A Pareto-Optimality-based approach for selecting the best Machine Learning models in mild cognitive impairment prediction

Paolo Sorino*¶, Vincenzo Paparella*¶, Domenico Lofù*, Tommaso Colafiglio*,
Eugenio Di Sciascio*, Fedelucio Narducci*, Rodolfo Sardone†‡ and Tommaso Di Noia*

* *Dept. of Electrical and Information Engineering (DEI), Politecnico di Bari, Bari (Italy),*
e-mail: {name.surname}@poliba.it

† *Unit of Epidemiology and Biostatistics - Local Health Authority of Taranto, Taranto (Italy),*
‡ *Institute of Life Course and Medical Sciences - University of Liverpool - Liverpool (UK),*
e-mail: {name.surname}@liverpool.ac.uk

¶ *Corresponding authors*

***Abstract*—Mild Cognitive Impairment (MCI) is a syndrome characterized by cognitive impairment that is greater than expected for a subject's age and level of education. Nevertheless, it does not interfere with daily activity. Prevalence in epidemiological and population-based studies ranges from 3% to 19% in adults older than 65 years. A very interesting approach in this area is related to the identification of an Artificial Intelligence (AI)-based model and a subset of relevant features to predict the MCI clinical outcome. In our study, we propose a Pareto-optimality-based approach to identify the best model for predicting MCI. In fact, the best model achieves an Accuracy and Recall on *Yes MCI* of 71% and 80% respectively. With this approach, it is possible to select the best model in order to predict *Yes MCI* (highest risk class). Our study presents a new best model selection approach that can be applied in identifying the best model that can be applied in various disease classification problems.**

***Index Terms*—MCI-Prediction, Machine Learning, Multi objective optimization.**

I. Introduction

The concept of MCI, introduced by Petersen *et al* at the end of the 1990s, covers the cognitive ageing discipline [1]. MCI describes a condition in which an individual develops an age-independent decline in cognitive abilities, but not severe enough to meet the criteria for dementia. MCI can affect memory, language, attention, and other cognitive functions and is a risk factor for dementia. According to the Alzheimer's Association, about $15 - 20\%$ of people aged 65 or older have MCI [1]. The criteria for MCI diagnosis, defined by Petersen [1] *et al*. include: (i) memory problems, (ii) objective memory disorder, (iii) absence of other cognitive disorders or repercussions on daily life, (iv) normal general cognitive function and (v) absence of dementia. In 1997, the emphasis was on the compulsory presence of memory problems and memory disorders. In 1999 these criteria were clarified, with MCI defined solely in clinical terms [2].

MCI corresponds to stage 0.5 on the Clinical Dementia Rating Scale [3]. The concept of MCI made it possible to define a group of patients at a high risk of developing dementia, particularly Alzheimer-type dementia. The definition by Petersen, however, has been criticized for being tautological. When the concept of MCI is restricted to only memory disorder, based on tests used for the early diagnosis of Alzheimer's disease, it leads to identifying people at a high risk of progression to Alzheimer's disease. Early identification and treatment of MCI may be essential to delay or prevent the onset of dementia [4].

In recent decades numerous attempts have been made to classify the boundaries between normal and pathological aging [5]. Research in MCI field is ongoing and focuses on understanding the mechanisms underlying the pathology, developing effective diagnostic tools, and exploring potential treatments. Recent years have seen promising developments in identifying biomarkers for MCI, such as brain structure and function alterations detected through neuroimaging techniques. As the population ages, the prevalence of MCI is expected to increase, making it a significant public health concern. Therefore, further research is needed to improve our understanding of MCI, its risk factors, and potential interventions to prevent cognitive decline.

The key task of predictive medicine is to predict the likelihood of contracting a particular disease, to prevent the disease, or reduce its impact on the patient. Some of the public costs devoted to health care can be reduced by monitoring subjects using machine learning techniques to perform fast screenings. Indeed, in recent years, Data Mining and Machine Learning (ML) techniques have been successfully applied to solve predictive medicine and bioinformatics problems. In particular, ML algorithms in clinical practice are used to improve risk class prediction as much as possible. On the other hand, AI has recently contributed to medicine due to its effectiveness in identifying high-level features deeply embedded in clinical data. Different studies have investigated the use of ML or Deep Learning (DL) algorithms to predict clinical outcomes [6]–[9]. Traditional methods select the best models to perform such predictions based usually on the overall accuracy of the candidate models and by estimating other metrics. However, correctly classifying the subjects who are likely to contract a particular disease is more crucial than having a satisfactory overall accuracy. Consequently, the selected model should maximize the number of correctly classified subjects at risk,

[1]https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment

even if it means sacrificing some overall accuracy. In other words, we need to identify the models that provide the best balance between these two aspects. For this reason, we propose a Pareto-optimality-based approach to identify the best ML model. Specifically, we test our approach by training several ML models on features of the *Salus in Apulia* elderly cohort [8] to predict MCI using simple data available on the clinical dataset that best describes the outcome.

