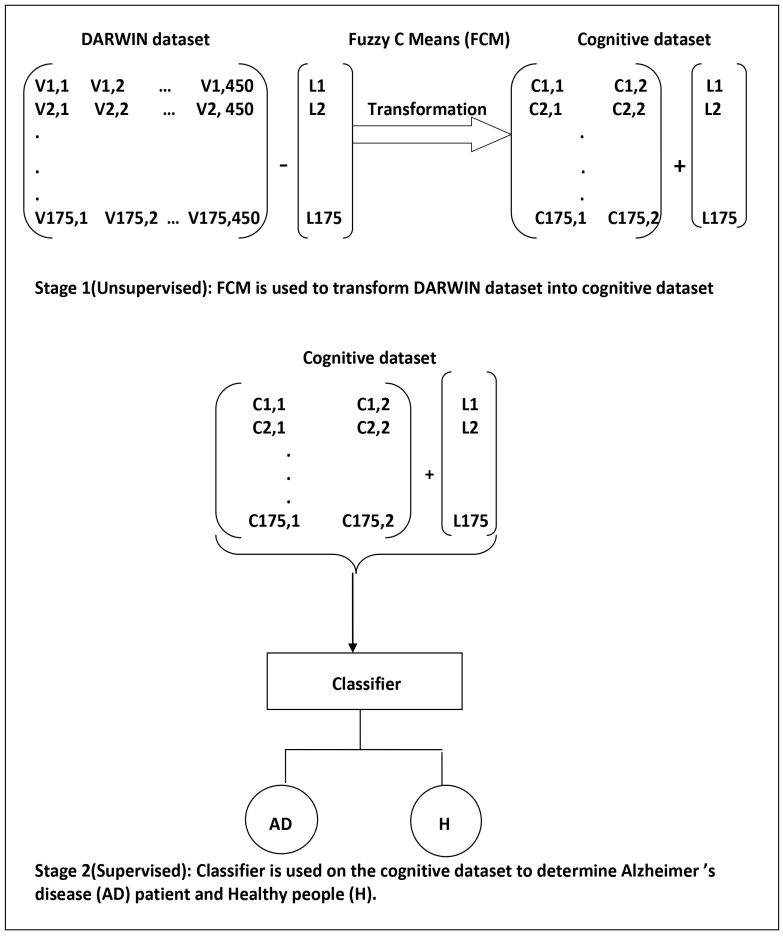


Figure 2. The data-centric model development process consisted of two stages. In Stage 1, unsupervised machine learning was employed, while Stage 2 involved the application of supervised machine learning. Within the context of the DARWIN dataset, **Vi,j** represents the **jth** feature value of the **ith** handwriting sample. **Ci,j** denotes the fuzzy membership value corresponding to the **jth** cluster center for the **ith** handwriting sample, while **Li** indicates the class label for the **ith** handwriting sample. It is important to note that in Stage 1, only the values **Vi,j** were input into the FCM algorithm. The labels were explicitly appended to the output of the FCM (Fuzzy C-Means) algorithm to train the supervised learning algorithms.

An important consideration when utilizing FCM is the determination of the optimal number of cluster centers, as this directly influences the data partitioning process. By selecting an appropriate number of cluster centers, we could effectively uncover significant patterns within the data and accurately perform a categorization. To quantitatively assess cluster quality and separation, we employed four widely recognized Clustering Valid-

ity Indices (CVIs): Silhouette Score [49], Davies–Bouldin Score [50], Calinski–Harabasz Score [51], and Dunn Score [52]. Clustering Validity Indices (CVIs) serve as metrics for evaluating the quality of clustering outcomes derived from clustering algorithms. Below, we elaborate on these four commonly utilized CVIs:

Silhouette Score: The Silhouette Score is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). It is calculated for each data point i using the following formula:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4}$$

where

- $a(i)$ is the average distance from point i to other points within the same cluster.
- $b(i)$ is the smallest average distance from point i to points in a different cluster.

The Silhouette Score for the entire dataset is the average of the Silhouette scores of all data points.

Davies–Bouldin Score: The Davies–Bouldin Score evaluates the average similarity between each cluster and its most similar cluster, relative to the average dissimilarity between points in different clusters. It is calculated as

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{\text{sim}(C_i, C_j) + \text{sim}(C_j, C_i)}{\text{sep}(C_i, C_j)} \right) \tag{5}$$

where

- k is the number of clusters.
- C_i and C_j are clusters.
- $\text{sim}(C_i, C_j)$ is a measure of similarity between clusters C_i and C_j.
- $\text{sep}(C_i, C_j)$ is a measure of dissimilarity between clusters C_i and C_j.

A lower Davies–Bouldin Score indicates better clustering.

Calinski–Harabasz Score: The Calinski–Harabasz Score, also known as the Variance Ratio Criterion, measures the ratio of between-cluster dispersion to within-cluster dispersion. It is given by

$$CH = \frac{B}{W} \times \frac{N - k}{k - 1} \tag{6}$$

where

- B is the between-cluster dispersion.
- W is the within-cluster dispersion.
- N is the total number of data points.
- k is the number of clusters.

A higher Calinski–Harabasz Score indicates denser and better-separated clusters.

Dunn Score: The Dunn Score assesses the compactness and separation of clusters. It is calculated as

$$D = \min_{1 \leq i \leq k} \left(\min_{j \neq i} \left(\frac{\text{sep}(C_i, C_j)}{\max_l (\text{diam}(C_l))} \right) \right) \tag{7}$$

where

- $\text{sep}(C_i, C_j)$ is the separation between clusters C_i and C_j.
- $\text{diam}(C_l)$ is the diameter of cluster C_l.

A higher Dunn Score indicates better clustering, with compact and well-separated clusters.

2.3.2. Stage 2: Prognostic Model Development Using Supervised Learning

In the second stage, we utilized a fuzzy membership matrix U, hereafter referred to as the cognitive dataset, to develop a predictor for Alzheimer's disease (AD) cases. Supervised machine learning involves leveraging a labeled training dataset to construct models capable of discerning patterns within data. These well-trained models are subsequently applied to an unlabeled testing dataset to accurately categorize its contents. During the training phase of our proposed supervised learning model, we employed the ground truths provided by the DARWIN dataset as labels for the cognitive dataset generated in stage 1. In this section, we present a succinct overview of ten state-of-the-art supervised machine learning algorithms [53] that have been employed to build models for the early detection of Alzheimer's disease.

Decision Tree (DT): A Decision Tree recursively splits data based on the most significant attribute. There is no specific mathematical equation for a Decision Tree itself, but it uses criteria like Gini impurity or entropy for splitting.

Random Forest (RF): Random Forest is an ensemble of decision trees. The final prediction is obtained by averaging or taking a majority vote from the individual trees.

Extra Trees (Extra Tree): Extra Trees, like Random Forest, is an ensemble of decision trees but with more randomness in feature selection and splitting. It uses the same mathematical rules for splitting nodes as traditional decision trees.

Logistic Regression (LR): Logistic Regression models the probability of an instance belonging to a particular class using the logistic function (sigmoid function).

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}} \tag{8}$$

Linear Discriminant Analysis (LDA): LDA finds the linear combinations of features that maximize the separation between different classes. Mathematical Equation for Linear Discriminant Analysis:

$$\mathbf{y} = \mathbf{W}^T \mathbf{x} \tag{9}$$

where $\mathbf{W}$ is the weight vector and $\mathbf{y}$ is the linear discriminant.

Gaussian Naive Bayes (GNB): GNB is based on Bayes' theorem and assumes that features are conditionally independent given the class. Mathematical Equation for Naive Bayes:

$$P(y|\mathbf{x}) \propto P(y) \prod_{i=1}^{n} P(x_i|y) \tag{10}$$

XGBoost (XGB): XGBoost is a gradient boosting algorithm that minimizes a loss function by adding decision trees iteratively. Objective Function in XGBoost:

$$\text{Obj}(\Theta) = \sum_{i=1}^{n} \mathcal{L}(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{11}$$

k-Nearest Neighbors (KNNs): KNNs assigns a class label to an instance based on the majority class among its k nearest neighbors. There is no specific mathematical equation for KNNs, but it uses distance metrics like Euclidean or Manhattan distance.

