Natural Resources for Human Health

eISSN: 2583-1194

Original Research

Received 03 June 2026
Revised 18 July 2026
Accepted 01 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (15s): 1353–1368

KEYWORDS:
Alzheimer disease Magnetic resonance imaging Clinical data Machine learning Model validation Study protocol

ADNI Imaging Inventory and Protocol for MRI and Clinical Classification of Alzheimer Disease

G. Gifta Jerith*[1], R. Poornima[2], G. Naveen Sundar[3], D. Narmadha[4], S. Joyson Abraham[5]

[1]*School of Engineering, Malla Reddy University, Hyderabad, India
[2]Department of Information Technology, RMK Engineering College, Chennai, India
[3]Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
[4]Division of Artificial Intelligence and Machine Learning, Karunya Institute of Technology and Sciences, Coimbatore, India
[5]Department of Mechanical Engineering, Malla Reddy College of Engineering and Technology, Hyderabad, India.

ABSTRACT: Combining magnetic resonance imaging (MRI) with clinical information requires reliable participant linkage and evaluation that controls diagnostic overlap and data leakage. This study aims to describe an available Alzheimer's Disease Neuroimaging Initiative (ADNI) imaging inventory and propose a comparison of MRI–clinical fusion with simpler models. We describe the supplied T1-weighted MRI catalogue and three separate ADNI4 illustrative linkages. The proposed study uses demographic and APOE predictors, with cognitive scores evaluated separately. Participant-level nested cross-validation will assess discrimination, calibration and explanation stability. The catalogue contains 7,263 MRI records from 824 participants: 1,325 screening, 274 baseline and 5,664 follow-up records. Phase labels comprise 7,213 ADNI1 and 50 ADNIGO records. The three ADNI4 examples illustrate differences in clinical timing and missing information. The final eligible cohort, quality-control exclusions and model-performance results are not yet available. The available evidence supports an imaging-inventory description and proposed analytical protocol. Completed participant-level linkage and validation are required before concluding diagnostic performance or clinical usefulness.

1. INTRODUCTION

Alzheimer disease causes progressive cognitive and functional impairment. Structural magnetic resonance imaging (MRI) can characterise anatomical changes, including medial temporal atrophy and ventricular enlargement, but clinical syndromes and underlying pathology are not interchangeable. A model trained on clinical diagnostic categories estimates agreement with those categories. It cannot, by itself, establish Alzheimer pathology or predict future progression. These distinctions are essential when defining the intended use of a diagnostic model. Evaluation must also identify which clinical information is available at the intended assessment and whether the model adds value beyond that information [7, 8, 15, 18].

Multimodal approaches combine imaging and clinical information through discriminant scoring, feature fusion, joint representations or feature selection. Published findings support investigation of these approaches, but differences in cohorts, endpoints and validation prevent direct transfer of reported accuracy to another study. Gated fusion therefore requires comparison with simpler models using identical participants and evaluation procedures [3, 4, 11, 23, 24].

Recent work in this journal has also proposed multimodal Alzheimer classification using MRI, PET and clinical information. NeuroFusion provides a conceptual precedent, but its numerical evaluation is explicitly illustrative and cannot establish a performance benchmark for this study [6]. The present protocol focuses on a narrower, testable comparison: the incremental contribution of MRI under participant-level separation and a clinical predictor set that limits direct overlap with diagnostic scoring.

Longitudinal imaging resources create additional challenges. Repeated scans can place information from the same participant in training and testing if partitioning occurs at the image level. Scanner differences, preprocessing decisions and model selection can also influence performance. Participant-level separation and transparent comparators are necessary for interpreting the incremental contribution of MRI [12, 13].

Explanation methods introduce a separate evaluation requirement. Score-CAM highlights image regions associated with model predictions, whereas LIME approximates predictions locally. Plausible-looking explanations do not establish causality or clinical validity; stability, fidelity and parameter-randomisation checks remain necessary [1, 20, 29].

Related machine-learning applications in glaucoma, diabetes and heart disease provide examples of model development across medical settings. Their findings do not validate an Alzheimer disease model, for which the population, target diagnosis, predictor timing and evaluation must be specified independently [10, 14, 21].

This report describes an available ADNI imaging inventory and three separate illustrative MRI–clinical linkages, then specifies a participant-level comparison of gated fusion, concatenation and unimodal models. The primary question is whether adding MRI improves held-out classification beyond age, sex, education and APOE information under identical participant partitions. A separate expanded analysis examines dependence on cognitive scores that may overlap with diagnostic assignment. Reporting is planned with reference to TRIPOD+AI, STROBE and PROBAST+AI. The immediate contribution is a traceable data inventory and an explicit evaluation protocol; benefits for diagnosis or community care remain untested [5, 16, 28].

MRI provides a non-invasive structural view of brain anatomy and can support the study of atrophy patterns, ventricular configuration and regional tissue contrast. In this protocol, the scans are treated as image inputs whose predictive contribution must be evaluated against clinical predictors under participant-level separation. A displayed scan should therefore be read as a documented acquisition example, not as a representative image of an entire diagnostic group. Differences in slice orientation, acquisition timing, scanner protocol, motion and preprocessing can alter visual appearance and may be mistaken for disease-related variation if they are not recorded.

