

Original Paper

# A Dynamic Adaptive Ensemble Learning Framework for Noninvasive Mild Cognitive Impairment Detection: Development and Validation Study

Aoyu Li<sup>1</sup>, PhD; Jingwen Li<sup>2</sup>, BD; Yishan Hu<sup>3</sup>, PhD; Yan Geng<sup>4</sup>, MD; Yan Qiang<sup>5</sup>, PhD; Juanjuan Zhao<sup>1,3</sup>, PhD

<sup>1</sup>School of Software, Taiyuan University of Technology, Jingzhong, China

<sup>2</sup>School of Computer Science, Xijing University, Xian, China

<sup>3</sup>College of Computer Science and Technology, Taiyuan University of Technology, Jinzhong, China

<sup>4</sup>Clinical Laboratory and Geriatric, Shanxi Provincial People's Hospital, Taiyuan, China

<sup>5</sup>School of Software, North University of China, Taiyuan, China

**Corresponding Author:**

Juanjuan Zhao, PhD

School of Software

Taiyuan University of Technology

No. 319, University Street, Yuji District

Jingzhong, 030600

China

Phone: 86 18636664123

Email: [zhaojuanjuan@tyut.edu.cn](mailto:zhaojuanjuan@tyut.edu.cn)

## Abstract

**Background:** The prompt and accurate identification of mild cognitive impairment (MCI) is crucial for preventing its progression into more severe neurodegenerative diseases. However, current diagnostic solutions, such as biomarkers and cognitive screening tests, prove costly, time-consuming, and invasive, hindering patient compliance and the accessibility of these tests. Therefore, exploring a more cost-effective, efficient, and noninvasive method to aid clinicians in detecting MCI is necessary.

**Objective:** This study aims to develop an ensemble learning framework that adaptively integrates multimodal physiological data collected from wearable wristbands and digital cognitive metrics recorded on tablets, thereby improving the accuracy and practicality of MCI detection.

**Methods:** We recruited 843 participants aged 60 years and older from the geriatrics and neurology departments of our collaborating hospitals, who were randomly divided into a development dataset (674/843 participants) and an internal test dataset (169/843 participants) at a 4:1 ratio. In addition, 226 older adults were recruited from 3 external centers to form an external test dataset. We measured their physiological signals (eg, electrodermal activity and photoplethysmography) and digital cognitive parameters (eg, reaction time and test scores) using the clinically certified Empatica 4 wristband and a tablet cognitive screening tool. The collected data underwent rigorous preprocessing, during which features in the time, frequency, and nonlinear domains were extracted from individual physiological signals. To address the challenges (eg, the curse of dimensionality and increased model complexity) posed by high-dimensional features, we developed a dynamic adaptive feature selection optimization algorithm to identify the most impactful subset of features for classification performance. Finally, the accuracy and efficiency of the classification model were improved by optimizing the combination of base learners.

**Results:** The experimental results indicate that the proposed MCI detection framework achieved classification accuracies of 88.4%, 85.5%, and 84.5% on the development, internal test, and external test datasets, respectively. The area under the curve for the binary classification task was 0.945 (95% CI 0.903-0.986), 0.912 (95% CI 0.859-0.965), and 0.904 (95% CI 0.846-0.962) on these datasets. Furthermore, a statistical analysis of feature subsets during the iterative modeling process revealed that the decay time of skin conductance response, the percentage of continuous normal-to-normal intervals exceeding 50 milliseconds, the ratio of low-frequency to high-frequency (LF/HF) components in heart rate variability, and cognitive time features emerged as the most prevalent and effective indicators. Specifically, compared with healthy individuals, patients with MCI exhibited a longer skin conductance response decay time during cognitive testing ( $P < .001$ ), a lower percentage of continuous normal-to-normal intervals exceeding 50 milliseconds ( $P < .001$ ), and higher LF/HF ( $P < .001$ ), accompanied by greater variability. Similarly, patients with MCI took longer to complete cognitive tests than healthy individuals ( $P < .001$ ).

**Conclusions:** The developed MCI detection framework has demonstrated exemplary performance and stability in large-scale validations. It establishes a new benchmark for noninvasive, effective early MCI detection that can be integrated into routine wearable and tablet-based assessments. Furthermore, the framework enables continuous and convenient self-screening within home or nonspecialized settings, effectively mitigating underresourced health care and geographic location constraints, making it an essential tool in the current fight against neurodegenerative diseases.

(*JMIR Med Inform 2025;13:e60250*) doi: [10.2196/60250](https://doi.org/10.2196/60250)

## KEYWORDS

mild cognitive impairment; ensemble learning; harmony search; combination optimization; digital cognitive assessment; physiological signal; cognitive impairment; detection; machine learning; cognitive metrics; photoplethysmography; neurodegenerative; Alzheimer; cognitive decline

## Introduction

### Background

Neurodegenerative conditions such as Alzheimer disease (AD) and related dementias precipitate accelerated cognitive deterioration, markedly impacting patients' daily lives and social engagement [1]. Current estimates suggest that approximately 50 million individuals worldwide suffer from dementia, with this number expected to soar to 152 million by 2050 [2]. Generally, patients diagnosed with mild cognitive impairment (MCI) are at a much higher risk of developing dementia [3]. MCI serves as an intermediate stage between normal cognitive aging and the severe pathological decline of dementia, influencing individuals' cognitive functions, social abilities, and mental health, and may lead to emotional disorders that disrupt daily life [4]. Epidemiological data reveal that the incidence of MCI is 6.7% among those aged 60-64 years, 8.4% for 65- to 69-year-olds, 10.1% for 70- to 74-year-olds, 14.8% for 75- to 79-year-olds, and 25.2% for 80- to 84-year-olds [5]. The annual transition rate from MCI to dementia or AD is about 10%-15% [6], significantly higher than the 1%-2% annual incidence of dementia in the general population. Despite various potential treatments for AD, including enzymes that inhibit the production of amyloid- $\beta$  and antibodies that clear amyloid- $\beta$  from the brain [7], no current medications can fully cure dementia or significantly alter its clinical course. Moreover, studies indicate that early intervention is effective, necessitating precise and sensitive diagnostic measures for MCI [8]. Thus, early identification of MCI is crucial as it enables timely interventions to slow cognitive decline and alleviate the burden of dementia [9].

Wearable devices provide a near-continuous, passive data collection method, offering a convenient and minimally invasive approach for the ongoing monitoring and tracking of cognitive decline in patients with MCI. Existing studies have demonstrated that various physiological indicators, such as heart rate variability [10], electrodermal activity [11], gait variability [12], skin temperature [13], respiratory rate [14], electroencephalography [15], eye movement [16], and electromyography [17], can be effectively used to assess cognitive function changes, providing an objective basis for the auxiliary diagnosis of early cognitive impairment. However, despite the availability of diverse physiological data from patients with MCI, challenges remain in the effective utilization of these data due to the complexity of high-dimensional

information (eg, feature redundancy, strong interfeature correlations, and noise interference) and the technical difficulties in multimodal data integration (eg, insufficient feature extraction and dimensionality reduction methods, challenges in aligning heterogeneous modalities, and limitations in handling noise and missing data).

In recent years, machine learning techniques have been increasingly applied to analyzing and processing complex, high-dimensional physiological data to facilitate the early detection of cognitive disorders, including MCI. Traditional algorithms such as naive Bayes [18], k-nearest neighbors (KNN) [19], support vector machines (SVM) [20], and logistic regression (LR) [21] have demonstrated a certain degree of effectiveness in identifying high-risk MCI populations. However, due to the limitations of single algorithms in modeling high-dimensional and multimodal data, such as insufficient representational capacity and unstable generalization performance, researchers have gradually shifted toward exploring ensemble methods, including bagging [22], boosting [23], and stacking [24]. These ensemble learning techniques integrate predictions from multiple models, effectively mitigating the limitations of single models and significantly improving overall predictive accuracy and robustness.

In addition, various swarm intelligence algorithms have been introduced for critical tasks such as feature selection and hyperparameter optimization to enhance the performance of machine learning models in high-dimensional data analysis. Swarm intelligence algorithms, including Harmony Search (HS) [25], Particle Swarm Optimization [26], and Genetic Algorithms [27], simulate cooperative behaviors observed in nature and have demonstrated outstanding potential in solving global optimization problems. The HS algorithm has gained attention as a metaheuristic optimization technique due to its simplicity, ease of implementation, and low-parameter adjustment requirements. Inspired by musical harmony improvisation, the HS algorithm iteratively adjusts the pitch of each instrument (analogous to high-dimensional features) to find the optimal feature combination. This approach offers an effective solution for feature selection involving physiological data and cognitive parameters of MCI patients, showing promising prospects in improving model prediction accuracy and computational efficiency.

**Objective**

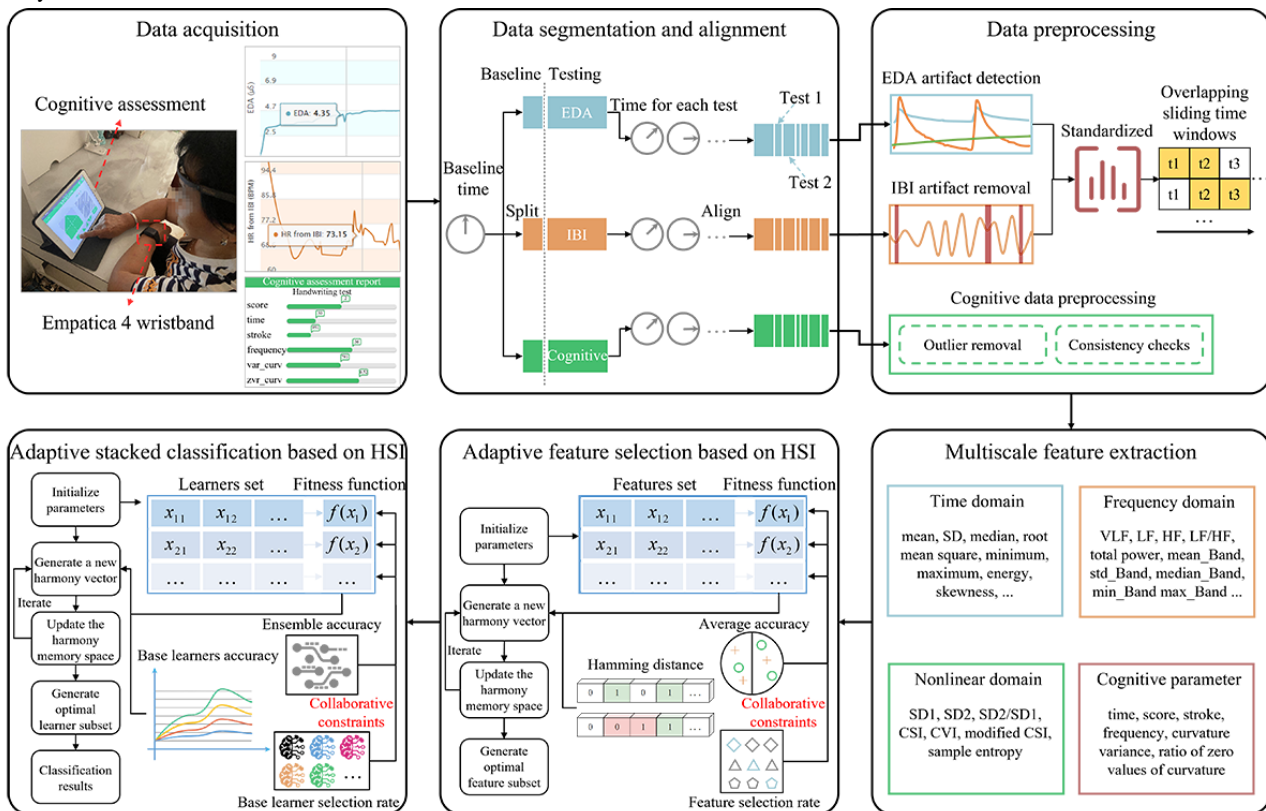
We propose a Dynamic Adaptive Ensemble Learning Framework based on an Improved Harmony Search (DAELF-HSI), designed to enhance MCI detection by addressing issues of feature redundancy and the inefficiency of multimodal data fusion. In contrast to previous studies, this research integrates multimodal physiological data collected through wearable wristbands (eg, heart rate variability and electrodermal activity) with cognitive assessment metrics recorded on tablet devices (eg, reaction time and test scores), aiming to exploit the potential value of multisource data comprehensively, thereby improving the accuracy and clinical utility of MCI detection. We hypothesize that the DAELF-HSI framework will not only effectively distinguish between patients with MCI and healthy individuals but also uncover critical discriminative information pertinent to MCI.