To summarize, the contributions of this work are:

- Our approach is based on easily available data in health datasets to predict MCI;
- Perform best model selection by exploiting a Pareto-based approach, in order to select the best model able to achieve an high overall accuracy and reducing the number of miss-classifications in predicting MCI subjects;

The remainder of this article is organized as follows: Section II introduces the rationale and background. Section III shows the proposed approach and experimental results in Section IV. Our article is concluded in Section V.

II. Rationale and Background

This section provides the background on MCI systems.

Alzheimer's disease (AD), a type of dementia primarily affecting the elderly, is prevalent in society. Detecting AD in its early stages poses a challenge for medical professionals, and currently, no biomarker is entirely reliable in detecting early-stage AD. Unfortunately, AD remains incurable, and clinical trials for AD drugs have had a high rate of failure. Many DL and supervised classification algorithms have been created to aid in AD identification in recent years. Nevertheless, these methods are still imperfect and are not yet able to identify AD with the highest accuracy.

MCI is a progressive and permanent neurological condition that frequently precedes AD and can lead to cognitive decline in older individuals. Early detection is crucial in implementing treatment approaches aimed at improving the quality of life for patients with MCI, as there are currently no known remedies or interventions to halt or reverse its progression.

In this regard, a different work will be discussed with a focus on the analysis of MCI, using Machine Learning and DL techniques.

Ansart *et al.* [10] seeks to address the issue concerning the significance of following proper guidelines for utilizing ML as a decision-making tool in clinical settings. The study investigated the automatic prediction of the clinical state progression of individuals with MCI. The results of this systematic and quantitative review, conducted on 234 experiments coming from 111 articles, raise doubts about the effectiveness of imaging, specifically Magnetic Resonance Imaging (MRI), in forecasting the advancement of individuals with MCI towards dementia. In particular, they showed that studies using cognitive variables or FDG PET reported significantly better results than studies that did not, and that including other feature types does not significantly improve performance compared to using cognition or FDG PET alone. To sum up, the authors demonstrated that even though a considerable number of techniques have been developed for $T1$ MRI, employing this imaging modality alone results in notably inferior performance.

Following the PRISMA guidelines, a comprehensive evaluation was carried out by Grueso *et al.* [11] on 452 studies that utilized ML techniques to analyze neuroimaging data. The primary objective was to determine whether individuals with MCI were prone to developing Alzheimer's disease dementia or maintaining their current state. According to the findings, the use of both MRI and PET in studies led to higher accuracy in classification compared to those using only one neuroimaging technique. The majority of the studies analyzed MRI and PET, while some also incorporated magnetoencephalography. The primary source of the datasets used in the research was the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Support Vector Machine (SVM) was the most prevalent and had an average accuracy of 75.4%. However, Convolutional Neural Network (CNN) had a higher average accuracy of 78.5%.

Alvi *et al.* [12] work on the detection of MCI using Electroencephalography (EEG) data. The authors proposed a DL approach that involved a Gated Recurrent Unit (GRU) model, aimed at accurately distinguishing MCI participants from Healthy Control (HC) subjects based on their EEG data. The stability and effectiveness of the proposed GRU model were assessed by comparing it with other classifiers such as Long Short-Term Memory (LSTM), SVM, and K-Nearest Neighbor (KNN) models. To evaluate the framework's stability, a $5-fold$ cross-validation was performed. The GRU-based deep learning model proposed in the study proved to be a dependable biomarker and assisted technicians in creating an innovative automated MCI detection approach. In fact, the GRU-based MCI detection system framework showcased its ability to differentiate between healthy and MCI subjects with high accuracy.