2. MATERIALS AND METHODS

2.1 Design and study status

This report combines a descriptive imaging-catalogue inventory, three illustrative MRI–clinical linkages and a proposed secondary-data analysis. ADNI is a longitudinal, multicentre resource for imaging, clinical and biological-marker research. Its clinical research assessments are described here as structured clinical data, rather than assumed to be routine electronic health records. The ADNI1/ADNIGO catalogue and the separately obtained ADNI4 examples have different roles and are not pooled in the proposed primary analysis [17, 19, 30].

This protocol specifies a planned secondary analysis as of 28 September 2026. Model development and validation have not yet been performed, and the protocol has not been registered. Before model fitting, the investigators will confirm feasibility, record a dated analytical freeze and retain any subsequent deviations. The completed work reported here comprises the imaging inventory and illustrative clinical linkages.

2.2 Data provenance and dates

The working imaging catalogue is dated 23 August 2026. The illustrative linkage uses September 2026 clinical extracts, including APOERES_22Sep2026.csv and ADAS_22Sep2026.csv. Acquisition dates, clinical assessment dates, file-download dates and the eventual analytical-freeze date will remain separate fields. The full-cohort manifest will retain participant, phase, visit, series and image identifiers, file hashes, source-table versions and a reconciliation log. The catalogue contains original MRI records; correction and scaling form part of the planned preprocessing.

Natural Resources for Human Health
eISSN: 2583-1194

2.3 Eligibility and selection of one MRI per participant

The candidate pool comprises the 824 participants reported in the supplied ADNI1/ADNIGO catalogue. Eligibility will require an original structural T1-weighted acquisition, an interpretable clinical diagnosis and successful image quality control. The index clinical assessment will be the first available enrolment baseline visit; screening will be used when no baseline diagnostic assessment exists. Later follow-up visits will not replace an unavailable enrolment assessment in the primary analysis. Participants with unresolved identity, missing target diagnosis, another dementia syndrome or unresolved diagnostic conflict will be excluded with an explicit reason.

The latest released diagnostic-summary source record will define cognitively normal (CN), mild cognitive impairment (MCI) or Alzheimer disease (AD) dementia after dictionary-based harmonisation across phases. ADNIMERGE2 will support linkage and consistency checks rather than silently override a contradictory source diagnosis. Discordant records for the same nominal visit will be checked against source forms; unresolved cases will be excluded. Diagnosis will not be reconstructed from MMSE, CDR, ADAS-Cog or FAQ thresholds.

Eligible MRI acquisitions must be coded screening or baseline and fall within 90 days before or after the index diagnostic assessment. After quality control, select the scan with the smallest absolute date difference; ties will favour baseline over screening, then the earlier acquisition and finally the lowest numeric image identifier. One full volume, not one slice, will represent each participant. Missing acquisition or diagnostic dates prevent primary inclusion. A 30-day matching-window sensitivity analysis will examine temporal dependence. Figure 1 describes this planned sequence.

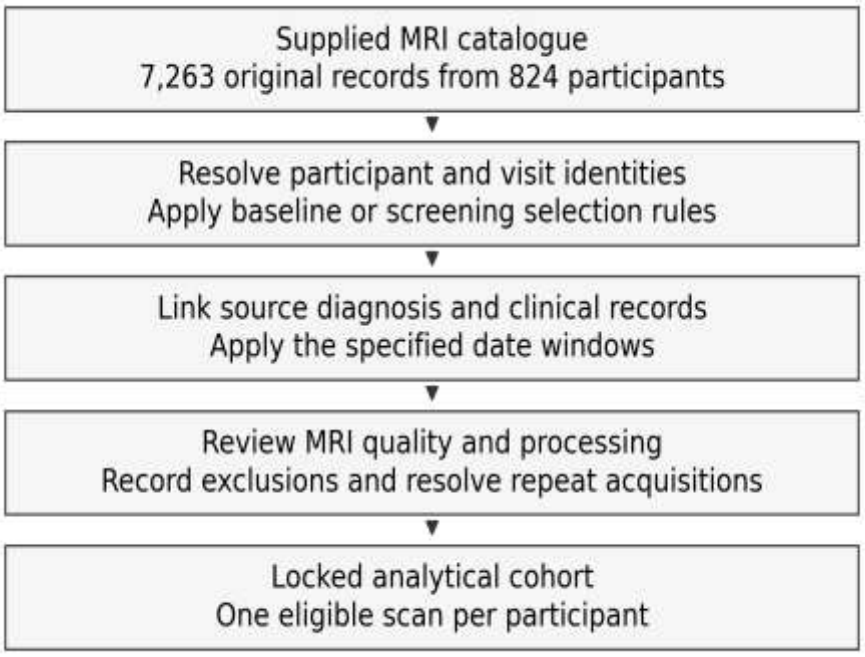


Figure 1. Proposed cohort-construction workflow. The catalogue counts are available; clinical eligibility, image exclusions and the final analytical sample remain to be established.

