Interdisciplinary Journal of Health, Environment and Computation
ISSN Online: 3108-1444 Vol.: 2, No.: 1 (July – September) 2026. PP. : 1-18

A STATISTICAL FRAMEWORK FOR BRAIN MRI DIAGNOSTIC CLASSIFICATION AND EVALUATION IN MEDICAL IMAGE PROCESSING

Israa Abdulameer Resen

Department of Mobile Communications and Computing Engineering, College of Engineering, University of Information Technology and Communications (UoITC), Baghdad, Iraq

Email: medical.insteng@gmail.com

https://doi.org/10.65983/ijhec.2026.01.0005

Received: August 17, 2026; Revised: September 21, 2026; Accepted : September 27, 2026

Abstract

The proposed work introduces the statistical framework of quantitative analysis and image processing methods for detecting structural brain changes and classifying the severity of dementia cases using $OASIS-1$ cross-sectional database. From $N_{\text{total}} = 437$ cases, only $N = 235$ individuals were included in a filtered analysis sample. Patients were divided into two groups according to CDR score. Cognitively Normal ($n_0 = 135$) and Demented ($n_1 = 100$). Continuous variables were tested for normal distribution using the Anderson-Darling test, followed by parametric (two-sample t-test) and non-parametric (Mann-Whitney U test) comparison between the groups. Further, a multivariable binary logistic regression model was developed based on structural MRI biomarkers of brain volume changes. Normalised Whole Brain Volume ($nWBV$), Estimated Total Intracranial Volume ($eTIV$), and Atlas Scaling Factor (ASF) were investigated. Other continuous covariates such as $MMSE$ score and age and a dichotomous gender covariate were created. Empirical results showed that $MMSE$ score ($\beta = -0.818, p < 0.001$, $OR = 0.4412$) and $nWBV$ ($\beta = -19.37, p = 0.007$) were independent, statistically significant covariates associated with dementia risk. Therefore, progressive cerebral atrophy and cognitive dysfunction are the main predictors of impairment. Other covariates (age, $eTIV$, ASF) turned out to be non-significant in the multivariable model, since their influence is incorporated in structural brain volume changes and cognitive function test scores. The constructed logistic regression model had good performance in terms of calibration and fit (Deviance $R^2 = 48.17\%$, Hosmer-Lemeshow $\chi^2(8) = 7.31,\ p = 0.503$). Evaluation of model classification performance revealed high accuracy of prediction of dementia presence with Area Under the ROC Curve (AUC) of 0.9159 (95% CI: $[0.878, 0.954]$]) and concordance index of 91.6%. Overall, a combination of quantitative $sMRI$ biomarkers and standard cognitive tests provides an efficient diagnostic framework for making clinical decisions in Alzheimer's disease and other forms of dementia.

Keywords: Applied Medical Statistics; Binary Logistic Regression; Diagnosis; Volumetric Markers; Anderson-Darling Normality Test; ROC/AUC Analysis; OASIS-1 Brain MRI.

1. Introduction

Increased global prevalence of neurodegenerative diseases, such as Alzheimer's disease (AD), and other forms of dementia constitutes a key challenge for contemporary healthcare systems, requiring effective and precise diagnostic methods. Quantitative structural Magnetic Resonance Imaging ($sMRI$) is now an essential part of clinical neuroimaging practice. It provides objective and non-invasive biomarkers of brain structure such as Normalised Whole Brain Volume ($nWBV$) and Estimated Total Intracranial Volume ($eTIV$). The foundation has already been laid for clinical neuroimaging and its use for assessing the progression of brain atrophy. These practices include the combination with other diagnostic practices, including Mini-Mental State Examination ($MMSE$) and Clinical Dementia Rating (CDR), which contribute to understanding the spectrum of cognitive decline (Folstein et al., 1975; Morris, 1993; Marcus et al., 2007). However, there is still a considerable gap in research on a fully developed computational framework for statistical classification. This framework could incorporate all the necessary criteria of rigorous multivariable parametric/non-parametric diagnosis, head-size correction, and reliable logit modelling without being subjected to multicollinearity or lacking clinical interpretation of early-stage dementia classification.

In order to solve the problems inherent in the conventional diagnostic procedures, numerous studies have focused on automatic image processing, segmentation, and computational intelligence for dementia staging. The earliest pioneering work by Slege et al., 1998, and the FreeSurfer processing stream by Fischl, 2012, contributed important automatic algorithms for (*N3*) intensity non-uniformity correction, surface-based cortical reconstructions, and volume-based segmentation of tissues. This pioneering computational effort led to enhanced automatic extraction of brain anatomy from $T1$-weighted structural MRI scans. Nevertheless, such work was predominantly conducted as standalone structural segmentation procedures rather than a complete predictive statistical method. There has been minimal integration of multi-variable modeling which would allow adjustment of volumetric biomarkers like ($nWBV$) and ($eTIV$) with respect to demographic covariates such as age, biological sex, and education.

Further advancements in the field of medical image analysis have seen the introduction of quantitative volumetric variables into statistical and machine learning models. For example, (Buckner et al., 2004; Klöppel et al., 2008; Cuingnet et al., 2011) assessed various structural neuroimaging biomarkers to distinguish healthy ageing from MCI and mild Alzheimer's disease. Also, current research has seen a move towards more complicated machine learning models and convolutional neural networks ($CNNs$), as exemplified by Suk et al., 2014 and Barnett et al., 2021. Although these sophisticated models have high raw classification scores, they have several methodological issues. First and foremost, they operate as black-box models that do not yield precise statistical measures such as the size of effect, odds ratios (OR), and confidence intervals of biological variables. This makes it difficult to use these models in a real-world clinical setting where clinicians require an understandable measure of diagnostic aid.