Forouzannezhad *et al.* [13] work on distinguishing between two groups, Normal Controls (CN) and early MCI (EMCI), to facilitate early diagnosis and treatment planning. They propose a ML approach that utilizes a Deep Neural Network (DNN) and combines MRI, Positron Emission Tomography (PET), and neuropsychological test scores to detect AD at an early stage. The algorithm achieves an exceptional accuracy of 84.0% in differentiating the cognitively normal control group from the EMCI group. The experimental results demonstrate the potential of the proposed deep learning algorithm for multiclass classification, including CN, late MCI, and AD. However, in the current literature, none of the reviewed papers use epidemiological data to predict MCI as most of the papers use imaging or EEG signal. In summary, the above discussion confirms that the current literature is lacking. Our work represents an advance in the state of the art by introducing a new approach to select the best model that can predict MCI using simple data readily available in healthcare datasets.

A. Pareto Optimality and Hypervolume

In this Section, we explain the concept of Pareto Optimality. Its definition is widely exploited in Multi-Objective Optimization Problems. Formally, a Multi-Objective Optimization Problem (MOOP) is defined as [14]:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \boldsymbol{f}(\mathbf{x}) = \{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x})\} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{1}$$

The vector $\mathbf{x} \in \mathbb{R}^n$ is formed by n independent *decison variables*. The set $\mathcal{X} \subseteq \mathbb{R}^n$, generally known as *feasible set*, is defined by a set of equality and inequality constraints. The

vector of functions $\boldsymbol{f(\cdot)}$ is composed by k scalar *objective functions* $f_i : \mathbb{R}^n \to \mathbb{R}^k$ with $i = 1, \ldots, k$; $k \geq 2$. The space $\mathbb{R}^k$ is known as *objective function space*.

In a MOOP, since typically there is no single global solution, it is usually adopted the concept of *Pareto optimality* which leverages on the *Pareto dominance* relation [15]. A vector $\mathbf{x}^\star$ Pareto-dominates vector $\mathbf{x}$, denoted by $\mathbf{x}^\star \prec \mathbf{x}$, if and only if: $f_i(\mathbf{x}^\star) \leq f_i(\mathbf{x}) \, \forall i \in \{1, \ldots, k\}$ and $\exists i \in \{1, \ldots, k\} \mid f_i(\mathbf{x}^\star) < f_i(\mathbf{x})$. Hence, a solution $\mathbf{x}^\star \in \mathcal{X}$ is Pareto optimal if there does not exist another solution $x \in \mathcal{X}$ such that $\mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathbf{x}^\star)$.

Then, solving the problem in Equation (1) means to find the solutions $\mathbf{x} \in \mathcal{X}$ such that their images $\mathbf{f}(\mathbf{x})$ are not Pareto-dominated by any other vector in the feasible set. The group of non Pareto-dominated solutions in the feasible set is called *Pareto optimal set*, whose image in the objective function space is known as *Pareto frontier* [16]. In essence, the Pareto frontier represents the solutions that accomodate the trade-offs between the different objective functions. However, sometimes there is a need to select a single solution from the Pareto frontier that best fits the considered task. In this regard, the *hypervolume* indicator helps to select a single solution from the Pareto frontier [17]. Given a Pareto optimal solution $\mathbf{x}^\star \in \mathbb{R}^k$, a reference point $\mathbf{r} \in \mathbb{R}^k$, and the Lebesgue measure λ, the hypervolume of $\mathbf{x}^\star$ with respect to $\mathbf{r}$ is $\mathcal{HV} = \lambda(\{\mathbf{x} \in \mathbb{R}^k \mid \mathbf{x}^\star \prec \mathbf{x} \prec \mathbf{r}\})$. The $\mathcal{HV}$ value is the volume of the hypercube determined by the solution $\mathbf{x}^\star$ and the reference point $\mathbf{r}$. Therefore, the solution with the highest hypervolume can be considered as the best solution among the Pareto optimal set.

