

Article

Sex Differences in Conversion Risk from Mild Cognitive Impairment to Alzheimer's Disease: An Explainable Machine Learning Study with Random Survival Forests and SHAP

Alessia Sarica *, Assunta Pelagi , Federica Aracri, Fulvia Arcuri, Aldo Quattrone, Andrea Quattrone and for the Alzheimer's Disease Neuroimaging Initiative †

Neuroscience Research Center, Department of Medical and Surgical Sciences, Magna Graecia University, 88100 Catanzaro, Italy; assunta.pelagi@studenti.unicz.it (A.P.)

* Correspondence: sarica@unicz.it

† Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:

http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Abstract: Alzheimer's disease (AD) exhibits sex-linked variations, with women having a higher prevalence, and little is known about the sexual dimorphism in progressing from Mild Cognitive Impairment (MCI) to AD. The main aim of our study was to shed light on the sex-specific conversion-to-AD risk factors using Random Survival Forests (RSF), a Machine Learning survival approach, and Shapley Additive Explanations (SHAP) on dementia biomarkers in stable (sMCI) and progressive (pMCI) patients. With this purpose, we built two separate models for male (M-RSF) and female (F-RSF) cohorts to assess whether global explanations differ between the sexes. Similarly, SHAP local explanations were obtained to investigate changes across sexes in feature contributions to individual risk predictions. The M-RSF achieved higher performance on the test set (0.87) than the F-RSF (0.79), and global explanations of male and female models had limited similarity (<71.1%). Common influential variables across the sexes included brain glucose metabolism and CSF biomarkers. Conversely, the M-RSF had a notable contribution from hippocampus, which had a lower impact on the F-RSF, while verbal memory and executive function were key contributors only in F-RSF. Our findings confirmed that females had a higher risk of progressing to dementia; moreover, we highlighted distinct sex-driven patterns of variable importance, uncovering different feature contribution risks across sexes that decrease/increase the conversion-to-AD risk.

Keywords: Alzheimer's disease; random survival forests; sex differences



Citation: Sarica, A.; Pelagi, A.; Aracri, F.; Arcuri, F.; Quattrone, A.; Quattrone, A.; for the Alzheimer's Disease Neuroimaging Initiative. Sex Differences in Conversion Risk from Mild Cognitive Impairment to Alzheimer's Disease: An Explainable Machine Learning Study with Random Survival Forests and SHAP. *Brain Sci.* **2024**, *14*, 201. <https://doi.org/10.3390/brainsci14030201>

Academic Editors: Luca Giancardo and Notger G. Müller

Received: 23 January 2024

Revised: 10 February 2024

Accepted: 20 February 2024

Published: 22 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative pathology that differentially affects women and men [1–5], where women have a higher prevalence than men, representing two-thirds of AD patients in the US [3]. Many hypotheses exist about sex differences in the progression from Mild Cognitive Impairment (MCI) to AD, but the literature reports heterogeneous findings [2,6]. Generally, the higher prevalence of AD in women has been associated with longer female life expectancy and with biases in patient enrollment [7]. However, other studies showed a more complex picture, focusing on the neurobiological vulnerability of women, probably related to sex hormones, like estrogen [5,7]. Regarding the psychosocial aspects, women are more prone to life stress, social isolation, and insomnia [8], and their vulnerability to stressful events is enhanced by genes like the APOE e4 allele [9]. In a non-cognitively impaired population, women demonstrate higher scores in verbal tasks and slower cognitive decline than men at all ages, while men perform better than

women in visuospatial and motor coordination tasks [2]. Differences in verbal memory tasks are lost at the early stage of AD [2], although other works found that these differences also persist at early stages [10,11]. However, the literature agrees in affirming that women lose their better verbal memory performance when dementia is diagnosed [2,10,11]. A recent study [4] that performed a sex-stratified analysis found that auditory verbal memory and difficulties in activities of daily living are stronger risk factors for women than men in predicting the progression from MCI to AD. Regarding neuroimaging evidence, women showed a faster rate of brain atrophy than men [2,4], and in particular, hippocampal volume changes in women compared to men had a more prominent contribution to the progression from a normal cognitive state to MCI or AD [5]. The rate of changes in white matter hyperintensities also showed sex-linked characteristics, affecting more men than women when progressing to AD [5]. Two other recent studies demonstrated the vulnerability of women to AD pathology: the first one [12] explored brain glucose metabolism and the plasma beta-amyloid 42/40 ratio, and the second one [13] investigated a combination of functional and structural markers.

On the other hand, several works showed that no sex differences in the progression to AD exist and that the risk is equal between males and females [6,14], contradicting the hypothesis of sexual dimorphism in dementia [15], and stating that those differences are due only to the longer life expectancy of women.

Given these discrepancies in the literature, we aimed to investigate the sex differences in the risk prediction of conversion from MCI to AD using a Machine Learning survival approach, Random Survival Forests (RSF) [16,17]. RSF is an adaptation of the Random Forests (RF) [18,19] algorithm to handle right-censored data and to provide the assessment of survival probability and risk, which is fully nonparametric and thus independent from data distribution; it can handle multicollinearity and it intrinsically provides feature selection [16,17]. RSF showed stability and robustness when trained on multi-modal data, and it had better performance on biomedical datasets [20,21] as well as on dementia data [22,23] compared to statistical approaches like the Cox Proportional Hazard [24] (CPH). Moreover, we demonstrated in [23] that the black-box nature of RSF and its poor explainability could be overcome through the model-agnostic method Shapley Additive Explanations (SHAP) [25]. SHAP provides a unified framework to interpret ML predictions based on game theory, which assigns to each feature a Shapley value that represents its average marginal contribution to the predicted risk across all possible feature coalitions [23,26].

In the present work, we used a dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) consisting of well-known dementia biomarkers [22,23], such as clinical, cognitive, cerebrospinal fluid (CSF) and imaging features, of stable MCI patients (sMCI) and progressive MCI patients (pMCI), who change their diagnosis to AD over time. In detail, we applied RSF and SHAP separately on the male and female MCI cohorts to predict the risk of conversion to AD, and more importantly to assess whether global and local explanations differ between the sexes. Differences in the explanations of male and female models were quantified using the Rank-Biased Overlap [27] (RBO), which has been used in survival analysis [22,23] to estimate the overlap between ML feature importance by varying the number of the top variables that are considered as important [28]. Finally, we investigated the individual predictions of male and female MCI patients, stratified by high-, medium-, and low-risk grades, using SHAP waterfall plots, which provide a highly intelligible overview of the variable contribution to the decrease or increase in the conversion-to-AD risk.

2. Materials and Methods

2.1. Dataset Preparation

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance

imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of Mild Cognitive Impairment (MCI) and early Alzheimer's disease (AD).

ADNI enrolls participants between the ages of 55 and 90 who are recruited at 57 sites in the United States and Canada. After obtaining informed consent, participants undergo a series of initial tests that are repeated at intervals over subsequent years, including a clinical evaluation, neuropsychological tests, genetic testing, lumbar puncture, and MRI and PET scans. Details about the inclusion/exclusion criteria and about the enrollment procedure can be found on the ADNI website.

Data table files (*csv*) from ADNI were downloaded on 5 June 2023, and they were as follows: DXSUM_PDXCONV_ADNIALL, ADNIMERGE, NEUROBAT, CDR, GDSCALE, FAQ, MMSE, ADASSCORES, UPENNBIOMK_MASTER_FINAL (9, 10, 12), and BAIPET-NMRC_04_12_18. The software KNIME 4.6.1 [29] was used to filter and join these tables. Details about the dataset preparation can be found elsewhere [22,23]. Briefly, the final dataset used for the ML analysis included patients whose diagnosis changed over time from MCI to AD (pMCI) and patients who maintained their baseline diagnosis as stable MCI (sMCI). The *event* or *ensorship* occurrence was a binary variable, where 1 (pMCI patient) represents the event of conversion from MCI to AD, and 0 represents censorship (sMCI patient). The *time* variable represented the number in months (m06, m12, m18, m24, m36, and m48) after the baseline visit in which the event/censorship occurred. The time interval ranged from 6 months to 36 months (3 years), which was different from [23] because data for month 48 were unusable due to the low sample size (16 males; 6 females). All data and all subjects were from the ADNI1 protocol and collected at baseline or the screening visit. The description of demographics, clinical, neuropsychological, and neuroimaging features are reported in Appendix A.

The final dataset consisted of 365 subjects that were split by sex: 233 males divided into 136 sMCI (M-sMCI) and 97 pMCI (M-pMCI), and 132 females divided into 62 sMCI (F-sMCI) and 70 pMCI (F-pMCI).

Categorical variables (PTETHCAT, PTRACCAT, PTMARRY) were converted to numerical data using the One-Hot Encoding approach [30,31] (python function *get_dummies()*). Missing data were imputed using the missForest algorithm [32] (python package *missingpy* 0.2.0), which demonstrated less error than statistical imputation methods on dementia [33] and Parkinson's disease [34] data. The descriptive statistics of the dataset stratified by sex are reported in Table 1.

2.2. Statistical Analysis

Statistical analyses were performed to compare features between male sMCI and male pMCI patients (M-sMCI vs. M-pMCI), and female sMCI and female pMCI patients (F-sMCI vs. F-pMCI). Moreover, we compared male sMCI with female sMCI patients (M-sMCI vs. F-sMCI), and male pMCI with female pMCI patients (M-pMCI vs. F-sMCI).

Analysis of variance (ANOVA) was employed to assess differences between groups in terms of age and years of education, while the Chi-square test was applied to evaluate differences in the distributions of categorical variables. Analysis of covariance (ANCOVA) with age as a covariate was employed for clinical and cognitive variables, while for neuroimaging features, ANCOVA had age and ICV (significant at $p < 0.05$). All statistical tests were implemented with Python 3.8 and the package *scikit-learn* 1.1.3.

Table 1. Demographic, clinical, cognitive, CSF, and imaging data of sMCI and pMCI patients stratified by sex.