Support Vector Machine (SVM): A SVM finds the hyperplane that best separates different classes with the maximum margin. Mathematical Equation for Linear SVM:

$$\begin{aligned} &\text{maximize } \frac{2}{\|\mathbf{w}\|} \text{ subject to } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \\ &\text{for } i = 1, \ldots, n \end{aligned} \tag{12}$$

Multi-Layer Perceptron (MLP): A MLP is a type of artificial neural network with multiple layers of interconnected nodes, including input, hidden, and output layers. Mathematical Equations for Forward Propagation:

$$z_j^{(l)} = \sum_i w_{ji}^{(l)} a_i^{(l-1)} \tag{13}$$

$$a_j^{(l)} = \sigma(z_j^{(l)}) \tag{14}$$

The weights $w_{ji}^{(l)}$ in MLPs are learned through an iterative optimization process using backpropagation, where the network minimizes the error by adjusting weights based on the gradient of a loss function with respect to these weights. This is typically achieved via gradient descent or its variants. While a full discussion of backpropagation is beyond the scope of this article, it plays a crucial role in training MLPs by propagating errors backward and updating the weights accordingly.

2.3.3. The Proposed Model with Optimal Setup

The proposed model stands out for its algorithms and their respective hyperparameters. In Stage 1, we iteratively applied the Fuzzy C-Means (FCM) algorithm to the DARWIN dataset to ascertain the optimal number of cluster centers, denoted as k, a pivotal hyperparameter in the model. We determined the preferred number of cluster centers to be 2, using an average scoring scheme derived from the amalgamation of four normalized CVI scores.

In Stage 2, we employed ten supervised learning algorithms, alongside the FCM algorithm, to identify the most effective classifier. A crucial step involved fine-tuning the hyperparameters, achieved through the widely adopted Grid-Search-CV method. Grid Search Cross-Validation (Grid Search CV) systematically explores a predefined grid of hyperparameters, assessing each combination's performance through cross-validation, typically k-fold cross-validation. The optimal configuration is selected based on the best performance metrics on the validation set. Furthermore, we utilized 10-fold cross-validation with the chosen optimal hyperparameter values for each classifier, to evaluate its performance. Notably, Random Forest (RF) emerged as the most effective approach, in conjunction with the FCM algorithm, for predicting Alzheimer's disease. For detailed parameter configurations, please refer to Table 4.

Table 4. Average demographic data of participants. Standard deviations are shown in parentheses.

Participant Class	Age (Mean ± SD)	# Male	# Female
Patients	71.5 (9.5)	44	46
Healthy People	68.9 (12)	39	51

2.4. Algorithm

The algorithmic representation of the proposed model comprises three parts: Algorithms 1–3. Stage 1 is delineated into two distinct procedures: "CogniGen", outlined in Algorithm 1, and the Fuzzy C-Means, detailed in Algorithm 2, whereas Stage 2 is represented in Algorithm 3.

Algorithm 1 CogniGen computes "cognitive features" from "clinical features"

Input: Dataset X, maximum number of clusters N, fuzziness parameter m, termination criterion ϵ, maximum number of iterations I.
Output: Optimal number of clusters k_{optimal}, Final Cluster Centers $\mathbf{M}_{\text{optimal}}$, Fuzzy Membership Matrix $\mathbf{U}_{\text{optimal}}$, Silhouette Scores array S, Davies–Bouldin Scores array DB, Calinski–Harabasz Scores array CH, Dunn Scores array D for $k = 1$ to N.

Initialize arrays S, DB, CH, D.
Initialize k_{optimal} to 1.
Initialize maximum average score max_avg_score to a small value.
for $k = 1$ to N **do**
    Call FCM algorithm on input parameters X, k, m, ϵ, I to obtain cluster centers $\mathbf{M}$ and membership matrix U.
    Calculate Silhouette Score S_k using the obtained clustering result.
    Calculate Davies–Bouldin Score DB_k using the obtained clustering result.
    Calculate Calinski–Harabasz Score CH_k using the obtained clustering result.
    Calculate Dunn Score D_k using the obtained clustering result.
    Update arrays $S[k]$, $DB[k]$, $CH[k]$, $D[k]$ with respective scores.
end for
for $k = 1$ to N **do**
    Calculate average score $avg_score_k = (S[k] + DB[k] + CH[k] + D[k])/4$.
    if $avg_score_k >$ max_avg_score **then**
        Update max_avg_score to avg_score_k.
        Update k_{optimal} to k.
    end if
end for
Call FCM algorithm on input parameters $X, k_{\text{optimal}}, m, \epsilon$, max_iterations to obtain final cluster centers $\mathbf{M}_{\text{optimal}}$ and membership matrix U_{optimal}.
Output $k_{\text{optimal}}, \mathbf{M}_{\text{optimal}}, \mathbf{U}_{\text{optimal}} S, DB, CH, D$.

Algorithm 2 Fuzzy C-Means Clustering

Input: Dataset X, number of clusters C, fuzziness parameter m, termination criterion ϵ, maximum number of iterations.
Output: Cluster centers $\mathbf{M}$, Membership matrix U.

Initialize C, the number of clusters.
Initialize $\mathbf{M}$, the matrix of cluster centers, with random values.
Choose the fuzziness parameter $m > 1$.
Initialize the termination criterion $\epsilon > 0$, and set the iteration counter $k = 0$.
repeat
    $k \leftarrow k + 1$.
    Update the membership u_{ij} for each data point x_i to each cluster j using:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - m_j\|}{\|x_i - m_k\|} \right)^{\frac{2}{m-1}}}$$

    Update the cluster centers $\mathbf{M}$ using:

$$m_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

    Calculate the objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \|x_i - m_j\|^2$$

until The change in J_m is less than ϵ or a maximum number of iterations is reached.
Output $\mathbf{M}, U$.

Algorithm 3 Random Forest Algorithm

Input: Training dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, number of trees T, number of features to consider at each split m
Output: Random forest model RF

function BUILDTREE(D, m)
    Randomly select m features from the dataset D
    Choose the best feature and split point based on impurity measure (e.g., Gini index, entropy)
    Split the dataset into two subsets D_{left} and D_{right} based on the selected feature and split point
    Create a tree node with the selected feature and split point
    if D_{left} is pure or stopping criterion is met **then**
        Attach a leaf node with majority class of D_{left} to the current node
    else
        Attach the result of BUILDTREE(D_{left}, m) to the current node
    end if
    if D_{right} is pure or stopping criterion is met **then**
        Attach a leaf node with majority class of D_{right} to the current node
    else
        Attach the result of BUILDTREE(D_{right}, m) to the current node
    end if
    return the tree node
end function

function RANDOMFOREST(D, T, m)
    $RF \leftarrow$ empty list to hold trees
    for $t \leftarrow 1$ to T **do**
        $D_t \leftarrow$ Bootstrap sample from D
        tree $\leftarrow$ BUILDTREE(D_t, m)
        Add *tree* to RF
    end for
    return RF
end function

2.5. Performance Evaluation Metrics

To identify the most precise set of classifiers, we analyzed and compared the scores of seven widely used performance metrics in machine learning: accuracy, sensitivity, specificity, precision, Matthews Correlation Coefficient (MCC), Cohen's Kappa, and AUC-ROC. In the subsequent section, we present a concise overview of the performance evaluation metrics employed in our study.

All the performance metrics in machine learning are based on the concept of a confusion matrix, which is a table (Table 5) used to assess the performance of a classification algorithm. The main components of a confusion matrix are

1. True Positive (TP): The number of instances correctly predicted as positive.
2. True Negative (TN): The number of instances correctly predicted as negative.
3. False Positive (FP): The number of instances incorrectly predicted as positive.
4. False Negative (FN): The number of instances incorrectly predicted as negative.

Table 5. Proposed Model and Hyperparameters.

Proposed Model	Hyper-Parameter
FCM + RF	Optimal cluster center = 2, {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50}

1. **Accuracy (Acc)**: The ratio of correctly predicted instances to the total number of instances in the dataset.

$$\text{Acc} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

2. **Sensitivity (SN)**: The proportion of actual positive cases that are correctly identified by the model.

$$\text{SN} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

3. **Specificity (SP)**: The proportion of actual negative cases that are correctly identified by the model.

$$\text{SP} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

4. **Precision**: The ratio of correctly predicted positive observations to the total predicted positive observations.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

5. **Matthews Correlation Coefficient (MCC)**: A measure of the quality of binary classifications, taking into account true and false positives and negatives.

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

6. **Cohen's Kappa**: A statistic measuring inter-rater agreement for qualitative items, accounting for agreements occurring by chance.

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

7. **Area Under the Receiver Operating Characteristic Curve (AUC ROC)**: A performance measurement for classification problems, representing the area under the ROC curve.

$$\text{AUC ROC} = \int_0^1 \text{TPR}(t)\, d(\text{FPR}(t))$$

2.6. Training of the Proposed Model

As a two-stage model, Stage 1 operates unsupervised, while Stage 2 is supervised. Setting up the working environment for the model required careful consideration. In Stage 1, we began by isolating the class label component from the DARWIN dataset and feeding it into the FCM algorithm. Since FCM is an unsupervised learning method, it autonomously extracted significant patterns from the data and did not require class information. Thus, in Stage 1, we observed the transformation of the DARWIN dataset into the cognitive dataset.

