Early mild cognitive impairment detection using cognitive-motor tasks and machine learning

1st Siwar Chaabene
MIRACL, University of sfax
Sfax, Tunisia
siwarchaabene@gmail.com
0000-0002-6458-9569

2nd Bassem Bouaziz
MIRACL, University of sfax
Sfax, Tunisia
bassem.bouaziz@isims.usf.tn
0000-0002-3692-9482

3rd Amal Boudaya
MIRACL, University of sfax
Sfax, Tunisia
amal.boudaya@gmail.com
0000-0003-0486-7923

4th Lotfi Chaari
University of Toulouse, IRIT
Toulouse, France
lotfi.chaari@toulouse-inp.fr
0000-0002-3590-0370

5th Anita Hökelmann
Otto-von-Guericke University
Magdeburg, Germany
anita.hoekelmann@ovgu.de
0000-0003-1383-9318

2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA) | 979-8-3503-3890-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/INISTA59065.2023.10310653

***Abstract*—Mild cognitive impairment (MCI) is a condition marked by impairment in one or more cognitive areas, but not necessarily all of them. It is frequently referred to as the stage between typical age-related cognitive decline and dementia. Recent studies had focused on different modalities to assess disorders such as dementia and Alzheimer's disease (AD). Heart rate variability (HRV) stands out among them as having the potential to identify MCI. In this paper, we propose a new MCI detection method using HRV signals. MCI patients were compared to age-matched healthy controls (HC) for the effect of performing additional cognitive and postural tasks. Twenty-four participants were enrolled to complete three tasks: a postural balance master task, two cognitive tasks called CERAD+ and Neurotrack, and baseline. HRV data were recorded during these experiments. Six machine learning (ML) models were examined for task classification including k-Nearest Neighbors, Decision tree, Random Forest, Extra Trees, Gradient Boosting, and XGBoost. To avoid over-fitting, cross-validation (CV) was employed to assess how well the built models performed. To boost accuracy, a voting ensemble classifier model is developed that combines the top ML models with the highest accuracy rates. The findings of this study demonstrated that MCI might be diagnosed with ML classifiers utilizing HRV signals, particularly when postural and cognitive functions are taken into account.**

***Index Terms*—MCI detection, HRV, healthy aging, Neurotrack, CERAD+, balance master, ML.**

I. Introduction

The prodromal stage of a condition that worsens and causes dementia and Alzheimer's disease (AD) is mild cognitive impairment (MCI). MCI is a condition that does not yet significantly affect a person's ability to carry out daily activities independently. MCI can be divided into amnestic (aMCI) and non-amnestic (nMCI) subtypes. The primary characteristic of aMCI is memory loss, and it is more likely to be an early sign of AD. On the other side, nMCI stands for non-memory domains, which is a decrease that may be caused by a variety of factors. Early monitoring and diagnosis of MCI may be on the horizon as a result of considerable advances in scientific study, which may be most useful in the early stages of the illness. Correspondingly, systems with high accuracy rates are increasingly needed to help with the early detection of MCI. Nearly 5-15% of MCI patients acquire dementia within one year, placing them at elevated risk for dementia development [1]. Moreover, older persons with MCI have been found to exhibit postural instability and balance issues. Falls are more common among elderly people with MCI than among older people who are cognitively well [2]. As a result, early MCI identification can be extremely important for early intervention, prevention, and the right therapies. The potential of ML to extract characteristics has great promise for aiding in illness. In particular, more focus has been placed in recent years on ML systems to assist in the diagnosis of MCI. Most works concentrate on neuroimaging such as magnetic resonance imaging (MRI) [3], Fluorodeoxyglucose-Positron Emission Tomography (FDG-PET) [4], and single-photon emission computerized tomography (SPECT) [5] as well as biomarker analysis [6]. Recently, non-invasive diagnostic techniques based on wearable technologies, such as EEG and HRV, have come to greater attention in MCI research. According to [7], the authors suggested employing the wearable CorSense device to measure HRV signals to identify differences between healthy and MCI participants based on statistical analysis. On the other side, cognitive and postural (i.e., motor) tasks are frequently employed in research investigations. These tasks may be more sensitive in identifying very specific alterations in brain function brought on by conditions including AD and MCI. Recently, studies [8], [9] have demonstrated that data acquisition with specific cognitive tasks may include helpful information for distinguishing AD and MCI patients. Furthermore, recent work [10] has used cognitive-motor dual-task paradigms to examine more subtle alterations in postural control linked to cognitive decline in older people in both healthy and MCI-affected individuals. To this regard, HRV data related to cognitive-motor tasks may provide relevant information that allows ML classifiers to classify MCI from HC patients. It is therefore worth looking into whether additional cognitive and postural tasks while HRV acquisition might increase ML-based classification accuracy in MCI detection. In this paper, we suggest a new method for MCI identification that makes use of baseline, postural, and cognitive HRV signals. The purpose of this study is to evaluate the classification performance of ML employing HRV signals to identify MCI patients from healthy controls at baseline, during cognitive and postural tasks.