2.4 Predictor timing and diagnostic circularity

The primary clinical set will comprise age at the selected MRI, recorded sex, completed years of education and APOE ε4 allele count. Sex and education will be obtained from participant-characteristic records at or before the index assessment; age will use the scan-specific value or a documented date-based calculation. APOE is a stable genotype: a valid result from any recorded date may be used for this retrospective classification study, with its test date and phase retained. Conflicting genotypes will be coded unknown. Later genotype ascertainment will not be represented as evidence of prospective availability.

Natural Resources for Human Health
eISSN: 2583-1194

A secondary expanded clinical set will add ADAS-Cog 13, CDR sum of boxes and FAQ. For time-varying scores, select the most recent assessment on or before the index diagnostic date within 90 days; later measurements are not predictors. The expanded set will be evaluated separately because these measures may overlap with diagnostic assignment. MMSE and global CDR in the illustrative table are descriptive and are not additional primary predictors. Primary and expanded analyses will use the same eligible participants and outer-fold assignments.

Table 1. Predictor sets and handling in the proposed analysis

Domain	Primary model	Secondary addition or handling
MRI	One eligible T1-weighted volume	Same image in every model requiring MRI
Demographic	Age, recorded sex, education	Training-fitted encoding and scaling
Genetic	APOE ε4 count: 0, 1 or 2	Unknown is a separate category
Cognitive	Excluded	ADAS-Cog 13
Clinical staging	Excluded	CDR sum of boxes
Functional	Excluded	FAQ score

2.5 Missing clinical information

Missingness will be reported by variable and diagnostic group before modelling. Continuous predictors will receive training-partition medians and missingness indicators, followed by training-fitted standardisation. If a variable has no observed training values, its imputed value and scale will be fixed to 0 and 1, respectively, with the missingness indicator retained. Recorded categorical values will be one-hot encoded using the documented dictionary; unknown or unseen values receive an explicit unknown category. APOE will not be assigned an inferred biological genotype. Outcomes and MRI volumes will not be imputed. Complete-case and no-APOE sensitivity analyses will assess dependence on missingness; their denominators will be reported.

2.6 MRI processing and quality control

DICOM series will be converted to NIfTI with orientation and voxel spacing checked against the source headers. The proposed fixed pipeline comprises N4 bias-field correction, FSL BET brain extraction and a 12-degree-of-freedom affine registration to the external MNI152 T1 brain template at 2-mm resolution. N4 will use a shrink factor of 2, a 200-mm B-spline distance and four convergence levels of 50, 50, 30 and 20 iterations with a tolerance of 10^{-6}. BET will start with a fractional threshold of 0.35 and vertical gradient 0; only 0.25 or 0.45 may be tried when blinded visual review identifies under- or over-extraction. FLIRT will use correlation ratio, trilinear resampling and nearest-neighbour mask resampling. These proposed settings draw on established processing methods [9, 22, 25].

Each registered volume will be clipped at its own within-brain 1st and 99th intensity percentiles and standardised using that volume's brain mean and standard deviation. These are image-specific operations, not cohort-fitted transforms. The 91 × 109 × 91 template grid will be zero-padded symmetrically to 96 × 112 × 96 voxels, placing an odd extra voxel on the high-index side. No study-derived template or global harmonisation will be fitted before partitioning. Package versions, the template checksum, commands and configuration will be frozen before processing the analytical cohort.

Two reviewers, blinded to diagnosis and model outputs, will assess coverage, orientation, motion, brain extraction and registration. Each item will be marked acceptable, requires permitted reprocessing, or unacceptable. Missing brain regions, severe artefact obscuring anatomy, persistent removal of brain tissue or unresolved gross misregistration will lead to exclusion. Disagreements will be resolved by joint review with a recorded reason. No processing option will be selected on the basis of model performance. Table 2 specifies acceptance checks; it is not a before-and-after results table.

Natural Resources for Human Health
eISSN: 2583-1194

Table 2. Prospective MRI quality-control criteria

Check	Acceptance criterion	Recorded output
Orientation and coverage	Consistent with DICOM orientation; full brain coverage	Pass or exclusion reason
Artefacts	No severe motion or other artefact obscuring anatomy	Reviewer decisions
Brain extraction	No substantial tissue loss or residual non-brain tissue	BET threshold and review
Registration	Brain boundaries and major anatomy align with template	Transform and review
Intensity and provenance	Finite values; non-zero brain variance; source traceable	Configuration and file hashes

2.7 Model architecture and comparator models

All neural models will use the same MRI or clinical branch when that modality is present. The MRI encoder will contain four 3D convolutional blocks with 32, 64, 128 and 256 channels, respectively. Each block uses a 3 × 3 × 3 kernel, stride 2, padding 1, eight-group normalisation and ReLU. Global average pooling followed by a dense projection produces a 128-dimensional MRI vector. The clinical encoder uses dense layers of 64 and 128 units with ReLU and dropout after each layer. Both branch outputs include dropout. Networks will be trained from random initialisation without external patient-data pretraining.