Moving to Stage 2, which operated under a supervised paradigm, we had to additionally merge the class label with the cognitive dataset generated in Stage 1. This amalgamated

cognitive dataset, inclusive of class labels, served as the foundation for training, cross-validation, and evaluation of the supervised learning algorithms.

The cognitive dataset was initially split into a training set (80%) and a held-out set (20%). The training set was then used for both hyperparameter tuning through GridSearchCV and subsequent model evaluation via 10-fold cross-validation. GridSearchCV optimized the model hyperparameters by internally employing cross-validation, and we used 5-fold cross validation. Following hyperparameter selection, 10-fold cross-validation assessed the model's performance across various subsets of the training data. Finally, the model's performance was validated on the held-out test set (20%), providing an unbiased estimate of its generalization to unseen data. This methodology effectively utilized the available data for model development and evaluation, while ensuring a robust performance estimation through separate test data evaluation. Furthermore, we repeated the whole procedure 10 times. This iterative approach enhanced the confidence in the model's predictive capabilities and guarded against overfitting or biases introduced by a single train–test split.

3. Results

Our research efforts spanned a broad spectrum of experiments geared towards crafting a reliable and precise tool for detecting Alzheimer's disease in its early stages. Using a two-stage approach, our experiments can be classified into those utilizing unsupervised learning techniques and others employing supervised learning methods. All experiments were carried out on a modest computational setup comprising an Intel(R) Core(TM) i5-5200U CPU @ 2.20 GHz, 4GB RAM, and a 1TB HDD, supplemented by computational resources from the freely accessible version of Google Colab.

3.1. Results from Unsupervised Machine Learning

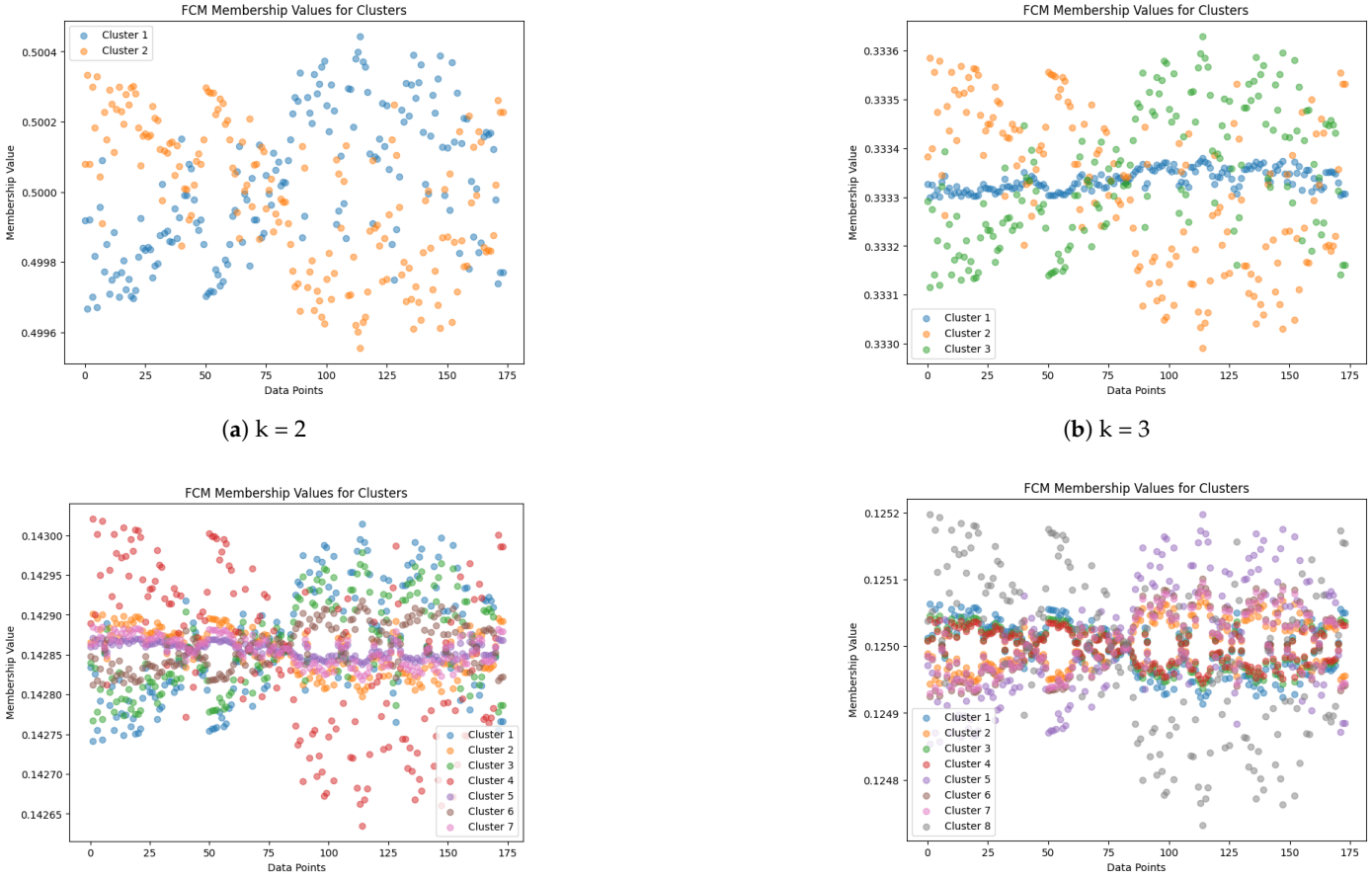
During the initial stage of our experimentation with the DARWIN dataset, we embarked on a meticulous process. Beginning by parsing the dataset, we segregated the class labels from the feature matrix component. With this separation in place, we commenced the application of the Fuzzy C Means algorithm on the feature matrix for our clustering objectives. This algorithm efficiently processed the handwriting samples encapsulated within the feature matrix, leveraging its flexibility to accommodate a user-defined number of cluster centers, denoted as k. As a result of this process, we obtained a crucial output—the fuzzy membership matrix U_k, intricately linked to the designated cluster center values k.

To proceed with informed decision-making in the subsequent stage, we undertook a comprehensive assessment of the clustering outcomes. Initially, we embarked on a systematic exploration, computing Cluster Validity Index (CVI) scores across a spectrum of k values, meticulously spanning a range of consecutive integers denoted by Ω. This endeavor provided us with a nuanced understanding of the clustering performance across varying degrees of granularity in cluster representation. Figure 3 visually encapsulates the diversity of clustering outcomes, portraying the fuzzy membership matrices corresponding to distinct k values, facilitating visual comprehension of the clustering dynamics.

Determining the optimal configuration of cluster centers constituted a pivotal aspect of our analysis. Through an iterative process guided by the CVI function, we aimed to pinpoint the value of k that maximized the clustering efficacy across the entire spectrum of Ω. This task necessitated the integration of multiple perspectives, requiring a judicious amalgamation of insights derived from a diverse set of metrics. Employing an average scoring mechanism, we distilled the complexities of the clustering performance into a succinct evaluation criterion, aggregating normalized CVI scores across the four key metrics

mentioned previously. This synthesized perspective served as a compass in our quest for the optimal cluster configuration.

Figure 4 serves as a testament to our analytical journey, encapsulating the culmination of our deliberations. At a glance, it reveals a pivotal revelation—the optimal cluster center configuration, delineated by $k = 2$. This pivotal insight not only underscores the efficacy of our methodology but also resonates with the ground truth embedded within the dataset. Remarkably, the optimal number of cluster centers aligned seamlessly with the inherent structure of the dataset, mirroring the underlying classes provided as a reference point. Thus, our journey through the intricacies of clustering analysis not only yielded tangible insights but also reaffirmed the congruence between analytical prowess and real-world phenomena.



(**a**) k = 2 (**b**) k = 3

(**c**) k = 7 (**d**) k = 8

Figure 3. Fuzzy membership matrices corresponding to cluster centers for different values of k: (**a**) $k = 2$, (**b**) $k = 3$, (**c**) $k = 7$, and (**d**) $k = 8$. The y-axis represents each of the handwriting samples, while the x-axis denotes their fuzzy membership values corresponding to each cluster center. Figures should be placed in the main text near the first time they are cited.