III. Material and Methods

A. Population

The population under study was recruited from the electoral rolls of Castellana Grotte (Apulia, Southern Italy). The sampling framework was the list of the health registry office until $31st$ December 2011, which included 19675 subjects, of which 4.021 aged $65+$ years. All the subjects belonged to the *"Salus in Apulia Study"*, a public health initiative funded by the Italian Ministry of Health and Apulia Regional Government and conducted at the National Institute of Gastroenterology IRCCS "Saverio De Bellis" Research Hospital. The mortality data were obtained from the Electronic Health Records of the Regione Puglia, updated until $31st$ May 2020. This study used data from a subpopulation of Salus in Apulia Study of 1929 older subjects that underwent all the assessments. All participants signed informed consent before their examination. The study was approved by the IRB of the head institution, the National Institute of Gastroenterology and Research Hospital "Saverio de Bellis" in Castellana Grotte (Apulia, Southern Italy). The study met the principles of the Helsinki Declaration and adhered to the Standards for Reporting Diagnostic Accuracy Studies (STARD) guidelines [2] and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines. The population under study is related to a subsample (1.929) of the one described above that contains observations of sick subjects with their respective diseases that were acquired during the interview stage.

B. Dataset Description

Subjects under study had 70 features containing continuous or categorical values of biochemical, anthropometric, and cognitive tests. The dataset is represented by tabular data composed of 1.929 observations and 70 features, including the target variable. To obtain a ready-to-use dataset, all categorical non-relevance features were removed from the analysis except sex. A new dataset was obtained containing 57 features.

C. Preprocessing

We parsed the dataset using pandas[3]. The data set was initially subsampled using the RandomUnderSampler technique in imblearn library [18] to balance the number of observations from both classes by randomly picking samples. After subsampling, the new data set contained 818 observations, 414 for the *No MCI* class and 414 for the *yes MCI* class. Subsequentially a dataset is divided into training and test sets using the standard $80/20$ method. Only continuous values of the observations are standardized according to the StandardScaler[4] technique, known in the ML literature [19] for reducing variance within datasets. Therefore, they are used to perform a feature selection analysis exploiting a SelectKBest method in scikit-learn. The best value of k was selected using a GridSearchCV(5-fold)[5] with a Random Forest Classifier [20] in which values from 1 to 57 were explored, consistent with the number of features in the dataset and max_depth and n_estimators for Random Forest. Afterwards, the best value of k was 9 Pure tone audiometry (PTA)$(0.5 - 2Khz)$left and right, Total Cholesterol, HDL Cholesterol, Mini Mental State Examination (MMSE), Frontal Assesment Battery (Frontal Assesment Battery) raw score, Vitamin D, the mean value of PTA$(0.5-2Khz)$ and mean value ofPTA$(8Khz)$ even though gender and sex were not selected, we added them to the feature set because in the medical domain are considered a characteristic variable [21]

D. Classification models

In order to determine the best classifier to predict the MCI status, we compared the following models: Multilayer Perceptron (MLP) [22]; Random Forest (RF) [20]; SVM [23]; Extreme Gradient Boosting (XGBoost) [24]; The models are developed with Python, using Scikit-learn library[6] [25]. The adopted classifiers, i.e. MLP, RF, SVM, and XGBoost are tuned using a grid search exploration strategy with a 5-fold cross-validation. The models maximize the Area Under the Receiver Operating Characteristic Curve (AUC) value in order to obtain the best predictive power in binary classification. For each model under study, the list of explored hyperparameter values is reported in Table I. It is worth mentioning that we aim to select the best machine learning model by exploiting the Pareto-optimality definition. However, the training of such models through Multi-Objective Optimization methods is out-of-scope.

[2]http://www.stard-statement.org/

[3]https://pandas.pydata.org/

[4]https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

[5]https://scikit- learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

[6]http://scikit-learn.org

TABLE I: Hyperparameter list, values, for the classification models reported in this work.

Algorithm	Hyperparameter	Values
Multilayer Perceptron	seed	{42}
	hidden_layer_sizes	{[(50, 50, 50), (50, 100, 50), (100,)]}
	activation	{tanh, relu}
	solver	{sgd, adam}
	alpha	{np.arange(0.1, 1, 0.2)}
	learning_rate	{constant, adaptive}
Random Forest	seed	{42}
	n_estimators	{np.arange(1, 50, 2)}
	max_features	{sqrt, log2}
	max_depth	{np.arange(1, 30, 1)}
	criterion	{gini, entropy}
Support Vector Machines	seed	{42}
	kernel	{rbf, poly}
	gamma	{0.1, 0.001, 0.3, 0.003, 0.5, 0.05}
eXtreme Gradient Boosting	seed	{42}
	n_estimators	{np.arange(1, 50, 2)}
	learning_rate	{np.arange(0.1, 1, 0.2)}
	scale_pos_weight	{1.000}

E. Evaluation Metrics

The metrics used to evaluate the trained models are the following.