Let h_{MRI} and h_C denote the projected MRI and clinical vectors. Both vectors and the gate g have d = 128 elements. The matrix W_g has dimensions d × 2d, and b_g has d elements. Concatenation yields a 256-dimensional comparator; gated fusion yields a 128-dimensional vector:

$$g = \sigma\left(W_g\,[h_{MRI}\,;\,h_C] + b_g\right) \tag{1}$$

$$h_{fusion} = g \odot h_{MRI} + (1 - g) \odot h_C \tag{2}$$

Here, σ is the element-wise sigmoid, [;] is concatenation, and ⊙ is element-wise multiplication. A linear classifier maps the selected representation to three logits z. For K = 3 classes ordered CN, MCI and AD, probabilities are:

$$p_k = \frac{\exp(z_k)}{\sum_{j=1}^{K}\exp(z_j)} \tag{3}$$

Neural models will minimise weighted cross-entropy. In each training partition, class k receives weight $w_k = N_{train}/(K\,n_k)$, where n_k is its participant count. These weights are recalculated inside each inner or outer training partition; test outcomes never influence them:

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} w_k\, y_{ik}\log(p_{ik}) \tag{4}$$

In equation (4), N denotes the training sample used in the objective, y_{ik} is a one-hot class indicator, and p_{ik} is the corresponding probability. The loss uses the participant average shown, without an additional normalisation by the sum of class weights. Table 3 defines the models. A regularised multinomial logistic regression benchmark provides a lower-complexity clinical comparator.

Natural Resources for Human Health
eISSN: 2583-1194

Table 3. Models compared under identical participant partitions

Model	Representation	Purpose
MRI only	128-dimensional MRI encoder output	Imaging baseline
Clinical neural network	128-dimensional clinical encoder output	Clinical baseline
Clinical logistic regression	Encoded clinical variables with L2 penalty	Low-complexity clinical baseline
Concatenation	256-dimensional joined branch outputs	Simple fusion comparator
Gated fusion	128-dimensional weighted branch combination	Proposed multimodal model

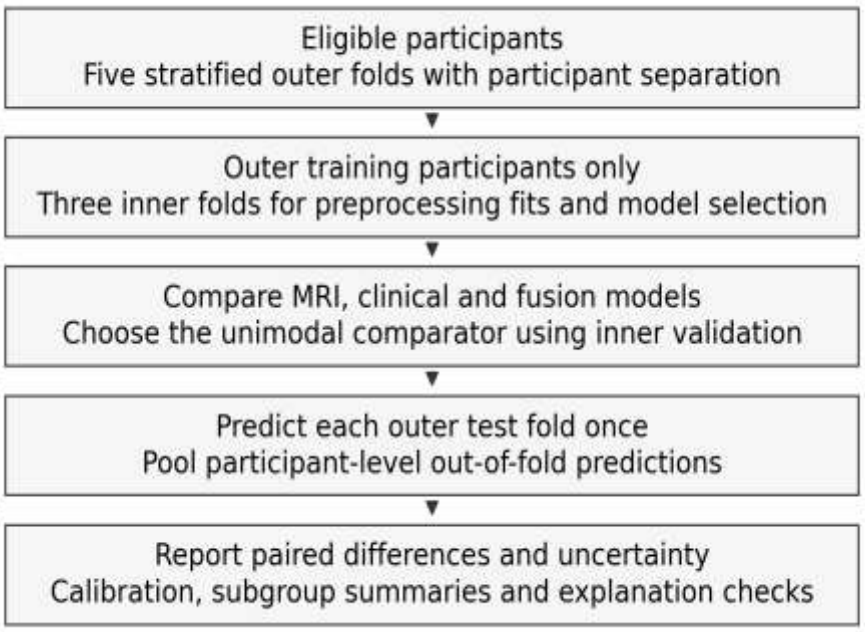


Figure 2. Proposed validation and reporting workflow. All fitted clinical transformations, tuning and comparator selection occur within the outer training participants.

2.8 Statistical analysis

2.8.1 Participant partitions and model selection

The proposed evaluation uses five stratified outer folds and three stratified inner folds. All data from an individual remain in the same fold. Sorting participants by identifier before allocation and using random seed 20260928 will make partitions reproducible. The same outer folds will be used for every model and predictor set. At least 20 eligible participants per diagnostic class will be required to run this design; this is an operational feasibility threshold, not a claim of adequate statistical power. If it is not met, analysis will pause for a dated protocol amendment before any performance comparison.

Within each outer training sample, the neural-model search comprises all eight combinations of learning rate {0.0001, 0.0003}, weight decay {0.00001, 0.0001} and dropout {0.2, 0.5}. AdamW will use $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$, with batch size 4, no data augmentation and a maximum of 100 epochs. Inner validation macro-F1 will determine early stopping

after 15 epochs without improvement; ties favour the earliest epoch. Each setting will use initialisation seeds 20260928, 20260929 and 20260930. Validation probabilities will be averaged across the three initialisations.