**Methods**

**Ethical Considerations**

The research was reviewed and approved by the Biomedical Ethics Review Committee of Taiyuan University of Technology

**Figure 1.** A dynamic adaptive ensemble learning framework for mild cognitive impairment detection. CSI: cardiac sympathetic index; CVI: cardiovascular index; EDA: electrodermal activity, HF: high frequency; HSI: harmony search improved; IBI: interbeat interval; LF: low frequency; VLF: very low frequency.



**Experimental Participants and Procedures**

The dataset used for machine learning modeling involves 843 participants aged 60 and older recruited from partner hospitals. The participants were randomly divided into a development dataset (674/843), and an independent testing dataset (169/843) in a ratio of 4:1. In addition, 226 older adults were recruited

(20240124). All methods were performed following relevant guidelines and regulations. Written informed consent was obtained from eligible participants under the principles of the Declaration of Helsinki. All participants signed an informed consent form. We provided US \$10 to eligible older adults as compensation for participation.

**Overview of the Proposed Detection Framework**

Figure 1 illustrates a Dynamic Adaptive Ensemble Learning Framework for MCI detection, integrating multimodal data that integrates individual physiological signals with cognitive tasks derived from serious games. The framework begins with data collection, followed by time series segmentation, alignment, and preprocessing. It then progresses to feature extraction and selection, culminating in constructing a classification model. Notably, the modules within the framework are interconnected and sequentially executed, forming a cohesive unit. The following sections will detail the stages, demonstrating the adaptability and effectiveness of the proposed MCI detection framework.

from 3 external centers to constitute an external testing dataset. Participants were identified using a purposive sampling method [28], with the process being meticulously overseen by experienced neurologists. The inclusion criteria for participants were (1) age ≥60 years; (2) normal hearing and vision, or corrected to normal; (3) completion of the Mini-Mental State Examination (MMSE) test; (4) completion of the Montreal

Cognitive Assessment (MoCA) test; (5) capability to engage in moderate activity without physical disabilities; (6) absence of severe depressive symptoms or other neurological disorders such as stroke or Parkinson disease; (7) ability to effectively use smart devices such as smartphones and tablets; and (8) informed consent signed by the participants or their guardians.

Neurologists contacted potential participants during their clinic visits, explaining the study's purpose, related procedures, and the possible impact of the research findings. Once potential participants expressed interest, neurologists conducted comprehensive medical evaluations, including detailed medical history collection, physical examinations, brain imaging (magnetic resonance imaging or computed tomography scans), and cognitive function assessments (using the MMSE and MoCA scales). The MCI group comprised 514 (48.1%) participants who scored below 26 on the MoCA, while the

healthy control (HC) group included 555 (51.9%) healthy individuals without symptoms of cognitive decline. Brain imaging scans revealed no structural abnormalities causing cognitive impairment. Furthermore, all patients with MCI met the criteria proposed by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association [29]. We also administered the habitual hand questionnaire [30], which consisted of 13 items, to all participants. The MCI and HC groups were matched for age, gender, hand preference, education, average sleep duration (in general), and years of smart device use. All participants completed cognitive tasks with the assistance of researchers, and there were no dropouts during the testing process. Table 1 summarizes the clinical and demographic characteristics of the 1069 participants enrolled across the development, internal, and external datasets.

**Table 1.** Characteristics of the 1069 participants enrolled across the development, internal, and external datasets.

	Training (n=674)		Testing (n=169)		External (n=226)		P value <sup>c</sup>
	MCI <sup>a</sup> (n=328)	HC <sup>b</sup> (n=346)	MCI (n=78)	HC (n=91)	MCI (n=108)	HC (n=118)	
Age (years), mean (SD)	70.39 (6.330)	69.64 (5.919)	69.85 (6.004)	69.46 (5.763)	74.29 (8.477)	73.9 (8.613)	.11 <sup>d</sup> ; .67 <sup>e</sup> ; .73 <sup>f</sup>
<b>Gender, n (%)</b>							.22 <sup>d</sup> ; .81 <sup>e</sup> ; .64 <sup>f</sup>
Women	181 (55.2)	207 (59.8)	44 (56.4)	53 (58.2)	61 (56.5)	63 (53.4)	—
Men	147 (44.8)	139 (40.2)	34 (43.6)	38 (41.8)	47 (43.5)	55 (46.6)	—
<b>Hand preference, n (%)</b>							.60 <sup>d</sup> ; .80 <sup>e</sup> ; .88 <sup>f</sup>
Left	41 (12.5)	48 (13.9)	6 (7.7)	8 (7.5)	13 (12.0)	15 (12.7)	—
Right	287 (87.5)	298 (86.1)	72 (92.3)	83 (91.2)	95 (88.0)	103 (87.3)	—
Education years, mean (SD)	5.81 (3.781)	6.32 (3.744)	6.22 (4.012)	6.48 (3.903)	5.34 (3.421)	5.22 (3.457)	.08 <sup>d</sup> ; .66 <sup>e</sup> ; .79 <sup>f</sup>
Hours of sleep, mean (SD)	7.53 (1.165)	7.40 (1.115)	7.51 (1.066)	7.82 (1.060)	7.13 (1.421)	7.01 (1.362)	.14 <sup>d</sup> ; .06 <sup>e</sup> ; .51 <sup>f</sup>
Smart device use (years), mean (SD)	4.97 (2.060)	5.07 (1.918)	4.32 (2.665)	4.79 (2.563)	4.71 (2.336)	4.90 (2.326)	.53 <sup>d</sup> ; .24 <sup>e</sup> ; .55 <sup>f</sup>

<sup>a</sup>MCI: mild cognitive impairment.

<sup>b</sup>HC: health control.

<sup>c</sup>P value: 2-tailed *t* tests (for continuous variables) and chi-square tests (for categorical variables).

<sup>d</sup>P value: statistical comparisons were performed between the MCI and HC groups within the training dataset.

<sup>e</sup>P value: statistical comparisons were performed between the MCI and HC groups within the testing dataset.

<sup>f</sup>P value: statistical comparisons were performed between the MCI and HC groups within the external dataset.

In the experimental setting, a well-trained experimenter instructed participants to sit on a comfortable chair and wear the Empatica 4 on their nondominant wrist. The Empatica 4 is a watch-like multisensor device that measures physiological data such as electrodermal activity (EDA), photoplethysmography, skin temperature, and accelerometer readings. It is compact, lightweight, and comfortable to wear, making it suitable for unobtrusive continuous monitoring during cognitive screening of older adults. The participants performed cognitive tasks on a 2019 iPad using the Brain Nursing mobile app developed by our team (Multimedia Appendix 1) [31], completing drawing-related tasks with an Apple Pencil. The

system includes 11 single tasks and 3 dual tasks, each taking only 1-3 minutes, designed to assess attention; short-term memory; working memory; scene recall and situational reconstruction; visual-conceptual and visual-motor tracking; orientation; executive function; language comprehension and expression; logical thinking; and fine motor control. To minimize interference from the nondominant hand during the painting tasks, the experimenter provided appropriate assistance, such as stabilizing the tablet. Upon completion of the testing, the Empatica 4 wristband was removed from the participant's wrist, physiological data were retrieved and downloaded through



the Empatica 4 Connect portal, and cognitive data were exported from the cloud.

## Data Segmentation and Alignment

### Overview

For the collected multisource data, such as EDA, interbeat interval (IBI) for describing heart rate variability (HRV), and cognitive data, ensuring the integrity, continuity, and temporal alignment of the data is crucial. As participants perform cognitive tasks, the tablet automatically records timestamps for each test, providing a reference for aligning physiological data. Thus, we align the EDA and IBI data using the start and end times of each test, as described in the procedures listed below.

### EDA Time Series Processing

Considering that the EDA.csv file downloaded from the cloud only contains information about the start time of the session and the sampling rate, it lacks the timestamps corresponding to each second of the signal. To address this deficiency, we generate a timestamp every 4 data points based on the session start time and the sensor sampling rate (4 Hz). Subsequently, we align the timestamps of the cognitive tests with the EDA series timestamps, thereby extracting the EDA signal segments corresponding to the specific cognitive tests.

### IBI Time Series Processing

Due to the automatic discarding of unidentifiable heartbeats by the Empatica 4 wristband during measurement, the IBI.csv file contains discontinuities that do not match the actual measuring intervals. It is crucial to accurately identify and fill these measurement gaps to ensure data integrity in the analysis of IBI time series for various test tasks. Following the suggestions of Rafi et al [32], this study limits the physiologically feasible range for IBI to within 2 seconds, and any data beyond the threshold was automatically labeled as a measurement gap. Subsequently, cubic spline interpolation [33] is used to estimate the missing values within these gaps. The overall continuity of the IBI dataset in the temporal dimension is optimized by using a curve-fitting method based on available data points. Finally, new timestamps are added to the IBI data, aligning the timestamps of cognitive tests with those of the IBI series.

## Data Preprocessing

### Overview

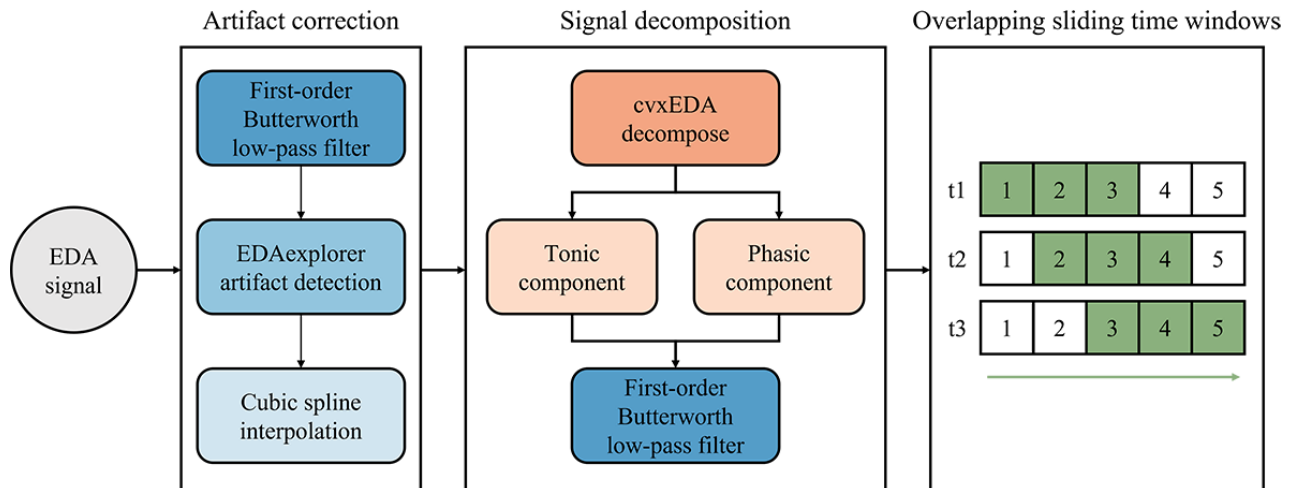
Commercial wearable devices are prone to artifacts, measurement gaps, or deviations from the measurement regime during data recording [34,35]. To ensure reliable information

is extracted from field-collected data, rigorous preprocessing is required to filter noise and artifacts and restore the original signal. Considering the differences among EDA, IBI, and digital parameters, we will detail the preprocessing methods for these metrics.

