Journal of
*Personalized
Medicine*

Article

Random Forest Model in the Diagnosis of Dementia Patients with Normal Mini-Mental State Examination Scores

Jie Wang [1], Zhuo Wang [2], Ning Liu [2], Caiyan Liu [1], Chenhui Mao [1], Liling Dong [1], Jie Li [1], Xinying Huang [1], Dan Lei [1], Shanshan Chu [1], Jianyong Wang [2,*] and Jing Gao [1,*]

[1] Department of Neurology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing 100730, China; wangjie_smu@163.com (J.W.); liucy-pumch@163.com (C.L.); maochenhui@pumch.cn (C.M.); sophie_d@163.com (L.D.); jielicathy@126.com (J.L.); hxypumch@163.com (X.H.); ld94616@163.com (D.L.); chuss9486@163.com (S.C.)

[2] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; wang-z18@mails.tsinghua.edu.cn (Z.W.); victorliucs@gmail.com (N.L.)

* Correspondence: jianyong@tsinghua.edu.cn (J.W.); gj107@163.com (J.G.); Tel.: +86-10-62789150 (J.W.); +86-13011809777 (J.G.)

Abstract: Background: Mini-Mental State Examination (MMSE) is the most widely used tool in cognitive screening. Some individuals with normal MMSE scores have extensive cognitive impairment. Systematic neuropsychological assessment should be performed in these patients. This study aimed to optimize the systematic neuropsychological test battery (NTB) by machine learning and develop new classification models for distinguishing mild cognitive impairment (MCI) and dementia among individuals with MMSE $\geq$ 26. **Methods:** 375 participants with MMSE $\geq$ 26 were assigned a diagnosis of cognitively unimpaired (CU) (n = 67), MCI (n = 174), or dementia (n = 134). We compared the performance of five machine learning algorithms, including logistic regression, decision tree, SVM, XGBoost, and random forest (RF), in identifying MCI and dementia. **Results:** RF performed best in identifying MCI and dementia. Six neuropsychological subtests with high-importance features were selected to form a simplified NTB, and the test time was cut in half. The AUC of the RF model was 0.89 for distinguishing MCI from CU, and 0.84 for distinguishing dementia from nondementia. **Conclusions:** This simplified cognitive assessment model can be useful for the diagnosis of MCI and dementia in patients with normal MMSE. It not only optimizes the content of cognitive evaluation, but also improves diagnosis and reduces missed diagnosis.

Keywords: machine learning; dementia; cognitive dysfunction; neuropsychological tests; mental status and dementia tests

check for updates

Citation: Wang, J.; Wang, Z.; Liu, N.; Liu, C.; Mao, C.; Dong, L.; Li, J.; Huang, X.; Lei, D.; Chu, S.; et al. Random Forest Model in the Diagnosis of Dementia Patients with Normal Mini-Mental State Examination Scores. *J. Pers. Med.* **2022**, *12*, 37. https://doi.org/10.3390/jpm12010037

Academic Editor: Niels Bergsland

Received: 23 November 2021
Accepted: 31 December 2021
Published: 4 January 2022

1. Introduction

The prevalence of dementia is rising with the aging of the population, affecting the quality of life and increasing the burden on society and the family [1]. Mild cognitive impairment (MCI) is considered a transitional stage between normal aging and dementia, with a higher risk of developing dementia. The diagnosis of MCI and dementia early has prognostic value [2,3].

The most widely used screening tool for dementia is the Mini-Mental State Examination (MMSE) [4], a 30-point instrument that assesses several domains including orientation, attention, language, memory, and executive function. MMSE has good sensitivity and specificity for detecting dementia. Creavin et al. reported that in the community, a pooled sensitivity of 0.85 and specificity of 0.90 at a cut point of 24, and sensitivity of 0.87 and specificity of 0.82 at a cut point of 25 [5]. Pooled estimates of 15 studies showed a sensitivity of 0.89 and specificity of 0.89 at a cut point of 23 or less or 24 or less [6]. However, the sensitivity (0.20–0.93) and specificity (0.48–0.93) to detect MCI vary significantly in different studies, meaning less consistent estimates for test accuracy [6]. Thus, its ability to distinguish between cognitively impaired subjects and cognitively unimpaired (CU) adults is limited [7–9], leading to the possibility that some patients with normal MMSE scores but cognitive impairment may be missed.

For these individuals with normal MMSE scores, a more comprehensive cognitive assessment is needed. The systematic neuropsychological test battery (NTB) designed by the Peking Union Medical College Hospital (PUMCH) consists of more than 20 subtests to evaluate five cognitive domains: executive function, visuospatial ability, language, memory, and abstract reasoning and calculation [10]. It takes into account Chinese culture and language and is suitable for the Chinese elderly to detect MCI and dementia. All these subtests have been used and validated in the Chinese population, and normative population data were available. However, administering such a comprehensive battery is time-consuming.

Recent studies had shown that machine learning (ML) exhibited excellent performance in identifying MCI and dementia [11–17], but these mostly used biomarker data such as neuroimaging and CSF components that were expensive technologies [12,13,16]. ML diagnostic models based on cognitive data were gradually being applied [11,15,18,19]. Random forest (RF), an ensemble ML method based on a set of decision trees, has positive significance in processing complex neuropsychological data and excellent predictive performance for the diagnosis of cognitive impairment [15]. Using the feature selection method in RF, we can determine the importance of features and delete insignificant ones, thereby reducing the complexity of the NTB.

Therefore, the purpose of this study was to use RF to simplify the NTB and shorten evaluation time. Several important neuropsychological subtests were selected, and new RF models were developed to classify CU, MCI, and dementia for people with normal MMSE scores.

2. Materials and Methods

2.1. Participants

375 (67 CU adults, 174 MCI patients and 134 dementia patients) participants were enrolled consecutively from the PUMCH dementia cohort, the Dementia Clinic of the Department of Neurology of PUMCH between May 2009 to April 2021. They received a detailed clinical evaluation that included medical history taking, physical and neurological examinations, a systemic of neuropsychological tests, laboratory testing, and neuroimaging studies (head CT or MRI). The inclusion criteria included MMSE score $\geq$ 26, with normal function in motor, sensory, balance, reflex, and ability to complete all neuropsychological tests. Patients with significant functional disabilities, a history of major psychiatric illness, or any other central nervous system disorders other than cognitive impairment were excluded.