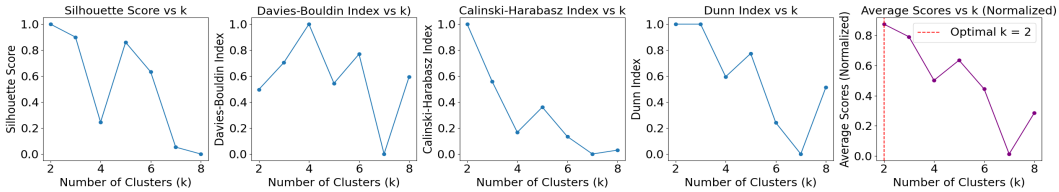


Figure 4. CVI scores versus the number of clusters. The optimal number of clusters was computed by taking the average of multiple CVI scores after normalization. Figures should be placed in the main text near the first time they are cited.

3.2. Results from Supervised Machine Learning

The fuzzy membership matrix U corresponding to the optimal cluster center $k = 2$ was combined with class label values from the DARWIN dataset to train supervised learning algorithms. We conducted two sets of experiments: in the first set, we directly applied all ten state-of-the-art supervised learning algorithms to the DARWIN dataset, without utilizing FCM as Stage 1. In the second experiment set, we applied the same algorithms after performing Stage 1 with FCM.

During the first experiment set, we investigated the performance of the supervised learning algorithms in isolation, without leveraging the preliminary clustering insights provided by FCM. This approach allowed us to discern the inherent capabilities and limitations of each algorithm in its raw form, providing valuable insights into their individual efficacy in handling the handwriting data.

In contrast, the second experiment set aimed to capitalize on the initial clustering performed by FCM in Stage 1. By incorporating the fuzzy membership matrix U as features alongside the class labels, we provided the supervised learning algorithms with a richer and more nuanced input space. This augmentation facilitated the algorithms in capturing subtle patterns and relationships within the data, potentially enhancing their predictive performance.

A comparative analysis between the two experiment sets enabled us to gain deeper insights into the interplay between clustering and supervised learning in the context of Alzheimer's disease diagnosis through handwriting analysis. By systematically evaluating the performance of the algorithms under different experimental conditions, we aimed to uncover optimal synergies between unsupervised clustering and subsequent supervised learning phases, with the ultimate goal of improving diagnostic accuracy and effectiveness.

3.2.1. Applying Classifiers Without Stage 1 (FCM)

In this series of experiments, we initiated our exploration with a grid-search cross-validation method, prioritizing accuracy as the primary performance criterion, to meticulously identify the optimal hyperparameter values for each of the supervised learning algorithms. Utilizing the DARWIN dataset enriched with clinical features, we embarked on an exhaustive exploration aimed at refining the algorithms' configurations. The detailed results of these experiments are meticulously documented in Table 6, encapsulating a total of 12,310 trials, executed with precision to pinpoint the optimal settings for the ten supervised learning algorithms under scrutiny.

Table 6. Structure and components of a typical confusion matrix.

Actual	**Predicted**	
	Positive	**Negative**
Positive	TP	FN
Negative	FP	TN

Of particular interest, it was revealed that the Support Vector Machine (SVM) algorithm attained the highest accuracy score, reaching an impressive 91%. This notable achievement was realized with a specific hyper-parameter configuration, with 'C' set to 1, 'gamma' set to 'scale', and 'kernel' set to 'rbf'. Following this discovery, we proceeded to conduct a rigorous comparison of all classifiers, employing seven performance evaluation metrics to comprehensively assess their efficacy.

To further validate the robustness of our findings and ensure the generalizability of the observed trends, we undertook a 10-fold cross-validation procedure. This meticulous validation process involved utilizing the specific optimal hyperparameter values identified

through the grid-search method. The outcomes of this 10-fold cross-validation experiment are meticulously outlined in Table 7, providing deeper insights into the classifiers' performance across multiple evaluation metrics.

Noteworthy among all classifiers considered were Random Forest (RF), Extra Trees (ET), Gaussian Naive Bayes (GNB), and XGBoost (XGB), which consistently exhibited superior performance across the evaluation metrics. These algorithms demonstrated remarkable robustness and efficacy in capturing the intricate patterns inherent in the DARWIN dataset. It is important to highlight that the XGBoost (XGB) classifier achieved the highest mean accuracy of 86%, even without employing Stage 1, the preliminary clustering with Fuzzy C Means.

This extensive experimentation and validation process not only contributed to refining the optimal hyperparameter configurations for each algorithm but also offered valuable insights into their comparative performance under stringent evaluation conditions. By meticulously assessing their performance using both grid-search cross-validation and 10-fold cross-validation, we ensured a comprehensive evaluation of their generalization capabilities and effectiveness in the complex task of handwriting-based Alzheimer's disease diagnosis.

Table 7. Determining optimal hyperparameter for different algorithms using Grid Search CV, without stage 1.

Algorithm	Param_Grid	Best_Param	Accuracy	No of Trials
DT	'max_depth': [None, 10, 20], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]	{ 'max_depth': 20, 'min_samples_leaf': 4, 'min_samples_split': 2}	85.71%	135
RF	'n_estimators': [50, 100, 150], 'max_depth': [None, 10, 20], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]	{ 'max_depth': 20, 'min_samples_leaf': 4, 'min_samples_split': 5, 'n_estimators': 150}	85.71%	405
Extra Tree	'n_estimators': [50, 100, 150], 'max_depth': [None, 10, 20], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]	{ 'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 150}	88.57%	405
LR	'penalty': ['l1', 'l2'], 'C': [0.001, 0.01, 0.1, 1, 10, 100]	{ 'C': 10, 'penalty': 'l2' }	88.57%	60
LDA	'solver': ['svd', 'lsqr', 'eigen'], 'shrinkage': [None, 'auto'], 'n_components': [None, 1, 2]	{ 'n_components': None, 'shrinkage': 'auto', 'solver': 'lsqr'}	88.57%	90
GNB	No tunable parameters	Not applicable	85.71%	Not applicable
XGB	'learning_rate': [0.05, 0.1, 0.3], 'max_depth': [3, 6, 9], 'min_child_weight': [1, 3, 5], 'gamma': [0, 0.1, 0.2], 'subsample': [0.6, 0.8, 1.0], 'colsample_bytree': [0.6, 0.8, 1.0], 'n_estimators': [50, 100, 150]	{ 'colsample_bytree': 0.8, 'gamma': 0, 'learning_rate': 0.1, 'max_depth': 6, 'min_child_weight': 1, 'n_estimators': 100, 'subsample': 1.0}	88.57%	10935
KNN	'n_neighbors': [3, 5, 7], 'weights': ['uniform', 'distance'], 'metric': ['euclidean', 'manhattan']	{ 'metric': 'manhattan', 'n_neighbors': 3, 'weights': 'uniform'}	48.57%	60
SVM	'C': [0.1, 1, 10], 'kernel': ['linear', 'rbf'], 'gamma': ['scale', 'auto']	{ 'C': 1, 'gamma': 'scale', 'kernel': 'rbf' }	91.43%	60
MLP	'hidden_layer_sizes': [(100,), (50, 100, 50)], 'activation': ['relu', 'tanh'], 'solver': ['adam', 'sgd'], 'alpha': [0.0001, 0.001], 'learning_rate': ['constant', 'adaptive']	{ 'activation': 'relu', 'alpha': 0.001, 'hidden_layer_sizes': (100,), 'learning_rate': 'constant', 'solver': 'sgd'}	85.71%	160

3.2.2. Applying Classifiers with Stage 1 (FCM)

In this comprehensive series of experiments, we meticulously employed all ten state-of-the-art supervised learning algorithms previously mentioned, leveraging the cognitive dataset generated in Stage 1 as the foundation for our analysis. To discern the optimal configuration of the algorithms' hyperparameters, we once again turned to the Grid-Search-CV method, meticulously exploring a myriad of parameter combinations to achieve optimal performance. The detailed results of this hyperparameter tuning process are thoughtfully presented in Table 8, offering a comprehensive overview of the optimized settings for each algorithm.

Building upon the insights gleaned from the hyperparameter tuning phase, we proceeded to conduct a rigorous evaluation of each algorithm's performance using 10-fold cross-validation. This meticulous validation approach allowed us to robustly assess the algorithms' predictive capabilities and generalization performance across different subsets of the dataset. The outcomes of these validation experiments are meticulously documented in Table 9, offering a detailed glimpse into each algorithm's performance metrics and highlighting their respective strengths and weaknesses. Figure 5 displays the AUC-ROC for each fold, as well as the micro-average and macro-average ROC curves for the Decision Tree (DT), Random Forest (RF), Extra Tree (ET), and K-nearest-neighbor (KNN) classifiers. AUC-ROC curves for all other classifiers are presented in the Supplementary Materials.