The first parameter considered to evaluate the performance of our models was the accuracy, defined as [26]:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of prediction}} \tag{2}$$

More specifically, the accuracy of a model is calculated with the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative. The Recall metric measures the ratio of correct positive classifications among the total number of positive samples:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

The Precision measures the ratio of correct positive classifications among the total positive classifications:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

The F1 score is the harmonic mean between recall and accuracy:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

The F1 score combines precision and recall into a single metric. The Area Under the Receiver Operating Characteristic Curve (AUC) is a metric that measures the capability of a classifier to separate the positive class from the negative one. It is formulated as follows:

$$AUC = \frac{\sum_{x^- \epsilon X^-} \sum_{x+ \epsilon X+} \left(1 \left(f\left(x^-\right) < f\left(x^+\right)\right)\right)}{X^- + X^+} \tag{7}$$

where $1(\cdot) = 1$ if $f\left(x^-\right) < f\left(x^+\right)$else $1(\cdot) = 0$ and, $X^+$is the set of positive samples, $X^-$is the set of negative samples, $f(\cdot)$ is the result of model prediction, and $1(\cdot)$ an indicator function [27].

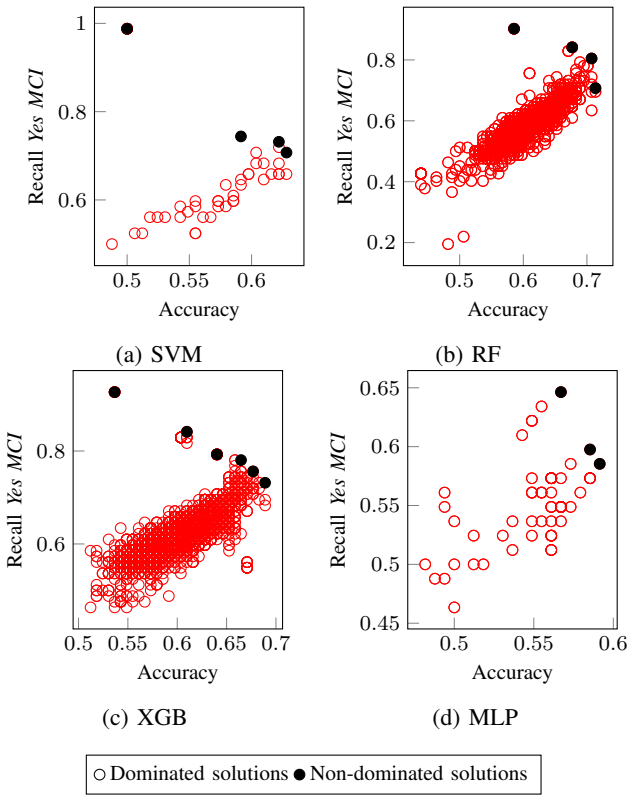

Fig. 1: Models' hyperparameter configurations in the objective function space *Accuracy/Recall* on *YES MCI* class. For each model, the black dots are on the Pareto frontier, while the red dots are dominated solutions.

F. Pareto Optimality-based Selection

In this Section, we describe the approach adopted to select the best classifier for our task. Oftentimes, traditional approaches consist in selecting the best models relying only on the accuracy of the candidate models. However, in MCI systems, it is of greater importance to classify MCI subjects rather than healthy subjects correctly. As a consequence, it is not sufficient to choose the model having the best overall accuracy. Indeed, the selected model should maximize the number of correctly classified MCI subjects (i.e., the Recall on

TABLE II: Pareto-optimal models' configurations and their performance. The best values for $\mathcal{HV}$ indicator are in bold. The models are characterized by the following hyperparameters configurations. RF_1: Entropy as *Criterion*, *Max depth*=3, *Max Features*=log2, *Number of estimators*=1; RF_2: Entropy as *Criterion*, *Max depth*=3, *Max Features*=sqrt, *Number of estimators*=9; RF_3: Entropy as *Criterion*, *Max depth*=1, *Max Features*=sqrt, *Number of estimators*=1; SVM_1 C=6, γ =0.003, poly as *Kernel*; $XGBoost_1$: *Learning Rate*=0.1, *Max depth*=1, *Number of estimators*=1; RF_4: Entropy as *Criterion*, *Max depth*=5, *Max Features*=log2, *Number of estimators*=13.