2.2. Neuropsychological Examinations

Cognitive tests included the Chinese version of the MMSE [20] and the PUMCH version of Montreal cognitive assessment (MoCA-P) [10]. Previous studies had shown that MMSE scores were influenced by age, gender, and particularly years of education [9]. Several studies that investigated the normative data of the MMSE in the Chinese population got different optimal cut-off points ranging from 19 to 26 for dementia screening [9,21,22]. In this study, we defined $\geq$26 points as normal MMSE scores. A Chinese version of ADL was used to determine impairment in everyday functioning [23], which was revised and supplemented according to the scale of Lawton and Brody [24], consisting of eight activities focused on instrumental ADL (IADL) (including using telephone, shopping, food preparation, housekeeping, laundry, transportation, managing medications, and handling finances) and 12 activities focused on the basic ADL (BADL) (e.g., dressing, bathing, eating, getting in or out of bed, using the toilet and so on). Each item of ADL range from 1 to 4 (1 = can do it myself, 2 = have some difficulty doing but can still do it by myself, 3 = need help to do it, 4 = cannot do it at all). The lowest ADL score was 20 points, indicating that the patient's ability was completely normal, and the highest was 80 points. The Hospital Anxiety and Depression (HAD) scale was used to screen for anxiety and depression among patients [25]. Participants were administered the above assessments as the diagnostic neuropsychological measures.

All subjects underwent the systemic NTB to evaluate five cognitive domains. These were: (1) Executive function: category verbal fluency [26], the digit symbol test (DST) [27], the trail making test A (TMT A) [28], the clock drawing test [8], paired-associate learning (PAL) of The Clinical Memory Test [29], the block design test of the Aphasia Battery of Chinese [30], and modified Luria three-step task [31]; (2) Visuospatial ability: the block design test and figure copying of the Aphasia Battery of Chinese [30], the copy of a modified Rey-Osterrieth figure [32], and gestures imitation; (3) Language: several subtests of the Aphasia Battery of Chinese including spontaneous speech, auditory comprehension, repetition, and naming [30]; (4) Memory: PAL, the logical memory test (LMT) of the modified Wechsler Memory Scale [33], and the auditory verbal learning test-Huashan version (AVLT-H) [34] were used to assess verbal memory. Nonverbal memory was measured by the modified Rey-Osterreith with a 10-min free recall; and (5) Abstract reasoning and calculation: subtests of the Wechsler Adult Intelligence Scale including similarities and calculations [27]. All subtests of NTB were not used to assist in making the clinical diagnosis of MCI or dementia, but as screening tests for machine learning.

2.3. Diagnostic Criteria

A clinical diagnosis of CU, MCI, or dementia was made based on all available information including clinical history and neuropsychological measures. MCI and dementia were diagnosed based on clinical judgment and/or on cognitive test performance according to the clinical criteria of the National Institute on Aging and the Alzheimer's Association (NIA-AA) guidelines [35–37]. Dementia diagnostic criteria included the following: evidence of decline from a previous level of cognitive performance; cognitive impairment diagnosed through history-taking and/or cognitive assessment; evidence of impairment in activities in daily living (ADL score > 23, IADL score > 11). MCI diagnostic criteria included the following: evidence of decline from a previous level of cognitive performance; no evidence of impairment in activities in daily living (ADL score $\leq$ 23, IADL score $\leq$ 11); not meeting the criteria for dementia. Subjects in the CU group had no or only mild cognitive decline, and neuropsychological tests were in the normal range.

2.4. Statistical Analysis

Continuous variables were described as mean ± standard deviation (M ± SD) and categorical variables as numbers and percentages (*n*, %). ANOVA with Bonferroni post-hoc tests or chi-square analysis was applied to detect significant differences between the different subgroups. A *p*-value of <0.05 was considered statistically significant. Statistical analysis was performed by SPSS version 24.0 software (Chicago, IL, USA).

2.5. Machine Learning

We manually extracted 64 features, including basic demographic information (sex, age, education years, etc.) and neuropsychological scores of NTB. All features were listed in Supplementary Table S1. At first, we used RF to calculate the importance of all features and perform feature selection. We tested all features with five-fold cross-validation and used mean area under the curves (AUC) as the performance metric. Different features had different importance in diagnosing dementia. Selecting the top-ranked features and filtering out the bottom-ranked features can simplify the classification process.

Next, other classification models, including logistic regression, decision tree, SVM, and XGBoost were trained and compared with RF. The performance of various models was evaluated by accuracy, precision, recall, F1 score, and AUC.

After selecting the features with high importance or the features we were interested in, 5-fold cross-validation was employed to train classification models, and the corresponding receiver operating characteristic (ROC) curves were also plotted. For each model, we got three ROC curves to distinguish CU, MCI, and dementia. The performance of each model effectiveness was evaluated using the mean ROC of the 5-fold cross-validation, the mean AUC, sensitivity, and specificity. AUC takes a value between 0 and 1, where AUC = 1 represents perfect diagnostic accuracy. Sensitivity is the true positive rate and specificity is the true negative rate. Sensitivity and specificity were calculated according to the maximal Youden's Index (sensitivity + specificity−1).

Classification models were built by using Python 3.7.9 with the package scikit-learn 0.23.2.

3. Results

3.1. Participants' Characteristics

375 participants, 161 men and 214 women, aged 65.51 ± 11.46 years, were recruited. Of these, 67 (17.9%) were CU, 174 (46.4%) had MCI, and 134 (35.7%) had dementia. Table 1 shows the baseline demographic and cognitive profiles of the three groups. The dementia group was significantly older than the MCI group, and years of education were significantly higher in the CUs than in the subjects with MCI and dementia. There was no significant gender difference between the three groups. For MMSE and MoCA-P scores, CU > MCI > dementia ($p < 0.001$); for ADL, IADL and BADL, CU = MCI < dementia.

Table 1. Comparison of demographic details and cognitive data among the groups.