The selected setting maximises mean inner-fold macro-F1; ties favour lower learning rate, higher weight decay and then higher dropout. For an outer refit, the epoch count is the rounded median of the selected setting's best inner-fold epochs, followed by refitting on all outer training participants. Logistic regression will use the same class weights and clinical preprocessing, an intercept, an L2 penalty and inverse-regularisation values C = {0.01, 0.1, 1, 10}; ties favour smaller C. The optimiser tolerance will be 10^{-6} with at most 5,000 iterations; non-convergence is logged as a failed candidate.

The unimodal comparator is a selection procedure: in each outer split, choose the MRI-only model, clinical neural model or clinical logistic model using inner-fold macro-F1 alone. Ties favour logistic regression, then the clinical neural model, then MRI. This selected comparator is then evaluated on the untouched outer test fold. Its model identity may differ across outer folds and will be reported. It is never selected using outer-test scores.

2.8.2 Primary estimand and uncertainty

The primary estimand is the difference in pooled out-of-fold macro-F1 between gated fusion and the inner-selected unimodal comparator using the primary clinical set. Every participant contributes exactly one held-out prediction from each procedure. Concatenation comparisons and the expanded clinical set are secondary. Initialisation-ensemble probabilities will be averaged before argmax classification; class-probability ties use the fixed order CN, MCI and AD. No probability threshold will be optimised on outer-test predictions.

Uncertainty will use 5,000 paired, diagnosis-stratified bootstrap resamples of participant-level out-of-fold predictions, with seed 20260928 and percentile 95% confidence intervals. Each resample will apply the same selected participants to all compared models. These intervals describe uncertainty conditional on the fitted cross-validation models; they do not capture all variation from repeating model development or transport to another setting. Fold scores will be shown descriptively and will not be treated as independent observations for a t-test. Secondary comparisons will be labelled exploratory rather than interpreted as confirmatory significance tests.

2.8.3 Performance and calibration

For each class, precision is TP/(TP + FP) and recall is TP/(TP + FN). An undefined precision or recall denominator will be assigned zero and flagged. If both precision and recall equal zero, class-specific F1 is defined as zero. Macro-F1 averages class-specific F1 values with equal weight:

$$\text{Macro-F1} = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{2\,\text{Precision}_k\,\text{Recall}_k}{\text{Precision}_k + \text{Recall}_k}\right] \quad (5)$$

Secondary metrics comprise balanced accuracy, one-versus-rest sensitivity and specificity, macro one-versus-rest AUROC, overall accuracy and the multiclass Matthews correlation coefficient. All will use participant-level held-out predictions and a consistent CN/MCI/AD class order. Undefined AUROC or specificity estimates will be reported as not estimable rather than replaced by zero. Probability quality will be evaluated using the unnormalised multiclass Brier score, which ranges from 0 to 2, and top-label expected calibration error:

$$\text{Brier} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(p_{ik} - y_{ik})^2 \quad (6)$$

$$\text{ECE} = \sum_{m=1}^{M}\frac{|B_m|}{N}\left|\text{acc}(B_m) - \text{conf}(B_m)\right| \quad (7)$$

In equations (5)–(7), N is the number of evaluated participants and K = 3. For ECE, M = 10 fixed equal-width confidence bins span 0 to 1, with each lower boundary included and the upper boundary excluded except in the last bin. Confidence is the maximum predicted probability; accuracy indicates whether the argmax class is correct. Bm contains the participants in bin m, and empty bins contribute zero. Reliability diagrams will display bin counts. Calibration must be assessed separately from rank discrimination [27].

Uncalibrated ensemble probabilities will be the primary probability outputs. In a secondary analysis, a single scalar temperature T between 0.05 and 10 will minimise unweighted negative log-likelihood using inner out-of-fold predictions from the outer training participants. For each candidate T, softmax will be applied to each initialisation's logits divided by T, then the three probability vectors will be averaged. The same transformation and averaging will be applied to the three refitted outer models. The outer test data will not determine T. This aligns calibration with the probability-ensemble prediction rule; averaging logits is not substituted for averaging probabilities. Both uncalibrated and calibrated Brier scores, ECE and reliability curves will be retained.

2.8.4 Explanation analysis

Score-CAM will be applied to all 256 channels of the final MRI convolutional activation layer, with the held-out case's predicted class fixed as the target. Channels will be trilinearly upsampled to the input grid and rescaled individually to [0, 1]; constant channels contribute zero. Each channel mask will modulate the MRI input while clinical inputs remain fixed. The implementation will retain the zero-image baseline, channel scores and positive weighted activation map. The released Score-CAM implementation and any changes needed for 3D inputs will be versioned before evaluation [29].

LIME will explain clinical contributions while holding MRI fixed. For each case, 5,000 perturbations will be generated in the original clinical-variable space using only the outer training data. Continuous observed values will be sampled from training-fitted truncated normal distributions within the observed range; categorical values, including unknown categories, will be sampled as whole categories using training frequencies. Missingness indicators will be derived from the perturbed raw values before applying the fitted imputation, scaling and encoding steps. One-hot components will not be sampled independently. Distances will be Euclidean in the transformed clinical space. The kernel weight will be $\exp(-\text{distance}^2/\text{kernel_width}^2)$, with $\text{kernel_width} = 0.75\sqrt{p}$ for p encoded features. A weighted ridge surrogate with penalty 1 will use all encoded features, with the five largest absolute contributions displayed. Weighted local R^2 below 0.8 will flag low fidelity. Independent marginal sampling can still produce clinically implausible combinations; these explanations describe model behaviour rather than causal effects [20].