Model	Accuracy	Recall *Yes MCI*	HV
RF_1	0.68	0.84	**0.57**
RF_2	0.71	0.80	**0.57**
RF_3	0.58	0.90	0.53
SVM_1	0.5	0.99	0.49
$XGBoost_1$	0.54	0.93	0.50
RF_4	0.71	0.71	0.50

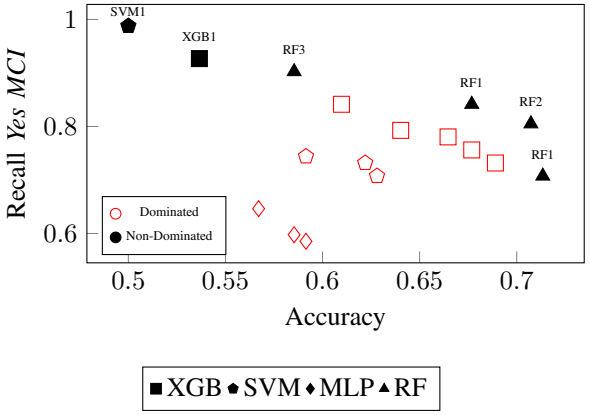

Fig. 2: Overall models' hyperparameter configurations in the objective function space *Accuracy/Recall* on *YES MCI* class. The black marks refer to Pareto-optimal solutions. The red marks represent dominated solutions. The marks' shapes indicate a particular class of model.

yes MCI class), even paying some trade-off with the overall accuracy of the system. In other words, we should perform a *multi-criteria* selection of the best model. To this end, we exploit the concept of Pareto Optimality to identify the models that best accommodate the trade-off between overall accuracy and Recall on *yes MCI* class.

Once all the models' configurations described above are trained, they represent different solutions in the objectives function space having accuracy and Recall on *yes MCI* class on the *x*-axis and *y*-axis, respectively. For each model, we identify the Pareto-optimal solutions lying on the Pareto frontier. To clarify this aspect, we plot the objective function spaces and the solutions on them for each model in Figure 1. Then, we take into account the Pareto frontier of each model, and we find the Pareto-optimal points among all the models (Figure 2). Finally, we exploit the widely-used Hypervolume indicator to select one–best solution from the final Pareto frontier.

IV. Results

In this section, we detail the results obtained from our experimental approach. Specifically, in Table II, we report the results

TABLE III: Results for the MCI prediction of the Best Classifier.

Model	Class	Precision	Recall	F1-score	Accuracy
RF_1	No MCI	0.76	0.51	0,61	0.68
	Yes MCI	0.63	0.84	0.72	
RF_2	No MCI	0.76	0.61	0,67	0.71
	Yes MCI	0.67	0.80	0.73	

obtained by computing the Hypervolume indicator for each model lying on the Pareto frontier in Figure 2. We observe that six models are Pareto-optimal. On the one hand, we notice that no configuration of MLP lies on the Pareto frontier. On the other hand, four out of six Pareto-optimal solutions are different configurations of Random Forest. Specifically, two Random Forest configurations achieve the highest $\mathcal{HV}$ value. Both of them are characterized by the Shannon entropy as information gain and 3 as the maximum depth of the tree. Hence, we have two candidates as the best model to perform the MCI prediction. To determine the best one, we also analyze the performance of such models in terms of Precision, Recall, and F1-score for each class. From Table III, we claim that RF_2 is the best model since it achieves better performance than RF_1 with respect to all the considered metrics except for Recall on *Yes MCI* class. To summarize, RF_2 performs better according to a *multi-criteria* evaluation. In fact, the performance obtained by RF_2 is the most balanced as it achieves an accuracy of 71%, a high value of recall (0.80), the best value of precision, in the prediction of *Yes MCI* (0.67), and the best value of F1-score in the prediction of *Yes MCI* (0.73).

V. Discussion and Conclusion

In this study, we propose a Pareto-based approach to identifying the best model for predicting MCI. Specifically, we trained and tested various machine-learning algorithms in order to choose the best one. From our perspective, the Pareto-based approach is relevant because it allows to select the best model according to multiple objectives. In this way, the identification of the best model does not solely depend on the overall accuracy of the model but also on other criteria tailored to the considered task. Practically, in MCI systems, we are able to select a model with both acceptable overall accuracy and effectiveness in classifying MCI subjects, which is crucial in medical tasks. Hence, this approach could be used to select the best machine learning algorithms for predicting high-risk classes that can be used in epidemiological studies or screening. The aim is to provide a smaller set of easy-to-find features in health datasets to predict MCI, thus reducing time and waiting lists for more extensive instrumental examinations.