Furthermore, a major flaw in methodology found throughout many contemporary neuroimaging studies based on statistical frameworks is a lack of comprehensive diagnostic procedures. Many medical classification studies use parametric hypothesis testing without prior validation of the normality assumption with the help of such statistics as the Anderson-Darling test or the Shapiro-Wilk test (Montgomery et al., 2007). Moreover, regression models tend to include collinear volumetric variables without checking the Variance Inflation Factor (VIF) (Navidi, 2008). As an illustration, concurrent inclusion of raw volume measurement ($eTIV$) along with spatial normalisation measures such as Atlas Scaling Factor (ASF) often leads to multicollinearity, which in turn leads to an increase in the variance of estimations and poor model calibration (Altman, 1990). Thus, there is a need for further research into building a classification framework which would effectively combine volumetric $sMRI$ measures with standardised clinical scores.

To address these shortcomings, the main goal of this research is to develop, implement and thoroughly analyse a full end-to-end quantitative statistical analysis methodology for the diagnosis of brain MRI images using the cross-sectional $OASIS-1$ dataset ($N = 235$). In particular, this research intends to: (1) perform full distribution diagnostics using the Anderson-Darling test to inform proper use of parametric (two-sample t-test) and non-parametric (Mann-Whitney U test) tests for group comparison; (2) build a multivariate binary logistic regression model utilizing normalized volumetric $sMRI$ markers ($nWBV, eTIV, ASF$) along with clinical and demographic variables ($MMSE, Age, Gender$) to find out independent predictors for dementia; and (3) conduct a thorough performance evaluation of the resulting classifier by means of Area Under the ROC Curve (AUC), Hosmer-Lemeshow test and collinearity diagnostics (VIF). By explicitly incorporating structural brain atrophy measures with standardised cognition testing in a well-vetted logit equation, this research addresses the problems of lack of model transparency and distributional consideration previously reported.

2. Methodology

This part provides an explanation of the quantitative statistical approach and image processing technique used to analyse structural changes in the brain and categorise dementia levels. This method incorporates cohort selection, variables' definition, diagnostics of distributions, bivariate parametric/non-parametric analysis, binary logistic regression modelling and classification validation.

2.1. Cohort Selection and Sample Characteristics

The empirical data used for this study have been collected from the $OASIS-1$ cross-sectional database ($OASIS_cross-sectional.csv$) (Marcus et al., 2007). The dataset contains cross-sectional data on $N_{\text{total}} = 437$ individuals in the age range of 18 to 96 years old. To maintain the validity of diagnosis and to rule out factors related to unexamined young patients, an inclusion/exclusion filter was implemented:

1. **Clinical Assessment Criterion:** Subjects with no Clinical Dementia Rating CDR scores (Morris, 1993) (mainly young controls below 60 years of age) were not included.
2. **Analytical Cohort ($N = 235$):** The analytical group includes $N = 235$ elderly participants, all clinically, cognitively and structurally assessed using $T1$-weighted MRI.

3. **Stratification for Modelling:** To create a binary classification model, CDR values (0, 0.5, 1, 2) were transformed into a binary dependent variable ($Binary_CDR$):
    - $Binary_CDR = 0$: Cognitively normal / control group ($n_0 = 135$ subjects, 57.45%)
    - $Binary_CDR = 1$: Demented / cognitively impaired group ($n_1 = 100$ subjects, 42.55%, includes early, mild and moderate stages of Alzheimer's disease).

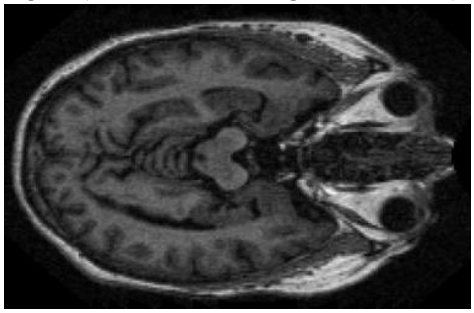
Statistical power calculations demonstrated that the final number of participants used in this study ($N = 235$) has more than 85% statistical power ($\beta = 0.15$) to reveal moderate effect sizes at $\alpha = 0.05$ (Cohen, 1988). $OASIS - 1$ dataset is a freely accessible, entirely anonymous open access dataset. All procedures for recruiting participants and conducting imaging experiments have been conducted following the Institutional Review Board IRB at Washington University.

2.2. Qualitative Inspection of Structural Brain MRI

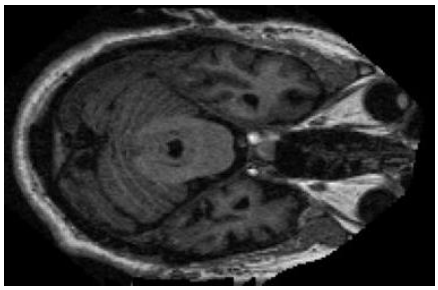
To set the visual and anatomical benchmark of the neurodegenerative process in different dementia stages, representative axial $T1$ −weighted structural MR images have been taken for the four aforementioned categories. The qualitative assessment of the aligned MR images presented in Figure 1 reveals the following continuum of neurodegeneration:

As the level of dementia severity progresses from Non-Demented ($CDR = 0$) to Moderate Dementia ($CDR = 2.0$), the shrinking of the cortex and enlargement of lateral ventricles are more pronounced.

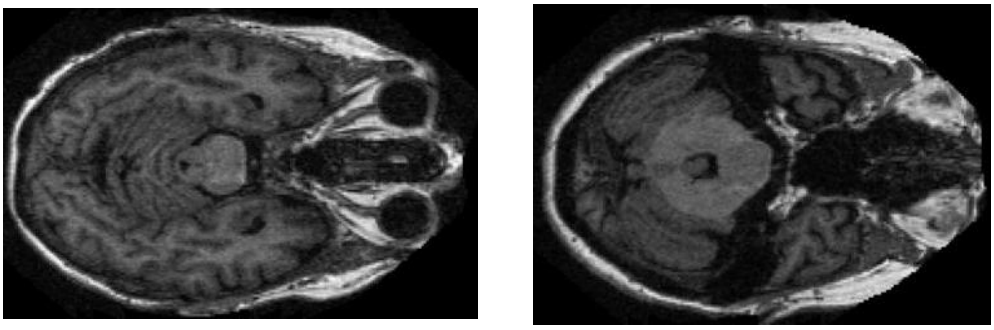
These changes confirm the clinical basis for extracting the volumetric biomarkers, Normalised Whole Brain Volume ($nWBV$) and Estimated Total Intracranial Volume ($eTIV$) (Marcus et al., 2007; Buckner et al., 2004), using automatic classification.
Both volumetric segmentation and gain field correction were done using the conventional OASIS protocol, which uses automated tissue segmentation (FreeSurfer/ Gain Field Correction) for deriving spatial intensity normalisation and tissue volume targets (Fischl, 2012; Slege et al., 1998).