Explanation stability will be summarised across the three initialisations and five repetitions of Gaussian image noise with standard deviation 0.01 in normalised intensity units. Within-brain Spearman correlation and Dice overlap of the highest-attribution 10% of voxels will describe image-map agreement. LIME will be repeated with seeds 0–19, reporting top-three-feature Jaccard agreement, sign consistency and local fidelity. Parameter-randomisation checks will replace the classification head and then all neural weights while holding the target class and example fixed. Agreement with the original map will be reported without presenting an arbitrary overlap value as proof of validity [1].

Quantitative explanation summaries will use every eligible outer-test case. Displayed cases will be selected deterministically by participant order within each true/predicted class combination, rather than by visual attractiveness. Failed or constant maps and low-fidelity local explanations will remain in the reporting denominator. Score-CAM and LIME will be described as model-behaviour analyses, not causal explanations or independent clinical validation.

2.8.5 Sensitivity analyses and reproducibility

Exploratory analyses will examine the expanded clinical set, complete cases, exclusion of APOE, a 30-day matching window and baseline-only MRI selection. Age groups below 70, 70–79 and at least 80 years, recorded sex, APOE status, field strength and acquisition phase will be reported when each required diagnostic class has at least 10 participants within the subgroup. Otherwise, report denominators without unstable class-specific performance estimates. Site-level results will be descriptive; participant-level cross-validation does not establish generalisation to unseen sites. A separate site-held-out evaluation would require a dated amendment and sufficient eligible participants.

The analytical package will retain eligibility and exclusion logs, source-table dictionaries, matching intervals, participant split files, configuration files, package versions, random seeds and out-of-fold predictions in the authorised research environment. The final report will distinguish planned, implemented and amended procedures. Publicly shareable code and aggregate outputs will be considered subject to ADNI conditions; no public repository or completed software implementation is claimed here.

3. Results and Discussion

Natural Resources for Human Health
eISSN: 2583-1194

3.1 Imaging catalogue inventory

Completed descriptive findings are limited to the supplied catalogue inventory and the three illustrative linkages. The catalogue reports 7,263 original T1-weighted MRI records from 824 unique participants, equivalent to 8.81 records per participant on average. This mean describes record multiplicity and does not imply equal scan availability across individuals. The phase field contains 7,213 ADNI1-labelled records and 50 ADNIGO-labelled records; phase labels therefore apply to records and do not define separate participant counts.

Table 4. Supplied MRI catalogue inventory before full-cohort eligibility assessment

Item	Count or status	Interpretation
Unique participants	824	Candidate pool, not final analytical sample
Original T1-weighted MRI records	7,263	Includes repeat acquisitions and visits
ADNI1 / ADNIGO records	7,213 / 50	Record-level phase labels
Screening records	1,325 (18.2%)	Potential baseline-adjacent acquisitions
Baseline records	274 (3.8%)	Not the number of eligible participants
Follow-up records	5,664 (78.0%)	Excluded from primary enrolment selection
Visits represented	Screening through month 48	Longitudinal catalogue structure
Eligible CN / MCI / AD counts	Not yet established	Requires clinical linkage and image QC

The visit totals reconcile exactly to 7,263 records. Follow-up records form the largest category, whereas screening and baseline together account for 1,599 records (22.0%). This distribution explains why catalogue size cannot be used as the model-development sample size. A participant may have multiple acquisitions in the same category, and some participants may lack a clinically matched, acceptable enrolment scan. No CN/MCI/AD prevalence or group-level demographic comparison is inferred from this inventory.

3.2 Illustrative ADNI4 MRI and clinical linkage

Three separately obtained ADNI4 MPRAGE series provide illustrative MRI–clinical examples. Their acquisition dates are 19 February, 3 June and 7 August 2025 for cases A, B and C, respectively. These are not established as the selected baseline scans of the ADNI1/ADNIGO catalogue and are not included in an inferred analytical denominator. Table 5 preserves the available values without imputing missing genotype or cognitive assessments.

Table 5. Clinical information for three illustrative ADNI4 cases

Case and diagnosis	Age/sex/education	MMSE	CDR global / sum of boxes	APOE genotype
A CN	95 / female / 18	30	0 / 0	ε3/ε3

Natural Resources for Human Health
eISSN: 2583-1194

Case and diagnosis	Age/sex/education	MMSE	CDR global / sum of boxes	APOE genotype
B MCI	77 / male / 18	29	0.5 / 1.5	Unavailable
C Probable AD dementia	82 / male / 16	20	0.5 / 4.0	ε3/ε4

Age and education are in years. MMSE denotes Mini-Mental State Examination; CDR denotes Clinical Dementia Rating. These selected cases do not estimate group means, class prevalence or model performance.