Of particular note, the tree-based classifiers, including Decision Trees (DT), Random Forests (RF), and Extra Trees (ET), emerged as the standout performers in terms of predictive accuracy and robustness. These algorithms showcased exceptional performance across the various evaluation metrics, underscoring their efficacy in handling the complexities inherent in the AD diagnosis task.

Interestingly, a notable observation emerged from our analysis: all classifiers, with the exception of Logistic Regression (LR) and Multi-layer Perceptron (MLP), demonstrated a significant enhancement in their predictive capabilities when leveraging the insights gained from Fuzzy C Means (FCM) clustering at Stage 1 (Figure 6). This finding highlights the invaluable role of unsupervised clustering in augmenting the predictive power of supervised learning algorithms, particularly in the context of AD diagnosis through handwriting analysis.

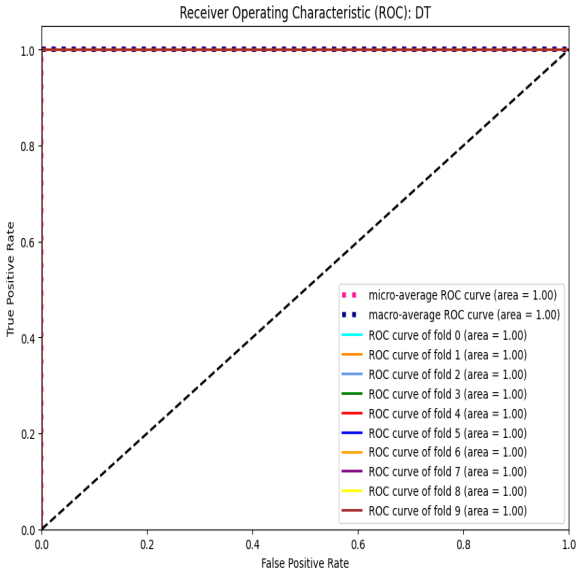
In summary, our meticulous experimentation and rigorous validation procedures provided valuable insights into the optimal configuration and performance of supervised learning algorithms in the domain of AD diagnosis. By systematically exploring different algorithmic configurations and leveraging the synergies between unsupervised clustering and supervised learning, we have laid a solid foundation for future advancements in this critical area of research.

Table 8. Performance metrics for all the classifiers using 10-fold cross validation, without stage 1.

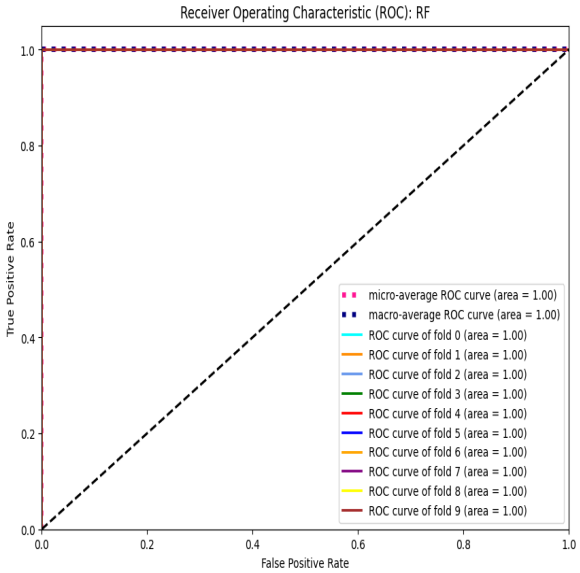
Model	Accuracy	Sensitivity	Specificity	Precision	MCC	Cohen's Kappa	AUC ROC
DT	0.7327 ± 0.1741	0.725 ± 0.2683	0.7306 ± 0.2041	0.7555 ± 0.1992	0.4748 ± 0.3496	0.4539 ± 0.3543	0.7278 ± 0.1768
RF	0.8376 ± 0.1255	0.900 ± 0.1606	0.7986 ± 0.1351	0.8319 ± 0.0902	0.7157 ± 0.1821	0.6981 ± 0.1921	0.8493 ± 0.0959
ET	0.8255 ± 0.1289	0.8875 ± 0.1649	0.8111 ± 0.1608	0.8398 ± 0.1209	0.7119 ± 0.2394	0.6983 ± 0.2473	0.8493 ± 0.1227
LR	0.7908 ± 0.1450	0.8194 ± 0.1745	0.7667 ± 0.2073	0.8071 ± 0.1603	0.6024 ± 0.2854	0.5837 ± 0.2889	0.7931 ± 0.1440
LDA	0.8144 ± 0.1232	0.8194 ± 0.1597	0.8097 ± 0.1924	0.8453 ± 0.1394	0.6449 ± 0.2441	0.6278 ± 0.2491	0.8146 ± 0.1245
GNB	0.8376 ± 0.1118	0.8653 ± 0.1632	0.8125 ± 0.1341	0.8330 ± 0.1009	0.6876 ± 0.2170	0.6761 ± 0.2226	0.8389 ± 0.1107
XGB	0.8611 ± 0.0885	0.9000 ± 0.1160	0.8208 ± 0.1713	0.8615 ± 0.1102	0.7371 ± 0.1735	0.7201 ± 0.1819	0.8604 ± 0.0924
KNN	0.6526 ± 0.1351	0.3444 ± 0.2406	0.9764 ± 0.0473	0.8667 ± 0.3055	0.3800 ± 0.2878	0.3163 ± 0.2670	0.6604 ± 0.1331
SVM	0.8141 ± 0.1462	0.8403 ± 0.1638	0.7847 ± 0.2345	0.8305 ± 0.1585	0.6397 ± 0.2974	0.6244 ± 0.2994	0.8125 ± 0.1499
MLP	0.4824 ± 0.0844	0.8000 ± 0.4000	0.2000 ± 0.4000	NaN	0.0000 ± 0.0000	0.0000 ± 0.0000	0.5000 ± 0.0000

Table 9. Determining optimal hyperparameters for the different algorithms using Grid-Search CV, with stage 1.

Algorithm	Param_Grid	Best_Param	Acc	No of Trials
FCM	Cluster center $k = 2$ to 8	$k = 2$	Best CVI score	7
DT	'max_depth': [None, 10, 20], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]	'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2	97.14%	135
RF	'n_estimators': [50, 100, 150], 'max_depth': [None, 10, 20], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]	'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50	100%	405
Extra Tree	'n_estimators': [50, 100, 150], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]	'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50	97.14%	135
LR	'C': [0.001, 0.01, 0.1, 1, 10, 100], 'penalty': ['l1', 'l2'], 'solver': ['liblinear']	'C': 0.001, 'penalty': 'l2', 'solver': 'liblinear'	54.29%	60
LDA	'solver': ['svd', 'lsqr', 'eigen']	'solver': 'svd'	91.43%	15
GNB	No tunable parameter	Not applicable	91.43%	NA
XGB	'learning_rate': [0.05, 0.1, 0.3], 'max_depth': [3, 6, 9], 'min_child_weight': [1, 3, 5], 'gamma': [0, 0.1, 0.2], 'subsample': [0.6, 0.8, 1.0], 'colsample_bytree': [0.6, 0.8, 1.0], 'n_estimators': [50, 100, 150]	'colsample_bytree': 0.6, 'gamma': 0, 'learning_rate': 0.05, 'max_depth': 3, 'min_child_weight': 1, 'n_estimators': 50, 'subsample': 0.8	94.29%	10935
KNN	'n_neighbors': [3, 5, 7], 'weights': ['uniform', 'distance'], 'p': [1, 2]	'n_neighbors': 3, 'p': 1, 'weights': 'uniform'	97.14%	60
SVM	'C': [0.1, 1, 10], 'gamma': ['scale', 'auto'], 'kernel': ['linear', 'rbf']	'C': 0.1, 'gamma': 'scale', 'kernel': 'rbf'	91.43%	60
MLP	'hidden_layer_sizes': [(100,), (50, 50), (50, 50, 50)], 'activation': ['relu', 'tanh'], 'solver': ['adam'], 'alpha': [0.0001, 0.001], 'learning_rate': ['constant', 'adaptive']	'activation': 'relu', 'alpha': 0.0001, 'hidden_layer_sizes': (100,), 'learning_rate': 'adaptive', 'solver': 'adam'	45.71%	120