(A) Non-Demented (CDR = 0)



(B) Very Mild Dementia (CDR = 0.5)



(C) Mild Dementia (CDR = 1.0) (D) Moderate Dementia (CDR = 2.0)

Figure 1: Typical slices of $T1$ structural brain MRI images obtained from the $OASIS-1$ data set for the four levels of severity of dementia (Marcus et al., 2007)

2.3. Variable Definitions and Data Dictionary

Table 1 gives an example of a representative excerpt from the raw data set that was structured and formatted before the data analysis and cohort selection. As seen from Table 1, young subjects (such as $OAS1_0004_MR1$, Age 28) do not have clinical rating scales ($CDR, MMSE, Educ, SES$). These observations were purposely eliminated from the data analysis because they did not meet inclusion criteria to form the final analytical clean cohort ($N = 235$).

Table 1: Raw Data Example from the $OASIS-1$ Cross Sectional Study ((Marcus et al., 2007)

ID	M/F	Hand	Age	Educ	SES	MMSE	CDR	eTIV	nWBV	ASF
OAS1_0001_MR1	F	R	74	2	3	29	0	1344	0.743	1.306
OAS1_0002_MR1	F	R	55	4	1	29	0	1147	0.810	1.531
OAS1_0003_MR1	F	R	73	4	3	27	0.5	1454	0.708	1.207
OAS1_0004_MR1	M	R	28					1588	0.803	1.105
OAS1_0005_MR1	M	R	18					1737	0.848	1.010
OAS1_0006_MR1	F	R	24					1131	0.862	1.551
OAS1_0007_MR1	M	R	21					1516	0.830	1.157
OAS1_0008_MR1	N/A									
OAS1_0009_MR1	F	R	20					1505	0.843	1.166
OAS1_0010_MR1	M	R	74	5	2	30	0	1636	0.689	1.073

OAS1_0011_MR1	F	R	52	3	2	30	0	1321	0.827	1.329
OAS1_0012_MR1	M	R	30					1574	0.842	1.115
OAS1_0013_MR1	F	R	81	5	2	30	0	1664	0.679	1.055
OAS1_0014_MR1	F	R	19					1525	0.856	1.151
OAS1_0015_MR1	M	R	76	2		28	0.5	1738	0.719	1.010

The variables used in this study and their mathematical units are provided in Table 2 in the form of a comprehensive data dictionary. The variables used in this research can be divided into three main categories, which are:

Demographics and Socioeconomic Status: Age (control variable for non-pathological ageing), biological sex (Male/Female, control for variation in initial brain volume), $Educ$ (years of education), and SES (index of socioeconomic status).
Examination and Cognitive Scores: $MMSE$ score (Folstein et al., 1975) (screening test for global cognition) and CDR rating (Morris, 1993) (staging dementia test) dichotomised to $Binary_CDR$ $(0 = Control, 1 = Demented)$ for binary classification models.
Volumetric MRI Biomarkers: Normalised Whole Brain Volume ($nWBV$, calculated as $nWBV = \frac{V_{\text{Brain}}}{eTIV}$, serving as the primary marker of cerebral atrophy), Estimated Total Intracranial Volume ($eTIV$), and Atlas Scaling Factor (ASF).

Table 2: Data Dictionary and Variable Specifications ($N = 235$)

Variable Name	Data Type	Unit / Scale	Statistical Role	Clinical / Anatomical Description
ID	Nominal	String Code	Subject Identifier	Reference code for MRI scanning session (for example, $OAS1_0001_MR1$).
CDR	Ordinal	Scale (0-2.0)	Clinical Gold Standard	Clinician's Dementia Rating scale for rating dementia severity (0 =Normal, 0.5 =Very mild, 1.0 =Mild, 2.0 =Moderate).
$Binary_CDR$	Binary	Code $(0/1)$	Primary Dependent (Y)	Target Variable for Classification (0: Cognitively Normal, 1: Impaired/Demented).
$nWBV$	Continuous	Ratio (0-1)	Primary Predictor (X_1)	**Whole Brain Volume Normalisation:** Ratio of the brain volume to the intracranial volume. The main indicator of brain atrophy.
$eTIV$	Continuous	mm^3	Secondary Predictor (X_2)	**Calculated intracranial volume estimate:** Intracranial volume estimate

				used to account for premorbid brain volume.
ASF	Continuous	Unitless Ratio	Alignment Metric	**Atlas Scaling Factor:** Scale factor of transformation necessary to normalise head shape to MNI atlas space.
Age	Continuous	Years	Control Variable (X_3)	**Age of subject at time of scan:** Added as a covariate to account for brain shrinkage with age.
M/F	Binary	Categorical (M/F)	Control Variable (X_4)	Biological Sex, when controlling for cranial volume capacity differences between men and women.
MMSE	Continuous	Score (0-30)	Cognitive Screening	Mini-Mental State Examination $MMSE$ score assessment of global cognitive function (Folstein et al., 1975).

2.4. Baseline Descriptive Statistics

Table 3 displays the demographics, cognition, and volume-based neuroimaging features of the study population based on the diagnosis. As mentioned in the baseline profile, the analysis set comprises $N = 235$ individuals, out of which $n_0 = 135$ belong to the control category (or are cognitively normal with $Binary_CDR = 0$) and $n_1 = 100$ are Demented with $Binary_CDR = 1$. The stratified comparison highlights:

1. **Cognitive Deficits:** Individuals in the Demented category have notably lower mean MMSE scores (24.32±4.17) when compared to the control group (29.10 ± 1.13).
2. **Volumetric Atrophy:** There is a clear reduction in Normalised Whole Brain Volume ($nWBV$) in individuals with impairment (0.722 ± 0.036) in comparison with those without dementia (0.769 ± 0.046), suggesting severe loss of neurological tissue.
3. **Demographics:** Individuals in the Demented group are on average older (76.76 ± 7.12) than individuals in the control group (69.07 ± 13.88), whereas intracranial volume measures ($eTIV$ and ASF) show uniformity in distribution for both categories.