Clinical timing was not identical across the examples. Case A had CDR and FAQ assessments on 10 February 2025, nine days before MRI, and MMSE and ADAS-Cog assessments on 28 February, nine days after MRI. Case B had MMSE and CDR assessments on 2 June, one day before MRI; the MCI label remains illustrative because its linked diagnostic entry lacks an assessment date, and no matched ADNI4 FAQ or ADAS-Cog result was available. Case C had MMSE and CDR assessments on 5 August, two days before MRI, and a probable AD dementia diagnostic record on 8 August, one day after MRI. Nearby baseline assessments on 14 August, seven days after MRI, recorded FAQ 11, ADAS-Cog 11 of 31.33 and ADAS-Cog 13 of 44.33. These timings describe the examples and do not establish their eligibility under the proposed primary predictor rules.

APOE entries in APOERES_22Sep2026.csv are participant-level genetic records. Case A's ε3/ε3 result was recorded at ADNI1 screening on 17 April 2006 and is historical rather than an ADNI4 measurement. Case C's ε3/ε4 result is labelled ADNI4 but has no test date. Case B has no available result in that extract. The AD diagnosis is taken from the diagnostic source, not inferred from the global CDR value.

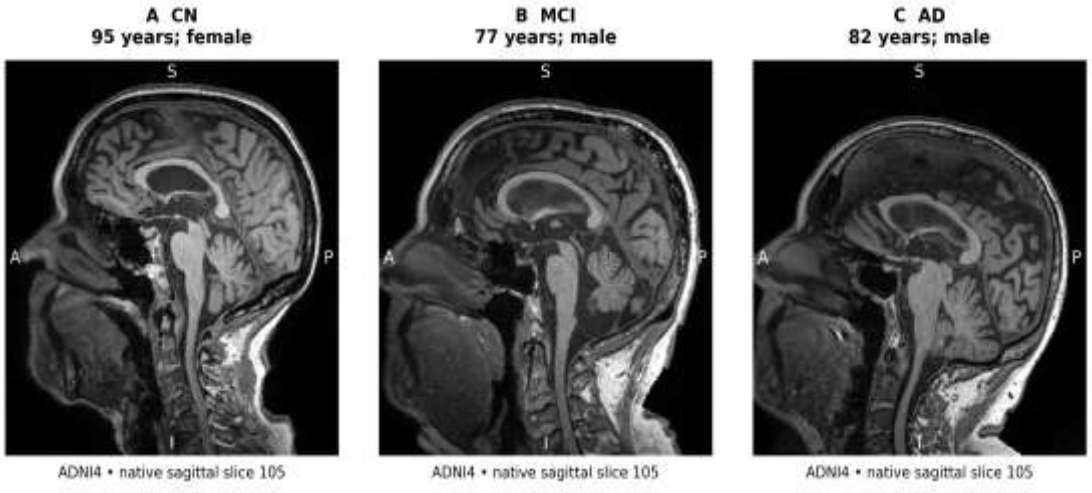


Figure 3. Illustrative ADNI4 T1-weighted MRI cases A–C. Native sagittal pixels are displayed with superior (S) above, inferior (I) below, anterior (A) left and posterior (P) right, using DICOM orientation metadata. The MCI acquisition is slightly oblique. These are source images, not processed-image comparisons or model outputs. The scan panels are provided for image-provenance illustration.

The three displayed scans are sagittal T1-weighted source images selected to demonstrate the available ADNI4 image format and the need for careful image provenance. Superior and inferior directions identify the vertical anatomical axis, while anterior and posterior markers identify the front-to-back orientation. The letters outside the image are orientation markers derived from the acquisition metadata; they are not diagnostic annotations. The scan labels CN, MCI and probable AD dementia are linked clinical categories for these illustrative cases. They do not establish disease pathology, group-level morphology or diagnostic accuracy. Because the cases have different clinical-assessment dates and the MCI example has incomplete diagnostic timing, visual comparison should not be interpreted as a controlled before-and-after or between-group experiment.

Natural Resources for Human Health
eISSN: 2583-1194

Each series contains 208 slices; DICOM instance 105 is shown. The same instance number does not ensure an identical anatomical plane across participants. Display limits were set independently at the 1st and 99.5th intensity percentiles, without spatial interpolation. No quantitative atrophy comparison, representative class morphology or diagnostic accuracy is inferred from these images.

3.3 Pending analytical outcomes

Representative MRI images have been acquired. Full-cohort image acquisition, clinical linkage and preprocessing remain incomplete, and the model experiment has not been performed. Figure 4 therefore reports a proposed participant-selection pathway with unresolved denominators, rather than a completed recruitment or exclusion flow. Figure 5 displays the available record-level visit distribution; Figure 6 shows the record-level phase composition, and Figure 7 summarises the timing of available clinical assessments around the three illustrative scans. Confusion matrices, AUROC curves, calibration plots and explanation maps cannot be supplied as results before the corresponding analyses are completed.