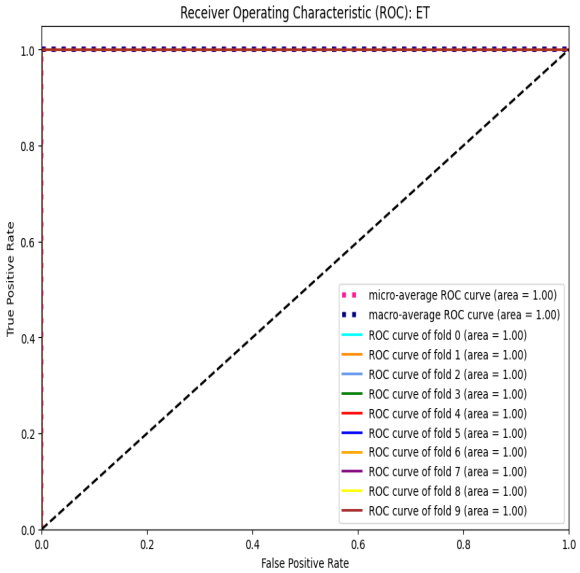


(**a**) DT

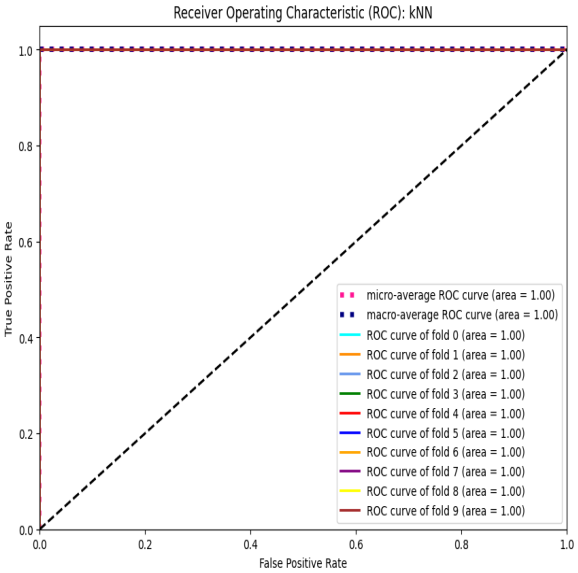


(**b**) RF

Figure 5. *Cont.*



(**c**) ET



(**d**) KNN

Figure 5. AUC-ROC curves for each fold, along with micro-average and macro-average ROC curves for (**a**) Decision Tree (DT), (**b**) Random Forest (RF), (**c**) Extra Trees (ET), and (**d**) k-Nearest Neighbors (KNNs).

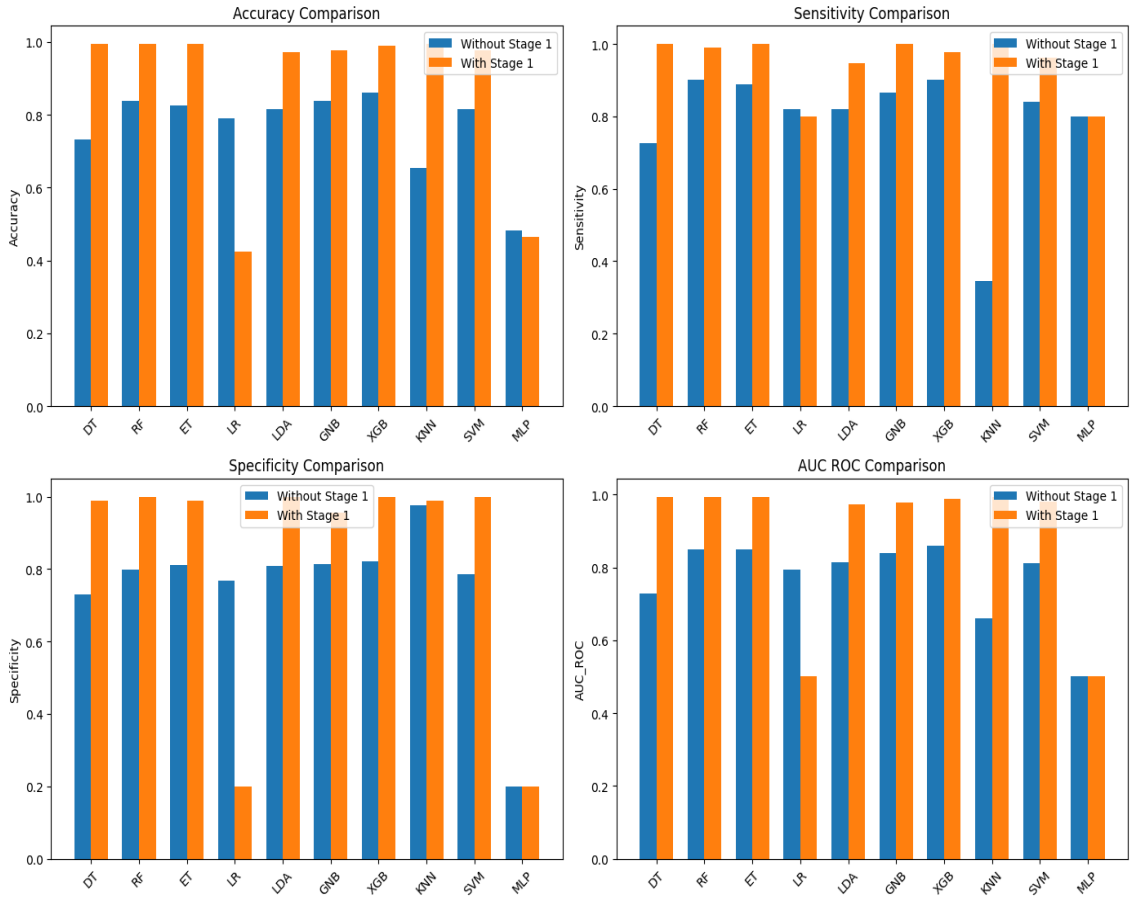


Figure 6. Comparative performance of the base line classifiers, with and without stage 1.

3.3. Comparative Performance

Furthermore, we conducted an extensive comparative analysis with previously reported approaches focused on predicting Alzheimer's disease utilizing the DARWIN dataset. The outcomes from these studies are meticulously detailed in Table 10, providing a comprehensive side-by-side comparison with the results obtained from our proposed models (Table 11). Notably, the integration of the unsupervised algorithm Fuzzy C Means (FCM) with the supervised Random Forest (RF) algorithm led to a significant enhancement in overall predictive performance. This noteworthy improvement underscored the effectiveness of synergizing the strengths of a hybrid model, harnessing both unsupervised and supervised learning paradigms to achieve superior predictive accuracy.

Table 10. Performance metrics for all the classifiers using 10-fold cross-validation, with stage 1.

Model	Accuracy	Sensitivity	Specificity	Precision	MCC	Cohen's Kappa	AUC ROC
DT	0.9944 ± 0.0167	1.0000 ± 0.0000	0.9889 ± 0.0333	0.9900 ± 0.0300	0.9894 ± 0.0317	0.9889 ± 0.0333	0.9944 ± 0.0167
RF	0.9944 ± 0.0167	0.9889 ± 0.0333	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9894 ± 0.0317	0.9889 ± 0.0333	0.9944 ± 0.0167
ET	0.9944 ± 0.0167	1.0000 ± 0.0000	0.9889 ± 0.0333	0.9900 ± 0.0300	0.9894 ± 0.0317	0.9889 ± 0.0333	0.9944 ± 0.0167
LR	0.4232 ± 0.0973	0.8000 ± 0.4000	0.2000 ± 0.4000	NaN	0.0000 ± 0.0000	0.0000 ± 0.0000	0.5000 ± 0.0000
LDA	0.9716 ± 0.0460	0.9471 ± 0.0837	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9473 ± 0.0847	0.9424 ± 0.0930	0.9735 ± 0.0419
GNB	0.9775 ± 0.0372	1.0000 ± 0.0000	0.9562 ± 0.0735	0.9584 ± 0.0667	0.9572 ± 0.0693	0.9540 ± 0.0752	0.9781 ± 0.0368
XGB	0.9889 ± 0.0333	0.9778 ± 0.0667	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9798 ± 0.0607	0.9778 ± 0.0667	0.9889 ± 0.0333
KNN	0.9944 ± 0.0167	1.0000 ± 0.0000	0.9889 ± 0.0333	0.9900 ± 0.0300	0.9894 ± 0.0317	0.9889 ± 0.0333	0.9944 ± 0.0167
SVM	0.9771 ± 0.0375	0.9609 ± 0.0656	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9570 ± 0.0692	0.9538 ± 0.0751	0.9805 ± 0.0328
MLP	0.4644 ± 0.1187	0.8000 ± 0.4000	0.2000 ± 0.4000	NaN	0.0000 ± 0.0000	0.0000 ± 0.0000	0.5000 ± 0.0000

The success observed in the final results reaffirmed the validity of our comprehensive methodology and underscored the potential for advanced techniques to enhance the predictive capabilities of machine learning models in medical research and diagnostic applications. By leveraging the complementary strengths of unsupervised clustering and supervised learning, our hybrid approach demonstrated potential for advancing the state-of-the-art in Alzheimer's disease prediction. This innovative methodology opens new avenues for research and underscores the importance of interdisciplinary collaboration in tackling complex medical challenges.