	M (233)		F (132)	
	sMCI (136)	pMCI (97)	sMCI (62)	pMCI (70)
Demographic:				
Age	74.7 ± 7.3	74.6 ± 7.5	75.8 ± 7.6	74.6 ± 6.0
Education level	15 ± 3.3	15.4 ± 2.9	16.2 ± 2.6	16.1 ± 2.8
Biomarker:				
APOE4 (0/1/2)	82/42/12	38/43/16	28/26/8	16/43/11
Clinical scale:				
CDRSB	1.4 ± 0.8	1.9 ± 0.9	1.6 ± 0.9	1.8 ± 1.1
FAQ	2.3 ± 3.2	5.6 ± 5.3	2.8 ± 4.1	5.6 ± 4.5
Neuropsychological assessment:				
ADAS11	10.5 ± 4.2	13.12 ± 3.8	10.5 ± 4.6	13.2 ± 4.5
ADAS13	16.8 ± 6.0	21.3 ± 5.0	16.9 ± 6.7	21.1 ± 6.2
ADASQ4	5.6 ± 2.2	7.13 ± 1.9	5.5 ± 2.4	7.1 ± 2.0
MMSE	27.3 ± 1.8	26.68 ± 1.7	27.2 ± 1.7	26.6 ± 1.8
RAVLT_immediate	33.1 ± 9.6	27.25 ± 6.9	33.1 ± 10.6	27.2 ± 6.2
RAVLT_learning	3.8 ± 2.3	2.74 ± 1.9	3.7 ± 2.3	3.0 ± 2.0
RAVLT_forgetting	4.5 ± 2.4	4.86 ± 2.09	4.5 ± 2.3	5.0 ± 2.2
RAVLT_perc_forgetting	59.9 ± 31.5	77.62 ± 27.6	63.8 ± 31.0	79.0 ± 28.5
LDELTOTAL	4.3 ± 2.7	2.8 ± 2.4	4.8 ± 2.5	3.3 ± 3.1
DIGITSCOR	38.9 ± 11.1	34.13 ± 11.2	37.0 ± 9.7	34.0 ± 10.5
TRABSCOR	116.4 ± 64.2	146.85 ± 79.9	125.6 ± 67.1	149.9 ± 80.2
mPACCdigit	−3.9 ± 3.9	−3.9 ± 4.8	−3.8 ± 3.8	−3.9 ± 4.9
mPACCtrailsB	−3.8 ± 4.0	−3.7 ± 4.8	−3.8 ± 3.8	−3.0 ± 4.8
GDTOTAL	1.6 ± 1.4	1.54 ± 1.3	1.6 ± 1.3	1.6 ± 1.4
COPYSCOR	4.7 ± 0.7	4.5 ± 1.2	4.6 ± 0.8	4.7 ± 0.6
BNTTOTAL	25.5 ± 4.0	24.55 ± 4.6	26.2 ± 3.2	25.7 ± 3.6
CSF:				
ABETA42	1027.5 ± 398.4	676.7 ± 224.3	881.4 ± 364.2	708.8 ± 309.2
TAU	307.2 ± 89.4	316.43 ± 73.2	307.5 ± 110.9	331.0 ± 90.2
PTAU	30.4 ± 10.8	31.76 ± 8.0	31.2 ± 14.0	33.3 ± 10.8
Neuroimaging:				
Ventricles	41,196.5 ± 24,245.5	44,937.2 ± 18,888.8	48,192.5 ± 26,633.5	50,164.0 ± 27,112.5
Hippocampus	6699.8 ± 987.2	5862.6 ± 923.9	6452.7 ± 963.8	6092 ± 1095.6
WholeBrain	1,005,263.4 ± 106,966.9	973,259.03 ± 115,953.1	1,013,898.5 ± 100,967.5	990,467.4 ± 111,353.5
Entorhinal	3480.9 ± 711.5	2997.0 ± 698.9	3475.5 ± 707.2	3031.5 ± 723.4
Fusiform	16,831.29 ± 2179.3	15,618.0 ± 2409.4	16,947.7 ± 2161.6	15,884.9 ± 2393.1
MidTemp	19,134.46 ± 2554.2	17,311.45 ± 3127.2	19,604.9 ± 2652.1	17,911.5 ± 2708.5
ICV	1,562,645.6 ± 163,009.0	1,554,468.3 ± 170,366.0	1,609,215.3 ± 162,431.2	1,587,939.1 ± 176,315.7
FDG	1.2 ± 0.1	1.07 ± 0.1	1.2 ± 0.1	1.1 ± 0.1
HCI	7.1 ± 2.8	9.54 ± 2.5	7.1 ± 2.7	9.9 ± 2.9
Occurrence of event (pMCI = 1) and censorship (sMCI = 0) per timepoint (in months):				
m06	17	14	3	8
m12	12	27	5	20
m18	14	20	7	15
m24	21	18	10	18
m36	72	18	37	8

Mean and standard deviation are calculated after imputation of missing data. For abbreviations see Appendix A.

2.3. Random Survival Forests

Various studies have assessed the efficacy of ML techniques for dementia survival analysis, especially for predicting the conversion risk from MCI to AD [31,35–39]. Most of them showed that Random Survival Forests [17] had better performance than the classical statistical approaches like Cox Proportional Hazard, or other methods based on Random Forests [18,19]. In particular, we demonstrated in [22] that RSF had higher accuracy than Conditional Survival Forest (CSF) [40] and Extra Survival Trees (XST) [41] in predicting the conversion-to-AD risk on dementia biomarkers from ADNI. Moreover, we showed in [23] that the clinical utility of RSF can be boosted through SHAP to enhance its interpretability. The strengths of RSF rely on the robustness to outliers, no convergence issues, preservation from overfitting thanks to out-of-bag (cross-validated) prediction, the reliable inference of training data, and particularly its intrinsic variable importance measure, which is fully nonparametric and independent from data distributions [17].

In detail, RSF follows the same principles of RF [18,19] for growing decision trees, and when splitting tree nodes, it applies bootstrapping and random feature selection. The rule for splitting a node is based on the log-rank test statistic to maximize the survival difference between daughter nodes. For each node in the tree, the null hypothesis that there is no difference between the two groups in the probability of an event is tested. The ensemble's cumulative hazard is estimated with cumulative hazard functions calculated for each tree, while out-of-bag (OOB) estimators are used to assess the prediction accuracy and the variable importance [16,17,22].

2.4. Machine Learning Analysis

A forked repository (<https://github.com/bacalfa/pysurvival/>, Bacalfa) from the python package *PySurvival* (<https://square.github.io/pysurvival/>, Fotso et al., 2019) was used to conduct survival analyses to have the compatibility of the RSF algorithm implementation with the sklearn package (accessed on 1 December 2023). The package *seaborn* (0.12.2) was employed to modify the original plotting functions of *PySurvival*.

Two RSF models were built separately, one trained only on male MCI patients (M-RSF) and the second trained only on female MCI patients (F-RSF), with the same procedure in [23] and as described below. Datasets were randomly split with a static seed into training and test sets (80–20%) stratified by the column *event* and *time* to maintain the original distribution of occurrences, obtaining 109 sMCI and 77 pMCI in the training set and 27 sMCI and 20 pMCI in the test set for the male group, and 50 sMCI and 55 pMCI in the training set and 12 sMCI and 15 pMCI in the test set for the female group. Hyperparameter tuning was applied to maximize the performance on the training set through a randomized search (RandomizedSearchCV) with a 3-fold cross-validation (*cv*) and 50 repetitions [20,22,23,31]. RSF hyperparameters were as follows: maximum depth (*max_depth*), minimum number of samples required to be at a leaf node (*min_node_size*), number of features to consider when looking for the best split (*max_features*), and percentage of original samples used in each tree building (*sample_size_pct*). As described in [22,23], the number of trees was kept static at 200, and importance mode (*importance_mode*) was set to permutation for both M-RSF and F-RSF to allow their comparison.

The performance of RSF models was evaluated using Harrell's concordance index (*c*-index) [42] on training sets (with a 5-fold cross-validation) and on test sets. The *c*-index was born to generalize the area under the ROC curve (AUC) in the presence of right-censored data and for the survival analysis; the model has an almost perfect discriminatory power if its value is close to 1, while it has no ability to discriminate between low- and high-risk subjects if it is close to 0.5 (random prediction) [22,23]. In addition to the *c*-index, we evaluated the accuracy of the predicted survival function on the test set across multiple timepoints with the Integrated Brier score (IBS) [43]. The IBS value is between 0 and 1, where 0 is for a perfect model, while a cut-off limit of 0.25 is considered as critical [43].

The estimated survival time curve of test sets was obtained using the Kaplan–Meier method (KM) [44] and visually compared with predicted survival curves determined by the M-RSF and F-RSF models. Deviations from KM curves were quantified using the Root Mean Square Error (RMSE) and median/mean absolute error.

2.4.1. Global Explanation

In addition to the feature importance provided intrinsically by the two models M-RSF and F-RSF, we evaluated the permutation importance [23], which is defined as an increase in the prediction error when a feature's value is randomly shuffled. Permutation importance was implemented using ELI5 [45] with 50 repetitions (python package scikit-learn 1.3.0) [23].

As a further global explanation, we employed Shapley Additive Explanation [25] (SHAP, python package *SHAP* 0.42.1), which is a model-agnostic unified framework based on game theory to interpret ML classification predictions and has also been recently applied for survival analysis [23]. Two SHAP explainers (*shap.Explainer*) were fit separately on predicted risk scores of training sets by M-RSF and by F-RSF (function *predict_risk* by *pysurvival*).

A pairwise similarity between the global explanations of the two models, M-RSF and F-RSF, was quantitatively evaluated using the Rank-Biased Overlap [27] (RBO, python package *rbo* v.0.1.2, <https://github.com/changyaochen/rbo> accessed on 1 December 2023), which can assume values in the range [0, 1], where 0 means disjoint and 1 means identical. RBO has been used in survival analysis [22,23] to estimate the overlap between ML feature rankings by varying the number of top variables considered as important (depths d) [28].

No feature selection was applied since the recent literature on survival analysis showed no improvement in performance [20,23,26]. In the same way, we kept correlated variables, since it has been demonstrated that multicollinearity did not perturb SHAP explanations of RSF [23].

2.4.2. Local Explanation

Local explanations of both M-RSF and F-RSF models were explored with SHAP on test sets. Individual predictions determined by M-RSF and F-RSF were used to manually stratify male and female pMCI test patients according to their conversion-to-AD risk score (low, medium, and high) [22]. Then, we estimated the cumulative density function of six randomly selected pMCI patients, one male and one female per risk grade (M-pMCI#1 and F-pMCI#1 high risk, M-pMCI#2 and F-pMCI#2 medium risk, M-pMCI#3 and F-pMCI#3 low risk), and one stable MCI test subject per sex (M-sMCI and F-sMCI with a numeric risk score lower than 1). These test subjects were finally studied with SHAP waterfall plots.

3. Results

The results of statistical analysis between male and female MCI patients are reported in Table 2. Regarding the analysis of the male group, sMCI and pMCI patients had significantly different values in almost all features, except for age, education level, RAVLT forgetting, mPACCdigit, mPACCtrailsB, GDTOTAL, BNTTOTAL, TAU, PTAU, Ventricles, WholeBrain, and ICV ($p > 0.05$). In the female group, sMCI and pMCI patients had a higher number of statistically insignificant comparisons than the male group. Features without differences between female sMCI and female pMCI patients were age, education level, CDRSB, RAVLT forgetting, DIGITSCOR, TRABSCOR, mPACCdigit, mPACCtrailsB, GDTOTAL, COPYSCOR, BNTTOTAL, TAU, PTAU, Ventricles, WholeBrain, and ICV ($p > 0.05$). In the comparison between male and female sMCI patients, only education level and ABETA42 were significantly different, while no other features showed differences in the comparison between male and female pMCI patients (Table 2).