	Total *n* = 375	**CU *n* = 67**	**MCI *n* = 174**	**Dementia *n* = 134**	**χ^2/F [a]**	**Post Hoc Tests [b,c]**
Age (years)	65.51 ± 11.46	63.24 ± 12.00	64.16 ± 11.61	68.41 ± 10.44	7.05 **	1 = 2 < 3
Gender (% female)	214 (57.1%)	43 (64.2%)	99 (56.9%)	72 (53.7%)	1.99	-
Education years	12.28 ± 3.91	13.88 ± 3.34	11.93 ± 3.98	11.96 ± 3.92	6.63 **	1 > 2 = 3
MMSE	27.80 ± 1.31	28.70 ± 1.17	27.95 ± 1.22	27.15 ± 1.17	40.42 **	1 > 2 > 3
MoCA-P	24.35 ± 3.08	27.18 ± 1.65	24.64 ± 2.77	22.54 ± 2.82	71.52 **	1 > 2 > 3
ADL	24.34 ± 4.57	21.78 ± 2.05	22.26 ± 2.53	28.31 ± 4.85	136.32 **	1 = 2 < 3
IADL	11.39 ± 3.30	9.45 ± 1.82	9.82 ± 1.99	14.39 ± 3.11	160.18 **	1 = 2 < 3
BADL	12.95 ± 1.92	12.33 ± 0.73	12.45 ± 1.01	13.93 ± 2.69	31.29 **	1 = 2 < 3
HAD-anxiety	4.66 ± 3.38	4.45 ± 3.15	4.48 ± 3.52	5.01 ± 3.29	1.06	-
HAD-depression	4.88 ± 3.48	4.50 ± 3.50	4.46 ± 3.44	5.64 ± 3.41	4.86 *	1 = 2 < 3

Data were shown as mean ± standard deviation (SD) or frequency (percentage, %). [a] Test statistic: F = one-way ANOVA value; χ^2 = chi-square test value. [b] 1: CU group; 2: MCI group; and 3: Dementia group. [c] Pairwise comparisons among the three groups of subjects were conducted using the Bonferroni post hoc tests. * $p < 0.05$; ** $p < 0.001$. Abbreviations: ADL = Activities of Daily Living; BADL = Basic ADL; CU = Cognitively Unimpaired; HAD = Hospital Anxiety and Depression; IADL = Instrumental ADL; MCI = Mild Cognitive Impairment; MMSE = Mini-Mental State Examination; MoCA-P = PUMCH version of Montreal Cognitive Assessment; PUMCH = Peking Union Medical College Hospital.

3.2. Assessment of Feature Importance

We extracted all features (64 features) into the RF classification model and calculated feature importance. ROC analysis for the detection of MCI and dementia and the top 20 features were shown in Figure 1. ROC-AUC of all features for distinguishing MCI from CU was 0.90 ± 0.04, sensitivity and specificity were 0.89 and 0.77 (Figure 1A), and the most important feature was PAL-T (total score of the three learning trials of PAL) (Figure 1B). ROC-AUC of all features for distinguishing dementia from MCI was 0.81 ± 0.07, sensitivity and specificity were 0.75 and 0.74 (Figure 1C), and the most important feature was AVLT N5 (the fifth long-delayed free recall trial of AVLT-H) (Figure 1D). ROC-AUC of all features for distinguishing dementia from non-dementia was 0.87 ± 0.04, sensitivity and specificity were 0.90 and 0.73 (Figure 1E), and the most important feature was AVLT N5 (Figure 1F).