### EDA Signal Preprocessing

EDA serves as a biosignal, mirroring the individual physiological and emotional states, and consists mainly of slowly varying tonic and rapidly fluctuating phasic activities. Tonic activity, also known as skin conductance level (SCL), primarily reflects the physiological activity level of an individual at rest, indicating the continuous regulation of the autonomic nervous system. In contrast, phasic activity, or SCR, is a rapid and transient physiological response to specific stimuli, revealing an individual's adaptability and reactivity to sudden events. To enhance EDA signal quality, we propose a multistage automatic artifact removal method, including artifact correction, signal decomposition, and overlapping sliding time windows, as shown in Figure 2. The specific steps involved are:

1. Low-pass filtering: EDA signal is filtered using a first-order Butterworth low-pass filter with a cutoff frequency of 0.6 Hz [36], which preserves its low-frequency components and eliminates high-frequency noise.
2. Artifact detection: EDAexplorer [37] is used to detect artifacts in the filtered signal, identifying and marking anomalies within the signal to provide a basis for data repair.
3. Cubic spline interpolation: apply cubic spline interpolation to the identified artifact segments, using segmented cubic polynomials to approximate missing data points while ensuring continuity in function values and their first and second derivatives, thereby smoothly completing the missing data.
4. Signal decomposition: by solving the convex optimization approach, cvxEDA [38] separates the signal into tonic and phasic components, enabling enhanced analysis and interpretation of the underlying physiological mechanisms within the EDA signal.
5. Component filtering: refilter the decomposed tonic and phasic components using a low-pass Butterworth filter to eliminate negative SCR and SCL values, enhancing the signal quality.
6. Time window segmentation: segment the processed tonic and phasic components into overlapping time windows of 60 seconds with a step size of 1 second to facilitate subsequent feature extraction.

**Figure 2.** Electrodermal activity signal preprocessing flow.

### IBI Signal Preprocessing

Wearable devices commonly use photoplethysmography sensors to monitor the continuous variations in interbeat or R-R intervals. However, obtaining raw photoplethysmography data from the Empatica 4 wristband presents challenges, so we shifted to analyzing IBI data, which is more readily accessible. Analyzing IBI data allows for calculating HRV, reflecting the variations in time between consecutive heartbeats. Although

Empatica 4 offers convenience and noninvasiveness for recording HRV, it still faces issues such as artifacts or measurement gaps [39,40]. Therefore, we initially adopted 4 artifact detection rules to identify artifacts, as shown in [Textbox 1](#). Subsequently, detected artifacts were interpolated using cubic spline interpolation to fill in missing values. Finally, the cleaned IBI data was segmented into overlapping time windows of 60 seconds with a 1-second step size, creating datasets for subsequent feature extraction.

**Textbox 1.** List of interbeat interval artifact detection rules.

Study and rule description

#### Rafi et al [32]

- Discard any interbeat intervals that do not fall within the physiological range of 250-2000 milliseconds (equivalent to a heart rate of 30-240 beats per minute).

#### Malik et al [41]

- Each interbeat interval should be at most 20% from the previous one.

#### Acar et al [42]

- Calculate the average of the 9 interbeat intervals preceding the current interbeat interval. It should be removed if the current interbeat interval differs from this average by more than 20%.

#### Karlsson et al [43]

- Remove any interbeat interval that differs by more than 20% from the average of its immediate preceding and succeeding interbeat intervals.

### Cognitive Data Preprocessing

Outlier removal and data consistency checks were performed manually to preprocess digital cognitive parameters. The specific steps include (1) format validation: ensuring that all data entries adhere to the required format specifications (eg, the time recorded in seconds and scores in numerical format) and correcting any inconsistencies; (2) range validation: checking that all data values fall within predefined acceptable ranges, such as ensuring reaction times are within a reasonable range of seconds; (3) continuity validation: assessing the continuity of the data, including verifying that the timestamps for each test are in sequential order and checking for any missing data points. Through these steps, we aim to identify and eliminate extreme

outliers caused by user errors or external interference while ensuring the logical consistency of data format, range, and time series, thereby improving the overall quality and reliability of the data.

### Multiscale Feature Extraction

In this study, we comprehensively analyzed data collected from the Empatica 4 wristband and tablet devices to explore participants' physiological and cognitive responses to various cognitive tasks. Specifically, we used the FLIRT toolkit [44] and NeuroKit2 [45] to extract 39 EDA-related features (including 17 SCL features and 22 SCR features) from the EDA signals and 23 features (including HR and HRV) from the IBI data. Furthermore, we collected several cognitive parameters,

including time, score, stroke, frequency, and curvature (variance and the ratio of 0 values) during each test. Detailed information regarding all these features is available in [Multimedia Appendix 1](#).

### Dynamic Adaptive Feature Selection Based on Improved Harmony Search

Feature selection plays a critical role in handling high-dimensional datasets, as not all features impact the outcome. An excessive number of features can result in the curse of dimensionality and increased model complexity. Therefore, this study proposes a feature selection algorithm based on HSI to sift through extracted physiological and cognitive features. Analogous to musical notes, each feature represents a note that may or may not be selected into a subset. Musicians repeatedly adjust the notes to achieve the best harmony effect until reaching a satisfactory harmony. Similarly, the feature selection algorithm based on HSI continuously tunes parameters and modifies the generated feature subset to ensure diversity and avoid convergence to local optima. In particular, we integrate Hamming distance into the HSI algorithm to gauge the disparity between the newly generated harmony vector and the optimal vector in the harmony memory. The Hamming distance is used to assess the similarity between these vectors to fine-tune the search probability. A high Hamming distance leads to a moderate reduction in the search probability to promote exploratory efforts, while a low Hamming distance results in a moderate increase in leveraging known information. Finally, to evaluate the quality of the harmony effect, we minimize the average classification error rate of all base learners and the feature selection rate as optimization objectives. The detailed algorithm is provided in [Multimedia Appendix 2](#).

### Dynamic Adaptive Stacking Classification Based on Improved Harmony Search

An essential goal of this study is to distinguish between healthy individuals and patients with MCI effectively. As mentioned, we opt for a subset of features demonstrating balanced performance across all base learners during the feature selection phase. Nevertheless, certain learners continue to demonstrate suboptimal performance, and merely stacking multiple base learners increases algorithmic complexity and computational demands. In essence, the selection of learners, akin to feature selection, constitutes a combinatorial optimization problem focused on enhancing classification performance. Thus, this study proposes using HSI to optimize the stacking of base learners. Unlike feature selection algorithms that use HSI, it

leverages the accuracy of the current base learners and their quantity to guide hyperparameter adjustments. The strategy aims to mitigate the adverse effects of underperforming learners on the overall model while simultaneously enhancing model efficiency and minimizing computational costs. Furthermore, we selected the KNN, decision tree (DT), random forest (RF), Gaussian naive Bayes, SVM, multilayer perceptron, LR, gradient boosting DT, and XGBoost as base learners, with LR serving as the meta-learner. The detailed algorithm is provided in [Multimedia Appendix 2](#).

### Statistical Analysis and Machine Learning Model

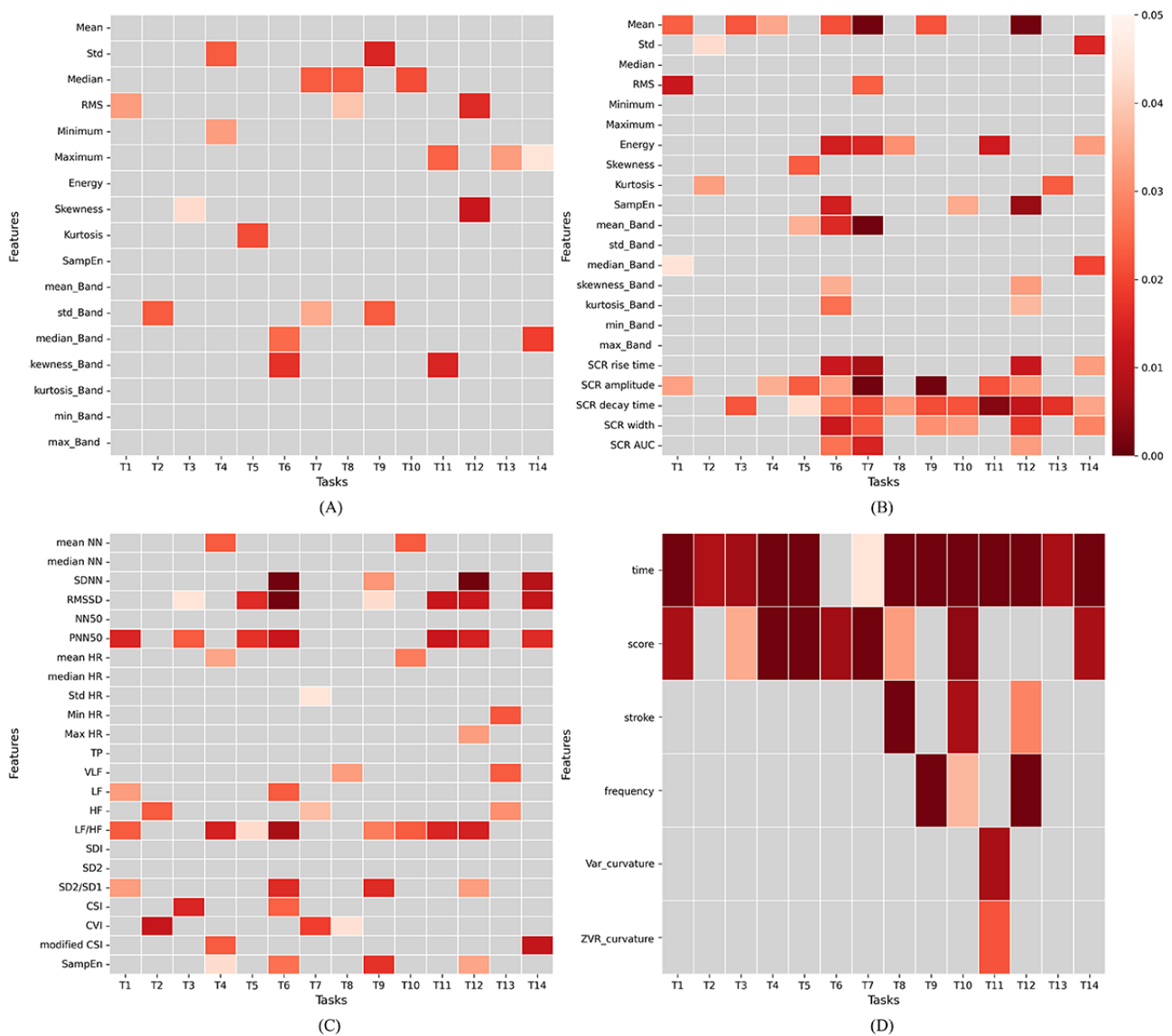
This study analyzed demographic characteristics, cognitive parameters, and physiological features using the Statistical Package for the Social Sciences (SPSS, version 22.0 for Windows, IBM). Initially, the Kolmogorov-Smirnov test was used to assess whether continuous variables such as age, years of education, hours of sleep, years of smart device usage, and cognitive data conformed to a normal distribution. Descriptive statistics were described using means (SD) for normally distributed variables. Subsequently, the *t* test was used for between-group comparisons to determine if there were significant differences between the HC group and the MCI group. In contrast, we used the nonparametric Mann-Whitney U test for nonnormally distributed variables to assess intergroup differences. For categorical variables such as gender and hand preference, data were described in terms of counts (percentages), and the chi-square test was used for group comparisons. The significance level for all statistical analyses was set at  $P < .05$ . Furthermore, to validate the performance of the proposed detection framework, experiments were conducted using a 5-fold cross-validation approach, using 4 evaluation metrics (accuracy, precision, recall, and  $F_1$ -score) to assess the classification outcomes. All learners used the HSI for feature selection, and the average values were used as the final classification results.