Table 2. Statistical analysis between male and female MCI patients.

	<i>p</i> -Value			
	M-sMCI vs. M-pMCI	F-sMCI vs. F-pMCI	M-sMCI vs. F-sMCI	M-pMCI vs. F-pMCI
Demographic:				
Age	0.95 ^a	0.32 ^a	0.35 ^a	0.96 ^a
Education level	0.27 ^a	0.81 ^a	0.006^a	0.14 ^a
Biomarker:				
APOE4 (0/1/2)	0.005^b	0.02^b	0.14 ^b	0.58 ^b
Clinical scale:				
CDRSB	< 0.001^c	0.12 ^c	0.11 ^c	0.90 ^c
FAQ	< 0.001^c	< 0.001^c	0.34 ^c	0.93 ^c
Neuropsychological assessment:				
ADAS11	< 0.001^c	< 0.001^c	0.95 ^c	0.93 ^c
ADAS13	< 0.001^c	< 0.001^c	0.97 ^c	0.78 ^c
ADASQ4	< 0.001^c	< 0.001^c	0.67 ^c	0.94 ^c
MMSE	0.01^c	0.03^c	0.82 ^c	0.76 ^c
RAVLT_immediate	< 0.001^c	< 0.001^c	0.87 ^c	0.92 ^c
RAVLT_learning	< 0.001^c	0.03^c	0.88 ^c	0.46 ^c
RAVLT_forgetting	0.24 ^c	0.3 ^c	0.82 ^c	0.64 ^c
RAVLT_perc_forgetting	< 0.001^c	0.003^c	0.41 ^c	0.76 ^c
LDELTOTAL	< 0.001^c	0.004^c	0.28 ^c	0.28 ^c
DIGITSCOR	0.001^c	0.06 ^c	0.31 ^c	0.93 ^c
TRABSCOR	0.001^c	0.06 ^c	0.43 ^c	0.81 ^c
mPACCdigit	0.94 ^c	0.97 ^c	0.86 ^c	0.98 ^c
mPACCtrailsB	0.89 ^c	0.21 ^c	0.89 ^c	0.35 ^c
GDTOTAL	0.77 ^c	0.98 ^c	0.85 ^c	0.87 ^c
COPYSCOR	0.03^c	0.76 ^c	0.72 ^c	0.07 ^c
BNTTOTAL	0.09 ^c	0.33 ^c	0.22 ^c	0.07 ^c
CSF:				
ABETA42	< 0.001^c	0.003^c	0.015^c	0.43 ^c
TAU	0.41 ^c	0.17 ^c	0.97 ^c	0.25 ^c
PTAU	0.29 ^c	0.32 ^c	0.67 ^c	0.29 ^c
Neuroimaging:				
Ventricles	0.067 ^d	0.11 ^d	0.50 ^d	0.33 ^d
Hippocampus	< 0.001^d	0.01^d	0.053 ^d	0.27 ^d
WholeBrain	0.02 ^c	0.10 ^c	0.38 ^c	0.32 ^c
Entorhinal	< 0.001^d	< 0.001^d	0.75 ^d	0.91 ^d
Fusiform	< 0.001^d	0.005	0.71 ^d	0.96 ^d
MidTemp	< 0.001^d	0.001^d	0.54 ^d	0.47 ^d
ICV	0.70 ^c	0.45 ^c	0.063 ^c	0.22 ^c
FDG	< 0.001^c	< 0.001^c	0.66 ^c	0.94 ^c
HCI	< 0.001^c	< 0.001^c	0.96 ^c	0.34 ^c

In bold: significant result at $p < 0.05$. ^a One-way ANOVA; ^b Chi-square test; ^c ANCOVA with age in covariates; ^d ANCOVA with age and ICV in covariates. For abbreviations, see Appendix A.

Table 3 reports the results of hyperparameter tuning obtained through a randomized search. Optimal hyperparameter values provided a *c*-index (mean of 3-fold *cv* with 50 repetitions) of 0.839 for M-RSF and 0.804 for the F-RSF. Regarding the performance of best models, M-RSF reached high values of the *c*-index both on the test set and on the training set (0.873, 5-fold *cv*: 0.823 ± 0.04), while F-RSF had lower performance (0.791, 5-fold *cv*: 0.803 ± 0.04). The IBS score was 0.10 for M-RSF and 0.12 for F-RSF.

Table 3. Hyperparameters of Random Survival Forests (RSF) trained on male MCI (M-RSF) and female MCI (F-RSF) patients. Optimal values of hyperparameters were obtained through a randomized search with 3-fold cross-validation and 50 repetitions.

Hyperparameter	Parameter Distribution	Optimal Value	
		M-RSF	F-RSF
<i>max_depth</i>	integer from a reciprocal continuous random distribution in range (5, 50)	26	42
<i>min_node_size</i>	integer from a reciprocal continuous random distribution in range (1, 40)	34	19
<i>max_features</i>	['all', 'sqrt', 'log2']	'sqrt'	'sqrt'
<i>sample_size_pct</i>	[0.60, 0.70, 0.80, 0.90]	0.70	0.60

Figure 1 depicts the plots comparing the KM and predicted survival curves of test subjects (Figure 1a, male MCI patients; and Figure 1b, female MCI patients). M-RSF and F-RSF models showed a large overlap with the KM as demonstrated by low values of RMSE and median and mean absolute error, although a slight decrease in accuracy occurred as time progressed. The bottom plots in Figure 1a,b represent the IBS prediction error per timepoint, where both M-RSF and F-RSF models showed a global maximum at the 24th month but never exceeded the IBS cut-off (dotted red line).

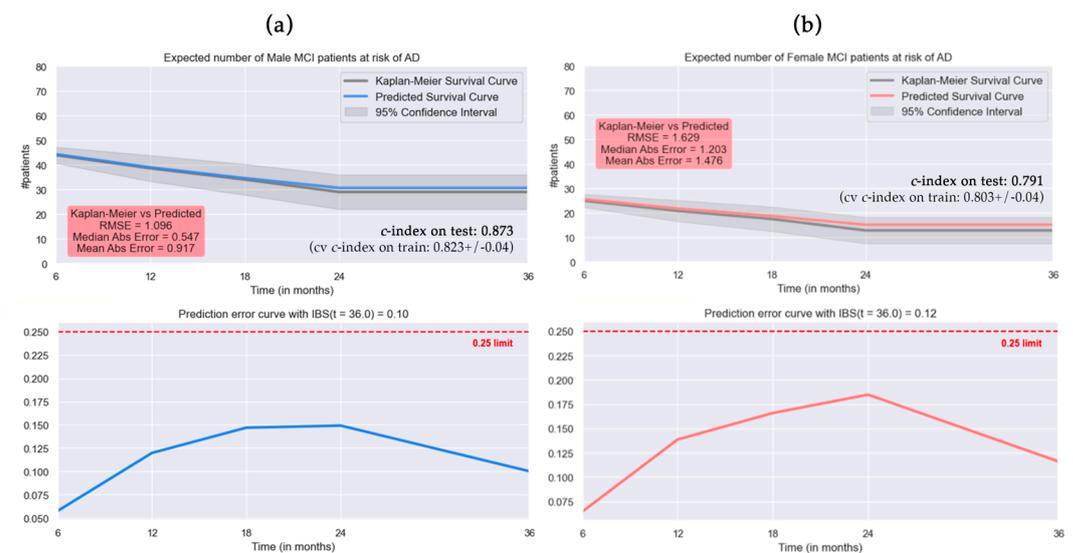


Figure 1. Performance per timepoint on the test set by RSF trained on (a) male MCI patients; (b) female MCI patients. Upper: plot over time of expected number of MCI patients at risk of conversion to AD, estimated survival curve by Kaplan-Meier in gray. Bottom: Integrated Brier error curve (IBS, critical cut-off limit of 0.25 in red). C-index on the test set, cross-validated (cv) c-index on the training set (mean \pm standard deviation), Root Mean Square Error (RMSE), and median and mean absolute error are also reported.

Global explanations on the male MCI training set and the female MCI training set are reported in Figures 2a and 2b, respectively. The rankings of features ordered by their prediction importance are—from the left to the right—RSF feature importance, permutation importance (mean value), and SHAP importance (mean absolute value). Regarding the M-RSF model (Figure 2a), the top three features in the three rankings were FDG, ABETA42, and HCI, while the top three of the F-RSF model were FDG, HCI, and FAQ. Figure 2c depicts the RBO curves of similarity between male and female rankings by increasing depth d (RSF M vs. F in plum, Perm M vs. F in violet, SHAP M vs. F in purple). All three pairwise comparisons had low overlap, with a maximum RBO value of 71.1% within the

top 12 variables for RSF M vs. F, 60.3% within the top 13 variables for Perm M vs. F, and 67.3% within the top 14 variables for SHAP M vs. F.