The Random Forest considered achieved an Accuracy = 0.71, Precision = 0.76, Recall = 0.61, F1-Score = 0.67 respectively in the class *No MCI* and Precision = 0.67, Recall = 0.80, F1-Score = 0.73 respectively in the class *Yes MCI*. These results appear to be good in the current state of the art [28], [29]. However, the metrics achieved refer to a small population sample. Possible implementations could be aimed at validating this system on a larger test sample. The model selected by the Pareto-based approach was able to well discriminate subjects at risk of MCI. In future work, DL algorithms such as Recurrent neural network (RNN), Long short-term memory

(LSTM) or Gated recurrent unit (GRU) will be evaluated and analyzed in MCI prediction. In addition, a best model selection mechanism may be applied to other datasets for MCI prediction or other disease prediction. The number of observations in the dataset under study will be increased, and a DL algorithm will be evaluated to increase predictive performance. Subsequently, a web-based application based on the ML/DL algorithm may be developed to provide a ready-to-use tool. In conclusion, this model may be considered valid support for medical decision-making as regards health policies, epidemiological studies, and screening.

ACKNOWLEDGMENT

This work was partial support of the projects: Italian P.O. Puglia FESR 2014 – 2020 (project code 6ESURE5) 'SECURE SAFE APULIA', Fincons CdP3, PASSPARTOUT, Servizi Locali 2.0, ERP4.0.

REFERENCES

[1] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. Kokmen, and E. G. Tangelos, "Aging, memory, and mild cognitive impairment," *International psychogeriatrics*, vol. 9, no. S1, pp. 65–69, 1997.

[2] R. C. Petersen, R. Doody, A. Kurz, R. C. Mohs, J. C. Morris, P. V. Rabins, K. Ritchie, M. Rossor, L. Thal, and B. Winblad, "Current concepts in mild cognitive impairment," *Archives of neurology*, vol. 58, no. 12, pp. 1985–1992, 2001.

[3] C. P. Hughes, L. Berg, W. Danziger, L. A. Coben, and R. L. Martin, "A new clinical scale for the staging of dementia," *The British journal of psychiatry*, vol. 140, no. 6, pp. 566–572, 1982.

[4] F. Portet, P. Ousset, P. Visser, G. Frisoni, F. Nobili, P. Scheltens, B. Vellas, J. Touchon, M. W. G. of the European Consortium on Alzheimer's Disease (EADC *et al.*, "Mild cognitive impairment (mci) in medical practice: a critical review of the concept and new diagnostic procedure. report of the mci working group of the european consortium on alzheimer's disease," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 77, no. 6, pp. 714–718, 2006.

[5] K. Ritchie and J. Touchon, "Mild cognitive impairment: conceptual basis and current nosological status," *The Lancet*, vol. 355, no. 9199, pp. 225–228, 2000.

[6] P. Sorino, M. G. Caruso, G. Misciagna, C. Bonfiglio, A. Campanella, A. Mirizzi, I. Franco, A. Bianco, C. Buongiorno, R. Liuzzi *et al.*, "Selecting the best machine learning algorithm to support the diagnosis of non-alcoholic fatty liver disease: A meta learner study," *PLoS One*, vol. 15, no. 10, p. e0240867, 2020.

[7] P. Sorino, A. Campanella, C. Bonfiglio, A. Mirizzi, I. Franco, A. Bianco, M. G. Caruso, G. Misciagna, L. R. Aballay, C. Buongiorno *et al.*, "Development and validation of a neural network for nafld diagnosis," *Scientific Reports*, vol. 11, no. 1, pp. 1–13, 2021.

[8] F. Castellana, L. Lampignano, I. Bortone, R. Zupo, M. Lozupone, C. Griseta, A. Daniele, G. De Pergola, G. Giannelli, R. Sardone *et al.*, "Physical frailty, multimorbidity, and all-cause mortality in an older population from southern italy: results from the salus in apulia study," *Journal of the American Medical Directors Association*, vol. 22, no. 3, pp. 598–605, 2021.