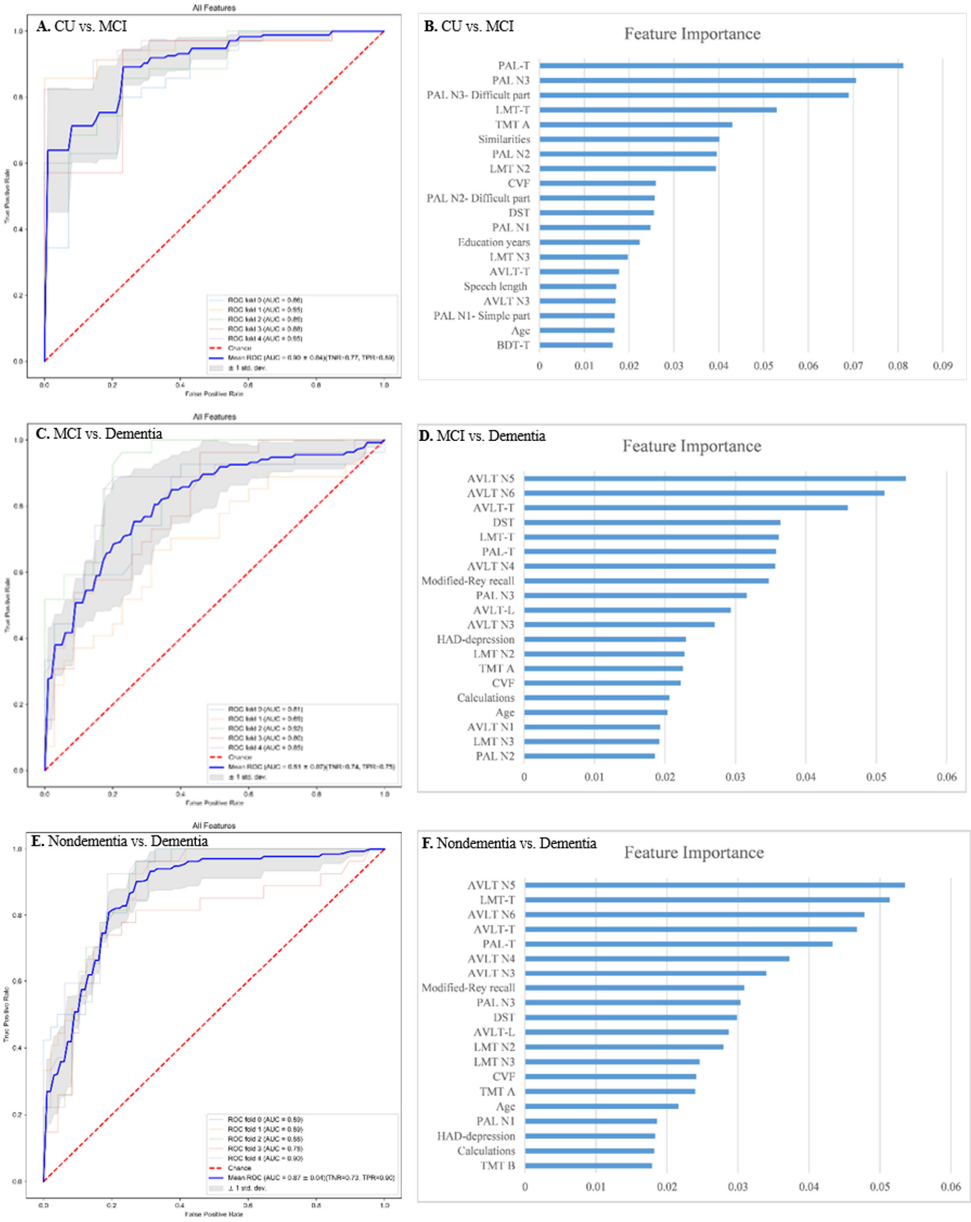

Figure 1. Receiver operating characteristic (ROC) curve analysis for the detection of MCI and dementia and the optimal 20 features. (**A**) ROC curve of all features for the detection of MCI from CU. (**B**) 20 top-ranked features for the detection of MCI from CU. (**C**) ROC curve of all features for the detection of dementia from MCI. (**D**) 20 top-ranked features for the detection of dementia from MCI. (**E**) ROC curve of all features for the detection of dementia from non-dementia. (**F**) 20 top-ranked features for the detection of dementia from non-dementia. Abbreviations: AVLT N1 = the first learning trial of AVLT-H (auditory verbal learning test-Huashan version); AVLT N3 = the third learning trial of AVLT-H; AVLT N4 = the fourth short delayed free recall trial of AVLT-H; AVLT N5 = the fifth long delayed free recall trial of AVLT-H; AVLT N6 = the sixth delayed category cue recall trial of AVLT-H; AVLT-L = total score of AVLT N1, N2,and N3; AVLT-T = total score of AVLT N1, N2, N3, N4 and N5; BDT-T = total score of the block design test; CVF = category verbal fluency; DST = Digit Symbol Test; HAD = hospital anxiety and depression; LMT N2 = the second story of logical memory test (LMT); LMT N3 = the third story of LMT; LMT-T = total score of LMT; PAL N1 = The first learning trial of PAL (paired-associate learning); PAL N1-Simple part = simple word pairs of PAL N1; PAL N2 = The second learning trial of PAL; PAL N2-Difficult part = difficult word pairs of PAL N2; PAL N3 = The third learning trial of PAL; PAL N3-Difficult part = difficult word pairs of PAL N3; PAL-T = total score of PAL N1, N2, and N3; TMT A = trail making test A; TMT B = trail making test B.

3.3. Performance of Various Classification Models

Table 2 shows the performance of various classification models. The accuracies of the logistic regression, decision tree, SVM, XGBoost, and RF models were 0.605, 0.597, 0.624, 0.664, and 0.680, while the AUCs were 0.796, 0.696, 0.809, 0.816, and 0.852. Among these methods, The RF classifier achieved the most stable performance with high accuracy compared with other classifiers.

Table 2. Performance of models trained by various methods.

	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Logistic Regression	60.53	60.80	60.08	60.12	79.62
Decision Tree	59.73	60.48	60.86	60.21	69.55
SVM	62.40	65.37	59.29	61.17	80.87
XGBoost	66.40	67.78	66.15	66.70	81.61
Random Forest	68.00	71.09	66.73	68.02	85.17

3.4. Selecting the Optimal Neuropsychological Tests to Establish Diagnostic Models

Finally, we selected six interested neuropsychological subtests with 22 high importance features (including AVLT-H, PAL, modified Rey figure, LMT, DST, and TMT A). The selected features contained in each neuropsychological subtest were listed in Supplementary Table S2. These features trained four new RF diagnosis models. The Performance (ROC AUC, sensitivity, and specificity) of these four models were shown in Table 3. If we selected three selected subtests (AVLT-H, PAL, and modified Rey figure) with 19 features to establish the diagnosis model, AUC to detect CU from MCI, MCI from dementia, dementia from nondementia was 0.86, 0.77, 0.84, respectively. If we selected four subtests (AVLT-H, PAL, modified Rey figure, and LMT) with 20 features, AUC to discriminate CU from MCI, MCI from dementia, dementia from non-dementia was 0.87, 0.79, 0.83. If we selected five subtests (AVLT-H, PAL, modified Rey figure, LMT, and DST) with 21 features, AUC to detect CU from MCI, MCI from dementia, dementia from nondementia was 0.86, 0.77, 0.84, respectively. When we chose all six important subtests with 22 selected features to establish the RF classification model, AUC to detect CU from MCI was 0.89 (sensitivity = 0.87 and specificity = 0.85), AUC to detect MCI from dementia was 0.79 (sensitivity = 0.84 and specificity = 0.63), and AUC to detect dementia from nondementia was 0.84 (sensitivity = 0.72 and specificity = 0.81). RF Model based on 22 neuropsychological features was almost equivalent to the model established using all 64 features. At the same time, the cognitive tests time was reduced from more than an hour to 30 min.

Table 3. Performance of the four new RF diagnosis models on the classification of CU, MCI, and Dementia.

New Diagnosis Models	Subtests of Interest	Number of Features	ROC AUC for CU vs. MCI (Sensitivity, Specificity)	ROC AUC for MCI vs. Dementia (Sensitivity, Specificity)	ROC AUC for Dementia vs. Nondementia (Sensitivity, Specificity)
Model-1	PAL, AVLT-H, Modified-Rey	19	0.86 (0.79, 0.84)	0.77 (0.68, 0.76)	0.84 (0.72, 0.81)
Model-2	PAL, AVLT-H, Modified-Rey, LMT	20	0.87 (0.78, 0.84)	0.79 (0.76, 0.66)	0.83 (0.70, 0.83)
Model-3	PAL, AVLT-H, Modified-Rey, LMT, DST	21	0.87 (0.83, 0.84)	0.79 (0.81, 0.65)	0.84 (0.84, 0.71)
Model-4	PAL, AVLT-H, Modified-Rey, LMT, DST, TMT A	22	0.89 (0.92, 0.74)	0.79 (0.84, 0.63)	0.84 (0.85, 0.73)

Abbreviations: AVLT-H = Auditory Verbal Learning Test-Huashan version; CU = Cognitively Unimpaired; DST = Digit Symbol Test; LMT = Logical Memory Test; MCI = Mild Cognitive Impairment; Modified-Rey = Modified Rey-Osterreith figure; PAL = Paired-Associate Learning.