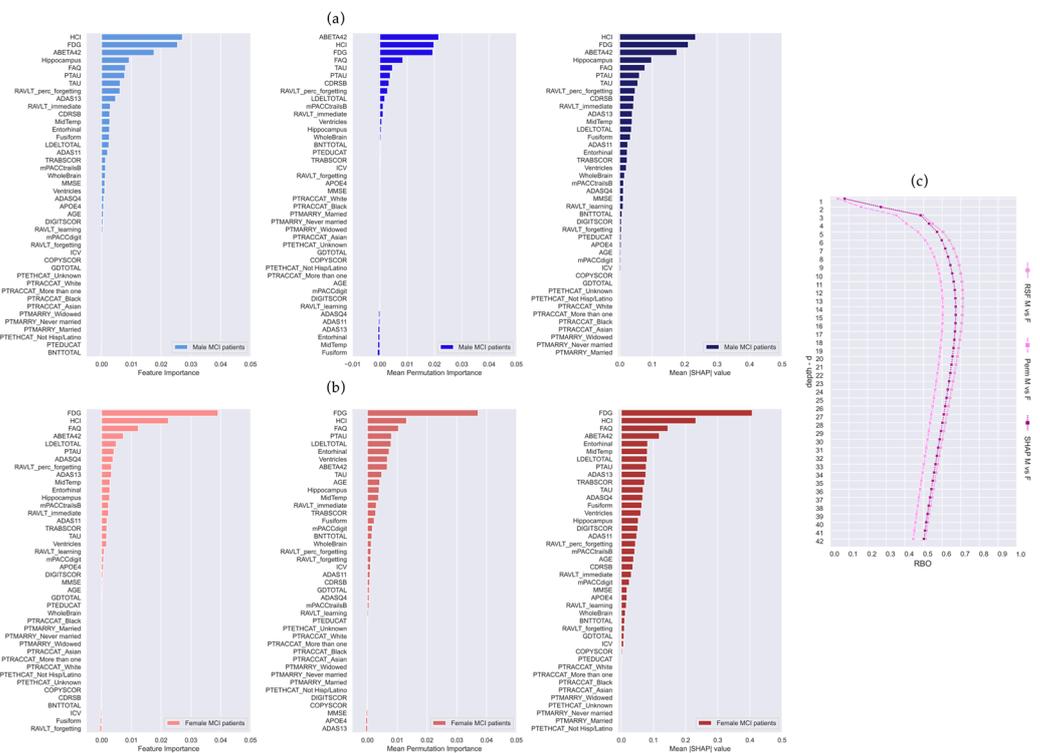


Figure 2. Random Survival Forests (RSF) global explanations trained on (a) male MCI (M-RSF) and (b) female MCI (F-RSF) patients. From left to right: RSF feature importance (VIMP), permutation importance (mean value), and SHAP importance (mean $|\text{SHAP}|$ value). (c) Rank-Biased Overlap (RBO) curves assessing the overlap between male and female (M vs. F) variable rankings at different numbers of the important features considered (depth d): RSF feature importance (in plum), mean permutation importance (in violet), and mean $|\text{SHAP}|$ importance (in purple).

Local explanations with SHAP on the test sets are reported in Figure 3. Similarly to global explanations, FDG and HCI were the top features in common between the M-RSF (Figure 3a) and F-RSF models (Figure 3b). The most evident differences in local explanations are the contributions of the hippocampus in M-RSF (+0.09) and RAVLT_perc_forgetting (+0.05), which were absent in F-RSF among the first features. On the contrary, the contribution of LDELTOTAL (+0.09) in the F-RSF model had a low impact on the M-RSF model (+0.03), and the TRABSCOR contribution (+0.08) in the F-RSF model was not among the most contributing variables in the M-RSF model.

The distributions of the risk score in progressing to AD predicted by the M-RSF on test MCI patients are reported as histograms in Figure 4a. Male pMCI test subjects were manually stratified into three risk grades: low range [1.39, 2] (in green), medium range [2, 2.6] (in orange), and high range [2.6, 3.47] (in red). The RSF survival functions of three randomly selected male pMCI subjects per risk grade are shown in Figure 4b. High-risk patient M-pMCI#1 had a risk score of 3.262, converted to AD at the 12th month, and the predicted survival probabilities at each timepoint were [0.89, 0.71, 0.57, 0.43, and 0.28]. Medium-risk patient M-pMCI#2 had a risk score of 1.962, converted to AD at the 24th month, and the predicted survival probabilities at each timepoint were [0.95, 0.83, 0.72, 0.63, and 0.50]. Low-risk patient M-pMCI#3 had a risk score of 1.395, converted to AD at the 36th month, and the predicted survival probabilities at each timepoint were [0.96, 0.87, 0.79, 0.73, and 0.62]. The M-sMCI subject—who does not convert to AD within 36 months—had a risk score of 0.459 and very high predicted survival probabilities per timepoint [0.98, 0.94, 0.92, 0.90, and 0.84].

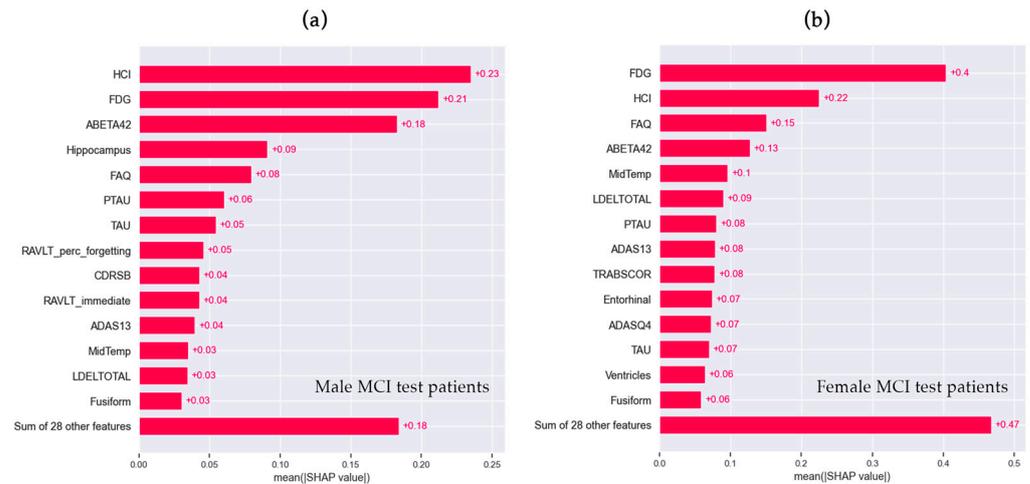


Figure 3. SHAP bar plots of (a) male MCI (M-RSF) and (b) female MCI test patients (F-RSF).

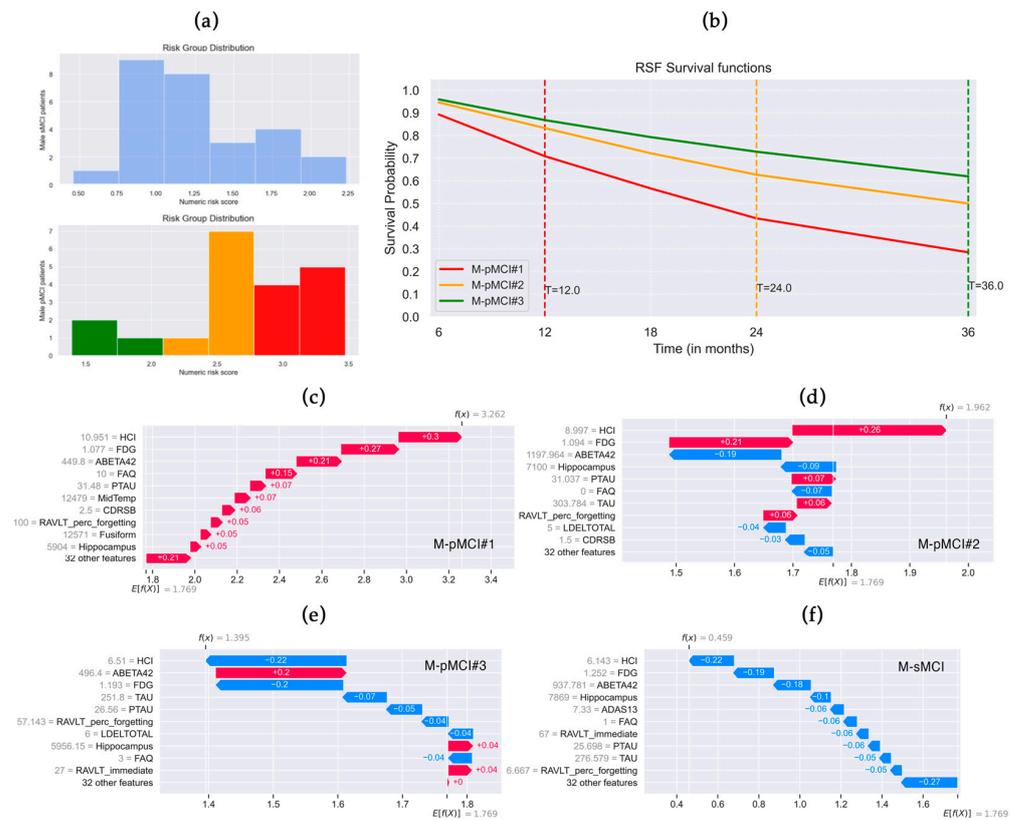


Figure 4. Random Survival Forests (RSF) local explanations trained on male MCI patients. (a) Histograms of male sMCI and pMCI patients' risk distribution predicted by M-RSF. Patients were stratified by risk grade: low (in green, range 1.39–2), medium (in orange, range 2–2.6), high (in red, range 2.6–3.47). (b) RSF survival functions of male pMCI patients per risk score: M-pMCI#1 high risk (score 3.262, converted to AD after 12 months), M-pMCI#2 medium risk (score 1.962, converted to AD after 24 months), M-pMCI#3 low risk (score 1.395, converted to AD after 36 months). SHAP waterfall plot of (c) patient M-pMCI#1, (d) patient M-pMCI#2, (e) patient M-pMCI#3, and (f) stable MCI patient who does not convert to AD within 36 months (M-sMCI, risk score 0.459). Features that decrease the risk are in blue, while those that increase it are in red. Average predicted risk $E[f(x)] = 1.769$. Actual value of the feature is in gray.

The distributions of the risk score in progressing to AD predicted by M-RSF on test patients are reported as histograms in Figure 4a. Male pMCI test subjects were manually stratified into three risk grades: low range [1.39, 2] (in green), medium range [2, 2.6] (in orange), and high range [2.6, 3.47] (in red). RSF survival functions of three randomly selected male pMCI subjects per risk grade are in Figure 4b. High-risk patient M-pMCI#1 had a risk score of 3.262, converted to AD at the 12th month, and the predicted survival probabilities at each timepoint were [0.89, 0.71, 0.57, 0.43, and 0.28]. Medium-risk patient M-pMCI#2 had a risk score of 1.962, converted to AD at the 24th month, and the predicted survival probabilities at each timepoint were [0.95, 0.83, 0.72, 0.63, and 0.50]. Low-risk patient M-pMCI#3 had a risk score of 1.395, converted to AD at the 36th month, and the predicted survival probabilities at each timepoint were [0.96, 0.87, 0.79, 0.73, and 0.62]. The M-sMCI subject—who does not convert to AD within 36 months—had a risk score of 0.459 and very high predicted survival probabilities per timepoint [0.98, 0.94, 0.92, 0.90, and 0.84].

In SHAP waterfall plots, a red arrow indicates that the feature increases the risk of conversion from MCI to AD, while a blue arrow indicates that the feature decreases it. The sum of all variable contributions provides the final SHAP value, which corresponds to the prediction risk score. SHAP waterfall plots of M-pMCI#1, M-pMCI#2, M-pMCI#3, and M-sMCI patients are reported in Figure 4c–f, where the actual value of each feature is also reported (in gray). Variables with the highest influence on risk prediction of M-pMCI#1, M-pMCI#2, M-pMCI#3, and M-sMCI subjects were FDG, ABETA42, and HCl (Figure 4c–f), as also found in global and local explanations (Figures 2a and 3a).