The remainder of this paper is as follows. Section II presents the background of HRV analysis. Section III details the proposed method. Section IV presents the results of the experiment and data analysis. In Section V, the findings are analyzed and interpreted, and the discussion concludes with recommendations for future research.

II. HRV Analysis

Based on portable sensor technology, HRV is frequently used in scientific and clinical studies to detect various anomalies. HRV is the interval variation between successive heartbeats (RR-intervals), which are known as interbeat intervals (IBIs). It may be analyzed in the time, frequency, and non-linear domains. Table I indicted the most often used indices for HRV analysis. Besides, the analysis may be performed with either a long-term (LT) duration lasting 24 hours or a short-term (ST) duration lasting 5-minutes (min). It is recommended to make recordings that last 5-min or more for HRV quantitative analysis. On the other hand, recent studies [11], [12] have shown that employing HRV characteristics in an ultra-short-term (UST) recording ($<$ 5-Min) makes HRV indices accurate for recordings under that time. Theoretically, it should be understood that several HRV metrics become insignificant if calculated in a UTS period [13]. The time, frequency, and non-linear domains indices used to describe LT, ST, and UST HRV recordings are the same, although their predictive abilities many vary [14]. Uncertainty still exists regarding the minimum time epoch at which all HRV features domains may be accurately documented. The next subsections discuss the HRV features and the relationship between epoch required to estimate recording values.

A. Temporal features

The degree of variability in measures of the IBI is quantified using time-domain HRV indices which are produced from straightforward statistical computations. HRV data collected during intervals ranging from 1-min to more than 24 hours is often assessed by HRV time-domain indices. Given that multiple time-domain HRV indices have been proven to be accurate, recent studies [11], [12] have advocated the use of ultra-short recordings. In addition to offering suitable minimum ST measurement intervals, the authors in [12] also put out recommendations for ultra-short measurement intervals.

TABLE I: HRV features.

Metric	Unit	Description
Time domain features		
MeanHR	ms	Mean heart Rate
MaxHR	ms	Max heart rate
MinHR	ms	Min heart rate
STDHR	ms	Standard deviation (STD) of heart rate
MeanRR	ms	Mean of RR-intervals
MedianRR	ms	Median Absolute values of the successive differences between the RR-intervals
RangeRR	ms	Difference between the Max and Min RR-interval
SDRR	ms	STD of RR-intervals
SDSD		STD of differences between adjacent RR-intervals
RMSSD	ms	Root mean square successive RR-interval differences
CVSD	%	Coefficient of variation of successive differences equal to the RMSSD divided by MedianRR
CVRR	%	Coefficient of variation equal to the ratio of SDRR divided by MeanRR
HTI		Integral of the density distribution divided by the maximum of the density distribution
Frequency domain features		
LF	ms^2	Absolute power(AP) of low-frequency(.04 to .15Hz)
HF	ms^2	AP of high-frequency (.15 to .40 Hz)
VLF	ms^2	AP of very-low-frequency (.003 to .04 Hz)
LF/HF	ms^2	Ratio of LF-to-HF power
TotalPower	%	Total power density spectral
LFNU	nu	Normalized LF power
HFNU	nu	Normalized HF power
Non linear domain features		
SD1	ms	Poincaré plot STD perpendicular the line of identity
SD2	ms	Poincaré plot STD along the line of identity
SD2/SD1	%	Ratio between SD2 and SD1
CSI		Cardiac Sympathetic Index
CVI		Cadiac Vagal Index
MCVI		Modified CVI
SampEn		Sample entropy of the data

The time indices include statistical measures such as MeanRR, SDNN, RMSSD, CVNNI, MaxHR, as well as geometric measures such as HTI, (see Table I). According to the literature, these metrics are often employed between 1 and 5-min epochs. Some of these metrics are calculated based on UST recordings.