## Results

### Statistical Comparison of EDA, HRV, and Cognitive Features Between Groups

We conducted statistical analyses on features extracted from EDA, HRV, and cognitive tasks to identify the key features distinguishing between healthy individuals and patients with MCI. As shown in [Figure 3](#), red squares indicate significant differences ( $P < .05$ ) between the 2 groups on specific features during certain cognitive tasks, and the depth of the color reflects the degree of significance of these differences.

**Figure 3.** Results of the t tests performed on features extracted from electrodermal activity, heart rate variability, and cognitive tasks between groups. HRV: heart rate variability; SCL: skin conductance level; SCR: skin conductance response.



Based on observations from Figure 3A, relatively few SCL features distinguish between patients with MCI and healthy individuals, including the SD, median, root mean square, and SD of spectral power. Conversely, Figure 3B reveals more significant differences in SCR features, including mean, energy, amplitude, rise time, delay time, and width in the time domain and mean in the frequency domain. These findings highlight several key points: (1) the SD and root mean square of SCL indicate variability and instability in physiological responses, with differences between patients with MCI and healthy individuals reflecting disparate levels of physiological variability; (2) statistical differences in the SD of SCL spectral power between the groups indicate that patients with MCI exhibit significantly different physiological responses within specific frequency ranges, possibly related to impaired cognition associated with MCI; (3) statistical differences in SCR time domain features between groups indicate that patients with MCI exhibit variations in the intensity and timing of physiological responses, suggesting impaired regulatory capabilities of their nervous systems; and (4) variations in the mean values in the

SCR frequency domain indicate that patients with MCI have different physiological response frequency distributions when processing stimuli compared with healthy individuals.

Figure 3C depicts the statistical analysis results applied to features extracted from HRV between patients with MCI and healthy controls. The analysis indicates statistically significant distinctions in HRV indices such as SDNN (SD of N-N intervals), RMSSD (root mean square of successive differences), PNN50 (percentage of successive R-R intervals > 50 ms), LF/HF (low-frequency to high-frequency ratio) ratio, SD2/SD1 (ratio of the SD2 and SD1 of Poincaré plot), and SampEn (sample entropy). These results highlight several key aspects which are (1) SDNN, RMSSD, and PNN50, which quantify overall and short-term heart rate variations, reveal disparities in autonomic nervous system functioning between the groups; (2) the LF/HF ratio reflects imbalances between sympathetic and parasympathetic nervous activities, indicating autonomic dysregulation in patients with MCI relative to controls; and (3) SD2/SD1 and SampEn focus on balance between long-term and short-term variability, as well as the complexity and irregularity



of HRV, illustrating differences in autonomic nervous system adaptability and complexity between groups. Finally, observations from Figure 3D reveal that (1) multiple cognitive tests using time and scores as indicators can significantly distinguish between healthy individuals and patients with MCI, and (2) features such as handwriting, frequency, and curvature show varying degrees of significant differences across different drawing tasks.

### Performance of the DAELF-HSI in Mild Cognitive Impairment Detection

Table 2 lists the classification outcomes of the DAELF-HSI compared with 6 machine learning models and the application of HSI for feature selection in these models. Following a thorough assessment involving 100 iterations of 5-fold cross-validation, the DAELF-HSI demonstrated an average accuracy of 88.5%, surpassing the performance of other

algorithms significantly. Moreover, it exhibited superior precision, recall, and  $F_1$ -score metrics, achieving 89.1%, 88.7%, and 88.9%, respectively, all at notably high levels. When HSI-based feature selection was not employed, the SVM model outperformed other machine learning algorithms with an accuracy of 79.6%. However, after integrating HSI feature selection, models like KNN-HSI and multilayer perceptron-HSI displayed improved performance, surpassing that of SVM-HSI. Noteworthy is the consistent enhancement in performance observed across all machine learning models upon the introduction of HSI feature selection, with accuracy improvements ranging from 3% to 5%, resulting in all models achieving accuracy levels exceeding 81%. It highlights the efficacy of the feature selection algorithm based on HSI in identifying crucial features that enhance the predictive capabilities of the machine learning models under investigation.

**Table 2.** Performance comparison of Dynamic Adaptive Ensemble Learning Framework based on an Improved Harmony Search with 6 machine learning models and applying Improved Harmony Search to model feature selection.

Methods	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	$F_1$ -score, mean (SD)
KNN <sup>a</sup>	0.792 (0.018)	0.775 (0.023)	0.806 (0.029)	0.790 (0.019)
SVM <sup>b</sup>	0.796 (0.024)	0.788 (0.025)	0.800 (0.020)	0.794 (0.027)
GNB <sup>c</sup>	0.771 (0.022)	0.787 (0.019)	0.796 (0.022)	0.791 (0.025)
DT <sup>d</sup>	0.786 (0.025)	0.791 (0.033)	0.783 (0.025)	0.787 (0.033)
MLP <sup>e</sup>	0.790 (0.017)	0.807 (0.025)	0.794 (0.026)	0.800 (0.019)
LR <sup>f</sup>	0.784 (0.020)	0.789 (0.034)	0.791 (0.028)	0.790 (0.038)
KNN-HSI <sup>g</sup>	0.842 (0.017)	0.848 (0.018)	0.830 (0.027)	0.839 (0.024)
SVM-HSI	0.831 (0.022)	0.833 (0.023)	0.839 (0.030)	0.825 (0.027)
GNB-HSI	0.815 (0.023)	0.815 (0.030)	0.800 (0.025)	0.807 (0.027)
DT-HSI	0.827 (0.019)	0.833 (0.022)	0.809 (0.028)	0.821 (0.031)
MLP-HSI	0.836 (0.022)	0.839 (0.024)	0.842 (0.025)	0.840 (0.018)
LR-HSI	0.817 (0.025)	0.815 (0.027)	0.823 (0.031)	0.819 (0.035)
DAELF-HSI <sup>h</sup> (ours)	0.885 (0.020)	0.891 (0.021)	0.887 (0.024)	0.889 (0.025)

<sup>a</sup>KNN: k-nearest neighbors.

<sup>b</sup>SVM: support vector machines.

<sup>c</sup>GNB: Gaussian naive Bayes.

<sup>d</sup>DT: decision tree.

<sup>e</sup>MLP: multilayer perceptron.

<sup>f</sup>LR: logistic regression.

<sup>g</sup>HSI: Improved Harmony Search.

<sup>h</sup>DAELF-HSI: Dynamic Adaptive Ensemble Learning Framework based on an Improved Harmony Search.

Similarly, Table 3 presents the effectiveness of DAELF-HSI in classification compared with 5 ensemble learning models, augmented by HSI feature selection. In line with the findings in Table 2, DAELF-HSI maintains its superior performance. Noteworthy is that the ensemble learning techniques outperform the individual machine learning models discussed earlier, with XGBoost achieving a commendable average accuracy of 81.9%.

The Bagging model, which uses KNN as a base learner, demonstrates improved performance compared with the stand-alone KNN model. However, the bagging model using ensemble SVM as a base learner falls short of expectations, slightly underperforming compared with the stand-alone SVM model, possibly due to the inherent instability advantages associated with SVM.

**Table 3.** Performance comparison of Dynamic Adaptive Ensemble Learning Framework based on an Improved Harmony Search with 5 ensemble learning models and applying HSI to model feature selection.

Methods	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	$F_1$ -score, mean (SD)
Bag (KNN <sup>a</sup> )	0.794 (0.022)	0.791 (0.031)	0.794 (0.026)	0.792 (0.027)
Bag (SVM <sup>b</sup> )	0.779 (0.011)	0.788 (0.021)	0.780 (0.017)	0.784 (0.021)
RF <sup>c</sup>	0.804 (0.015)	0.802 (0.037)	0.816 (0.027)	0.809 (0.025)
GBDT <sup>d</sup>	0.813 (0.019)	0.816 (0.028)	0.813 (0.027)	0.814 (0.019)
XGBoost	0.819 (0.015)	0.825 (0.024)	0.810 (0.026)	0.817 (0.024)
Bag (KNN)-HSI <sup>e</sup>	0.833 (0.024)	0.842 (0.029)	0.821 (0.034)	0.831 (0.036)
Bag (SVM)-HSI	0.811 (0.016)	0.812 (0.024)	0.819 (0.027)	0.815 (0.024)
RF-HSI	0.854 (0.017)	0.858 (0.018)	0.847 (0.021)	0.852 (0.024)
GBDT-HSI	0.848 (0.027)	0.839 (0.025)	0.852 (0.023)	0.845 (0.016)
XGBoost-HSI	0.854 (0.022)	0.860 (0.020)	0.849 (0.028)	0.854 (0.029)
DAELF-HSI <sup>f</sup> (ours)	0.885 (0.020)	0.891 (0.021)	0.887 (0.024)	0.889 (0.025)

<sup>a</sup>KNN: k-nearest neighbors.

<sup>b</sup>SVM: support vector machines.

<sup>c</sup>RF: random forest.

<sup>d</sup>GBDT: gradient boosting decision tree.

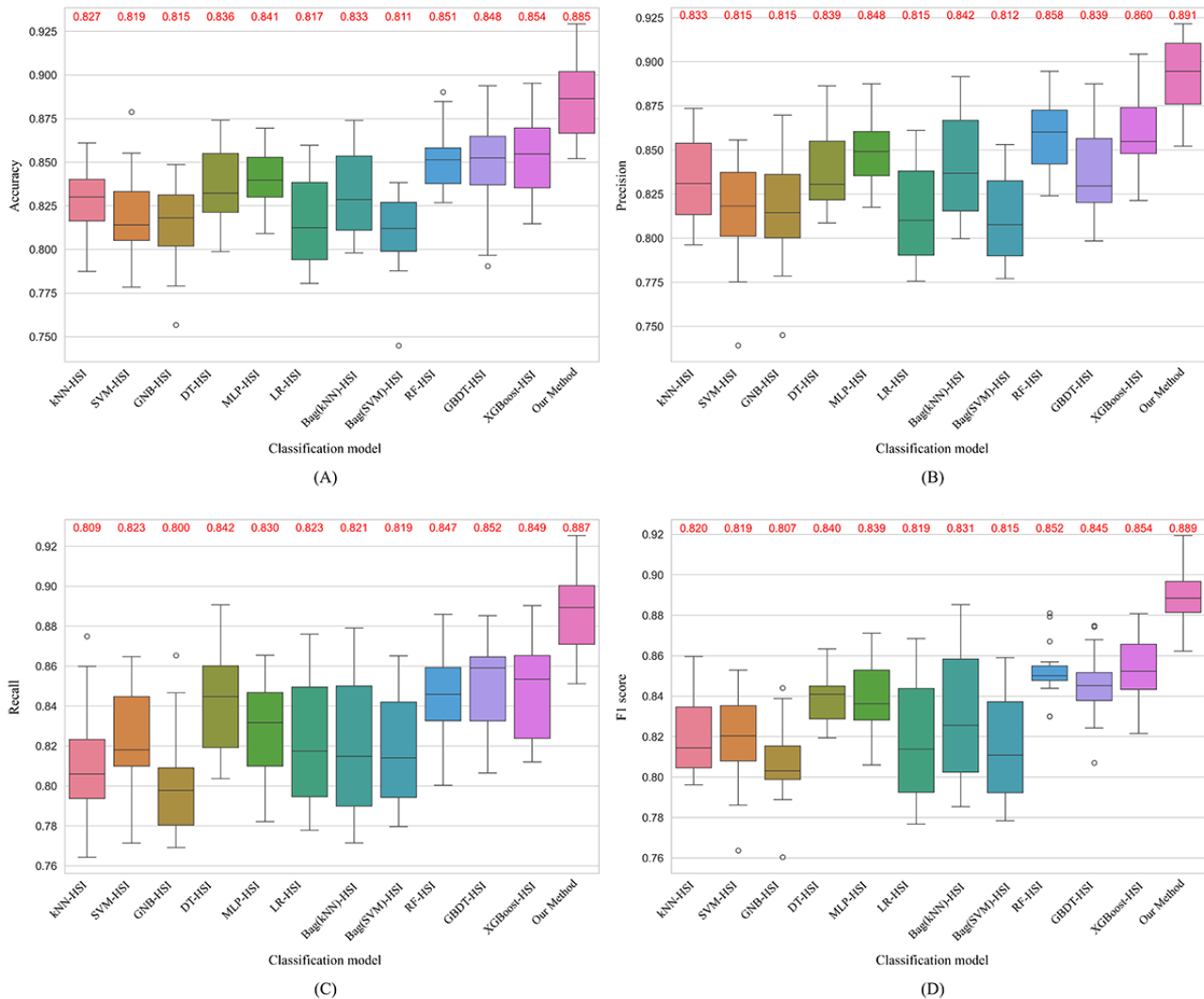
<sup>e</sup>HSI: Improved Harmony Search.