Regarding the risk prediction by the F-RSF on female MCI patients, histograms of its distributions are reported in Figure 5a. The stratification per risk grade of female pMCI test was as follows: low range [1.51, 2.3] (in green), medium range [2.3, 3.7] (in orange), and high range [3.7, 5.05] (in red). RSF survival functions of three randomly selected female pMCI subjects per risk grade are depicted in Figure 5b. High-risk patient F-pMCI#1 had a risk score of 4.683, converted to AD at the 6th month, and the predicted survival probabilities at each timepoint were [0.89, 0.62, 0.43, 0.30, and 0.22]. Medium-risk patient F-pMCI#2 had a risk score of 2.799, converted to AD at the 12th month, and the predicted survival probabilities at each timepoint were [0.95, 0.79, 0.67, 0.50, and 0.42].

Low-risk patient F-pMCI#3 had a risk score of 1.51, converted to AD at the 24th month, and the predicted survival probabilities at each timepoint were [0.98, 0.89, 0.84, 0.67, and 0.59]. F-sMCI subject—who does not convert to AD within 36 months—had a risk score of 0.528 and very high predicted survival probabilities per timepoint [0.99, 0.96, 0.94, 0.86, and 0.83].

From Figure 5c–f, which show the SHAP waterfall plots of F-pMCI#1, F-pMCI#2, F-pMCI#3, and F-sMCI patients, it can be noted that variables with the highest influence on risk prediction, such as FDG, HCl, and FAQ, similarly to global and local explanations (Figures 2b and 3b), as well as LDELTTOTAL, have a particularly evident contribution in high- and medium-risk patients (Figure 5c,d).

It is worth noting that the average predicted risk by SHAP for male MCI patients ($E[f(x)] = 1.769$) was lower than the average predicted risk for female MCI patients ($E[f(x)] = 2.534$).

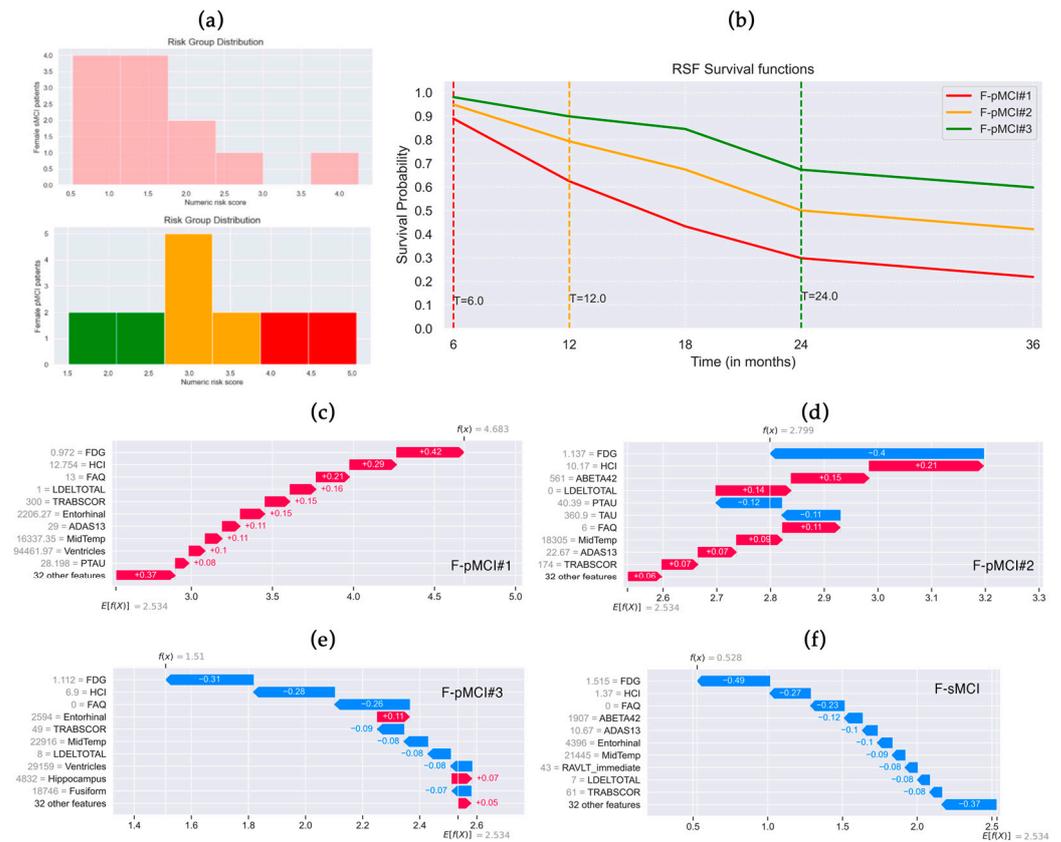


Figure 5. Random Survival Forests (RSF) local explanations trained on female MCI patients. **(a)** Histograms of female sMCI and pMCI patients' risk distribution predicted by F-RSF. Patients were stratified by risk grade: low (in green, range 1.51–2.3), medium (in orange, range 2.3–3.7), high (in red, range 3.7–5.05). **(b)** RSF survival functions of female pMCI patients per risk score: F-pMCI#1 high risk (score 4.683, converted to AD after 6 months), F-pMCI#2 medium risk (score 2.799, converted to AD after 12 months), F-pMCI#3 low risk (score 1.51, converted to AD after 24 months). SHAP waterfall plot of **(c)** patient F-pMCI#1, **(d)** patient F-pMCI#2, **(e)** patient F-pMCI#3, and **(f)** stable MCI patient who does not convert to AD within 36 months (F-sMCI, risk score 0.528). Features that decrease the risk are in blue, while those that increase it are in red. Average predicted risk $E[f(x)] = 2.534$. Actual value of the feature is in gray.

4. Discussion

The present study explored sex-specific differences in predicting the conversion risk from Mild Cognitive Impairment (MCI) to Alzheimer's Disease (AD) within 3 years using Random Survival Forests (RSF) and SHAP, a model-agnostic approach to boost explainability. The model trained only on male MCI patients (M-RSF) demonstrated optimal performance on both test and training sets, with an accuracy of 0.873 on the test set, while the female model (F-RSF) exhibited slightly lower performance (c-index of 0.791 on the test set), probably due to the lower sample size. Both models displayed low Integrated Brier Scores (IBS), indicating a precise prediction per timepoint. The comparison between Kaplan–Meier and RSF predicted survival curves revealed robust model performance, with high overlap. Despite some common influential features, differences in global and local explanations suggested sex-specific variations in the feature contribution to conversion-to-AD risk prediction. Of note, the average predicted risk for male MCI patients was observed to be lower than that for female MCI patients, underscoring potential sex-specific variations in the risk of conversion from MCI to AD.

Global and local explanations revealed four main common features as influential in both male and female models: FDG, HCI, ABETA42, and FAQ. Very few works have explored the differences between male and female brain glucose metabolism, and findings

about gender effect on FDG hypometabolism in normal aging as well as in AD progression are controversial [46]. Overall, a correlation between age and education and brain glucose metabolism was found in temporal and medial frontal regions in healthy adult subjects, but without any significant changes across sexes [46]. This correlation has also been confirmed in AD patients, although males and females exhibited different degrees of association involving different anatomical regions [46]. In our study, feature FDG is calculated as the mean average counting of angular, temporal, and posterior cingulate regions and HCI is a single measurement of FDG-PET hypometabolism (Appendix A), which cannot catch the specific anatomical regional differences between male and female cohorts, as in the past literature [46]. This could explain why FDG and HCI had the same influence in M-RSF and F-RSF models (Figures 2 and 3).

Beta-amyloid (1–42) (feature ABETA42) is a protein that decreases in both the plasma and cerebrospinal fluid (CSF) of dementia patients, which generally does not differ between the sexes [2], in cognitively unimpaired subjects, MCI, or AD patients. Our ML findings confirmed this absence of sex-driven changes in CSF biomarkers, where ABETA42 was among the first four features in the M-RSF and F-RSF, and PTAU and TAU were within the fifteen variables in global and local explanations (Figures 2 and 3).

Feature FAQ provides an assessment of daily living instrumental activities [47] and it is usually administered to the caregiver. In Berezuk et al. [4], FAQ resulted as a significant risk factor for both sexes, although in women it had a stronger effect, which was similarly evident in our results in local explanations (Figure 3). The slightly higher SHAP value of FAQ in the female model (+0.15) than in the male model (+0.08) to the increase in conversion risk score could be associated, as reported by Berezuk et al. [4], with different cognitive reserves across sexes.

In the neural network of learning and memory, the hippocampus plays the role of the central hub, thus its pathological changes contribute to memory impairment [48] resulting in dementia [5]. Differences in hippocampal volume between biological sexes were found in a study that explored amnesic MCI patients, where men had a larger hippocampal volume [49]. Burke et al. investigated the progression from normal cognition to MCI and to probable AD, and they showed that higher hippocampal volumes, or in other words, less hippocampal atrophy, decreased the risk of conversion to AD in women, but it had a more significant role in men [5]. In contrast to the previous studies that highlight hippocampus volume as an important risk factor in the progression from MCI to AD for both sexes [4], we found the hippocampus to have high SHAP values in the male model, while it was completely absent within the first fifteen most important variables in the female model (Figures 2 and 3).

Our two models, M-RSF and F-RSF, also differ in the contribution of two cognitive measures, LDELTOTAL and TRABSCOR, for which SHAP values were higher in the female model than in the male one. LDELTOTAL assesses verbal memory [50], and a plethora of works investigated sex differences in this cognitive domain in normal cognition as well as in dementia [2]. In detail, women score better than men in verbal tasks at all ages, and two studies [10,11] demonstrated that this sex-linked difference persists in the early stages of AD, although other works found that superior female performance in verbal memory tasks is lost in early stages [2]. However, the literature supports that women and men affected by dementia had similar verbal memory scores [2,10,11].

From the local explanations of female MCI patients (Figure 5), it could be noticed how low scores in LDELTOTAL were associated with higher SHAP values in the waterfall plot of high- and medium-risk pMCI patients (Figure 5c,d), while in the low-risk female pMCI and sMCI patient (Figure 5e,f), values of LDELTOTAL higher than 7 decreased the risk to progress to AD within 3 years. Similarly to LDELTOTAL, TRABSCOR, which evaluates executive functions [51], had a higher contribution in the female model than in the male one, where higher scores increased the risk and lower scores decreased it, as depicted in the explanation of female pMCI patients (Figure 5). These results regarding LDELTOTAL

and TRABSCOR broadly support the evidence about a stronger decline in memory and executive function in women who progress from MCI to AD compared with men [13].