Table 3: Stratified Descriptive Statistics for the Analytical Sample Size ($N = 235$)

Characteristic / Variable	Total Cohort ($N = 235$)	Cognitively Normal ($Binary_CDR = 0, n_0 = 135$)	Demented ($Binary_CDR = 1, n_1 = 100$)	Statistical Metric
Age (Years)	72.34 ± 12.11	69.07 ± 13.88	76.76 ± 7.12	Mean $\pm$ SD
$Gender$ (Male / Female)	88/147	43/92	45/55	Count (n)
$MMSE$ Score	27.06 ± 3.78	29.10 ± 1.13	24.32 ± 4.17	Mean $\pm$ SD
$nWBV$ (Normalised Volume)	0.749 ± 0.048	0.769 ± 0.046	0.722 ± 0.036	Mean $\pm$ SD
$eTIV$ (mm^3)	1459.49 ± 160.41	1441.53 ± 152.36	1483.75 ± 168.32	Mean $\pm$ SD

ASF Factor	1.217 ± 0.129	1.231 ± 0.127	1.197 ± 0.131	Mean $\pm$ *SD*

2.5. Statistical Framework and Workflow Execution

The methodology comprises a rigorous four-step analysis pipeline, as shown in Figure 2 below:

1. **Data Preprocessing & Imputation:** Missing cognitive and demographic data are accounted for, and the dataset is reduced to create the sample of interest ($N = 235$). In case of missing continuous data (e.g., SES) for the sample of interest ($N = 235$), median-matching imputation (Little & Rubin, 2019) has been employed according to the subgroups of diagnostics matched by age.
2. **Tests for Normal Distribution & Hypothesis Tests:** Continuous volumetric biomarkers are checked for normality using the Anderson-Darling test; the choice between parametric test ($2-$sample $t-$test) and non-parametric test (Mann-Whitney U test) is made after checking results of distribution test.
3. **Logistic Regression Model Building:** Logit models are built to predict the likelihood of developing dementia based on $nWBV$, $eTIV$, and demographic covariates ($Age, Gender$).
4. **Model Diagnostic Evaluation:** The model's discrimination and classification performance is evaluated by means of Sensitivity, Specificity, Accuracy and Area Under the ROC Curve (AUC).

All calculations, tests for normality, modelling of binary logistic regression, and ROC curve analysis were done via Minitab® (Version 21.0) (Mathews, 2004), where the statistical significance level was set to $\alpha = 0.05$.

2.5.i. Normality Diagnostics and Bivariate Hypothesis Testing

Continuous volumetric measures ($nWBV, eTIV, ASF$) were tested for their normal distribution based on the Anderson-Darling test (Stephens, 1974; Risan et al., 2024). The null hypothesis (H_0) states that the feature is Gaussian in nature. The test for normally distributed features is done through the independent $2-$sample $t-$test. Non-normally distributed features are tested using the non-parametric Mann-Whitney U test (Altman, 1990; Bethea, 1995) for significance of the difference in volume between the control group ($n_0 = 135$) and the demented group ($n_1 = 100$).

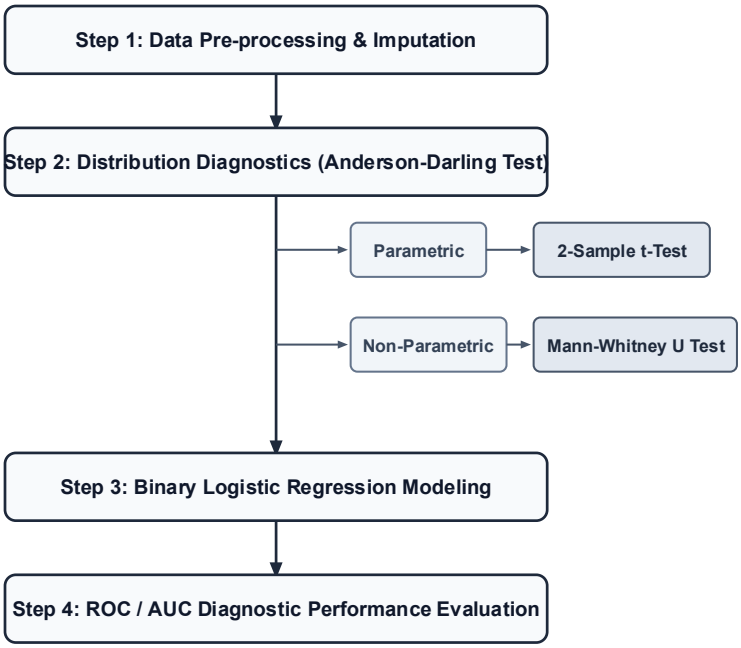


Figure 2: Total Process of Analysis

2.5.ii. Binary Logistic Regression Formulation

The Binary Logistic Regression model is developed to determine the likelihood of diagnosing dementia using the volumetric features from an *MRI* scan and clinical covariates (Hosmer et al., 2000). The mathematical formulation of the logit link function is represented by:

$$\ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 \cdot \text{nWBV} + \beta_2 \cdot \text{eTIV} + \beta_3 \cdot \text{Age} + \beta_4 \cdot \text{Gender} \quad (1)$$

By applying the sigmoid function, we obtain the posterior probability $P(Y=1|X)$ of classifying into dementia:

$$P(Y=1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1\cdot\text{nWBV}+\beta_2\cdot\text{eTIV}+\beta_3\cdot\text{Age}+\beta_4\cdot\text{Gender})}} \quad (2)$$

The categorisation of the subject was operationalised through a decision point of $p = 0.5$:

$$\hat{Y} = \begin{cases} 1 \quad \text{(Demented)}, & \text{if } P(Y=1|X) \geq 0.5 \\ 0 \quad \text{(Control)}, & \text{if } P(Y=1|X) < 0.5 \end{cases}$$

The $P(Y=1)$ is the probability of a patient being diagnosed with dementia ($Binary_CDR = 1$), β_0 refers to the constant term, and $\beta_1, \beta_2, \beta_3$ and β_4 refer to the regression coefficient estimates of $nWBV$, $eTIV$, Age, and Sex, respectively.