<sup>f</sup>DAELF-HSI: Dynamic Adaptive Ensemble Learning Framework based on an Improved Harmony Search.

Figure 4 illustrates the box plots of 12 algorithms after 100 independent experiments across 4 evaluation metrics. The quartiles within each box plot depict algorithmic performance, while the red numbers above each box plot represent the mean values for the respective metrics. DAELF-HSI outstrips competing models across all metrics, demonstrating superior stability and an absence of significant outliers. Conversely, SVM-HSI and GNB-HSI exhibit more outliers, indicating

challenges in achieving precise model fits for specific data distributions, leading to notable performance fluctuations. Although LR-HSI and Bag (KNN)-HSI do not display outliers, their wide IQR suggests instability across diverse feature distributions. Notably, RF-HSI demonstrates significant outliers in the  $F_1$ -score, highlighting the model's vulnerability to certain data distributions or feature sets.

**Figure 4.** Comparison of evaluation metrics between Dynamic Adaptive Ensemble Learning Framework based on an Improved Harmony Search and various machine learning models applying harmony search improved feature selection. DT: decision tree; GBDT-HSI: gradient boosting decision tree-harmony search improved; GNB: Gaussian naïve Bayes; LR: logistic regression; KNN: k-nearest neighbors; RF-HSI: random forest-harmony search improved; SVM: support vector machines.

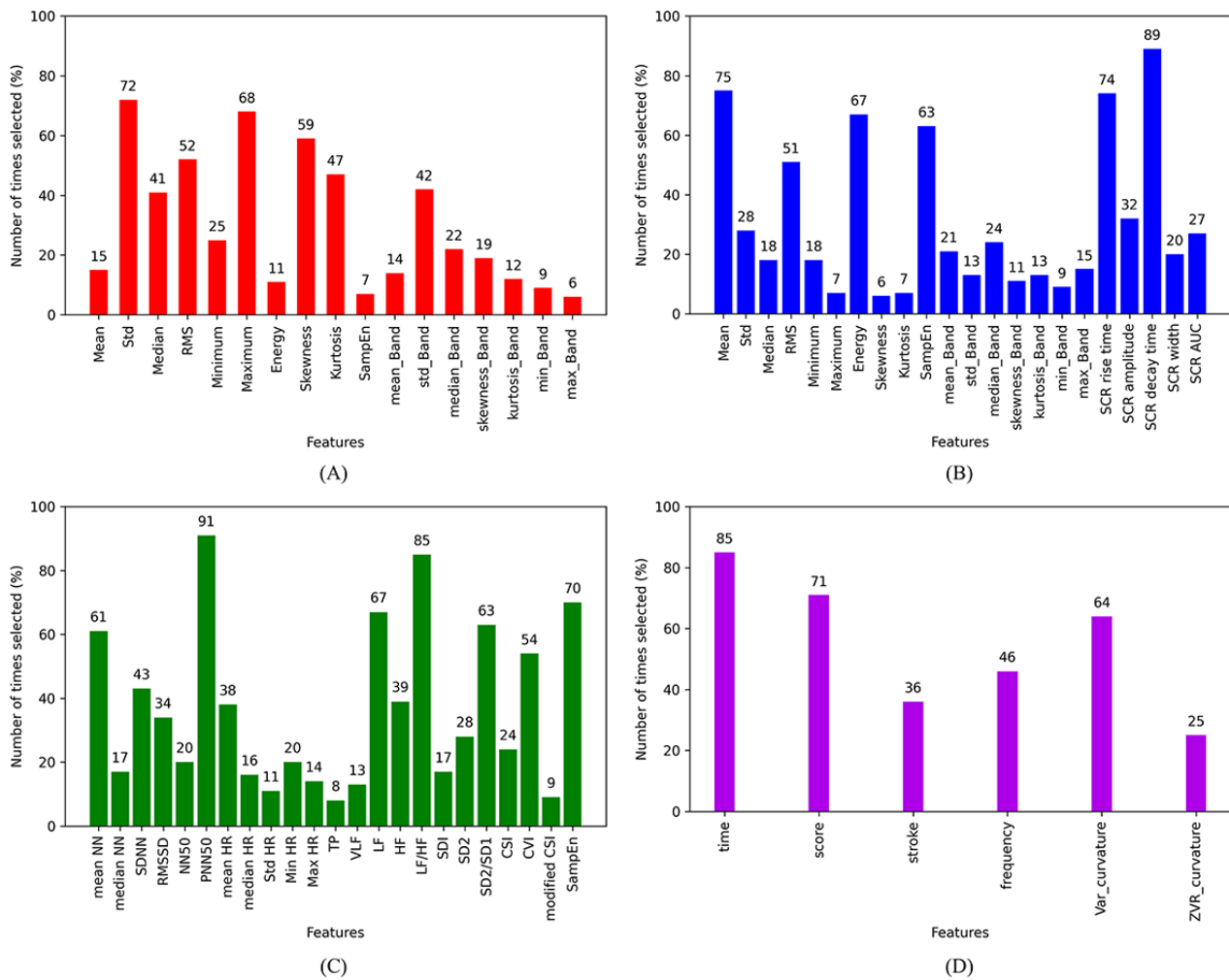


### Analysis of the Optimization Module of the DAELF-HSI

Figure 5 presents the frequency analysis of physiological features (including SCL, SCR, and HRV) and digital cognitive parameters during 100 iterations of the DAELF-HSI model. Notably, SCR decay time (SCR feature set), PNN50, LF/HF (HRV feature set), and time (cognitive feature set) are highlighted for their importance in distinguishing between the 2 groups, appearing in over 80% of the selections. However,

certain SCR features (eg, SCR amplitude, SCR width, and mean band) and HRV features (eg, SDNN, RMSSD, and mean HR), despite showing statistically significant differences ( $P < .05$ ; Figure 3) in group differentiation, were infrequently chosen. This infrequency suggests that these parameters are highly correlated with features previously selected, leading the HSI feature selection algorithm to deem them redundant. Overall, the proposed HSI feature selection optimization algorithm can identify critical features, address redundancy, and accurately detect patients with MCI with limited features.

**Figure 5.** The number of times each skin conductance level, skin conductance response, heart rate variability, and cognitive feature was selected for mild cognitive impairment detection.

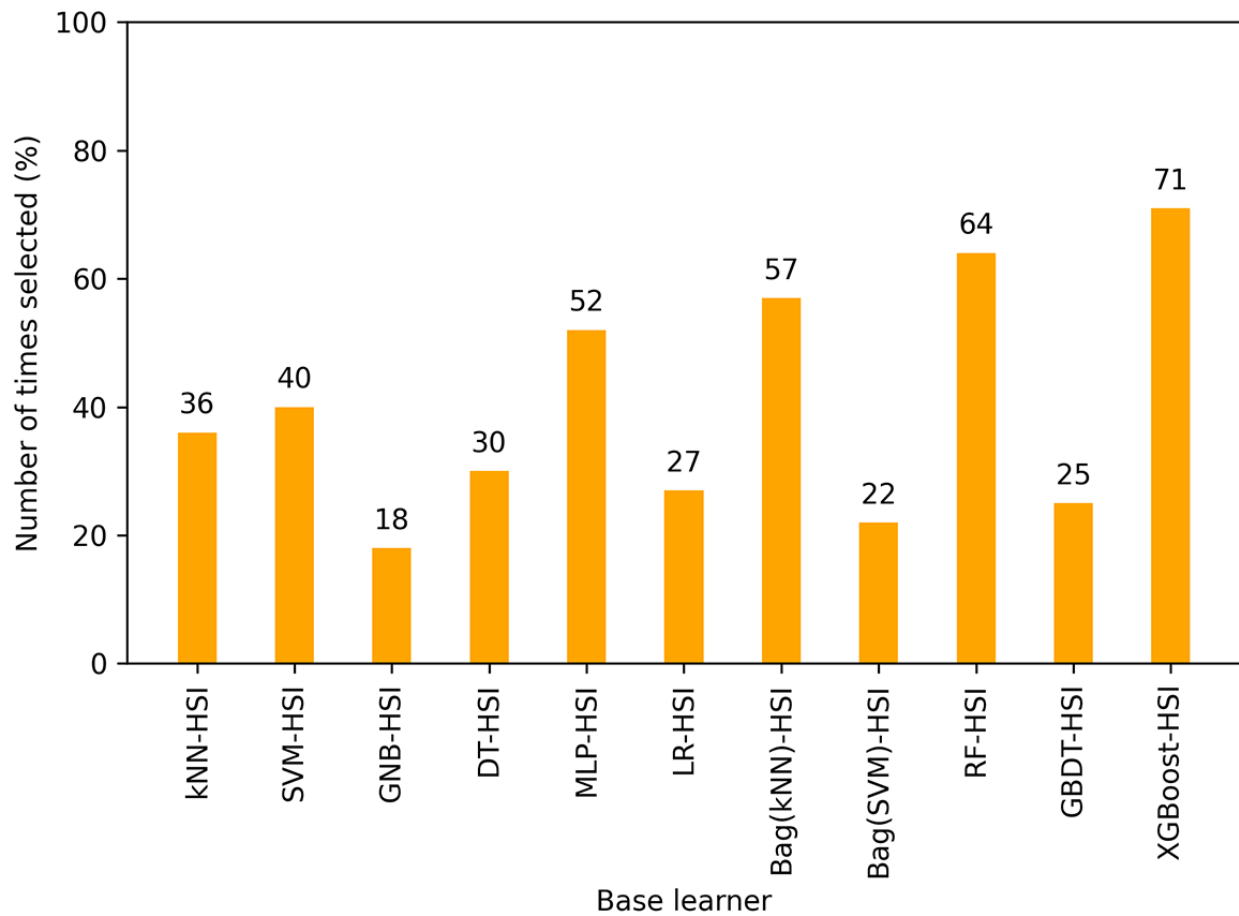


In addition, Figure 6 illustrates the frequency distribution of various base learners across 100 iterations within the DAELF-HSI model. Specifically, RF-HSI and XGBoost-HSI are the predominant selections, highlighting their substantial contributions to model efficacy and demonstrating the model’s proficiency in selecting optimal learners. Conversely, GNB-HSI, Bag (SVM)-HSI, and GBDT-HSI are chosen less frequently. This bias in selection can be primarily attributed to 2 factors:

first, these base learners inherently exhibit lower accuracy, leading the adaptive hyperparameter strategy of the DAELF-HSI to assign them reduced weights; second, a high level of correlation exists between these learners and other high-performing ensemble members, making their inclusion less impactful due to redundant error distributions and similarities in learner characteristics with more effective alternatives.



**Figure 6.** The number of times each base learner was selected for mild cognitive impairment detection. DT: decision tree; GBDT-HSI: gradient boosting decision tree-harmony search improved; GNB: Gaussian naïve Bayes; LR: logistic regression; KNN: k-nearest neighbors; RF-HSI: random forest-harmony search improved; SVM: support vector machines.



Finally, we substantiate the importance of feature selection and stacking optimization stages within the DAELF-HSI through ablation experiments. As delineated in Table 4, a total of 3 experimental models were structured to assess the impact of omitting 1 or both optimization stages, where “✓” denotes the inclusion of that stage. The models are configured as follows:

1. Model A uses 11 machine learning models as base learners and LR as the meta-learner, incorporating all features.
2. Model B uses 11 machine learning models as base learners with LR as the meta-learner, but only includes features selected through the HSI algorithm.
3. Model C integrates the HSI for stacking with 11 machine learning models, incorporating all features.