Differences across sexes in the progression to AD have often been associated with genetic risk factors, like APOE- ϵ 4 allele [7,15]. For example, it has been found that the odds ratio of AD in women with one copy of the APOE- ϵ 4 allele is greater than that in men, as reported in a review about sexual dimorphism in Alzheimer's disease [15]. However, other studies did not support the role of APOE in differentiating men and women when progressing to dementia [15]. In our study, the feature APOE4 did not have high importance in the male or female model, resulting in a low ranking, as depicted in the global explanations (Figure 2). This is in accordance with the work of Burke et al. [5], who found that APOE- ϵ 4 did not contribute differentially to the progression to MCI or AD among men and women, and with Sohn et al. [3], who found no interaction effects between sex and APOE- ϵ 4.

Summarizing our findings, we find that, similarly to Berezuk et al. [4], the male RSF model had a greater number of CSF and neuroimaging features contributing to the risk prediction, and the female RSF model showed a greater number of neuropsychological tests involved in the progression to AD.

4.1. Limitations

We must recognize several limitations in the current work. The study relies on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), and as such, the findings may be specific to the characteristics of this particular cohort. Extrapolating the results to broader populations may require caution, as demographic and clinical characteristics can vary across different settings and populations. In detail, we must highlight that the ADNI sample is highly educated, thus we cannot state that in a less educated sample, the same variables would have the same influence. The study considers a fixed timeframe for prediction (36 months). The progression of MCI to AD may exhibit temporal dynamics that extend beyond this timeframe. The study includes a diverse set of biomarkers and features, but the exclusion of certain relevant biological factors or the absence of specific genetic markers might limit the comprehensiveness of the risk prediction models. Another limitation is that this study is not longitudinal, and for this reason, we cannot make causal inferences. Future research endeavors may benefit from collaborative efforts, standardized methodologies, and the integration of diverse datasets to unravel the underlying complexities in this field. Furthermore, our results should be confirmed on independent cohorts of men and women MCI patients to ensure reproducibility.

From a methodological point of view, disparities in the number of stable and progressive MCI patients within the two male and female datasets may have introduced bias. If one class is overrepresented or underrepresented, the model's performance could be influenced, potentially leading to feature importance rankings that are biased toward the majority class. Efforts to balance the dataset or explore alternative techniques to handle imbalances should be considered, including focusing on the distribution of event occurrences per timepoint. Another methodological issue relies on the fact that the RSF algorithm assumes proportional hazards over time, implying that the hazard ratios remain constant. However, this assumption may not hold in all situations, and violations could impact the accuracy of predictions. A more nuanced examination of time-dependent effects could provide additional insights, for example using model-agnostic methods tailored for survival analysis, such as survLIME [52] or survSHAP [53].

4.2. Clinical Implications

Despite the above-described limitations, we believe that our work has important clinical implications. Indeed, we proposed a novel ML approach to investigate sex-specific patterns in AD progression, and as far as we know, it is the first work of this type. Given our optimal performance and robust and stable explanations, we can state that the use of RSF, together with SHAP, represents a valuable tool for personalized interventions and

treatment strategies. The different features that SHAP revealed in the study reflect the multifactorial complexity of Alzheimer's disease, highlighting the importance of considering the interactions between genetic, environmental, and sex-related risk factors.

A detailed understanding of how these features differentially influence disease progression in men and women can aid in identifying individuals at high risk of progression from Mild Cognitive Impairment to AD, significantly enhance therapeutic and diagnostic approaches, and enable the formulation of targeted and personalized preventive interventions.

A key element for both sexes has been identified as the FAQ, which can clinically facilitate the identification of patients requiring support in daily living activities. According to Berezuk et al. [4], women tend to have more experience in these activities, thus developing a greater functional reserve. In line with this theory, one could hypothesize that a personalized medicine approach aimed at enhancing these capabilities and slowing functional decline in both sexes.

For males, important features have emerged as the hippocampus and the RAVLT. The critical role of the hippocampus in memory abilities is well-known, and the RAVLT also assesses this specific cognitive function. Burke et al. [5] found that a reduced hippocampal volume increases the risk of conversion and that stress can accelerate the process of atrophy. Consequently, targeted interventions could be considered to provide tools for managing stressful events in order to prevent atrophy and enhance mnemonic functions as protection against the decline of this cognitive ability.

In women, various cognitive domains such as executive functions, processing speed, and mental flexibility, assessed using the Trail Making Test (TRABSCOR), are relevant. In this case too, it could be useful to enhance these faculties with specific cognitive rehabilitation.

Finally, in addition to genetic factors and biomarkers, various cognitive functions appear to play a significant role; future efforts could aim at implementing targeted and personalized interventions to strengthen those abilities that seem to play a role in the conversion from MCI to AD, in order to prevent or slow cognitive decline.

5. Conclusions

In the present work, we applied Random Survival Forests, a Machine Learning technique for survival analysis, to shed light on how and whether feature contributions change according to sex when predicting the risk of progressing to Alzheimer's disease from Mild Cognitive Impairment. Our results confirmed that women have a higher risk of progressing to dementia; moreover, we highlighted peculiar sex-driven patterns of feature importance with different contributions to the decrease/increase in conversion-to-AD risk. In conclusion, the consistency of these findings with the existing literature underscores established trends, while discrepancies highlight the intricate and multifactorial nature of sex-specific differences in AD progression.

Author Contributions: Conceptualization, A.S., A.P., A.Q. (Aldo Quattrone) and A.Q. (Andrea Quattrone); methodology, A.S., F.A. (Federica Aracri) and F.A. (Fulvia Arcuri); software, A.S. and F.A. (Federica Aracri); formal analysis, A.S. and F.A. (Fulvia Arcuri); investigation, A.S. and A.P.; writing—original draft preparation, A.S. and A.P.; writing—review and editing, A.S., A.P., A.Q. (Aldo Quattrone) and A.Q. (Andrea Quattrone); visualization, A.S. and A.P.; supervision, A.Q. (Aldo Quattrone) and A.Q. (Andrea Quattrone); project administration, A.Q. (Aldo Quattrone) and A.Q. (Andrea Quattrone). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The data for this article were collected from ADNI. The ADNI study was conducted according to Good Clinical Practice guidelines, the Declaration of Helsinki, US 21CFR Part 50—Protection of Human Subjects, and Part 56—Institutional Review Boards, and pursuant to state and federal HIPAA regulations. Each participating site obtained ethical approval from their Institutional Review Board before commencing subject enrolment. More details can be found at adni.loni.usc.edu.

Informed Consent Statement: Informed written consent was obtained from all participants at each site. More details can be found at adni.loni.usc.edu.

Data Availability Statement: Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Accessed on 5 June 2023.

Acknowledgments: Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

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

Appendix A

List of demographics, clinical, neuropsychological, and neuroimaging features from ADNI:

- **Demographic variables:** Age, gender (PTGENDER), education levels (PTEDUCAT), ethnicity (PTETHCAT) and race (PTRACCAT) [54], and marital status (PTMARRY) [55].
- **Biomarker:** APOE4 allele genotype, i.e., presence of APOE gene that makes the APOE4 protein, associated with late-stage AD [56].
- **Clinical scales:**
 - Clinical Dementia Rating Sum of Boxes (CDRSB) is the sum score of the six domains used to accurately stage the severity of Alzheimer’s disease, dementia, and Mild Cognitive Impairment [57].
 - Functional Activities Questionnaire (FAQ): an informant-based clinician-administered questionnaire that assesses the functional daily living impairment in dementia [47]. The total score ranges from a minimum of 0 to a maximum of 30. A recommended cut-off of 9, indicating dependence on the caregiver in three or more activities, is suggested to identify impaired function and potential cognitive impairment [47].
- **Neuropsychological assessment:**
 - Alzheimer’s Disease Assessment Scale (ADAS), items 11 and 13, and delayed word recall (Q4) for assessing the memory, language, and praxis domains with 11 tasks, both subject-completed tests and observer-based assessments [58]. Total scores can range from 0 to 70, and higher scores (≥ 18) suggest more significant cognitive impairment [59].
 - Mini-Mental State Examination (MMSE): 30 questions on orientation, short-term memory retention, attention, short-term recall, and language to measure cognitive impairment and stage of the severity level [60]. The MMSE scores range from 0 to 30, and lower scores suggest a greater level of cognitive impairment [61].

- Rey Auditory Verbal Learning Test (RAVLT) [62] is a tool designed for assessing various aspects of cognitive function, including episodic declarative memory, immediate memory span, verbal learning, susceptibility to proactive and retroactive interferences, retention of information, and abilities related to recall and memory recognition. In detail, RAVLT_immediate evaluates immediate memory span (the sum of scores from the first five trials, i.e., Trials 1 to 5), RAVLT_learning measures learning ability and memorization of new information within a given time period (the score of Trial 5 minus the score of Trial 1), RAVLT_forgetting (the score of Trial 5 minus the score of the delayed recall) and RAVLT_percent_forgetting (RAVLT Forgetting divided by the score of Trial 5) estimate the amount of forgotten information [48].
- The total delayed recall score of the Logic Memory subtest of the Wechsler Memory Scale-Revised (LDELTOTAL), which assesses verbal memory. The correct responses to the items are summed, and the maximum score assigned is 25, for both immediate and delayed recall. Higher scores reflect greater verbal memory ability [50].
- Digit Symbol Substitution (DIGITSCOR) to evaluate attention, processing speed, and executive function [63]. The score is given by the total number of correct symbols executed within the allotted time.
- Trails B (TRABSCOR): time to complete part B of the Trail Making Test [64] that assesses different cognitive domains such as processing speed, sequencing, mental flexibility, and visual-motor skills. Higher scores indicate worse performance (i.e., longer completion times) [51].
- ADNI-modified Preclinical Alzheimer’s Cognitive Composite (PACC) with Digit Symbol Substitution (mPACCdigit), and with Trails B (mPACCtrailsB) that measure the first signs of cognitive decline [65].
- Geriatric Depression Scale (GDTOTAL) to identify depression in elderly subjects [66]. A higher score on the GDS indicates a higher level of depressive symptoms.
- Total score of the Clock Test (COPYSCOR) [66]. The Clock Drawing test evaluates various cognitive functions, including verbal understanding, memory, spatial knowledge, abstract thinking, planning, concentration, and visuoconstructive skills [67].
- Boston Naming Test (BNTTOTAL) assesses naming ability using 30 items [66].
- **Cerebrospinal fluid (CSF) biomarker:** $A\beta_{1-42}$ (ABETA42), total tau (TAU), phosphorylated tau (PTAU) concentrations [68].
- **Neuroimaging measures:** MRI volumes of the ventricles, hippocampus, whole brain, entorhinal cortex, fusiform, middle temporal gyrus (MidTemp), and total intracranial volume (ICV), calculated with Freesurfer [69]; average fluorodeoxyglucose positron emission tomography of angular, temporal, and posterior cingulate (FDG) [70]; hypometabolic convergence index (HCI) [71], an FDG-PET index that provides a single measurement of cerebral hypometabolism compared to the AD patients group.