B. Frequency features

Using Fast Fourier Transformation (FFT) or autoregressive (AR) modeling, we may separate HRV into ultra-low-frequency (ULF), very-low-frequency (VLF), low-frequency (LF), and high-frequency (HF) bands. According to the literature, frequency characteristics must be calculated using a sample of 1 to 5-min. Besides, among these metrics (per Table I) can be used in UST recording epochs without providing any loss information [12].

C. Non linear features

Non-linear features allow us to assess a time series' unpredictability. The most commonly used non-linear features are the Poincare plot and Cardiac Sympathetic Index (CSI)& Cadiac Vagal Index (CVI), which may be measured on ST and UST recordings [12]. A common nonlinear metric for assessing the complexity of any time series is sample entropy or SampEn. This metric is used in HRV analysis to determine the rate of entropy in the ST [12].

III. Materials and Methods

This section describes the participants criteria, the experimental paradigm, physiological measurement, signal processing, feature extraction, feature selection, and classification. Figure 1 illustrates the pipeline of the proposed MCI detection method.

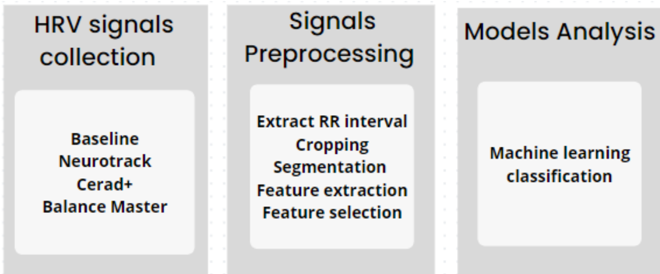

Fig. 1: Pipeline of the proposed MCI detection method.

A. Participants

Experiments were conducted in 2021 at the Otto von Guericke University in Magdeburg, Germany. A total of 24 participants—13 MCI patients and 11 healthy controls (HC)—were involved in the research. A qualified neurologist identified the individuals as having MCI. Subject ages ranged from 65 to 85. There are hardly any age, gender, or educational disparities between the MCI and HC groups. The demographic details of the participants are shown in Table II.

TABLE II: Demographic information of all subjects.

Characteristic	HC	MCI
Ages (years)	72±6	72.75±6
Gender (M/F)	1/10	8/5
Education (years)	14.85±1.63	14.25±1.66

M: Male and F: Female.

B. Experimental paradigm

All participants underwent the following assessments: (i) a clinical evaluation; (ii) socioeconomic assessment questionnaires; and (iii) neurophysiological and neuropsychological assessments using HRV and EEG modalities, including resting, cognitive tasks using (i) Neurotrack, which includes the following cognitive domains: attention, processing speed, memory, associative learning, inhibition, and executive function and (ii) the Consortium to Establish a Registry for AD+ (CERAD+) [15] which includes the following cognitive domains: visual perception, learning ability, memory, language, in addition to executive function, and attention tests, as well as a postural task using (iii) NeuroCom® Smart Balance Master which was used to measure postural balance, detect gait issues, and assess fall risk across the board in seniors. The task includes a dynamic force plate and a visual surround that can be moved independently, as well as an overhead connection for a safety harness strap, and a computer with software.
The EEG and HRV data are gathered at the beginning of the experiment when the subjects are at 5-min rest. The following 20 to 30-min are spent by participants utilizing the Neurotrack test, followed by a 5-min rest. After that, participants complete 20 to 30-min of CERAD+ assessments before taking a 5-min rest. Finally, participants complete 10 to 15-min of the Balance Master task. In this work, we focus only on HRV recording while subjects do the rest, Neurotrack, CERAD-Plus and balance Master.