4. Discussion

The present study found that 35.7 percent of subjects with MMSE scores $\geq$ 26 had evidence of dementia. Similar results have been obtained from previous studies [38,39]. This suggests that MMSE, as the only cognitive testing tool, is not sufficient to diagnose cognitive impairment. According to the 2011 NIA-AA criteria of "dementia", when clinical history and bedside cognitive tests cannot provide evidence of cognitive impairment, neuropsychological tests should be performed [36]. In this study, we applied the RF algorithm to determine the contribution of different cognitive tests and to screen out efficient neuropsychological features for better diagnosis of cognitive impairment. Our results showed that the RF algorithm has satisfactory performance in the task of diagnosing MCI (AUC = 0.89) and dementia (AUC = 0.84). The ML method helped develop a simplified version of NTB for CU, MCI, and dementia classification in patients with MMSE scores $\geq$ 26. The diagnostic model finally included six neuropsychological tests with highly important features, and other low-importance tests were deleted, thus greatly shortening the evaluation time.

The NTB is suitable for the Chinese cultural background and language habits, but the normative data of its subtests have not been updated for a long time. As the education level and living conditions of the Chinese have improved significantly in recent decades, the clinical value of the norms has been limited. Reestablishing the norms for large samples is time-consuming and requires organization and resources to conduct. In addition, the norms are influenced by many factors such as age, gender, education level, and residence (rural or urban). ML has the potential to solve the above problems by allowing multi-dimensional interactions between variables [15]. It also can rank variables that are critical to assessing cognitive impairment, which can be used to optimize neuropsychological testing [40,41]. RF can handle both linear and non-linear data and offers an advanced method to deal with outliers or missing values [42]. It has been used to solve classification and regression problems and can serve as a powerful tool to distinguish MCI and dementia [43]. Studies have found that the RF algorithm has excellent efficiency in diagnosing dementia based on neuropsychological testing [15]. Kleiman et al. reported that RF two-class classification showed greater clinical utility compared to the three-class approach in classifying cognitive impairment [44]. Therefore, our two-class models for distinguishing MCI from CU, dementia from MCI, or dementia from nondementia.

One review [45] that included 59 studies indicated that MMSE, as a global cognitive screening tool, showed the highest discrimination coefficient in the ML automatic classification of cognitive impairment. However, previous studies did not focus on people with normal MMSE scores when developing diagnostic models or optimizing neuropsychological tests using ML methods [45]. In these studies, subjects with MCI and mild dementia had significantly lower baseline scores on the bedside cognitive tests than our sample [11,41,44,46,47]. For example, Quintana et al. [47] reported that the mean MMSE score of the MCI group and dementia group was 25.77 ± 2.22, 20.37 ± 3.98, respectively. In the Chiu et al. [11] study, the mean MMSE and MoCA scores in the very mild dementia group were 19.7 ± 4.7, 12.4 ± 6.0, respectively. Lower MMSE scores indicate more severe impairment of cognition, and the diagnostic accuracy of the ML model developed based on this situation will be higher, which means that it is more difficult to detect dementia in people with normal MMSE. Classification models using ML on demographical and neuropsychological data in the literature showed wide heterogeneity in performance metrics. Weakley et al. [48] reported a sensitivity and specificity of 0.84 and 0.89 for differentiating MCI from CU, and 0.95 and 0.97 for dementia and CU, and Battista et al. [41] with 0.98 and 0.81 for MCI, and 1.00 and 0.96 for dementia. In this work, the selected sample were subjects whose MMSE was higher than the cut-off value. This is the first time to address the question that classifies people with normal MMSE. Our results showed that the RF model has good sensitivity (0.87) and specificity (0.85) for differentiating MCI from CU, as well as good sensitivity (0.85) and specificity (0.73) for dementia from nondementia.

RF had also been proven to be more effective in feature selection. Previous studies that focused on ML and cognitive measures had the disadvantage of having fewer neuropsychological features [47,49], or they just focused on the comparison between MCI and CU or CU and dementia [50,51]. Our study included 20 neuropsychological tests and compared CU, MCI, and dementia groups. The most frequent optimal neuropsychological tests reported in the literature were episodic memory [41,47,49] (like AVLT, logical memory test) and semantic fluency [46,47,52]. However, these neuropsychological measures mainly focus on Alzheimer's disease and dementia and cannot examine the damage of multiple cognitive domains. In our research, the combination of six tests is sufficient to cover multiple cognitive domains including executive function, visual perception function, language, memory, and attention, which can help diagnose all-cause dementia. AVLT-H and LMT, which assess both immediate and delayed recall, are popular methods for detecting episodic memory impairment [53,54]. PAL measures the strength of memory binding of twelve word-pairs [29]. The word pairs are presented verbally, one pair at a time. Then the participant hears the first word of each word-pair and is asked to answer the last word. PAL assesses episodic memory and executive function and could successfully detect MCI and dementia [55,56]. Modified Rey includes copy and delayed recall of the complex figure, assessing visuospatial ability and nonverbal memory. Good performance of DST and TMT A requires intact motor speed, attention, and visual perception functions, which is an important executive domain involved in semantic information processing [57]. The 2011 NIA-AA staging criteria also suggests some neuropsychological tests that are considered to be predictors of conversion from MCI to dementia [33]. These tests are generally consistent with those selected in our study.

In addition, the RF algorithm could be used not only to optimize the NTB but also to simplify individual subtests. For example, AVLT-H begins with three learning trials, followed by the fourth short delayed free recall trial, the fifth long-delayed free recall trial, the sixth category cue recall trial, and the recognition trial [53]. When ranking variables' importance, we found that AVLT N5 was the most important feature. Therefore, we choose to administer the first five trials of AVLT-H in the future practical application and delete the sixth category cue recall trial and the recognition trial. The second story of LMT was the best predictor among the three stories, so only the second story needs to be completed when performing this neuropsychological test.