**Table 4.** Performance comparison of different ablation modules applied to the DAELF-HSI<sup>a</sup>.

Methods	HSI <sup>b</sup> features selection	HSI learners stacking	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	$F_1$ -score, mean (SD)
Model A			0.826 (0.023)	0.819 (0.026)	0.831 (0.024)	0.825 (0.029)
Model B	✓		0.837 (0.027)	0.831 (0.028)	0.839 (0.030)	0.835 (0.035)
Model C		✓	0.854 (0.025)	0.858 (0.019)	0.849 (0.022)	0.853 (0.025)
DAELF-HSI	✓	✓	0.885 (0.020)	0.891 (0.021)	0.887 (0.024)	0.889 (0.025)

<sup>a</sup>DAELF-HSI: Dynamic Adaptive Ensemble Learning Framework based on an Improved Harmony Search.

<sup>b</sup>HSI: Improved Harmony Search.

According to the results in Table 4, model A shows lackluster classification performance in the absence of the 2 optimization stages, primarily due to the poor performance of certain base learners and the inclusion of numerous redundant features during training. Model B, which incorporates HSI-selected features,

demonstrates a slight performance improvement of around 1%, suggesting that merely reducing feature redundancy is insufficient to enhance the output of underperforming base learners significantly. Furthermore, Model C, which solely incorporates the base learner stacking optimization stage,

improves performance by nearly 3%, underscoring the importance of high-quality base learners in stacking algorithms. Notably, the simultaneous implementation of both optimization stages substantially enhances the model's performance, highlighting the essential contribution of each stage to the overall efficacy of the algorithm.

### Performance Evaluation of DAELF-HSI Across Different Datasets

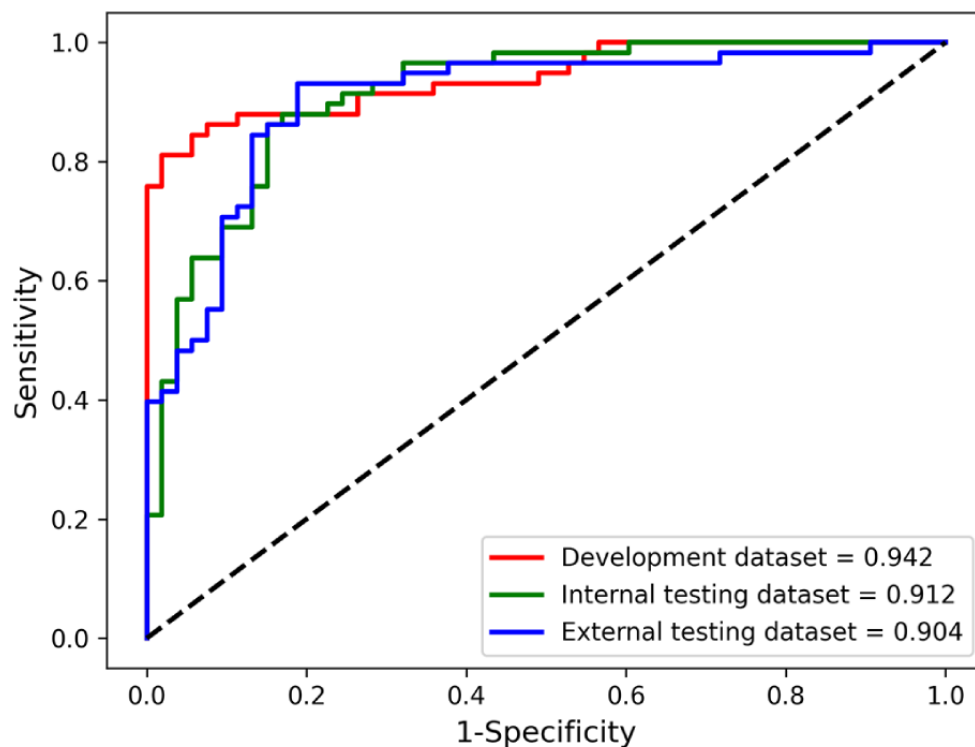
As shown in Table 5, we also used 4 metrics (including accuracy, sensitivity, specificity, and area under the curve [AUC]) to evaluate the binary classification performance of the classification model for patients with MCI versus healthy individuals in internal and external test datasets. Specifically, the model demonstrated excellent performance in the development dataset, with accuracy, sensitivity, and specificity at 88.4%, 86.1%, and 90.9%, respectively. In the internal test dataset, the model's accuracy slightly decreased to 85.5%, while sensitivity remained unchanged at 86.1%, and specificity

dropped to 84.9%, suggesting that a small number of healthy individuals were misclassified as patients with MCI, but overall performance remained satisfactory. However, in the external test dataset, the model's accuracy, sensitivity, and specificity decreased by 3.9%, 0.4%, and 8.1%, respectively. Nevertheless, these results demonstrate the model's effectiveness, especially when evaluated on new and diverse samples. Furthermore, the AUC, which serves as an indicator of the model's validity, was recorded at 0.945 (95% CI 0.903-0.986), 0.912 (95% CI 0.859-0.965), and 0.904 (95% CI 0.846-0.962) across the 3 datasets, as illustrated in Figure 7. The AUC value reflects the classification performance of the model on different datasets. The DAELF-HSI model has an AUC value above the threshold of 0.9 on these datasets, demonstrating excellent sensitivity and specificity in detecting patients with MCI. In other words, the model cannot only effectively identify patients with MCI but also reduce the probability of misjudgment, enhancing reliability in clinical applications.

**Table 5.** The diagnostic value of the classification model in differentiating between healthy older adults and patients with mild cognitive impairment was assessed using internal and external testing datasets.

Dataset	Accuracy	Sensitivity	Specificity	Area under the curve (95% CI)	P value
Development dataset	0.884	0.861	0.909	0.945 (0.903-0.986)	<.001
Internal testing dataset	0.855	0.861	0.848	0.912 (0.859-0.965)	<.001
External testing dataset	0.845	0.857	0.828	0.904 (0.846-0.962)	<.001

**Figure 7.** The receiver operating characteristic curve of our model across different datasets.



### Clinical Utility Analysis

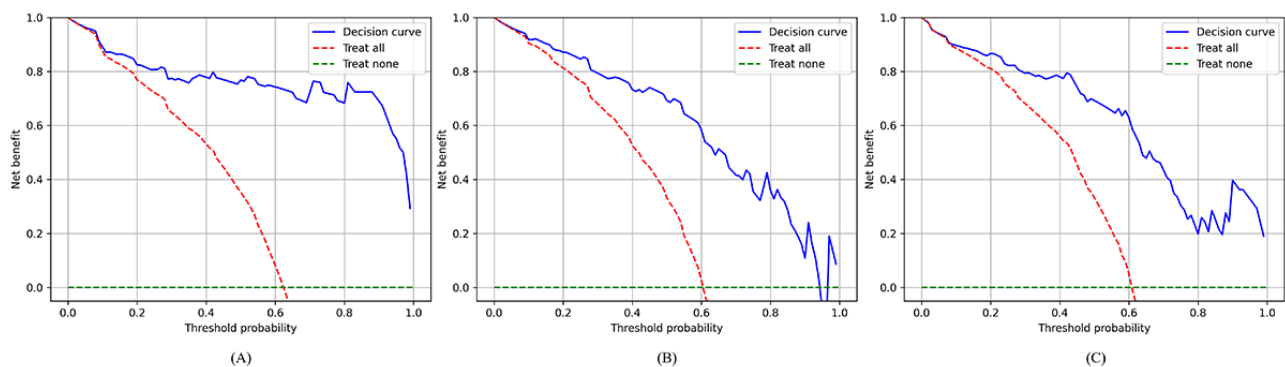
To assess the practical clinical value of the proposed model, we performed decision curve analysis (DCA) on the development, internal testing, and external testing sets, presenting the model's

decision curve along with the "Treat all" and "Treat none" strategies, as shown in Figure 8. Specifically, the model's decision curve demonstrated clear clinical benefits in the development set, with its net benefit consistently exceeding that of the "Treat all" and "Treat none" strategies. Notably, in the

0 to 0.85 threshold range, the model's net benefit declined from 1 to 0.75 and further decreased to 0.3 in the higher threshold range, indicating that the model effectively avoids overtreatment in low-risk patient groups. In the internal testing set, the decision curve remained stable, with a gradual decline in net benefit that aligned with expectations, suggesting that as the threshold increased, the model opted for fewer treatment decisions, and validated the model's effectiveness in high-risk patient populations and its alignment with the clinical need to reduce unnecessary treatments. However, the model's decision curve showed slight variations in the external testing set compared with the other sets. The net benefit decreased from 1 to 0.2 in

the 0-0.8 threshold range, indicating reduced efficacy in this range. However, in the 0.8-0.9 range, the net benefit rose from 0 to 0.4, suggesting that adjusting the threshold appropriately enhanced the model's predictive accuracy. In the 0.9-1 threshold range, the net benefit dropped again to 0.2, which may indicate cautious predictions in the high-risk zone, thus limiting its clinical utility. In conclusion, the model demonstrated a net benefit consistently higher than both the "Treat all" and "Treat none" strategies across the 3 sets, particularly in the mid-to-low threshold ranges, highlighting that the model can effectively guide clinical decision-making, reduce unnecessary treatments, and improve early disease detection and intervention efficiency.

**Figure 8.** Decision curve analysis on different datasets, showing the model's decision curves for the binary classification task, along with the "Treat all" and "Treat none" strategies. (A) Development set, (B) internal testing set, and (C) external testing set.



## Discussion

### Interpretability of Physiological Features

Previous studies demonstrate that HRV is an essential physiological assessment indicator that could differentiate physiological responses between patients with MCI and HCs under diverse stressors and conditions [46,47]. In this study, compared with healthy counterparts, patients with MCI exhibited decreased HR, prolonged N-N intervals, and an increased ratio of low to high-frequency components, consistent with trends observed in past research [10,48]. Moreover, higher levels of HRV have been associated with improved cognitive function, whereas lower HRV correlates with cognitive impairment [49]. During cognitive tasks, abnormal responses in the autonomic nervous system of patients with MCI were observed through monitoring HRV, highlighting the efficacy of HRV as a sensitive physiological indicator during cognitive engagement.

On the other hand, this research assessed the applicability of EDA features for identifying MCI, advancing upon sparse existing research in this field. Specifically, the EDA signal was segregated into tonic and phasic components, from which 39 features encompassing time, frequency, and SCR time domains were derived. Comparative statistical analysis revealed that, during cognitive tasks, patients with MCI exhibited significantly enhanced SCR amplitudes, elongated SCR widths, and protracted SCR decay time relative to healthy counterparts, indicating a more intense or unexpectedly elevated physiological response to stimuli. These pronounced variations in SCR parameters effectively discriminate between patients with MCI and healthy individuals, highlighting the sensitivity of EDA

features in detecting autonomic nervous system dysfunctions, a common correlate of cognitive impairments [50].

### Predictive Accuracy of the Proposed Model

A challenge in enhancing the classification accuracy for discriminating patients with MCI involves identifying the most discriminative feature set. Thus, we proposed a feature selection algorithm based on HSI to mitigate the presence of redundant features and alleviate the adverse effects of high-dimensional data on classification learning, as high dimensionality can impede effective target classification [11]. Furthermore, we explored how to balance the diversity of base learners and classification quality in stacking for a specific dataset, treating the selection of appropriate base learners as a combinatorial optimization problem and solving it using heuristic algorithms. The superior performance of the DAELF-HSI model has been validated on development, internal testing, and external testing datasets, achieving accuracy of 88.4%, 85.5%, and 84.5%. Furthermore, the model demonstrated excellent sensitivity and specificity in detecting patients with MCI, with corresponding AUC of 0.945 (95% CI 0.903-0.986), 0.912 (95% CI 0.859-0.965), and 0.904 (95% CI 0.846-0.962) for these datasets.