C. Data Measurement

HRV signals were recorded with a Polar H10 wearable device. These signals were sampled at a frequency of 10Hz. The flow of the experiment is shown in Figure 2. The recorded signals were separated into six segments: Rest1 (5-min), Neurotrack (15 to 20-min), Rest2 (5-min), Cerad+ (20 to 30-min), Rest3 (5-min), and Balance Master (15-min). The time for Rests 1, 2, and 3 were merged to create a longer segment of (15-min) for the same assessment. We extract the RR interval from the raw HR signals as follows: RR interval (sec) = 60/HR [16]. Quantitative data analysis for each HRV recording was done at 1-min intervals (i.e., window time) during Rests, Neurotrack, CERAD+, and Balance Master tasks.

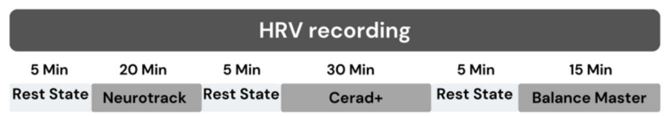

Fig. 2: Experiment flow.

D. Feature Extraction and Selection

The used 27 characteristics are summarized in Table I. Using the features derived from time, frequency, and nonlinear spaces, is expected to help retrieving high separability between two classes (i.e., Healthy people and people at MCI). However, some redundancies may be involved in the retained feature vectors. Moreover, the computational time necessary to fit high-dimensional classifiers may also increase. The models generalization capacity may also decrease. This highlights the significance of choosing the best discriminative features when building classifiers. A univariate feature selection (FS) method based on mutual information (MI) is used in the feature space to find the most discriminant features for classification. This method is a well-known pertinent criterion for choosing feature subsets from the input dataset. In practice, the MI feature selection method was applied to find key aspects of a classification challenge. In this regard, we have computed the MI between our extracted HRV feature and the class label in our dataset based on the entropy estimation technique. Following that, we sort these features according to their MI rankings. Higher MI levels reflect a stronger link between the feature and the target variable, implying that the feature contains more discriminative information. Finally, we choose how many top-ranked features we want to keep for the classification step.

E. Machine Learning Analysis

For classification tasks, there are several supervised ML techniques. However, aggregation of many classifiers (ensemble learning) is also an efficient strategy that has been considered in many studies [17]. Building an ensemble model may be done in five different ways including (i) *Bagging*, where the same kind of classifiers are employed and vote to determine the result, (ii) *Boosting* where a succession of classifiers is used, so that the performance of one model influences the performance of the next, (iii) *stacking*, where different models are mixed to provide a more accurate prediction by ”stacking” individual predictions, (v) *Voting Classifier*, which incorporates many classifiers and bases its conclusion on votes, and (vi) *Averaging*, where predictions of various models are included and averaged. In this work, different ML models are used: k-nearest neighbors (KNN) [18], Decision Tree (DT) [19], Random Forest (RF) [20], Extra Trees (ExT) [21], Gradient Boosting (GB) [22], XGBoost (XGB) [23] and Voting classifier (Vhard) [24].

IV. Experimental validation

A. Evaluation metrics

The model’s classification performance is first assessed using the confusion matrix [25]. As shown in Table III, the matrix’s rows and columns each correspond to occurrences of the actual class and the predicted class, respectively.

TABLE III: An example of a two-classes (MCI and HC) confusion matrix.

Actual \ **Predicted**	*MCI*	*HC*
MCI	True Positives (TP)	False Positives (FP)
HC	False Negatives (FN)	True Negatives (TN)

where True Positive (TP) denotes predictions of MCI as MCI, False Positive (FP) denotes predictions of HC as MCI, False Negative (FN) denotes predictions of MCI as HC, and True Negative (TN) denotes predictions of HC as HC. These metrics are evaluated, leading to accuracy, precision, recall and F1-scores.