There were two main limitations to this study. First, this study was a retrospective, single-center, observational study with inherent selection bias. Prospective, multi-centered, large-scale studies are therefore warranted. A second limitation is that we did not subclassify dementia. Subjects in the dementia group were patients with all-cause dementia, most of which is Alzheimer's disease and vascular dementia, and other dementia subtypes such as frontotemporal dementia and dementia with Lewy body were rare. This might cause some features to become less important. For example, language-related features such as repetition and naming were removed. Future research needs to consider dementia subtypes.

5. Conclusions

The present study showed that the RF algorithm can be a useful tool to classify CU, MCI, and dementia among a population with normal MMSE. We found that the optimized NTB, consisting of six neuropsychological tests (AVLT-H, PAL, modified Rey figure, LMT, DST, and TMT A), enables detection of MCI and dementia with good sensitivity and specificity. As cognitive markers, neuropsychological assessments have the excellent performance to identify cognitive disorders. For low- and middle-income countries, this has advantages over using classifiers based on more invasive, expensive, and time-consuming methods such as cerebrospinal fluid markers.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/jpm12010037/s1, Table S1: 64 features that extracted, Table S2: Six neuropsychological subtests and they selected features.

Author Contributions: Conceptualization, J.W. (Jie Wang) and J.G.; Data curation, J.W. (Jie Wang), Z.W., N.L., C.L., C.M., L.D., J.L., X.H., D.L., S.C. and J.G.; Formal analysis, J.W. (Jie Wang), Z.W., N.L. and J.W. (Jianyong Wang); Funding acquisition, J.G.; Investigation, J.W. (Jie Wang), C.L., C.M., L.D., J.L., X.H., D.L., S.C. and J.G.; Methodology, J.W. (Jie Wang), Z.W., N.L., J.W. (Jianyong Wang) and J.G.; Project administration, J.G.; Resources, J.W. (Jianyong Wang) and J.G.; Software, Z.W., N.L. and J.W. (Jianyong Wang); Writing—original draft, J.W. (Jie Wang); Writing—review & editing, Z.W. and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China [2020YFA0804500, 2016YFC1306300], CAMS Innovation Fund for Medical Sciences (CIFMS) (2020-I2MC&T-B-010, 2016-I2M-1-004), National Natural Science Foundation of China [81550021, 30470618], and the strategic priority research program (Pilot study) 'biological basis of aging and therapeutic strategies' of the Chinese Academy of Sciences [grant XDPB10].

Informed Consent Statement: All subjects gave their informed consent for inclusion before they participated in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors are deeply grateful to the patients and clinicians who participated in this work.

Conflicts of Interest: The authors report no competing interests.

References

1. Feigin, V.L.; Nichols, E.; Alam, T.; Bannick, M.S.; Beghi, E.; Blake, N.; Culpepper, W.J.; Dorsey, E.R.; Elbaz, A.; Ellenbogen, R.G.; et al. Global, regional, and national burden of neurological disorders, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **2019**, *18*, 459–480. [CrossRef]
2. Roberts, R.O.; Knopman, D.S.; Mielke, M.M.; Cha, R.H.; Pankratz, V.S.; Christianson, T.J.; Geda, Y.E.; Boeve, B.F.; Ivnik, R.J.; Tangalos, E.G.; et al. Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. *Neurology* **2014**, *82*, 317–325. [CrossRef] [PubMed]
3. Olazarán, J.; Reisberg, B.; Clare, L.; Cruz, I.; Peña-Casanova, J.; Del Ser, T.; Woods, B.; Beck, C.; Auer, S.; Lai, C.; et al. Nonpharmacological Therapies in Alzheimer's Disease: A Systematic Review of Efficacy. *Dement. Geriatr. Cogn. Disord.* **2010**, *30*, 161–178. [CrossRef] [PubMed]
4. 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]
5. Creavin, S.T.; Wisniewski, S.; Noel-Storr, A.H.; Trevelyan, C.M.; Hampton, T.; Rayment, D.; Thom, V.M.; Nash, K.J.; Elhamoui, H.; Milligan, R.; et al. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database Syst. Rev.* **2016**, *1*, CD011145. [CrossRef] [PubMed]
6. Patnode, C.D.; Perdue, L.A.; Rossom, R.C.; Rushkin, M.C.; Redmond, N.; Thomas, R.G.; Lin, J.S. Screening for Cognitive Impairment in Older Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* **2020**, *323*, 764–785. [CrossRef]
7. Breton, A.; Casey, D.; Arnaoutoglou, N.A. Cognitive tests for the detection of mild cognitive impairment (MCI), the prodromal stage of dementia: Meta-analysis of diagnostic accuracy studies. *Int. J. Geriatr. Psychiatry* **2019**, *34*, 233–242. [CrossRef]
8. Nasreddine, Z.S.; Phillips, N.A.; Bédirian, V.; Charbonneau, S.; Whitehead, V.; Collin, I.; Cummings, J.L.; Chertkow, H. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *J. Am. Geriatr. Soc.* **2005**, *53*, 695–699. [CrossRef] [PubMed]
9. Li, H.; Jia, J.; Yang, Z. Mini-Mental State Examination in Elderly Chinese: A Population-Based Normative Study. *J. Alzheimer's Dis.* **2016**, *53*, 487–496. [CrossRef] [PubMed]
10. Tan, J.P.; Li, N.; Gao, J.; Wang, L.-N.; Zhao, Y.-M.; Yu, B.-C.; Du, W.; Zhang, W.-J.; Cui, L.-Q.; Wang, Q.-S.; et al. Optimal Cutoff Scores for Dementia and Mild Cognitive Impairment of the Montreal Cognitive Assessment among Elderly and Oldest-Old Chinese Population. *J. Alzheimer's Dis.* **2014**, *43*, 1403–1412. [CrossRef]
11. Chiu, P.Y.; Tang, H.; Wei, C.Y.; Zhang, C.; Hung, G.U.; Zhou, W. NMD-12: A new machine-learning derived screening instrument to detect mild cognitive impairment and dementia. *PLoS ONE* **2019**, *14*, e0213430. [CrossRef] [PubMed]
12. Davatzikos, C. Machine learning in neuroimaging: Progress and challenges. *NeuroImage* **2019**, *197*, 652–656. [CrossRef]