References

1. Kim, S.; Kim, M.J.; Kim, S.; Kang, H.S.; Lim, S.W.; Myung, W.; Lee, Y.; Hong, C.H.; Choi, S.H.; Na, D.L.; et al. Gender differences in risk factors for transition from mild cognitive impairment to Alzheimer’s disease: A CREDOS study. *Compr. Psychiatry* **2015**, *62*, 114–122. [[CrossRef](#)] [[PubMed](#)]
2. Ferretti, M.T.; Iulita, M.F.; Cavado, E.; Chiesa, P.A.; Schumacher Dimech, A.; Santuccione Chadha, A.; Baracchi, F.; Girouard, H.; Misoch, S.; Giacobini, E.; et al. Sex differences in Alzheimer disease—The gateway to precision medicine. *Nat. Rev. Neurol.* **2018**, *14*, 457–469. [[CrossRef](#)] [[PubMed](#)]
3. Sohn, D.; Shpanskaya, K.; Lucas, J.E.; Petrella, J.R.; Saykin, A.J.; Tanzi, R.E.; Samatova, N.F.; Doraiswamy, P.M. Sex Differences in Cognitive Decline in Subjects with High Likelihood of Mild Cognitive Impairment due to Alzheimer’s disease. *Sci. Rep.* **2018**, *8*, 7490. [[CrossRef](#)]
4. Berezuk, C.; Khan, M.; Callahan, B.L.; Ramirez, J.; Black, S.E.; Zakzanis, K.K.; Alzheimer’s Disease Neuroimaging Initiative. Sex differences in risk factors that predict progression from mild cognitive impairment to Alzheimer’s dementia. *J. Int. Neuropsychol. Soc.* **2023**, *29*, 360–368. [[CrossRef](#)] [[PubMed](#)]

5. Burke, S.L.; Hu, T.; Fava, N.M.; Li, T.; Rodriguez, M.J.; Schuldiner, K.L.; Burgess, A.; Laird, A. Sex differences in the development of mild cognitive impairment and probable Alzheimer's disease as predicted by hippocampal volume or white matter hyperintensities. *J. Women Aging* **2019**, *31*, 140–164. [[CrossRef](#)]
6. Eliot, L.; Ahmed, A.; Khan, H.; Patel, J. Dump the “dimorphism”: Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neurosci. Biobehav. Rev.* **2021**, *125*, 667–697. [[CrossRef](#)]
7. Lin, K.A.; Choudhury, K.R.; Rathakrishnan, B.G.; Marks, D.M.; Petrella, J.R.; Doraiswamy, P.M.; Alzheimer's Disease Neuroimaging Initiative. Marked gender differences in progression of mild cognitive impairment over 8 years. *Alzheimer's Dement.* **2015**, *1*, 103–110. [[CrossRef](#)] [[PubMed](#)]
8. Artero, S.; Ancelin, M.-L.; Portet, F.; Dupuy, A.; Berr, C.; Dartigues, J.-F.; Tzourio, C.; Rouaud, O.; Poncet, M.; Pasquier, F. Risk profiles for mild cognitive impairment and progression to dementia are gender specific. *J. Neurol. Neurosurg. Psychiatry* **2008**, *79*, 979–984. [[CrossRef](#)]
9. Liew, T.M. Subjective cognitive decline, APOE e4 allele, and the risk of neurocognitive disorders: Age- and sex-stratified cohort study. *Aust. N. Z. J. Psychiatry* **2022**, *56*, 1664–1675. [[CrossRef](#)] [[PubMed](#)]
10. Sundermann, E.E.; Biegon, A.; Rubin, L.H.; Lipton, R.B.; Mowrey, W.; Landau, S.; Maki, P.M.; Alzheimer's Disease Neuroimaging Initiative. Better verbal memory in women than men in MCI despite similar levels of hippocampal atrophy. *Neurology* **2016**, *86*, 1368–1376. [[CrossRef](#)] [[PubMed](#)]
11. Sundermann, E.E.; Maki, P.; Biegon, A.; Lipton, R.B.; Mielke, M.M.; Machulda, M.; Bondi, M.W.; Alzheimer's Disease Neuroimaging Initiative. Sex-specific norms for verbal memory tests may improve diagnostic accuracy of amnesic MCI. *Neurology* **2019**, *93*, e1881–e1889. [[CrossRef](#)] [[PubMed](#)]
12. Park, J.C.; Lim, H.; Byun, M.S.; Yi, D.; Byeon, G.; Jung, G.; Kim, Y.K.; Lee, D.Y.; Han, S.H.; Mook-Jung, I. Sex differences in the progression of glucose metabolism dysfunction in Alzheimer's disease. *Exp. Mol. Med.* **2023**, *55*, 1023–1032. [[CrossRef](#)] [[PubMed](#)]
13. Fernández, A.; Cuesta, P.; Marcos, A.; Montenegro-Peña, M.; Yus, M.; Rodríguez-Rojo, I.C.; Bruña, R.; Maestú, F.; López, M.E. Sex differences in the progression to Alzheimer's disease: A combination of functional and structural markers. *GeroScience* **2023**, *46*, 2619–2640. [[CrossRef](#)] [[PubMed](#)]
14. Hebert, L.E.; Scherr, P.A.; McCann, J.J.; Beckett, L.A.; Evans, D.A. Is the risk of developing Alzheimer's disease greater for women than for men? *Am. J. Epidemiol.* **2001**, *153*, 132–136. [[CrossRef](#)] [[PubMed](#)]
15. Fisher, D.W.; Bennett, D.A.; Dong, H. Sexual dimorphism in predisposition to Alzheimer's disease. *Neurobiol. Aging* **2018**, *70*, 308–324. [[CrossRef](#)]
16. Ishwaran, H.; Kogalur, U.B. Consistency of Random Survival Forests. *Stat. Probab. Lett.* **2010**, *80*, 1056–1064. [[CrossRef](#)]
17. Ishwaran, H.; Kogalur, U.B.; Blackstone, E.H.; Lauer, M.S. Random survival forests. *Ann. Appl. Stat.* **2008**, *2*, 841–860. [[CrossRef](#)]
18. Sarica, A.; Cerasa, A.; Quattrone, A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Front. Aging Neurosci.* **2017**, *9*, 329. [[CrossRef](#)]
19. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
20. Jung, J.O.; Crnovrsanin, N.; Wirsik, N.M.; Nienhuser, H.; Peters, L.; Popp, F.; Schulze, A.; Wagner, M.; Muller-Stich, B.P.; Buchler, M.W.; et al. Machine learning for optimized individual survival prediction in resectable upper gastrointestinal cancer. *J. Cancer Res. Clin. Oncol.* **2022**, *149*, 1691–1702. [[CrossRef](#)]
21. Chen, Z.; Xu, H.; Li, Z.; Zhang, Y.; Zhou, T.; You, W.; Pan, K.; Li, W. Random survival forest: Applying machine learning algorithm in survival analysis of biomedical data. *Zhonghua Yu Fang Yi Xue Za Zhi Chin. J. Prev. Med.* **2021**, *55*, 104–109.
22. Sarica, A.; Aracri, F.; Bianco, M.G.; Vaccaro, M.G.; Quattrone, A.; Quattrone, A. Conversion from Mild Cognitive Impairment to Alzheimer's disease: A comparison of tree-based Machine Learning algorithms for Survival Analysis. In Proceedings of the International Conference on Brain Informatics, Hoboken, NJ, USA, 1–3 August 2023; pp. 179–190.
23. Sarica, A.; Aracri, F.; Bianco, M.G.; Arcuri, F.; Quattrone, A.; Quattrone, A.; Alzheimer's Disease Neuroimaging Initiative. Explainability of random survival forests in predicting conversion risk from mild cognitive impairment to Alzheimer's disease. *Brain Inform.* **2023**, *10*, 31. [[CrossRef](#)]
24. Cox, D.R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B Methodol.* **1972**, *34*, 187–202. [[CrossRef](#)]
25. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
26. Moncada-Torres, A.; van Maaren, M.C.; Hendriks, M.P.; Siesling, S.; Geleijnse, G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Sci. Rep.* **2021**, *11*, 6968. [[CrossRef](#)] [[PubMed](#)]
27. Webber, W.; Moffat, A.; Zobel, J. A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst. TOIS* **2010**, *28*, 1–38. [[CrossRef](#)]
28. Sarica, A.; Quattrone, A.; Quattrone, A. Introducing the Rank-Biased Overlap as Similarity Measure for Feature Importance in Explainable Machine Learning: A Case Study on Parkinson's Disease. In Proceedings of the Brain Informatics: 15th International Conference, BI 2022, Padua, Italy, 15–17 July 2022; pp. 129–139.
29. Sarica, A.; Di Fatta, G.; Cannataro, M. K-Surfer: A KNIME extension for the management and analysis of human brain MRI FreeSurfer/FSL data. In Proceedings of the Brain Informatics and Health: International Conference, BIH 2014, Warsaw, Poland, 11–14 August 2014; pp. 481–492.
30. Hancock, J.T.; Khoshgoftaar, T.M. Survey on categorical data for neural networks. *J. Big Data* **2020**, *7*, 28. [[CrossRef](#)]
31. Spooner, A.; Chen, E.; Sowmya, A.; Sachdev, P.; Kochan, N.A.; Trollor, J.; Brodaty, H. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Sci. Rep.* **2020**, *10*, 20410. [[CrossRef](#)] [[PubMed](#)]