- Accuracy (ACC): proportion of correct predictions to the total predictions given by:

$$ACC = (TP + TN)/(TP + TN + FP + FN). \quad (1)$$

- Precision (PRE): proportion of predicted positive cases that were correct, defined as:

$$PRE = TP/(TP + FP). \quad (2)$$

- Recall (REC) also known as sensitivity: proportion of true positives to the total positives, defined as:

$$REC = TP/(TP + FN). \quad (3)$$

- F1-score (F1s): combines the precision and recall scores to count the number of times a model correctly predicted the whole dataset, given by:

$$F1s = (2 \times PRE \times REC)/(PRE + REC). \quad (4)$$

Furthermore, to assess the effectiveness of the ML models and avoid over-fitting, cross-validation (CV) with K-Fold [26] was performed. The data was split into 10 separate folds. In each fold, 90% of the data were utilized for training and 10% for testing. All proposed ML models are trained, and performance metrics are shown per fold for each split of the data. This guarantees that all data is used to train and test the model. It is averaged over all folds for assessment reasons.

B. MCI detection results

To identify MCI patients, six supervised ML algorithms were employed to evaluate the performance of cognitive and postural activities. Our data was segmented, 10-fold CV was performed, separate models were generated and evaluated, and the higher-performing models were then integrated using the voting ensemble approach. The present section introduces the study’s key findings, including selected features from Baseline, Neurotrack, CERAD+, and Balance Master tasks, as well as the results of the individual and ensemble models of the HRV throughout each task. The results are then compared to other recent works that have been documented in the literature.

1) Selected Features: The effectiveness of learning approaches is increased by choosing relevant characteristics for ML models. The most crucial features for each specific recording were obtained using the MI feature selection method. Table IV illustrates the 10 most important features for each recording task. We can see that some features as MinHR, MaxHR, MeanHR, MediamRR, MeanRR and CVSD were shared by the different used tasks. Those features are outlined in bold font in Table IV. It indicated that a collection of features may have a high likelihood of accurately predicting the MCI vs HC classification problem’s solution.

TABLE IV: 10 best selected features for each task.

Task	**Selected features**
Baseline	**MinHR**, **MaxHR**, **MeanHR**, VLF, RangeRR, CVRR, **MedianRR**, **meanRR**, CVI, **CVSD**
Neurotrack	**MedianRR**, **MinHR**, **MeanHR**, **MaxHR**, **MeanRR**, VLF, HTI, **CVSD**, TotalPower, sampEn
Cerad+	**MinHR**, **MeanRR**, **MeanHR**, **MedianRR**, **MaxHR**, RangeRR, HTI, CVRR, MCVI,**CVSD**
Balance Master	VLF, **MeanHR**, **MeanRR**, **maxHR**, **MedianRR**, TotalPower, **CVSD**, HF, RangeRR, **minHR**

2) Cross-validation results of ML-based classifiers: The effectiveness of the built-in ML models was assessed using 10-fold CV to avoid over-fitting. The MCI vs HC categorization was compared using both all features and just the selected features for each task. In terms of average accuracy, Table V illustrates the 10-fold CV classification performance for the six ML models for each task with and without features selection. According to classification without the FS method, the boosting classifiers, including GB and XGB, offer the maximum level of accuracy for the proposed cognitive and postural tasks. Furthermore, the balance Master task has the best classifier’s highest percentage at 78.93% by using the GB model. According to classification involving a feature selection step, it can be noticed that the classification accuracy increases for all ML models. Furthermore, it should be mentioned that the GB, XGB, and ExT classifiers outperform all the other models, with the maximum average accuracy obtained in the classification by GB being 81.40% during the Balance Master task. Further, we discovered that the chosen characteristics were quite effective in classifying the desired value. However, the integration of time and frequency domain information with the use of nonlinear metrics might potentially increase the accuracy of HRV MCI classification. Based on the findings of the 10-fold CV of ML classifiers with the feature selection method, we propose a voting classifier approach to increase accuracy. The voting classifier strategy incorporates the ML algorithms RF, ExT, GB, and XGB, which deliver the highest accuracy rates in comparison to other ML algorithms. These results are presented in the following subsection.

TABLE V: 10-fold CV accuracy results of ML classifiers with and without feature selection (FS).