To test for classification generalizability and overfitting, parameters were validated using $a\ 5-fold\ cross$ −validation technique (Hastie et al., 2009) (or using a $70/30$ training-testing stratification technique) among the analytic sample ($N = 235$).

2.5.iii. Performance Metrics and Classification Evaluation

The performance and goodness of fit of the binary logistic regression model were tested using $ROC - AUC$ (area under the receiver operating characteristic curve) to determine the general discrimination ability, as seen in Table 4 below. Furthermore, Somers' D rank correlation and McFadden's Pseudo R^2 have been determined to test the concordance and proportion of variance explained by the model, respectively (Fawcett, 2006).

Table 4: Metrics for Classification Performance

Performance Metric	Primary Analytical Evaluation Purpose
Area Under ROC (AUC)	Measurement of the global discriminative power of the model.
Somers' D Correlation	Measures the degree of ordinal correlation and agreement between the predictions and the observations.
McFadden's Pseudo R^2	Represents the share of variance explained by the logit regression model.

3. Results

3.1. Distribution Diagnostics and Bivariate Analysis

Diagnosing normality through probability plots was conducted using the raw baseline data set ($N = 436$), while bivariate analysis and logistic regression were conducted using the filtered data set ($N = 235$) after removing subjects that were not assessed.

In order to identify the required bivariate statistical methods for comparing the groups, the empirical distributions of the continuous volumetric MRI biomarkers and demographic variables were checked using Shapiro-Wilk and Anderson-Darling tests. Variables $MMSE$, $nWBV$ and ASF failed to meet the conditions of normal distribution ($p < 0.05$) and therefore required the use of non-parametric analysis techniques. On the other hand, variables age and $eTIV$ met the requirements for parametric analysis.

1. **Estimated Total Intracranial Volume ($eTIV$):** The empirical values of $eTIV$ matched well to a theoretical normal distribution ($N = 436$, Mean $= 1482$ mm^3, StDev $= 158.7$). The value of the Anderson-Darling AD statistic was obtained as 0.669 with a p −value of 0.080, as illustrated in Figure 3. Since $p > 0.05$, the hypothesis of normality could not be rejected, which implies that $eTIV$ meets the criteria of normal distribution and, therefore, parametric tests of differences in independent samples can be applied further.

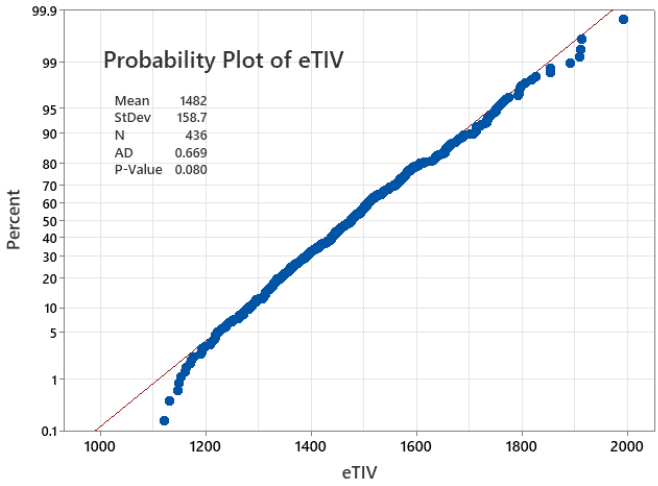
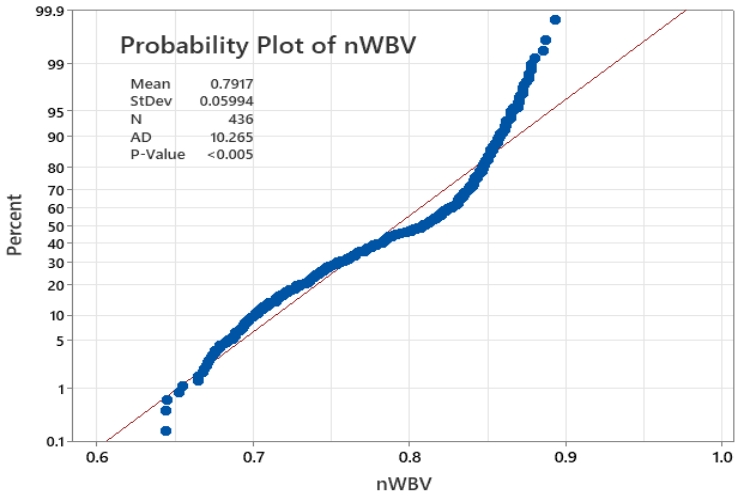


Figure 3. Normal Probability Plot of eTIV

2. **Whole brain volume nWBV:** However, contrary to this, $nWBV$ ($N = 436$, Mean $= 0.7917$, StDev $= 0.05994$) showed extreme deviation from the assumption of normality. The value for AD from the Anderson-Darling test was found to be 10.265 with $p < 0.005$. As $p < 0.05$, therefore, the assumption of normality was violated, as illustrated in Figure 4. For this reason, non-parametric Mann-Whitney U tests were used to compare the atrophy in cerebral matter between the two groups.



2.

Figure 4. Normal Probability Plot of nWBV

3. **Atlas Scaling Factor (ASF):** The pattern of distribution for ASF ($N = 436$, Mean $= 1.199$, StDev $= 0.1287$) indicated marginal rejection of normality with an AD value of 0.773 and p −value 0.044 as depicted in Figure 5. Given that the p −value was just under the nominal significance level ($p < 0.05$), the ASF distribution was considered non-normally distributed; hence the use of non-parametric Mann-Whitney U tests.