Table 11. Comparative performance with the proposed model.

Model	Acc (%)	Sn (%)	Sp (%)	Precision (%)	AUC-ROC (%)
De Gregorio et al. [40]	91	83	100	-	-
Cilia et al. [41]	94.28	88.24	100	-	-
Parziale et al. [42]	97.12	94.23	100	-	-
Subha et al. [43]	90	92	-	88	90
Gattulli et al. [44]	88	90	86	-	-
Onder et al. [45]	85	-	-	-	-
Hakan et al. [46]	97.14	-	90	95	-
Erdogmus et al. [48]	90.4	-	-	-	-
Mitra and Rehman [47]	97.14	95	100	100	94.21
Proposed Model	99.44	100	100	100	99.44

4. Conclusions, Limitations, and Future Work

This study successfully developed a hybrid predictor model for Alzheimer's disease (AD) using handwriting analysis and machine learning techniques, ensuring both robustness and accuracy. The model operates in two stages: initially employing unsupervised clustering, specifically the Fuzzy C-Mean algorithm, to cluster handwriting samples. Fuzzy logic, chosen for its robust ability to handle data uncertainties, is particularly well-suited for medical diagnosis, where precision, adaptability, and the ability to manage ambiguous or overlapping symptoms are crucial for accurate decision-making. The study employed

four widely used Cluster Validity Indices to qualify the clustering from FCM and determine optimal cluster centers. The fuzzy membership matrix, treating each cluster as a feature with membership degrees as values, transitioned to the supervised stage for predictor model development. This two-stage model significantly enhanced the performance in detecting AD patients, achieving a mean accuracy of 99%, mean sensitivity of 98%, mean specificity of 100%, mean precision of 100%, and mean MCC and Cohen's Kappa of 98%, along with a mean AUC-ROC score of 99%. Notably, integrating unsupervised cognition with supervised learning using FCM in the first stage improved the classifier performance by 12% to 26%, highlighting the importance of this integration.

Despite these promising results, this study has certain limitations. First, handwriting variability due to external factors such as age, education level, or underlying health conditions was not thoroughly analyzed. These factors may influence the model's effectiveness and should be explored in future research. Additionally, the model was tested on a relatively small group of 174 participants, raising concerns about its generalizability to a larger population. A more extensive and diverse dataset is required to validate the model's performance across different demographic groups.

To address these limitations, future work will focus on expanding the dataset by implementing the protocol proposed by Cilia et al. [7], collecting additional handwriting samples from both healthy individuals and AD patients. This will improve the model's robustness and enhance its applicability in real-world clinical settings. Furthermore, the model could be deployed as software for mobile devices and embedded systems, facilitating easy access for practitioners and supporting early diagnosis in non-clinical environments.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/info16030249/s1, Table S1: List of tasks performed. Task categories are memory and dictation (M), Graphic (G), and Copy (C). Table S2: List of features associated with each tasks.

Author Contributions: Conceptualization, S.U.R. and U.M.; methodology, S.U.R. and U.M.; software, S.U.R. and U.M.; validation, S.U.R. and U.M.; formal analysis, S.U.R. and U.M.; investigation, S.U.R. and U.M.; resources, S.U.R. and U.M.; data curation, U.M.; writing—original draft preparation, S.U.R. and U.M.; writing—review and editing, S.U.R. and U.M.; visualization, S.U.R. and U.M.; supervision, S.U.R.; project administration, S.U.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: This work was supported in part by Kingdom University, Bahrain, under Grant KU–SRU–2024–13.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Available online: https://www.who.int/news-room/fact-sheets/detail/dementia (accessed on 13 November 2023).
2. Singhal, A.K.; Naithani, V.; Bangar, O.P. Medicinal plants with a potential to treat Alzheimer and associated symptoms. *Int. J. Nutr. Pharmacol. Neurol. Dis.* **2012**, *2*, 84–91. [CrossRef]
3. Armstrong, M.J.; Litvan, I.; Lang, A.E.; Bak, T.H.; Bhatia, K.P.; Borroni, B.; Boxer, A.L.; Dickson, D.W.; Grossman, M.; Hallett, M.; et al. Criteria for the diagnosis of corticobasal degeneration. *Neurology* **2013**, *80*, 496–503. [CrossRef] [PubMed]
4. Duranti, E.; Villa, C. From Brain to Muscle: The Role of Muscle Tissue in Neurodegenerative Disorders. *Biology* **2024**, *13*, 719. [CrossRef]

5. Saxton, J.; Lopez, O.L.; Ratcliff, G.; Dulberg, C.; Fried, L.P.; Carlson, M.C.; Newman, A.B.; Kuller, L. Preclinical Alzheimer disease: Neuropsychological test performance 1.5 to 8 years prior to onset. *Neurology* **2004**, *63*, 2341–2347. [CrossRef]
6. Rasmussen, J.; Langerman, H. Alzheimer's disease–why we need early diagnosis. *Degener. Neurol. Neuromuscul. Dis.* **2019**, *9*, 123–130. [CrossRef] [PubMed]
7. Cilia, N.D.; De Stefano, C.; Fontanella, F.; Di Freca, A.S. An experimental protocol to support cognitive impairment diagnosis by using handwriting analysis. *Procedia Comput. Sci.* **2018**, *141*, 466–471. [CrossRef]
8. Nichols, E.; Steinmetz, J.D.; Vollset, S.E.; Fukutaki, K.; Chalek, J.; Abd-Allah, F.; Abdoli, A.; Abualhasan, A.; Abu-Gharbieh, E.; Akram, T.T.; et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **2022**, *7*, e105–e125. [CrossRef]
9. Available online: https://pubmed.ncbi.nlm.nih.gov/38689398/ (accessed on 15 March 2025).
10. Available online: https://pubmed.ncbi.nlm.nih.gov/33667416/ (accessed on 15 March 2025).
11. Reisberg, B.; Ferris, S.H.; de Leon, M.J.; Crook, T. The Global Deterioration Scale for assessment of primary degenerative dementia. *Am. J. Psychiatry* **1982**, *139*, 1136–1139.
12. Precup, R.E.; Teban, T.A.; Albu, A.; Borlea, A.B.; Zamfirache, I.A.; Petriu, E.M. Evolving fuzzy models for prosthetic hand myoelectric-based control. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 4625–4636. [CrossRef]
13. Myszczynska, M.A.; Ojamies, P.N.; Lacoste, A.M.; Neil, D.; Saffari, A.; Mead, R.; Hautbergue, G.M.; Holbrook, J.D.; Ferraiuolo, L. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat. Rev. Neurol.* **2020**, *16*, 440–456. [CrossRef]
14. Belić, M.; Bobić, V.; Badža, M.; Šolaja, N.; Đurić-Jovičić, M.; Kostić, V.S. Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease—A review. *Clin. Neurol. Neurosurg.* **2019**, *184*, 105442. [CrossRef] [PubMed]
15. Scott, I.; Carter, S.; Coiera, E. Clinician checklist for assessing suitability of machine learning applications in healthcare. *BMJ Health Care Inform.* **2021**, *28*, e100251. [CrossRef] [PubMed]
16. Ranjan, N.; Kumar, D.U.; Dongare, V.; Chavan, K.; Kuwar, Y. Diagnosis of Parkinson disease using handwriting analysis. *Int. J. Comput. Appl.* **2022**, *184*, 13–16.
17. Pozna, C.; Precup, R.E. Applications of signatures to expert systems modelling. *Acta Polytech. Hung.* **2014**, *11*, 21–39.
18. Pirlo, G.; Diaz, M.; Ferrer, M.A.; Impedovo, D.; Occhionero, F.; Zurlo, U. Early diagnosis of neurodegenerative diseases by handwritten signature analysis. In *New Trends in Image Analysis and Processing—Proceedings of the ICIAP 2015 Workshops: ICIAP 2015 International Workshops, BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, 7–8 September 2015*; Proceedings 18; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 290–297.
19. Senatore, R.; Marcelli, A. A paradigm for emulating the early learning stage of handwriting: Performance comparison between healthy controls and Parkinson's disease patients in drawing loop shapes. *Hum. Mov. Sci.* **2019**, *65*, 89–101. [CrossRef]
20. Xie, L.; Das, S.R.; Wisse, L.E.; Ittyerah, R.; de Flores, R.; Shaw, L.M.; Yushkevich, P.A.; Wolk, D.A.; Alzheimer's Disease Neuroimaging Initiative. Baseline structural MRI and plasma biomarkers predict longitudinal structural atrophy and cognitive decline in early Alzheimer's disease. *Alzheimer's Res. Ther.* **2023**, *15*, 79. [CrossRef]
21. Abbas, S.Q.; Chi, L.; Chen, Y.P.P. Transformed domain convolutional neural network for alzheimer's disease diagnosis using structural MRI. *Pattern Recognit.* **2023**, *133*, 109031. [CrossRef]
22. Wang, J.; Jin, C.; Zhou, J.; Zhou, R.; Tian, M.; Lee, H.J.; Zhang, H. PET molecular imaging for pathophysiological visualization in Alzheimer's disease. *Eur. J. Nucl. Med. Mol. Imaging* **2023**, *50*, 765–783. [CrossRef]
23. Rallabandi, V.S.; Seetharaman, K. Deep learning-based classification of healthy aging controls, mild cognitive impairment and Alzheimer's disease using fusion of MRI-PET imaging. *Biomed. Signal Process. Control* **2023**, *80*, 104312.
24. Kaya, M.; Çetin-Kaya, Y. A Novel Deep Learning Architecture Optimization for Multiclass Classification of Alzheimer's Disease Level. *IEEE Access* **2024**, *12*, 46562–46581. [CrossRef]
25. Rehman, S.U.; Tarek, N.; Magdy, C.; Kamel, M.; Abdelhalim, M.; Melek, A.; Mahmoud, L.N.; Sadek, I. AI-based tool for early detection of Alzheimer's disease. *Heliyon* **2024**, *10*, e29375. [CrossRef]
26. Ahmed, H.; Soliman, H.; Elmogy, M. Early detection of Alzheimer's disease using single nucleotide polymorphisms analysis based on gradient boosting tree. *Comput. Biol. Med.* **2022**, *146*, 105622. [CrossRef] [PubMed]
27. Yin, W.; Yang, T.; Wan, G.; Zhou, X. Identification of image genetic biomarkers of Alzheimer's disease by orthogonal structured sparse canonical correlation analysis based on a diagnostic information fusion. *Math. Biosci. Eng.* **2023**, *20*, 16648–16662. [CrossRef] [PubMed]
28. Chen, Y.; Sun, Y.; Luo, Z.; Chen, X.; Wang, Y.; Qi, B.; Lin, J.; Lin, W.W.; Sun, C.; Zhou, Y.; et al. Exercise modifies the transcriptional regulatory features of monocytes in Alzheimer's patients: A multi-omics integration analysis based on single cell technology. *Front. Aging Neurosci.* **2022**, *14*, 881488. [CrossRef]
29. Jiao, B.; Li, R.; Zhou, H.; Qing, K.; Liu, H.; Pan, H.; Lei, Y.; Fu, W.; Wang, X.; Xiao, X.; et al. Neural biomarker diagnosis and prediction to mild cognitive impairment and Alzheimer's disease using EEG technology. *Alzheimer's Res. Ther.* **2023**, *15*, 32. [CrossRef] [PubMed]