		KNN	*DT*	*RF*	*ExT*	*GB*	*XGB*
Accuracy (%)							
Baseline	without FS	59.38	60.97	68.55	***69.08***	67.73	68.01
	with FS	66.33	62.90	**69.08**	69.73	68.75	69.07
Neurotrack	without FS	66.6	68.15	74.05	73.12	73.33	***74.98***
	with FS	71.88	70.18	75.28	**77.66**	74.15	75.84
CERAD+	without FS	64.86	67.25	69.79	69.96	***70.92***	70.04
	with FS	65.49	69.31	70.26	68.99	**71.85**	70.91
Balance Master	without FS	56.65	75.7	75.28	76.96	***78.93***	75.3
	with FS	77.35	77.73	80.98	78.13	**81.40**	78.16

3) Final classifier: Our suggested ensemble learning models included four ML models: RF, ExT, GB, and XGB. A hard voting ensemble classifier for early MCI detection is presented in this study. The suggested voting model's classification results are compared with those of the other four classification models. However, the initial dataset was split into two sets for each experiment, with training including 80% of the subjects, testing comprising 20%. Table VI illustrates the classification performance results for MCI vs. HC classification using the HRV signal during various tasks. We can notice that all ML models showed the greatest similarity in classification scores when postural and cognitive tasks were compared to the resting state. In particular, the proposed voting method achieved the highest classification accuracy of 86%, 82.07% and 78.57% during respectively the Balance Master, Neurotrack, and Cerad+ tasks. Using a number of evaluation metrics for cognitive and postural activities, the suggested Vhard model performs better than the individual models.

4) Comparison: In [7], the authors developed an logistic regression (LR) model to distinguish between HC and MCI participants during cognitive task. The features used in this work are SDRR and RMSSD from the time domain and HF from the frequency domain. The accuracy obtained by this work is 76.5%. We applied this technique to our HRV data, and for the prediction challenge, we employed the proposed Vhard approach. According to [27], two ML algorithms, including gradient boosting decision tree (GBDT) and XGBoost, were used to classify HC and MCI participants based on HRV signals. MeanHR, MeanRR, RMSSD, LF, HF, and LF/HF are the most relevant HRV characteristics employed in this work. The highest F1 values for both GBDT and XGBoost are, respectively, 84.04% and 78.02%. Table VII displays the accuracy values of the testing set using the competing method. Ten of the 27 features are only applied with the proposed technique. As a result, the findings in this table are consistent with the above-mentioned better performance even with a subset of features.

TABLE VI: Performance Metrics.

task	Metrics	*RF*	*ExT*	*GB*	*XGB*	*Vhard*
Baseline	Accuracy (%)	70.66	66.66	69.33	77.33	72
	Precision (%)	69.91	65.46	68.31	76.75	71.16
	Recall (%)	70.23	64.91	67.66	77.34	70.41
	F1-score (%)	70.02	65.07	67.87	76.92	70.66
Neurotrack	Accuracy (%)	77.35	77.35	79.24	80.18	**82.07**
	Precision (%)	78.32	77.75	79.43	81.37	**84.47**
	Recall (%)	74.30	74.67	77	77.42	**79.01**
	F1-score (%)	75.08	75.38	77.68	78.33	**80.15**
Cerad+	Accuracy (%)	74.60	68.25	76.98	78.57	**78.57**
	Precision (%)	73.92	67.5	76.53	78.32	**78.17**
	Recall (%)	74.37	67.82	77.26	79.18	**77.18**
	F1-score (%)	74.07	67.59	76.65	78.34	**77.54**
Balance Master	Accuracy (%)	84	86	80	76	**86**
	Precision (%)	84	85.89	79.87	76.02	**86.66**
	Recall (%)	84.21	86.07	79.87	75.52	**85.42**
	F1-score (%)	83.97	85.94	79.87	75.64	**85.72**

TABLE VII: Comparison of accuracy with recent related work.