[9] C. Ardito, I. Bortone, T. Colafiglio, T. D. Noia, E. D. Sciascio, D. Lofù, F. Narducci, R. Sardone, and P. Sorino, "Brain computer interface: Deep learning approach to predict human emotion recognition," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2022, pp. 2689–2694.

[10] M. Ansart, S. Epelbaum, G. Bassignana, A. Bône, S. Bottani, T. Cattai, R. Couronné, J. Faouzi, I. Koval, M. Louis *et al.*, "Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review," *Medical Image Analysis*, vol. 67, p. 101848, 2021.

[11] S. Grueso and R. Viejo-Sobera, "Machine learning methods for predicting progression from mild cognitive impairment to alzheimer's disease dementia: a systematic review," *Alzheimer's research & therapy*, vol. 13, pp. 1–29, 2021.

[12] A. M. Alvi, S. Siuly, H. Wang, K. Wang, and F. Whittaker, "A deep learning based framework for diagnosis of mild cognitive impairment," *Knowledge-Based Systems*, vol. 248, p. 108815, 2022.

[13] P. Forouzannezhad, A. Abbaspour, C. Li, M. Cabrerizo, and M. Adjouadi, "A deep neural network approach for early diagnosis of mild cognitive impairment using multiple features," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. IEEE, 2018, pp. 1341–1346.

[14] R. Marler and J. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and Multidisciplinary Optimization*, vol. 26, pp. 369–395, 04 2004.

[15] C. von Lücken, B. Barán, and C. A. Brizuela, "A survey on multi-objective evolutionary algorithms for many-objective problems," *Comput. Optim. Appl.*, vol. 58, no. 3, pp. 707–756, 2014. [Online]. Available: https://doi.org/10.1007/s10589-014-9644-1

[16] V. Paparella, "Pursuing optimal trade-off solutions in multi-objective recommender systems," in *RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022*, J. Golbeck, F. M. Harper, V. Murdock, M. D. Ekstrand, B. Shapira, J. Basilico, K. T. Lundgaard, and E. Oldridge, Eds. ACM, 2022, pp. 727–729. [Online]. Available: https://doi.org/10.1145/3523227.3547425

[17] E. Zitzler, D. Brockhoff, and L. Thiele, "The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration," in *Evolutionary Multi-Criterion Optimization, 4th International Conference, EMO 2007, Matsushima, Japan, March 5-8, 2007, Proceedings*, ser. Lecture Notes in Computer Science, S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, Eds., vol. 4403. Springer, 2007, pp. 862–876. [Online]. Available: https://doi.org/10.1007/978-3-540-70928-2_64

[18] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017. [Online]. Available: http://jmlr.org/papers/v18/16-365.html

[19] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, and G. Srivastava, "Deep neural networks to predict diabetic retinopathy," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.

[20] M. Pal, "Random forest classifier for remote sensing classification," *International journal of remote sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[21] Y. Yuan, J. Li, N. Zhang, P. Fu, Z. Jing, C. Yu, D. Zhao, W. Hao, and C. Zhou, "Body mass index and mild cognitive impairment among rural older adults in china: the moderating roles of gender and age," *Bmc Psychiatry*, vol. 21, pp. 1–11, 2021.

[22] W. G. Baxt, "Application of artificial neural networks to clinical medicine," *The lancet*, vol. 346, no. 8983, pp. 1135–1138, 1995.

[23] W. S. Noble, "What is a support vector machine?" *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[24] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[26] A. K. Biswas, N. Noman, and A. R. Sikder, "Machine learning approach to predict protein phosphorylation sites by incorporating evolutionary information," *BMC bioinformatics*, vol. 11, no. 1, pp. 1–17, 2010.

[27] T. Calders and S. Jaroszewicz, "Efficient auc optimization for classification," in *European conference on principles of data mining and knowledge discovery*. Springer, 2007, pp. 42–53.

[28] B. Bratić, V. Kurbalija, M. Ivanović, I. Oder, and Z. Bosnić, "Machine learning for predicting cognitive diseases: methods, data sources and risk factors," *Journal of medical systems*, vol. 42, pp. 1–15, 2018.

[29] J. J. Gomar, M. T. Bobes-Bascaran, C. Conejero-Goldberg, P. Davies, T. E. Goldberg, A. D. N. Initiative *et al.*, "Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to alzheimer disease in patients in the alzheimer's disease neuroimaging initiative," *Archives of general psychiatry*, vol. 68, no. 9, pp. 961–969, 2011.