13. Shigemizu, D.; Akiyama, S.; Asanomi, Y.; Boroevich, K.; Sharma, A.; Tsunoda, T.; Sakurai, T.; Ozaki, K.; Ochiya, T.; Niida, S. A comparison of machine learning classifiers for dementia with Lewy bodies using miRNA expression data. *BMC Med. Genom.* **2019**, *12*, 150. [CrossRef]
14. Shehzad, A.; Rockwood, K.; Stanley, J.; Dunn, T.; Howlett, S.E. Use of Patient-Reported Symptoms from an Online Symptom Tracking Tool for Dementia Severity Staging: Development and Validation of a Machine Learning Approach. *J. Med. Internet Res.* **2020**, *22*, e20840. [CrossRef]
15. Yim, D.; Yeo, T.Y.; Park, M.H. Mild cognitive impairment, dementia, and cognitive dysfunction screening using machine learning. *J. Int. Med. Res.* **2020**, *48*, 300060520936881. [CrossRef]
16. Yilmaz, A.; Ustun, I.; Ugur, Z.; Akyol, S.; Hu, W.T.; Fiandaca, M.S.; Mapstone, M.; Federoff, H.; Maddens, M.; Graham, S.F. A Community-Based Study Identifying Metabolic Biomarkers of Mild Cognitive Impairment and Alzheimer's Disease Using Artificial Intelligence and Machine Learning. *J. Alzheimer's Dis.* **2020**, *78*, 1381–1392. [CrossRef] [PubMed]
17. Khatri, U.; Kwon, G.R. An Efficient Combination among sMRI, CSF, Cognitive Score, and APOE ε4 Biomarkers for Classification of AD and MCI Using Extreme Learning Machine. *Comput. Intell. Neurosci.* **2020**, *2020*, 8015156. [CrossRef] [PubMed]
18. Bougea, A.; Efthymiopoulou, E.; Spanou, I.; Zikos, P. A Novel Machine Learning Algorithm Predicts Dementia with Lewy Bodies Versus Parkinson's Disease Dementia Based on Clinical and Neuropsychological Scores. *J. Geriatr. Psychiatry Neurol.* **2021**, 891988721993556. [CrossRef]
19. Gurevich, P.; Stuke, H.; Kastrup, A.; Stuke, H.; Hildebrandt, H. Neuropsychological Testing and Machine Learning Distinguish Alzheimer's Disease from Other Causes for Cognitive Impairment. *Front. Aging Neurosci.* **2017**, *9*, 114. [CrossRef]
20. Zhang, Z.; Hong, X.; Li, H.; Zhao, J.H.; Huang, J.B.; Wei, J.; Wang, J.M.; Li, S.W.; Yang, E.L.; Wu, J.X. The mini-mental state examination in the Chinese residents population aged 55 years and over in the urban and rural areas of Beijing. *Chin. J. Neurol.* **1999**, *32*, 149–153.
21. Katzman, R.; Zhang, M.Y.; Ouang, Y.Q.; Wang, Z.; Liu, W.T.; Yu, E.; Wong, S.-C.; Salmon, D.P.; Grant, I. A Chinese version of the mini-mental state examination; Impact of illiteracy in a Shanghai dementia survey. *J. Clin. Epidemiol.* **1988**, *41*, 971–978. [CrossRef]
22. Xu, G.; Meyer, J.S.; Huang, Y.; Du, F.; Chowdhury, M.; Quach, M. Adapting Mini-Mental State Examination for dementia screening among illiterate or minimally educated elderly Chinese. *Int. J. Geriatr. Psychiatry* **2003**, *18*, 609–616. [CrossRef] [PubMed]
23. Zhang, M.; Yu, E.; He, Y. Tools for dementia epidemiological investigations and their applications. *Shanghai Arch. Psychiatry* **1995**, *7*, 1–62.
24. Lawton, M.P.; Brody, E.M. Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist* **1969**, *9*, 179–186. [CrossRef]
25. Zigmond, A.S.; Snaith, R.P. The Hospital Anxiety and Depression Scale. *Acta Psychiatr. Scand.* **1983**, *67*, 361–370. [CrossRef] [PubMed]
26. Chan, A.S.; Poon, M.W. Performance of 7- to 95-year-old individuals in a Chinese version of the category fluency test. *J. Int. Neuropsychol. Soc.* **1999**, *5*, 525–533. [CrossRef] [PubMed]
27. Gong, Y. *Manual of Modified Wechsler Adult Intelligence Scale (WAIS-RC)*; Hunan Med College: Changsha, China, 1982; pp. 45–48.
28. Gong, Y. The Chinese revision of Halstead-Reitan Neuropsychological Test Battery for Adults. *Acta Psychol. Sin.* **1986**, *18*, 433–442.
29. Xu, S.; Wu, Z. The construction of "The Clinical Memory Test". *Acta Psychol. Sin.* **1986**, *18*, 100–108.
30. Gao, S.; Zhu, Y.; Shi, S.; Peng, Y. Standard Aphasia Battery of Chinese. *Chin. Ment. Health J.* **1992**, *6*, 125–128.
31. Luria, A.R. *Higher Cortical Functions in Man*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
32. Fogel, B.S.; Schiffer, R.B.; Rao, S.M. *Synopsis of Neuropsychiatry*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2000.
33. Gong, Y.; Jiang, D.; Deng, J. *Manual of Modified Wechsler Memory Scale (WMS)*; Hunan Med College: Changsha, China, 1989; Volume 19.
34. Guo, Q.H.; Sun, Y.T.; Yu, P.M.; Hong, Z.; Lv, C.Z. Norm of auditory verbal learning test in the normal aged in Chinese community. *Chin. J. Clin. Psychol.* **2007**, *15*, 132–135.
35. Albert, M.S.; DeKosky, S.T.; Dickson, D.; Dubois, B.; Feldman, H.H.; Fox, N.C.; Gamst, A.; Holtzman, D.M.; Jagust, W.J.; Petersen, R.C.; et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* **2011**, *7*, 270–279. [CrossRef]
36. McKhann, G.M.; Knopman, D.S.; Chertkow, H.; Hyman, B.T.; Jack, C.R., Jr.; Kawas, C.H.; Klunk, W.E.; Koroshetz, W.J.; Manly, J.J.; Mayeux, R.; et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* **2011**, *7*, 263–269. [CrossRef]
37. Jack, C.R., Jr.; Bennett, D.A.; Blennow, K.; Carrillo, M.C.; Dunn, B.; Haeberlein, S.B.; Holtzman, D.M.; Jagust, W.; Jessen, F.; Karlawish, J.; et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's Dement.* **2018**, *14*, 535–562. [CrossRef]
38. Friedman, T.W.; Yelland, G.W.; Robinson, S.R. Subtle cognitive impairment in elders with Mini-Mental State Examination scores within the 'normal' range. *Int. J. Geriatr. Psychiatry* **2011**, *27*, 463–471. [CrossRef]
39. Votruba, K.L.; Persad, C.; Giordani, B. Cognitive Deficits in Healthy Elderly Population with "Normal" Scores on the Mini-Mental State Examination. *J. Geriatr. Psychiatry Neurol.* **2016**, *29*, 126–132. [CrossRef]