### Clinical Implications

The proposed framework offers significant clinical potential for the early detection of MCI through a noninvasive and cost-effective approach, providing a viable solution for widespread screening, particularly in home settings. Heart rate variability, a sensitive marker of autonomic nervous system dysfunction, can serve as a reliable indicator of cognitive impairment, helping clinicians identify at-risk individuals early in the progression of MCI. Similarly, electrodermal activity

features, such as skin conductance responses, effectively differentiate patients with MCI from healthy individuals, further highlighting the potential of autonomic measures in cognitive health assessments. Furthermore, the framework enables more convenient and personalized MCI screening, especially in remote or underserved areas with limited access to specialized medical expertise. Its adaptability allows seamless integration into various health care environments without significant infrastructural changes, reducing the cost and logistical barriers typically associated with traditional diagnostic methods. Moreover, the model's high sensitivity and specificity, as demonstrated by its AUC performance, indicate its reliability in identifying patients with MCI, which is critical for timely intervention and monitoring the progression of dementia to more severe stages. In conclusion, it provides clinicians with a powerful tool for early detection, improving patient outcomes and ultimately alleviating the burden of cognitive disorders on health care systems.

### Limitations and Future Work

For our study, it is important to acknowledge several potential limitations and proposed solutions. First, the developed detection model relies on data from electrodermal activity, photoplethysmography, and cognitive parameters and untested generalizability across different datasets, such as electroencephalogram and eye-tracking data. Future research plans to expand the dataset under laboratory conditions through electroencephalogram and eye-tracking tests to enhance the algorithm's adaptability to various data modalities. In addition, the clinical applicability of the Empatica 4 wristband is limited by its high cost, which may impede its widespread adoption for

household use. Subsequent studies will explore the potential and accuracy of more cost-effective commercial wearables (such as Xiaomi Mi Bands, Huawei Bands, and Garmin Smart 5) in screening for MCI.

### Conclusion

In this study, we proposed a Dynamic Adaptive Ensemble Learning Framework based on Improved Harmony Search, designed to consolidate data collected from wearable wristbands and tablet devices to establish a comprehensive and reliable method for auxiliary diagnosis of MCI. The framework uses the Empatica 4 wristband to collect EDA and photoplethysmography from participants, optimizing these physiological signals through a designed multistage automatic artifact removal algorithm. In addition, based on Improved Harmony Search, a dynamic adaptive feature selection algorithm deepens the analysis of the time, frequency, and nonlinear domain features, enhancing the model's predictive performance. An experimental study involving 1069 participants aged 60 and above demonstrated that DAELF-HSI achieved classification accuracies of 88.4%, 85.5%, and 84.5% on development, internal testing, and external testing datasets, respectively, proving its effectiveness in identifying discriminative information related to MCI. In summary, this study combines an iPad-based cognitive assessment tool with wearable devices, providing an innovative solution for developing a portable, home-friendly early detection and screening system for MCI, offering vital support to clinicians in early detection, improving patient outcomes, and alleviating the burden of cognitive impairments on health care systems.

### Acknowledgments

This research received funding from the National Natural Science Foundation of China (grants 62476190, 62376183, and U21A20469), the Central Government-Guided Science and Technology Development Program (grant YDZJSX2022C004), and Shanxi Provincial Administration of Traditional Chinese Medicine Scientific Research Project Plan (grant 2024ZYY2A033). We sincerely thank the volunteers whose invaluable participation made the successful completion of this study possible.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Detailed description of physiological features and cognitive parameters.

[\[PDF File \(Adobe PDF File\), 621 KB-Multimedia Appendix 1\]](#)

### Multimedia Appendix 2

Detailed description of feature selection and stacked classification algorithms.

[\[PDF File \(Adobe PDF File\), 215 KB-Multimedia Appendix 2\]](#)

### References

1. Murman DL. The impact of age on cognition. *Semin Hear*. 2015;36(3):111-121. [[FREE Full text](#)] [doi: [10.1055/s-0035-1555115](https://doi.org/10.1055/s-0035-1555115)] [Medline: [27516712](https://pubmed.ncbi.nlm.nih.gov/27516712/)]
2. Zeisel J, Bennett K, Fleming R. *World Alzheimer Report 2020: Design, dignity, dementia: dementia-related design and the built environment*. London, UK. Alzheimer's Disease International; 2020.
3. Prince M, Wimo A, Guerchet M, Ali GC, Wu YT, Prina M. *The Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost and Trends*. In: *World Alzheimer Report 2015*. London, UK. Alzheimer's Disease International; 2015.



4. Yates JA, Clare L, Woods RT, Wales Cognitive Function Ageing Study. What is the relationship between health, mood, and mild cognitive impairment? *J Alzheimers Dis*. 2017;55(3):1183-1193. [FREE Full text] [doi: [10.3233/JAD-160611](https://doi.org/10.3233/JAD-160611)] [Medline: [27792011](https://pubmed.ncbi.nlm.nih.gov/27792011/)]
5. Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, et al. Practice guideline update summary: mild cognitive impairment: report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. *Neurology*. 2018;90(3):126-135. [FREE Full text] [doi: [10.1212/WNL.0000000000004826](https://doi.org/10.1212/WNL.0000000000004826)] [Medline: [29282327](https://pubmed.ncbi.nlm.nih.gov/29282327/)]
6. Petersen RC, Roberts RO, Knopman DS, Boeve BF, Geda YE, Ivnik RJ, et al. Mild cognitive impairment: ten years later. *Arch Neurol*. 2009;66(12):1447-1455. [FREE Full text] [doi: [10.1001/archneurol.2009.266](https://doi.org/10.1001/archneurol.2009.266)] [Medline: [20008648](https://pubmed.ncbi.nlm.nih.gov/20008648/)]
7. Karran E, De Strooper B. The amyloid hypothesis in Alzheimer disease: new insights from new therapeutics. *Nat Rev Drug Discov*. 2022;21(4):306-318. [doi: [10.1038/s41573-022-00391-w](https://doi.org/10.1038/s41573-022-00391-w)] [Medline: [35177833](https://pubmed.ncbi.nlm.nih.gov/35177833/)]
8. Liss JL, Seleri Assunção S, Cummings J, Atri A, Geldmacher DS, Candela SF, et al. Practical recommendations for timely, accurate diagnosis of symptomatic Alzheimer's disease (MCI and dementia) in primary care: a review and synthesis. *J Intern Med*. 2021;290(2):310-334. [FREE Full text] [doi: [10.1111/joim.13244](https://doi.org/10.1111/joim.13244)] [Medline: [33458891](https://pubmed.ncbi.nlm.nih.gov/33458891/)]
9. Tong T, Thokala P, McMillan B, Ghosh R, Brazier J. Cost effectiveness of using cognitive screening tests for detecting dementia and mild cognitive impairment in primary care. *Int J Geriatr Psychiatry*. 2017;32(12):1392-1400. [FREE Full text] [doi: [10.1002/gps.4626](https://doi.org/10.1002/gps.4626)] [Medline: [27874210](https://pubmed.ncbi.nlm.nih.gov/27874210/)]
10. Alharbi EA, Jones JM, Alomainy A. Non-invasive solutions to identify distinctions between healthy and mild cognitive impairments participants. *IEEE J Transl Eng Health Med*. 2022;10:2700206. [FREE Full text] [doi: [10.1109/JTEHM.2022.3175361](https://doi.org/10.1109/JTEHM.2022.3175361)] [Medline: [35711336](https://pubmed.ncbi.nlm.nih.gov/35711336/)]
11. Rasmussen KW, Salgado S, Daustrand M, Berntsen D. Using nostalgia films to stimulate spontaneous autobiographical remembering in Alzheimer's disease. *J Appl Res Mem Cogn*. 2021;10(3):400-411. [doi: [10.1016/j.jarmac.2020.11.001](https://doi.org/10.1016/j.jarmac.2020.11.001)]
12. Montero-Odasso M, Muir SW, Speechley M. Dual-task complexity affects gait in people with mild cognitive impairment: the interplay between gait variability, dual tasking, and risk of falls. *Arch Phys Med Rehabil*. 2012;93(2):293-299. [doi: [10.1016/j.apmr.2011.08.026](https://doi.org/10.1016/j.apmr.2011.08.026)] [Medline: [22289240](https://pubmed.ncbi.nlm.nih.gov/22289240/)]
13. Eggenberger P, Bürgisser M, Rossi RM, Annaheim S. Body temperature is associated with cognitive performance in older adults with and without mild cognitive impairment: a cross-sectional analysis. *Front Aging Neurosci*. 2021;13:585904. [FREE Full text] [doi: [10.3389/fnagi.2021.585904](https://doi.org/10.3389/fnagi.2021.585904)] [Medline: [33643019](https://pubmed.ncbi.nlm.nih.gov/33643019/)]
14. Yaffe K, Laffan AM, Harrison SL, Redline S, Spira AP, Ensrud KE, et al. Sleep-disordered breathing, hypoxia, and risk of mild cognitive impairment and dementia in older women. *JAMA*. 2011;306(6):613-619. [FREE Full text] [doi: [10.1001/jama.2011.1115](https://doi.org/10.1001/jama.2011.1115)] [Medline: [21828324](https://pubmed.ncbi.nlm.nih.gov/21828324/)]
15. Xue C, Li A, Wu R, Chai J, Qiang Y, Zhao J, et al. VRNPT: a neuropsychological test tool for diagnosing mild cognitive impairment using virtual reality and EEG signals. *Int J Hum-Comput Int*. 2023;40(20):6268-6286. [doi: [10.1080/10447318.2023.2250605](https://doi.org/10.1080/10447318.2023.2250605)]
16. Cheng S, Wang J, Sheng D, Chen Y. Identification with your mind: a hybrid bci-based authentication approach for anti-shoulder-surfing attacks using EEG and eye movement data. *IEEE Trans Instrum Meas*. 2023;72:1-14. [doi: [10.1109/tim.2023.3241081](https://doi.org/10.1109/tim.2023.3241081)]
17. Marinelli L, Trompetto C, Puce L, Monacelli F, Mori L, Serrati C, et al. Electromyographic patterns of paratonia in normal subjects and in patients with mild cognitive impairment or Alzheimer's disease. *J Alzheimers Dis*. 2022;87(3):1065-1077. [doi: [10.3233/JAD-215526](https://doi.org/10.3233/JAD-215526)] [Medline: [35431243](https://pubmed.ncbi.nlm.nih.gov/35431243/)]
18. Bhagya Shree SR, Sheshadri HS. Diagnosis of Alzheimer's disease using naive Bayesian classifier. *Neural Comput Applic*. 2016;29(1):123-132. [doi: [10.1007/s00521-016-2416-3](https://doi.org/10.1007/s00521-016-2416-3)]
19. Elgammal YM, Zahran MA, Abdelsalam MM. A new strategy for the early detection of Alzheimer disease stages using multifractal geometry analysis based on k-nearest neighbor algorithm. *Sci Rep*. 2022;12(1):22381. [FREE Full text] [doi: [10.1038/s41598-022-26958-6](https://doi.org/10.1038/s41598-022-26958-6)] [Medline: [36572791](https://pubmed.ncbi.nlm.nih.gov/36572791/)]
20. Syaifullah AH, Shiino A, Kitahara H, Ito R, Ishida M, Tanigaki K. Machine learning for diagnosis of AD and prediction of MCI progression from brain MRI using brain anatomical analysis using diffeomorphic deformation. *Front Neurol*. 2020;11:576029. [FREE Full text] [doi: [10.3389/fneur.2020.576029](https://doi.org/10.3389/fneur.2020.576029)] [Medline: [33613411](https://pubmed.ncbi.nlm.nih.gov/33613411/)]
21. Kato S, Homma A, Sakuma T, Nakamura M. Detection of mild Alzheimer's disease and mild cognitive impairment from elderly speech: binary discrimination using logistic regression. *Annu Int Conf IEEE Eng Med Biol Soc*. 2015;2015:5569-5572. [doi: [10.1109/EMBC.2015.7319654](https://doi.org/10.1109/EMBC.2015.7319654)] [Medline: [26737554](https://pubmed.ncbi.nlm.nih.gov/26737554/)]
22. Dai P, Gwadry-Sridhar F, Bauer M, Borrie M. Bagging ensembles for the diagnosis and prognostication of Alzheimer's disease. 2016. Presented at: AAI'16: Proceedings of the Thirtieth AAI Conference on Artificial Intelligence; 2016; Phoenix, AZ. [doi: [10.1609/aaai.v30i1.9915](https://doi.org/10.1609/aaai.v30i1.9915)]
23. Ianculescu M, Paraschiv E, Alexandru A. Addressing mild cognitive impairment and boosting wellness for the elderly through personalized remote monitoring. *Healthcare (Basel)*. 2022;10(7):1214. [FREE Full text] [doi: [10.3390/healthcare10071214](https://doi.org/10.3390/healthcare10071214)] [Medline: [35885741](https://pubmed.ncbi.nlm.nih.gov/35885741/)]
24. Khan A, Zubair S. Development of a three tiered cognitive hybrid machine learning algorithm for effective diagnosis of Alzheimer's disease. *J King Saud Univ Comput Inf Sci*. 2022;34(10):8000-8018. [doi: [10.1016/j.jksuci.2022.07.016](https://doi.org/10.1016/j.jksuci.2022.07.016)]