Method	Baseline	Neurotrack	Cerad+	Balance Master
Alharbi et al. [7]	64%	64.28%	55.66%	58%
Liu et al. [27]	65.33%	71.42%	73.58%	78%
Pro. Method	**72**%	**82.07**%	**78.57**%	**86**%

Based on the above findings, we may conclude that using cognitive and postural tasks yields better outcomes than the baseline to detect MCI using HRV signals. On the other hand, the proposed cognitive and postural activities may be added into HRV recording, to determine MCI. This finding warrants further investigation into developing cognitive and postural tasks for MCI. Furthermore, these findings illustrate the efficiency of the MI feature selection technique employed in our suggested MCI detection method in comparison to other current methods.

V. Discussion and Conclusion

This study aimed to compare individuals with MCI to healthy older adults to assess the effects of cognitive and postural activities. Additionally, our research aimed to explore the potential of a wearable Polar H10 device recording for assessing and distinguishing between MCI and HC patients based on HRV. In the current study, we investigated the use of HRV to extract potential physiological biomarker for identifying subjects who are more likely to develop MCI. We extracted then HRV characteristics from various domains and identified the optimal combination of features before selecting the best ML algorithms for MCI identification. We found that time features, as well as some frequency and nonlinear features, were the most significant risk factors for MCI identification. After that, a classification model was created using various ML algorithms, including KNN, DT, and LR, GB, ExT, and XGB. Classification tests were conducted using both whole and selected feature sets. The experimental findings show that by using a FS approach, we can properly identify MCI with a limited set of characteristics. Moreover, the top ML models with the highest accuracy rates are combined into a voting ensemble classifier model, which allowed to increase accuracy rate. In future work, we will attempt to increase prediction accuracy using deep learning (DL) approaches to find the most practical model for MCI detection. Furthermore, we will subdivide each task (i.e., Cerad+, Neurotrack, and Balance Master) into its sub-tests to avoid those that might decrease the accuracy rate as well as to provide an ultra-short term recording for MCI detection. We will also attempt to employ a variety of FS techniques in search of HRV feature subsets that accurately predict MCI.

Acknowledgment

This work is a part of the *ICT-rollator* project (Innovative ICT-based Rollator to Promote Active and Healthy Aging in Preventive and Therapeutic Conditions) funded by the German Ministry of Education and Research (BMBF Project No. 01DH21023A) and the Tunisian Ministry of Higher Education and Research.

References

[1] R. A Dunne, D. Aarsland, J. T O'Brien, C. Ballard, S. Banerjee, N. C Fox, J. D Isaacs, B. R Underwood, R. J Perry, D. Chan, T. Dening, A. J Thomas, J. Schryer, A.-M. Jones, A. R Evans, Ch. Alessi, E. J Coulthard, J. Pickett, P. Elton, R. W Jones, S. Mitchell, N. Hooper, C. Kalafatis, J. G C Rasmussen, H. Martin, J. M Schott, and A. Burns. Mild Cognitive Impairment: the Manchester consensus. *Age and Ageing*, 50(1):72–80, November 2020.

[2] S. Tyrovolas, A. Koyanagi, E. Lara, Z. Ivan Santini, and J. M. Haro. Mild cognitive impairment is associated with falls among older adults: Findings from the irish longitudinal study on ageing (TILDA). *Experimental Gerontology*, vol. 75:pp. 42–47, 2016.

[3] M. Pihlajamaki, A. Jauhiainen, and H. Soininen. Structural and functional MRI in mild cognitive impairment. *Current Alzheimer Research*, vol. 6:pp. 179–185, April 2009.

[4] M. Silveira and J. Marques. Boosting alzheimer disease diagnosis using PET images. pages 2556–2559. IEEE, August 2010.

[5] K. Hirao, T. Ohnishi, Y. Hirata, F. Yamashita, T. Mori, Y. Moriguchi, H. Matsuda, K. Nemoto, E. Imabayashi, M. Yamada, T. Iwamoto, K. Arima, and T. Asada. The prediction of rapid conversion to Alzheimer's disease in mild cognitive impairment using regional cerebral blood flow SPECT. *NeuroImage*, vol. 28:pp. 1014–1021, December 2005.

[6] R. Santangelo, F. Masserini, F. Agosta, A. Sala, S. P. Caminiti, G. Cecchetti, F. Caso, V. Martinelli, P. Pinto, G. Passerini, D. Perani, G. Magnani, and M. Filippi. CSF p-tau/AB42 ratio and brain FDG-PET may reliably detect MCI "imminent" converters to AD. *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47:pp. 3152–3164, December 2020.