40. Graham, S.A.; Lee, E.E.; Jeste, D.V.; Van Patten, R.; Twamley, E.W.; Nebeker, C.; Yamada, Y.; Kim, H.-C.; Depp, C.A. Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review. *Psychiatry Res.* **2019**, *284*, 112732. [CrossRef]
41. 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]
42. 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]
43. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
44. Kleiman, M.J.; Barenholtz, E.; Galvin, J.E. Screening for Early-Stage Alzheimer's Disease Using Optimized Feature Sets and Machine Learning. *J. Alzheimer's Dis.* **2021**, *81*, 355–366. [CrossRef]
45. Battista, P.; Salvatore, C.; Berlingeri, M.; Cerasa, A.; Castiglioni, I. Artificial intelligence and neuropsychological measures: The case of Alzheimer's disease. *Neurosci. Biobehav. Rev.* **2020**, *114*, 211–228. [CrossRef]
46. Lins, A.; Muniz, M.T.C.; Garcia, A.N.M.; Gomes, A.V.; Cabral, R.M.; Bastos-Filho, C.J.A. Using artificial neural networks to select the parameters for the prognostic of mild cognitive impairment and dementia in elderly individuals. *Comput. Methods Programs Biomed.* **2017**, *152*, 93–104. [CrossRef]
47. Quintana, M.; Guàrdia, J.; Sánchez-Benavides, G.; Aguilar, M.; Molinuevo, J.L.; Robles, A.; Barquero, M.S.; Antúnez, C.; Martínez-Parra, C.; García, A.F.; et al. Using artificial neural networks in clinical neuropsychology: High performance in mild cognitive impairment and Alzheimer's disease. *J. Clin. Exp. Neuropsychol.* **2012**, *34*, 195–208. [CrossRef]
48. Weakley, A.; Williams, J.A.; Schmitter-Edgecombe, M.; Cook, D.J. Neuropsychological test selection for cognitive impairment classification: A machine learning approach. *J. Clin. Exp. Neuropsychol.* **2015**, *37*, 899–916. [CrossRef]
49. Tunvirachaisakul, C.; Supasitthumrong, T.; Tangwongchai, S.; Hemrunroj, S.; Chuchuen, P.; Tawankanjanachot, I.; Likitchareon, Y.; Phanthumchinda, K.; Sriswasdi, S.; Maes, M. Characteristics of Mild Cognitive Impairment Using the Thai Version of the Consortium to Establish a Registry for Alzheimer's Disease Tests: A Multivariate and Machine Learning Study. *Dement. Geriatr. Cogn. Disord.* **2018**, *45*, 38–48. [CrossRef]
50. Lv, S.; Wang, X.; Cui, Y.; Jin, J.; Sun, Y.; Tang, Y.; Bai, Y.; Wang, Y.; Zhou, L. Application of attention network test and demographic information to detect mild cognitive impairment via combining feature selection with support vector machine. *Comput. Methods Programs Biomed.* **2009**, *97*, 11–18. [CrossRef]
51. Reverberi, C.; Cherubini, P.; Baldinelli, S.; Luzzi, S. Semantic fluency: Cognitive basis and diagnostic performance in focal dementias and Alzheimer's disease. *Cortex* **2014**, *54*, 150–164. [CrossRef]
52. Clark, D.G.; Kapur, P.; Geldmacher, D.S.; Brockington, J.; Harrell, L.; DeRamus, T.; Blanton, P.; Lokken, K.; Nicholas, A.; Marson, D. Latent information in fluency lists predicts functional decline in persons at risk for Alzheimer disease. *Cortex* **2014**, *55*, 202–218. [CrossRef]
53. Zhao, Q.; Lv, Y.; Zhou, Y.; Hong, Z.; Guo, Q. Short-Term Delayed Recall of Auditory Verbal Learning Test Is Equivalent to Long-Term Delayed Recall for Identifying Amnestic Mild Cognitive Impairment. *PLoS ONE* **2012**, *7*, e51157. [CrossRef]
54. Yu, H.; Guo, Q.; Hong, Z.; Lv, C. Logic Memory Test in early detection of Alzheimer's disease. *Nerve Dis. Ment. Hygeine* **2005**, *5*, 89–91.
55. Curiel, R.E.; Crocco, E.; Rosado, M.; Duara, R.; Greig, M.T.; Raffo, A.; Loewenstein, D.A. A Brief Computerized Paired Associate Test for the Detection of Mild Cognitive Impairment in Community-Dwelling Older Adults. *J. Alzheimer's Dis.* **2016**, *54*, 793–799. [CrossRef]
56. Duchek, J.M.; Cheney, M.; Ferraro, F.R.; Storandt, M. Paired Associate Learning in Senile Dementia of the Alzheimer Type. *Arch. Neurol.* **1991**, *48*, 1038–1040. [CrossRef]
57. Wang, L.; Nie, K.; Zhao, X.; Feng, S.; Xie, S.; He, X.; Ma, G.; Wang, L.; Huang, Z.; Huang, B.; et al. Characteristics of gray matter morphological change in Parkinson's disease patients with semantic abstract reasoning deficits. *Neurosci. Lett.* **2018**, *673*, 85–91. [CrossRef]