25. Moayedikia A, Ong K, Boo YL, Yeoh WG, Jensen R. Feature selection for high dimensional imbalanced class data using harmony search. *Eng Appl Artif Intell*. 2017;57:38-49. [doi: [10.1016/j.engappai.2016.10.008](https://doi.org/10.1016/j.engappai.2016.10.008)]
26. Esmin AAA, Coelho RA, Matwin S. A review on particle swarm optimization algorithm and its variants to clustering high-dimensional data. *Artif Intell Rev*. 2013;44(1):23-45. [doi: [10.1007/s10462-013-9400-4](https://doi.org/10.1007/s10462-013-9400-4)]
27. da S. Bohrer J, Dorn M. Enhancing classification with hybrid feature selection: a multi-objective genetic algorithm for high-dimensional data. *Expert Syst Appl*. 2024;255:124518. [doi: [10.1016/j.eswa.2024.124518](https://doi.org/10.1016/j.eswa.2024.124518)]
28. Polit DF, Beck CT. *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. United States. Lippincott Williams & Wilkins; 2016.
29. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimers Dement*. 2011;7(3):263-269. [FREE Full text] [doi: [10.1016/j.jalz.2011.03.005](https://doi.org/10.1016/j.jalz.2011.03.005)] [Medline: [21514250](https://pubmed.ncbi.nlm.nih.gov/21514250/)]
30. Chapman LJ, Chapman JP. The measurement of handedness. *Brain Cogn*. 1987;6(2):175-183. [doi: [10.1016/0278-2626\(87\)90118-7](https://doi.org/10.1016/0278-2626(87)90118-7)] [Medline: [3593557](https://pubmed.ncbi.nlm.nih.gov/3593557/)]
31. Li A, Xue C, Wu R, Wu W, Zhao J, Qiang Y. Unearthing subtle cognitive variations: a digital screening tool for detecting and monitoring mild cognitive impairment. *Int J Hum-Comput Interact*. 2024;1-21. [doi: [10.1080/10447318.2024.2327179](https://doi.org/10.1080/10447318.2024.2327179)]
32. Rafi H, Benezeth Y, Reynaud P, Arnoux E, Song FY, Demonceaux C. Personalization of AI models based on federated learning for driver stress monitoring. 2022. Presented at: Computer Vision – ECCV 2022 Workshops; October 23-27, 2022:575-585; Tel Aviv, Israel.
33. McKinley S, Levine M. Cubic spline interpolation. *College of the Redwoods*. 1998;45(1):1049-1060.
34. Smital L, Haider CR, Vitek M, Leinveber P, Jurak P, Nemcova A, et al. Real-time quality assessment of long-term ECG signals recorded by wearables in free-living conditions. *IEEE Trans Biomed Eng*. 2020;67(10):2721-2734. [doi: [10.1109/TBME.2020.2969719](https://doi.org/10.1109/TBME.2020.2969719)] [Medline: [31995473](https://pubmed.ncbi.nlm.nih.gov/31995473/)]
35. Larradet F, Niewiadomski R, Barresi G, Caldwell DG, Mattos LS. Toward emotion recognition from physiological signals in the wild: approaching the methodological issues in real-life data collection. *Front Psychol*. 2020;11:1111. [FREE Full text] [doi: [10.3389/fpsyg.2020.01111](https://doi.org/10.3389/fpsyg.2020.01111)] [Medline: [32760305](https://pubmed.ncbi.nlm.nih.gov/32760305/)]
36. Kalimeri K, Saitis C. Exploring multimodal biosignal features for stress detection during indoor mobility. 2016. Presented at: Proceedings of the 18th ACM International Conference on Multimodal Interaction; 2016 October 31:53-60; Tokyo, Japan.
37. Taylor S, Jaques N, Chen W, Fedor S, Sano A, Picard R. Automatic identification of artifacts in electrodermal activity data. *Annu Int Conf IEEE Eng Med Biol Soc*. 2015;2015:1934-1937. [FREE Full text] [doi: [10.1109/EMBC.2015.7318762](https://doi.org/10.1109/EMBC.2015.7318762)] [Medline: [26736662](https://pubmed.ncbi.nlm.nih.gov/26736662/)]
38. Greco A, Valenza G, Lanata A, Scilingo E, Citi L. cvxEDA: a convex optimization approach to electrodermal activity processing. *IEEE Trans Biomed Eng*. 2016;63(4):797-804. [doi: [10.1109/TBME.2015.2474131](https://doi.org/10.1109/TBME.2015.2474131)] [Medline: [26336110](https://pubmed.ncbi.nlm.nih.gov/26336110/)]
39. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digit Med*. 2020;3:18. [FREE Full text] [doi: [10.1038/s41746-020-0226-6](https://doi.org/10.1038/s41746-020-0226-6)] [Medline: [32047863](https://pubmed.ncbi.nlm.nih.gov/32047863/)]
40. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *NPJ Digit Med*. 2020;3:90. [FREE Full text] [doi: [10.1038/s41746-020-0297-4](https://doi.org/10.1038/s41746-020-0297-4)] [Medline: [32613085](https://pubmed.ncbi.nlm.nih.gov/32613085/)]
41. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, et al. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur Heart J*. 1996;17(3):354-381. [Medline: [8737210](https://pubmed.ncbi.nlm.nih.gov/8737210/)]
42. Acar B, Savelieva I, Hemingway H, Malik M. Automatic ectopic beat elimination in short-term heart rate variability measurement. *Comput Methods Programs Biomed*. 2000;63(2):123-131. [doi: [10.1016/s0169-2607\(00\)00081-x](https://doi.org/10.1016/s0169-2607(00)00081-x)] [Medline: [10960745](https://pubmed.ncbi.nlm.nih.gov/10960745/)]
43. Karlsson M, Hörnsten R, Rydberg A, Wiklund U. Automatic filtering of outliers in RR intervals before analysis of heart rate variability in Holter recordings: a comparison with carefully edited data. *Biomed Eng Online*. 2012;11:2. [FREE Full text] [doi: [10.1186/1475-925X-11-2](https://doi.org/10.1186/1475-925X-11-2)] [Medline: [22236441](https://pubmed.ncbi.nlm.nih.gov/22236441/)]
44. Föll S, Maritsch M, Spinola F, Mishra V, Barata F, Kowatsch T, et al. FLIRT: a feature generation toolkit for wearable data. *Comput Methods Programs Biomed*. 2021;212:106461. [FREE Full text] [doi: [10.1016/j.cmpb.2021.106461](https://doi.org/10.1016/j.cmpb.2021.106461)] [Medline: [34736174](https://pubmed.ncbi.nlm.nih.gov/34736174/)]
45. Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, et al. NeuroKit2: a Python toolbox for neurophysiological signal processing. *Behav Res Methods*. 2021;53(4):1689-1696. [doi: [10.3758/s13428-020-01516-y](https://doi.org/10.3758/s13428-020-01516-y)] [Medline: [33528817](https://pubmed.ncbi.nlm.nih.gov/33528817/)]
46. Arsalan A, Majid M. Human stress classification during public speaking using physiological signals. *Comput Biol Med*. 2021;133:104377. [doi: [10.1016/j.combiomed.2021.104377](https://doi.org/10.1016/j.combiomed.2021.104377)] [Medline: [33866254](https://pubmed.ncbi.nlm.nih.gov/33866254/)]
47. Tanev G, Saadi DB, Hoppe K, Sorensen HBD. Classification of acute stress using linear and non-linear heart rate variability analysis derived from sternal ECG. *Annu Int Conf IEEE Eng Med Biol Soc*. 2014;2014:3386-3389. [doi: [10.1109/EMBC.2014.6944349](https://doi.org/10.1109/EMBC.2014.6944349)] [Medline: [25570717](https://pubmed.ncbi.nlm.nih.gov/25570717/)]

48. Saif N, Yan P, Niotis K, Scheyer O, Rahman A, Berkowitz M, et al. Feasibility of using a wearable biosensor device in patients at risk for Alzheimer's disease dementia. *J Prev Alzheimers Dis.* 2020;7(2):104-111. [FREE Full text] [doi: [10.14283/jpad.2019.39](https://doi.org/10.14283/jpad.2019.39)] [Medline: [32236399](https://pubmed.ncbi.nlm.nih.gov/32236399/)]
49. Forte G, Favieri F, Casagrande M. Heart rate variability and cognitive function: a systematic review. *Front Neurosci.* 2019;13:710. [FREE Full text] [doi: [10.3389/fnins.2019.00710](https://doi.org/10.3389/fnins.2019.00710)] [Medline: [31354419](https://pubmed.ncbi.nlm.nih.gov/31354419/)]
50. Rasmussen KW, Salgado S, Daustrand M, Berntsen D. Using nostalgia films to stimulate spontaneous autobiographical remembering in Alzheimer's disease. *J Appl Res Mem Cogn.* 2021;10(3):400-411. [doi: [10.1016/j.jarmac.2020.11.001](https://doi.org/10.1016/j.jarmac.2020.11.001)]

## Abbreviations

**AD:** Alzheimer disease

**AUC:** area under the curve

**DAELF-HSI:** Dynamic Adaptive Ensemble Learning Framework based on an Improved Harmony Search

**DCA:** decision curve analysis

**DT:** decision tree

**EDA:** electrodermal activity

**HC:** healthy control

**HRV:** heart rate variability

**IBI:** interbeat interval

**KNN:** k-nearest neighbors

**LF/HF:** low-frequency to high-frequency ratio

**LR:** logistic regression

**MCI:** mild cognitive impairment

**MLP:** multilayer perceptron

**MoCA:** Montreal Cognitive Assessment

**PNN50:** percentage of successive R-R intervals > 50 ms

**RF:** random forest

**RMSSD:** root mean square of successive differences

**SampEn:** sample entropy

**SCL:** skin conductance level

**SCR:** skin conductance response

**SD2/SD1:** ratio of the SD2 and SD1 of Poincaré plot

**SDNN:** SD of N-N intervals

**SVM:** support vector machine

*Edited by C Lovis; submitted 06.05.24; peer-reviewed by L Yao, N Xiao, Y Huusi, T Ramsey; comments to author 25.07.24; revised version received 12.08.24; accepted 15.11.24; published 20.01.25*

*Please cite as:*

*Li A, Li J, Hu Y, Geng Y, Qiang Y, Zhao J*

*A Dynamic Adaptive Ensemble Learning Framework for Noninvasive Mild Cognitive Impairment Detection: Development and Validation Study*

*JMIR Med Inform 2025;13:e60250*

*URL: <https://medinform.jmir.org/2025/1/e60250>*

*doi: [10.2196/60250](https://doi.org/10.2196/60250)*

*PMID:*

©Aoyu Li, Jingwen Li, Yishan Hu, Yan Geng, Yan Qiang, Juanjuan Zhao. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org/>), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.