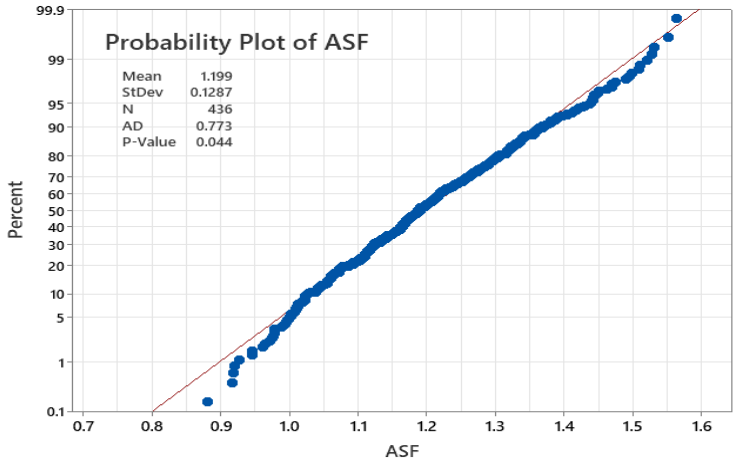


Figure 5. Normal Probability Plot of ASF

Table 5 displays the summary of the normality diagnostics for Structural MRI biomarkers ($N = 436$). The univariate diagnostics for comparisons between groups are displayed in Table 6 ($N = 235$). As shown in Table 6, continuous variables that met the condition of normality were subjected to two samples t −test and Mann Whitney U test.

Table 5: Summary of Normality Tests for MRI Biomarkers ($N = 436$)

Biomarker	Mean ± SD	AD Statistic	p −value
$eTIV$ (mm^3)	1482 ∓ 158.7	0.669	0.080
$nWBV$ (Ratio)	0.7917 ∓ 0.0599	10.265	< 0.005
ASF (unitless)	1.199 ∓ 0.1287	0.773	0.044

Table 6: Comparison of Univariate Diagnostics Among Groups ($N = 235$)

Variable / Biomarker	Demented ($n_1 = 100$)	Control ($n_0 = 135$)	Test Statistic (t / W)	p-value	Significance
Age (years)	76.76 ± 7.12	69.07 ± 13.88	$t = 5.62$	< 0.001	Significant
$MMSE$ (score)	24.32 ± 4.17	29.10 ± 1.13	$W = 5241.5$	< 0.001	Significant
$eTIV$ (mm^3)	1483.75 ± 168.32	1441.53 ± 152.36	$t = 1.99$	0.048	Significant
$nWBV$ (ratio)	0.722 ± 0.036	0.769 ± 0.046	$W = 5822.0$	< 0.001	Significant
ASF (unitless)	1.197 ± 0.131	1.231 ± 0.127	$W = 14102.5$	0.045	Significant

3.2. Multivariable Binary Logistic Regression Modelling

Independent determinants of dementia status ($1 = demented/CDR > 0$ compared to $0 = control/CDR = 0$) were determined using binary logistic regression analysis on $N = 235$ records (100 demented subjects and 135 controls).

3.2.i. Model Fit and Diagnostic Goodness-of-Fit

The overall multivariable model proved to be highly statistically significant and had good parameters for model fit:

- **Overall Model Significance:** The Wald test revealed that overall model fit is significant ($\chi^2(5) = 50.88, p < 0.0001$).

- **Model Explanatory Power:** The model obtained the Deviance R^2 of 48.17% (Adjusted $R^2 = 46.61\%$). Good information criteria values were obtained as $AIC = 178.14$ and $BIC = 198.89$.
- **Goodness-of-Fit:** The Hosmer-Lemeshow test showed $\chi^2(8) = 7.31, p = 0.503$. As $p > 0.05$, there is no lack of fit in the model, which confirms excellent model calibration.
- **Discriminating Power of the Model:** The Area Under the ROC curve (AUC) equals 0.9159.

3.2..ii. Predictor Effects and Multicollinearity Analysis

Table 7 reveals that out of all the considered continuous predictors in this study, only $MMSE$ and $nWBV$ were found to be independent and statistically significant predictors of dementia:

1. **Mini-Mental State Examination ($MMSE$):** Exhibited a strong negative association with the risk of dementia ($\beta = -0.818, SE = 0.132, Z = -6.18, p < 0.001$). With each one-point increment in the $MMSE$ score, the odds of being demented decreased by about 55.9% ($OR = 0.4412, 95\%\ CI: [0.3403, 0.5720]$).
2. **Normalised Whole Brain Volume ($nWBV$):** Demonstrated a statistically significant protective effect from dementia ($\beta = -19.37$, SE $= 7.15$, $Z = -2.71, p = 0.007$). A decrease in the normalised whole brain volume significantly increased the dementia risk (OR < 0.001, 95% CI: $[0.0000, 0.0047]$).
3. **Other Predictors (Age, $eTIV$, ASF):** Age ($\beta = -0.0185, p = 0.460$), $eTIV$ ($\beta = 0.00736, p = 0.387$), and ASF ($\beta = 7.5, p = 0.478$) failed to attain statistical significance when adjusted for $MMSE$ and $nWBV$. The Variance Inflation Factors (VIF) for Age (2.06), $MMSE$ (1.07), and $nWBV$ (2.11) do not suggest any multicollinearity problem, while the very high VIF values for $eTIV$ (51.23) and ASF (51.11) are due to mathematical collinearity.

Table 7: Logistic Regression Estimates, Odds Ratio and Collinearity Tests ($N = 235$)

Predictor Term	Coef (β)	SE Coef	Z-Value	p −value	Odds Ratio (OR)	95% CI for OR	VIF
Constant (β_0)	18.3	26.1	0.70	0.483	—	—	—
Age (years)	−0.0185	0.0251	−0.74	0.460	0.9817	[0.9346,1.0311]	2.06
$MMSE$ (score)	−0.818	0.132	−6.18	< 0.001	0.4412	[0.3403,0.5720]	1.07
$eTIV$ (mm^3)	0.00736	0.00851	0.86	0.387	1.0074	[0.9907,1.0243]	51.23
$nWBV$ (ratio)	−19.37	7.15	−2.71	0.007	< 0.001	[0.0000,0.0047]	2.11
ASF (unitless)	7.5	10.6	0.71	0.478	1848.34	[0.0000,1.92 × 10^{12}]	51.11

Note: Event level modelled = 1 (Demented).