30. Hou, Q.; Guan, Y.; Liu, X.; Xiao, M.; Lü, Y. Development and validation of a risk model for cognitive impairment in the older Chinese inpatients: An analysis based on a 5-year database. *J. Clin. Neurosci.* **2022**, *104*, 29–33. [CrossRef]
31. Odusami, M.; Maskeliūnas, R.; Damaševičius, R.; Misra, S. Machine learning with multimodal neuroimaging data to classify stages of Alzheimer's disease: A systematic review and meta-analysis. *Cogn. Neurodyn.* **2023**, *18*, 775–794. [CrossRef]
32. Qiu, S.; Miller, M.I.; Joshi, P.S.; Lee, J.C.; Xue, C.; Ni, Y.; Wang, Y.; De Anda-Duran, I.; Hwang, P.H.; Cramer, J.A.; et al. Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat. Commun.* **2022**, *13*, 3404. [CrossRef]
33. Accardo, A.; Chiap, A.; Borean, M.; Bravar, L.; Zoia, S.; Carrozzi, M.; Scabar, A. A device for quantitative kinematic analysis of children's handwriting movements. In Proceedings of the 11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007: MEDICON 2007, Ljubljana, Slovenia, 26–30 June 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 445–448.
34. De Stefano, C.; Fontanella, F.; Impedovo, D.; Pirlo, G.; di Freca, A.S. Handwriting analysis to support neurodegenerative diseases diagnosis: A review. *Pattern Recognit. Lett.* **2019**, *121*, 37–45. [CrossRef]
35. Vessio, G. Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review. *Appl. Sci.* **2019**, *9*, 4666. [CrossRef]
36. Singh, P.; Yadav, H. Influence of neurodegenerative diseases on handwriting. *Forensic Res. Criminol. Int. J.* **2021**, *9*, 110–114. [CrossRef]
37. Impedovo, D.; Pirlo, G.; Vessio, G.; Angelillo, M.T. A handwriting-based protocol for assessing neurodegenerative dementia. *Cogn. Comput.* **2019**, *11*, 576–586. [CrossRef]
38. Impedovo, D.; Pirlo, G.; Vessio, G. Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis. *Information* **2018**, *9*, 247. [CrossRef]
39. Cavallo, F.; Moschetti, A.; Esposito, D.; Maremmani, C.; Rovini, E. Upper limb motor pre-clinical assessment in Parkinson's disease using machine learning. *Park. Relat. Disord.* **2019**, *63*, 111–116. [CrossRef]
40. De Gregorio, G.; Desiato, D.; Marcelli, A.; Polese, G. A multi classifier approach for supporting Alzheimer's diagnosis based on handwriting analysis. In Proceedings of the Pattern Recognition. ICPR International Workshops and Challenges, Virtual Event, 10–15 January 2021; Proceedings, Part I; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 559–574.
41. Cilia, N.D.; De Gregorio, G.; De Stefano, C.; Fontanella, F.; Marcelli, A.; Parziale, A. Diagnosing Alzheimer's disease from on-line handwriting: A novel dataset and performance benchmarking. *Eng. Appl. Artif. Intell.* **2022**, *111*, 104822. [CrossRef]
42. Parziale, A.; Della Cioppa, A.; Marcelli, A. Investigating one-class classifiers to diagnose Alzheimer's disease from handwriting. In Proceedings of the International Conference on Image Analysis and Processing, Lecce, Italy, 23–27 May 2022; Springer International Publishing: Cham, Switzerland, 2022; pp. 111–123.
43. Subha, R.; Nayana, B.R.; Selvadass, M. Hybrid Machine Learning Model Using Particle Swarm Optimization for Effectual Diagnosis of Alzheimer's Disease from Handwriting. In Proceedings of the 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 21–23 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 491–495.
44. Gattulli, V.; Impedovo, D.; Semeraro, G.P.G. Handwriting Task-Selection based on the Analysis of Patterns in Classification Results on Alzheimer Dataset. *DSTNDS* **2023**, 18–29.
45. Önder, M.; Şentürk, Ü.; Polat, K.; Paulraj, D. Diagnosis of Alzheimer's Disease Using Boosting Classification Algorithms. In Proceedings of the 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 1–2 November 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
46. Hakan, Ö.C.A.L. A Novel Approach to Detection of Alzheimer's Disease from Handwriting: Triple Ensemble Learning Model. *Gazi Univ. J. Sci. Part C Des. Technol.* **2024**, *12*, 214–223.
47. Mitra, U.; Rehman, S.U. ML-Powered Handwriting Analysis for Early Detection of Alzheimer's Disease. *IEEE Access* **2024**, *12*, 69031–69050. [CrossRef]
48. Erdogmus, P.; Kabakus, A.T. The promise of convolutional neural networks for the early diagnosis of the Alzheimer's disease. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106254. [CrossRef]
49. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
50. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, 224–227. [CrossRef]
51. Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun.-Stat.-Theory Methods* **1974**, *3*, 1–27. [CrossRef]
52. Bezdek, J.C.; Pal, N.R. Some new indexes of cluster validity. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **1998**, *28*, 301–315. [CrossRef] [PubMed]
53. Nasteski, V. An overview of the supervised machine learning methods. *Horizons b* **2017**, *4*, 56. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.