[7] Alharbi EA., Jones JM., and Alomainy A. Non-invasive solutions to identify distinctions between healthy and mild cognitive impairments participants. *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 16:pp. 2700206, May 2022.

[8] Y. Hong, R. L. Alvarado, A. Jog, D. N. Greve, and D. H. Salat. Serial reaction time task performance in older adults with neuropsychologically defined mild cognitive impairment. *Journal of Alzheimer's Disease*, vol. 74(2):pp. 491—500, 2020.

[9] S. Valladares-Rodriguez, M. J. Fernández-Iglesias, L. Anido-Rifón, D. Facal, C. Rivas-Costa, and R. Pérez-Rodríguez. Touchscreen games to detect cognitive impairment in senior adults. A user-interaction pilot study. *International Journal of Medical Informatics*, vol. 127:pp. 52–62, 2019.

[10] S. W. Hunter, A. Divine, C. Frengopoulos, and M. M. Odasso. A framework for secondary cognitive and motor tasks in dual-task gait testing in people with mild cognitive impairment. *BMC Geriatrics*, vol. 18, 2018.

[11] F. Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, vol. 5, 2017.

[12] S. Ishaque, N. Khan, and S. Krishnan. Trends in heart-rate variability signal analysis. *Frontiers in Digital Health*, vol. 3, 2021.

[13] M. Malik, J. Thomas Bigger, A. John Camm, Robert E. Kleiger, Alberto Malliani, Arthur J. Moss, and Peter J. Schwartz. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, vol. 17(3):pp. 354–381, March 1996.

[14] F. Shaffer, Zachary M. Meehan, and Christopher L. Zerr. A critical review of ultra-short-term heart rate variability norms research. *Frontiers in Neuroscience*, vol. 14, 2020.

[15] GG. Fillenbaum, G. van Belle, JC. Morris, RC. Mohs, SS. Mirra, PC. Davis, PN. Tariot, JM. Silverman JM., CM. Clark, KA. Welsh-Bohmer, and A. Heyman. Consortium to establish a registry for Alzheimer's disease (CERAD): the first twenty years. *Alzheimer's & Dementia*, vol. 4(2):pp. 96–109, 2008.

[16] K. Dalmeida and G. Masala. HRV features as viable physiological markers for stress detection using wearable devices. *Sensors*, vol. 21:pp. 2873, 2021.

[17] T. G. Dietterich. Ensemble methods in machine learning. In *Multiple Classifier Systems*, pages 1–15, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.

[18] J.-Sh. Wang, Ch.-W. Lin, and Y-T Yang. A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. *Neurocomputing*, vol. 116:pp. 136–143, September 2013.

[19] J.R. Quinlan. Induction of decision trees. *Machine Learning*, vol. 1:81–106, 1986.

[20] L Breiman. Random forests. *Machine Learning*, vol. 45:pp. 5–32, October 2001.

[21] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, vol. 63:pp. 3–42, April 2006.

[22] A. Mayr, H. Binder, O. Gefeller, and M. Schmid. The evolution of boosting algorithms from machine learning to statistical modelling. *Methods of information in medicine*, vol. 53, March 2014.

[23] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. page 785–794, New York, NY, USA, 2016. Association for Computing Machinery.

[24] N. Peppes, E. Daskalakis, T. Alexakis, E. Adamopoulou, and K Demestichas. Performance of machine learning-based multi-model voting ensemble methods for network threat detection in agriculture 4.0. *Sensors*, vol. 21:pp. 7475, 2021.

[25] Kai Ming Ting. *Confusion Matrix*, pages pp. 260–260. Springer US, Boston, MA, 2017.

[26] T.-T. Wong. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, vol. 48(9):pp. 2839–2846, 2015.

[27] Z. Liu, L. Zhang, J. Wu, Z. Zheng, J. Gao, Y. Lin, Y. Liu, H. Xu, and Y. Zhou. Machine learning-based classification of circadian rhythm characteristics for mild cognitive impairment in the elderly. *Frontiers in Public Health*, 10(036886), October 2022.