3.2.iii. Predictive Logit Model Equation

The equation for empirical logistic regression of the log-odds (Y') and probability (P) of dementia is shown below:

$$Y' = 18.3 - 0.0185(\text{Age}) - 0.818(\text{MMSE}) + 0.00736(\text{eTIV}) - 19.37(\text{nWBV}) + 7.5(\text{ASF}) \quad (4)$$

$$\hat{P}(\text{Dementia}) = \frac{\exp(Y\prime)}{1+\exp(Y\prime)} \quad (5)$$

3.3. Diagnostic Performance and ROC Curve Analysis

The discriminative power and efficiency of diagnosis of the fitted multiple binary logistic regression model to classify the two groups of patients into either demented patients ($CDR > 0$) or non-demented controls ($CDR = 0$) was assessed by conducting the Receiver Operating Characteristic (ROC) curve analysis. The ROC curve of the adjusted model is shown in Figure 6. The Area Under the ROC Curve (AUC) is used as a comprehensive measure of diagnostic accuracy, ranging from 0.50 (no better than chance classification discrimination) to 1.00 (perfect discrimination).

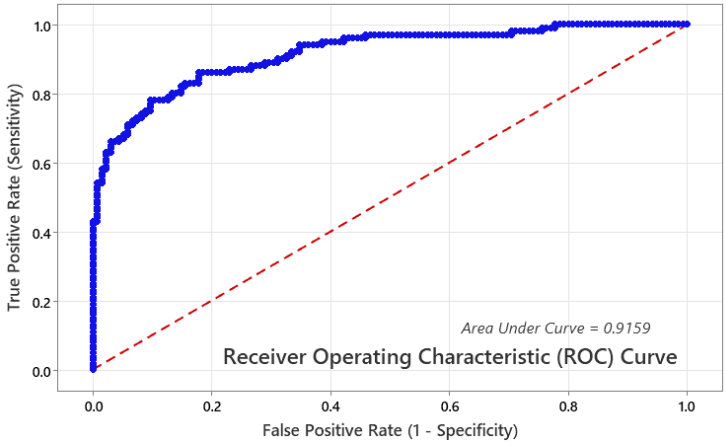


Figure 6. Receiver Operating Characteristic (ROC) Curve for the Multivariable Binary Logistic Regression Model (AUC = 0.9159)

3.3.i. Diagnostic Discrimination Ability

A. **Area Under the Curve (AUC):** The multivariate logit model showed excellent classification ability of the model, as it had an observed AUC value of 0.9159 (95% CI: [0.878,0.954]) $b$ased on well-known diagnostic criteria, $AUC >$ 0.90 is indicative of excellent discrimination ability (Table 8)

B. ***ROC* Curve Trajectory:** As shown in Figure 6 below, the ROC curve shows an almost vertical rise from the origin towards the upper left coordinate. This indicates a high True Positive Rate (Sensitivity) with respect to low False Positive Rates (1 −Specificity).

C. **Pairwise Association Statistics:** Further validation of the model was provided by analysis of the distribution of concordant and discordant pairs. The results show that there were high proportions of concordant pairs ($\approx 91.6\%$) and very few discordant pairs (Table 8).

Table 8: Performance of Diagnostic Classification and Tests of Association

Metric / Parameter	Empirical Value	Standard Classification Criteria / Interpretation
Area Under ROC Curve (AUC)	0.9159	Outstanding Discrimination Performance ($AUC > 0.90$)
95% Confidence Interval for AUC	$[0.878, 0.954]$	Statistically Significant Classification Accuracy ($p < 0.001$)
Concordance Index (C)	91.6%	High Pairwise Diagnostic Classification Accuracy
Somers' D	0.832	Strong Ordinal Predictive Power
Goodman-Kruskal Gamma	0.835	Excellent Monotonic Model Association
Kendall's Tau-a	0.407	Positive Diagnostic Association
Overall Classification Status	Highly Discriminative	Excellent Clinical Decision Support Potential

Note: AUC = Area Under the Receiver Operating Characteristic Curve; CDR = Clinical Dementia Rating.

4. Discussion

The main aim of this research work is the development of a scientifically proven statistical model for diagnosing dementia state ($CDR > 0$ versus $CDR = 0$) based on volumetric markers of $sMRI$ in addition to standardised clinical and demographic covariates using the $OASIS - 1$ cross-sectional dataset ($N = 235$). This analysis demonstrates the diagnostic significance of both cerebral atrophy and cognition assessment measures by the use of nonparametric diagnostics of distributions, bivariate statistical hypothesis testing, and multivariable binary logistic regression.

4.1. Evaluation of Volumetric Atrophy and Cognitive Predictors

The empirical results indicate that the variables $nWBV$ and $MMSE$ scores represent the key independent predictors of the existence of dementia. According to Table 7, $nWBV$ is significantly associated with protection from dementia ($\beta = -19.37, p = 0.007$). The reduction of normalised brain volume is directly correlated with an increase in the odds ratio of cognitive impairments. This is due to the ongoing process of neurodegenerative losses and shrinkage of the cerebral cortex. This fact perfectly correlates with the results of the qualitative examination of the axial $T1$ −weighted MRI (Figure 1).

On the other hand, $MMSE$ scores exhibited an extremely strong inverse relationship with the probability of dementia ($\beta = -0.818, p < 0.001, OR = 0.4412$). One unit increase in $MMSE$ score is equivalent to a decline in the odds of developing dementia by roughly 55.9%. It is noteworthy that although demographics and measures of cranial volume (Age, $eTIV$ and ASF) displayed statistically significant differences between the groups in bivariate univariate analysis (Table 6), their effect became statistically insignificant in the multivariate logit regression model (Table 7). This suggests that the variance of dementia associated with the factors of age and initial head volume is mostly accounted for by brain atrophy ($nWBV$) and cognitive decline ($MMSE$).

4.2. Model Goodness-of-Fit and Diagnostic Discrimination

The developed multivariable binary logistic regression model had strong statistical fit and a good explanatory power. Statistical significance of the whole model was shown using the Wald test ($\chi^2(5) = 50.88, p < 0.0001$), explaining 48.17% of deviance (Deviance$R^2 = 48.17\%$). Model calibration was additionally verified using the Hosmer-Lemeshow test (χ^2 $(8) = 7.31, p = 0.503$), where the insignificant p −value ($p > 0.05$) indicates a very good fit between observed outcomes and the predicted probabilities.