32. Stekhoven, D.J.; Buhlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **2012**, *28*, 112–118. [[CrossRef](#)] [[PubMed](#)]
33. Aracri, F.; Bianco, M.G.; Quattrone, A.; Sarica, A. Imputation of missing clinical, cognitive and neuroimaging data of Dementia using missForest, a Random Forest based algorithm. In Proceedings of the 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), L'Aquila, Italy, 22–24 June 2023; pp. 684–688.
34. Aracri, F.; Bianco, M.G.; Quattrone, A.; Sarica, A. Impact of imputation methods on supervised classification: A multiclass study on patients with parkinson's disease and subjects with scans without evidence of dopaminergic deficit. In Proceedings of the 2023 International Workshop on Biomedical Applications, Technologies and Sensors (BATS), Catanzaro, Italy, 28–29 September 2023; pp. 28–32.
35. Orozco-Sanchez, J.; Trevino, V.; Martinez-Ledesma, E.; Farber, J.; Tamez-Peña, J. Exploring survival models associated with MCI to AD conversion: A machine learning approach. *bioRxiv* **2019**, 836510. [[CrossRef](#)]
36. Nakagawa, T.; Ishida, M.; Naito, J.; Nagai, A.; Yamaguchi, S.; Onoda, K.; Alzheimer's Disease Neuroimaging Initiative. Prediction of conversion to Alzheimer's disease using deep survival analysis of MRI images. *Brain Commun.* **2020**, *2*, fcaa057. [[CrossRef](#)]
37. Mirabnahrzham, G.; Ma, D.; Beaulac, C.; Lee, S.; Popuri, K.; Lee, H.; Cao, J.; Galvin, J.E.; Wang, L.; Beg, M.F.; et al. Predicting time-to-conversion for dementia of Alzheimer's type using multi-modal deep survival analysis. *Neurobiol. Aging* **2023**, *121*, 139–156. [[CrossRef](#)]
38. Musto, H.; Stamate, D.; Pu, I.; Stahl, D. Predicting Alzheimers Disease Diagnosis Risk over Time with Survival Machine Learning on the ADNI Cohort. *arXiv* **2023**, arXiv:2306.10326.
39. Song, S.; Asken, B.; Armstrong, M.J.; Yang, Y.; Li, Z. Predicting Progression to Clinical Alzheimer's Disease Dementia Using the Random Survival Forest. *J. Alzheimer's Dis.* **2023**, *95*, 535–548. [[CrossRef](#)]
40. Wright, M.N.; Dankowski, T.; Ziegler, A. Unbiased split variable selection for random survival forests using maximally selected rank statistics. *Stat. Med.* **2017**, *36*, 1272–1284. [[CrossRef](#)] [[PubMed](#)]
41. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
42. Uno, H.; Cai, T.; Pencina, M.J.; D'Agostino, R.B.; Wei, L.-J. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat. Med.* **2011**, *30*, 1105–1117. [[CrossRef](#)]
43. Steyerberg, E.W.; Vickers, A.J.; Cook, N.R.; Gerdts, T.; Gonen, M.; Obuchowski, N.; Pencina, M.J.; Kattan, M.W. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* **2010**, *21*, 128–138. [[CrossRef](#)]
44. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [[CrossRef](#)]
45. Arya, V.; Bellamy, R.K.; Chen, P.-Y.; Dhurandhar, A.; Hind, M.; Hoffman, S.C.; Houde, S.; Liao, Q.V.; Luss, R.; Mojsilović, A. AI Explainability 360 Toolkit. In Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD), Bangalore, India, 2–4 January 2021; pp. 376–379.
46. Malpetti, M.; Ballarini, T.; Presotto, L.; Garibotto, V.; Tettamanti, M.; Perani, D.; Alzheimer's Disease Neuroimaging Initiative (ADNI) Database; Network for Efficiency and Standardization of Dementia Diagnosis (NEST-DD) Database. Gender differences in healthy aging and Alzheimer's Dementia: A (18) F-FDG-PET study of brain and cognitive reserve. *Hum. Brain Mapp.* **2017**, *38*, 4212–4227. [[CrossRef](#)]
47. Pfeffer, R.I.; Kurosaki, T.T.; Harrah, C.H., Jr.; Chance, J.M.; Filos, S. Measurement of functional activities in older adults in the community. *J. Gerontol.* **1982**, *37*, 323–329. [[CrossRef](#)]
48. Sarica, A.; Vasta, R.; Novellino, F.; Vaccaro, M.G.; Cerasa, A.; Quattrone, A.; Alzheimer's Disease Neuroimaging Initiative. MRI asymmetry index of hippocampal subfields increases through the continuum from the mild cognitive impairment to the Alzheimer's disease. *Front. Neurosci.* **2018**, *12*, 576. [[CrossRef](#)]
49. Bai, F.; Zhang, Z.; Watson, D.R.; Yu, H.; Shi, Y.; Zhu, W.; Wang, L.; Yuan, Y.; Qian, Y. Absent gender differences of hippocampal atrophy in amnesic type mild cognitive impairment. *Neurosci. Lett.* **2009**, *450*, 85–89. [[CrossRef](#)] [[PubMed](#)]
50. Bolognani, S.A.P.; Miranda, M.C.; Martins, M.; Rzezak, P.; Bueno, O.F.A.; de Camargo, C.H.P.; Pompeia, S. Development of alternative versions of the Logical Memory subtest of the WMS-R for use in Brazil. *Dement. Neuropsychol.* **2015**, *9*, 136–148. [[CrossRef](#)] [[PubMed](#)]
51. Bowie, C.R.; Harvey, P.D. Administration and interpretation of the Trail Making Test. *Nat. Protoc.* **2006**, *1*, 2277–2281. [[CrossRef](#)] [[PubMed](#)]
52. Kovalev, M.S.; Utkin, L.V.; Kasimov, E.M. SurvLIME: A method for explaining machine learning survival models. *Knowl.-Based Syst.* **2020**, *203*, 106164. [[CrossRef](#)]
53. Krzyżiński, M.; Spytek, M.; Baniecki, H.; Biecek, P. SurvSHAP (t): Time-dependent explanations of machine learning survival models. *Knowl.-Based Syst.* **2023**, *262*, 110234. [[CrossRef](#)]
54. Wright, C.B.; DeRosa, J.T.; Moon, M.P.; Strobino, K.; DeCarli, C.; Cheung, Y.K.; Assuras, S.; Levin, B.; Stern, Y.; Sun, X. Race/ethnic disparities in mild cognitive impairment and dementia: The Northern Manhattan Study. *J. Alzheimer's Dis.* **2021**, *80*, 1129–1138. [[CrossRef](#)]
55. Parra Bautista, Y.J.; Messeha, S.S.; Theran, C.; Aló, R.; Yedjou, C.; Adankai, V.; Babatunde, S.; Alzheimer's Disease Prediction of Longitudinal Evolution. Marital Status of Never Married with Rey Auditory Verbal Learning Test Cognition Performance Is Associated with Mild Cognitive Impairment. *Appl. Sci.* **2023**, *13*, 1656. [[CrossRef](#)]

56. Lambert, J.C.; Ibrahim-Verbaas, C.A.; Harold, D.; Naj, A.C.; Sims, R.; Bellenguez, C.; DeStafano, A.L.; Bis, J.C.; Beecham, G.W.; Grenier-Boley, B.; et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. *Nat. Genet.* **2013**, *45*, 1452–1458. [[CrossRef](#)]
57. O’Bryant, S.E.; Lacritz, L.H.; Hall, J.; Waring, S.C.; Chan, W.; Khodr, Z.G.; Massman, P.J.; Hobson, V.; Cullum, C.M. Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer’s coordinating center database. *Arch. Neurol.* **2010**, *67*, 746–749. [[CrossRef](#)] [[PubMed](#)]
58. Grassi, M.; Rouleaux, N.; Caldirola, D.; Loewenstein, D.; Schruers, K.; Perna, G.; Dumontier, M.; Alzheimer’s Disease Neuroimaging Initiative. A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion From Mild Cognitive Impairment to Alzheimer’s Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures. *Front. Neurol.* **2019**, *10*, 756. [[CrossRef](#)]
59. Rockwood, K.; Fay, S.; Gorman, M.; Carver, D.; Graham, J.E. The clinical meaningfulness of ADAS-Cog changes in Alzheimer’s disease patients treated with donepezil in an open-label trial. *BMC Neurol.* **2007**, *7*, 26. [[CrossRef](#)]
60. Folstein, M.F.; Folstein, S.E.; McHugh, P.R. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **1975**, *12*, 189–198. [[CrossRef](#)]
61. Chun, C.T.; Seward, K.; Patterson, A.; Melton, A.; MacDonald-Wicks, L. Evaluation of Available Cognitive Tools Used to Measure Mild Cognitive Decline: A Scoping Review. *Nutrients* **2021**, *13*, 3974. [[CrossRef](#)]
62. Estévez-González, A.; Kulisevsky, J.; Boltes, A.; Otermín, P.; García-Sánchez, C. Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer’s disease: Comparison with mild cognitive impairment and normal aging. *Int. J. Geriatr. Psychiatry* **2003**, *18*, 1021–1028. [[CrossRef](#)] [[PubMed](#)]
63. Jaeger, J. Digit Symbol Substitution Test: The Case for Sensitivity Over Specificity in Neuropsychological Testing. *J. Clin. Psychopharmacol.* **2018**, *38*, 513–519. [[CrossRef](#)] [[PubMed](#)]
64. Reitan, R.M. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept. Mot. Ski.* **1958**, *8*, 271–276. [[CrossRef](#)]
65. Donohue, M.C.; Sperling, R.A.; Salmon, D.P.; Rentz, D.M.; Raman, R.; Thomas, R.G.; Weiner, M.; Aisen, P.S. The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. *JAMA Neurol.* **2014**, *71*, 961–970. [[CrossRef](#)] [[PubMed](#)]
66. Battista, P.; Salvatore, C.; Castiglioni, I. Optimizing Neuropsychological Assessments for Cognitive, Behavioral, and Functional Impairment Classification: A Machine Learning Study. *Behav. Neurol.* **2017**, *2017*, 1850909. [[CrossRef](#)] [[PubMed](#)]
67. Pinto, E.; Peters, R. Literature review of the Clock Drawing Test as a tool for cognitive screening. *Dement. Geriatr. Cogn. Disord.* **2009**, *27*, 201–213. [[CrossRef](#)]
68. Shaw, L.M.; Vanderstichele, H.; Knapik-Czajka, M.; Clark, C.M.; Aisen, P.S.; Petersen, R.C.; Blennow, K.; Soares, H.; Simon, A.; Lewczuk, P.; et al. Cerebrospinal fluid biomarker signature in Alzheimer’s disease neuroimaging initiative subjects. *Ann. Neurol.* **2009**, *65*, 403–413. [[CrossRef](#)] [[PubMed](#)]
69. Dale, A.M.; Fischl, B.; Sereno, M.I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **1999**, *9*, 179–194. [[CrossRef](#)]
70. Landau, S.M.; Harvey, D.; Madison, C.M.; Reiman, E.M.; Foster, N.L.; Aisen, P.S.; Petersen, R.C.; Shaw, L.M.; Trojanowski, J.Q.; Jack, C.R., Jr.; et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* **2010**, *75*, 230–238. [[CrossRef](#)] [[PubMed](#)]
71. Chen, K.; Ayutyanont, N.; Langbaum, J.B.; Fleisher, A.S.; Reschke, C.; Lee, W.; Liu, X.; Bandy, D.; Alexander, G.E.; Thompson, P.M.; et al. Characterizing Alzheimer’s disease using a hypometabolic convergence index. *Neuroimage* **2011**, *56*, 52–60. [[CrossRef](#)] [[PubMed](#)]

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