In terms of discrimination ability, the approach proved to be very efficient, producing an Area Under the ROC Curve (AUC) of 0.9159 (95% CI: $[0.878, 0.954]$)(Figure 6, Table 8). Moreover, the strong discriminating power of the model is additionally confirmed by such indices as the Concordance Index (C), which amounts to 91.6%, and Somers' D rank correlation equaling 0.832.

4.3. Collinearity Diagnostics and Methodological Insights

An important methodological issue that has been taken into consideration in this context is that of assessing multicollinearity. Variance Inflation Factors for Age (2.06), $MMSE$ (1.07) and $nWBV$ (2.11) have all been found to be well within the acceptable range ($VIF < 5.0$), thus indicating no problem with multicollinearity between the predictor variables. On the contrary, the extreme values of VIF found between $eTIV$ (51.23) and ASF (51.11) result from the inherent mathematical relationship between these two variables since ASF is calculated to normalise individual brain structure to atlas space.

4.4. Limitations and Future Scope

However, despite the impressive measurement metrics, several weaknesses deserve mention:

A. **Observational Study:** The study is limited by its cross-sectional nature, which makes it hard to create a temporal causal pathway of neurodegenerative progression.
B. **Volumetric Precision:** The current model uses global volumetric measurements ($nWBV$). Regional subcortical segmentation of volumes, such as the hippocampus and ventricles, can improve early-stage classification ($CDR = 0.5$).
C. **Multi-Modality Analysis:** Future models must consider the integration of $sMRI$ biomarkers, clinical data, fluid biomarkers, and deep learning architectures.

5. Conclusions

The proposed research successfully developed and validated the multiple variable statistical model for the automated diagnosis of dementia based on $T1$ weighted MRI biomarkers ($nWBV$, $eTIV$, ASF) and covariates ($MMSE, Age$) in the $OASIS-1$ dataset ($N = 235$). The main results of the current research are:

1. **Biomarker Significance:** The two most significant independent predictors for dementia diagnosis are normalised whole brain volume ($nWBV$) and Mini-Mental State Examination ($MMSE$) with a p −value of 0.007 and $p < 0.001$, respectively.

2. **Diagnostic Power:** The multivariate logit model shows high discrimination power, which is proved by AUC value of 0.9159 (95% CI: $[0.878,0.954]$), which gives us a 91.6% concordance.
3. **Statistical Model Fitting:** The model shows high goodness-of-fit and calibration, which is confirmed by the value of pseudo R^2 equal to 48.17% (McFadden's) and a non-significant Hosmer-Lemeshow test ($\chi^2(8) = 7.31, p = 0.503$).

In conclusion, the incorporation of $sMRI$ −based volumetric markers together with standardised cognitive assessments yields an objective, reproducible, and highly accurate approach in computational neuroimaging for diagnosing and treating Alzheimer's disease and related forms of dementia.

Conflict of interest **:** Authors are declared that there is no conflict of interest regarding this study.

Ethics approval **:** Not applicable

Code availability **:** Available from the author on reasonable request.

AI-Assistance Disclosure **:** Drafting and editing of this manuscript were assisted by a general-purpose language model. The authors independently reviewed and revised all AI-assisted content, ensuring that the final text reflects the authors' own academic voice, interpretation, and responsibility for the accuracy and integrity of the work.

Availability of data and materials **:**

Authors' contributions **:** The author has been written and revised the whole paper.

References

I. Altman, D. G. (1990). *Practical statistics for medical research*. CRC Press.
II. Barnett, A. S., Irfanoglu, M. O., Landman, B., Rogers, B., & Pierpaoli, C. (2021). Mapping gradient nonlinearity and miscalibration using diffusion-weighted MR images of a uniform isotropic phantom. *Magnetic Resonance in Medicine*, *86*(6), 3259–3273.
III. Bethea, R. M. (1995). *Statistical methods for engineers and scientists* (Vol. 144). CRC Press.
IV. Buckner, R. L., Head, D., Parker, J., Fotenos, A. F., Marcus, D., Morris, J. C., & Snyder, A. Z. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalisation: reliability and validation against manual measurement of total intracranial volume. *Neuroimage*, *23*(2), 724–738.
V. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences. 2nd edn Lawrence Erlbaum: Hillsdale*. NJ.
VI. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.-O., Chupin, M., Benali, H., Colliot, O., & Initiative, A. D. N. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage*, *56*(2), 766–781.

VII. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874.
VIII. Fischl, B. (2012). FreeSurfer. *Neuroimage*, *62*(2), 774–781.
IX. Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.
X. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.
XI. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2000). *Applied logistic regression* (Vol. 2). wiley New York.
XII. Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., Fox, N. C., Jack Jr, C. R., Ashburner, J., & Frackowiak, R. S. J. (2008). Automatic classification of MR scans in Alzheimer's disease. *Brain*, *131*(3), 681–689.
XIII. Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.
XIV. Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, *19*(9), 1498–1507.
XV. Mathews, P. G. (2004). *Design of Experiments with MINITAB*. Quality press.
XVI. Montgomery, D. C., Runger, G. C., & Hubele, N. F. (2007). Engineering statistics. John Wiley & Sons. *Inc: Hoboken, NJ, USA*.
XVII. Morris, J. C. (1993). The Clinical Dementia Rating (CDR) current version and scoring rules. *Neurology*, *43*(11), 2412.
XVIII. Navidi, W. C. (2008). *Statistics for engineers and scientists* (Vol. 20). McGraw-Hill Higher Education New York.
XIX. Risan, H. K., Al-Azzawi, A. A., & Al-Zwainy, F. M. S. (2024). *Statistical Analysis for Civil Engineers*. Elsevier.
XX. Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, *17*(1), 87–97.
XXI. Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, *69*(347), 730–737.
XXII. Suk, H.-I., Lee, S.-W., Shen, D., & Initiative, A. D. N. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*, *101*, 569–582.