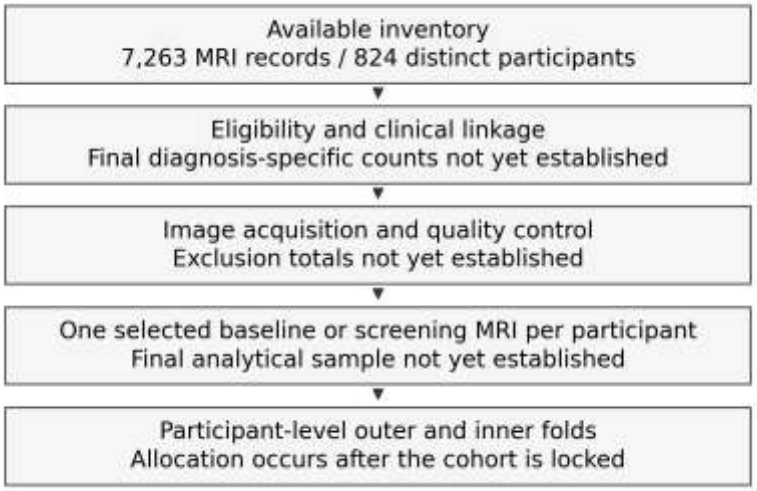


Figure 4. Proposed selection pathway with available catalogue counts and explicitly unresolved analytical denominators. The boxes after the inventory represent planned work, not completed exclusions.

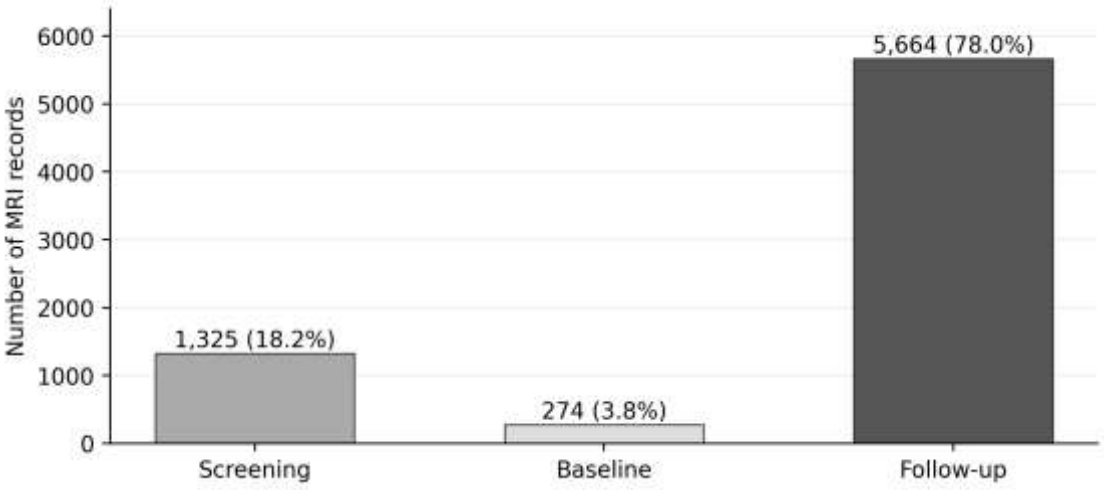


Figure 5. MRI record distribution in the supplied catalogue. Counts and percentages describe records, not independent participants. Screening and baseline records still require clinical matching, quality control and one-scan-per-participant selection.

Natural Resources for Human Health
eISSN: 2583-1194

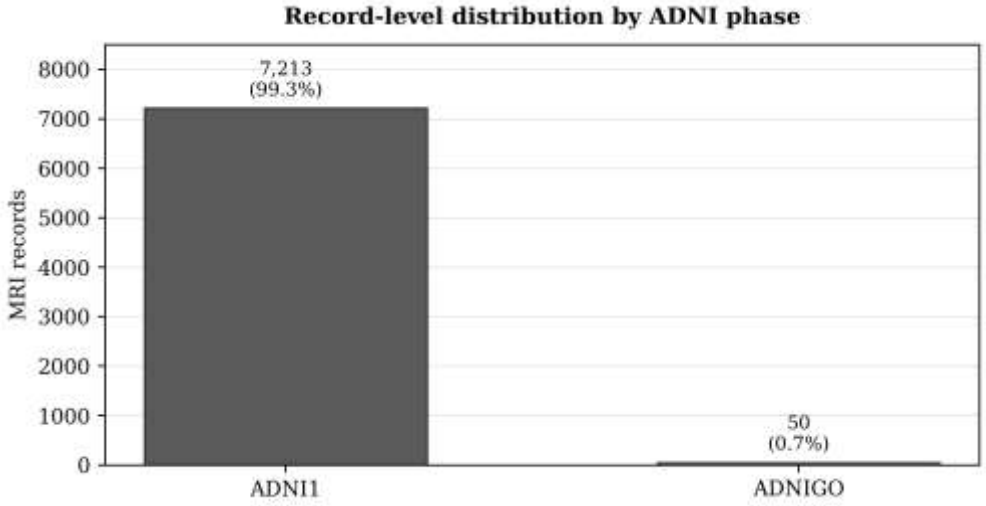


Figure 6. Record-level distribution by ADNI phase. Percentages use the 7,263 MRI records in the supplied catalogue; they do not represent independent participants or eligible analytical observations.

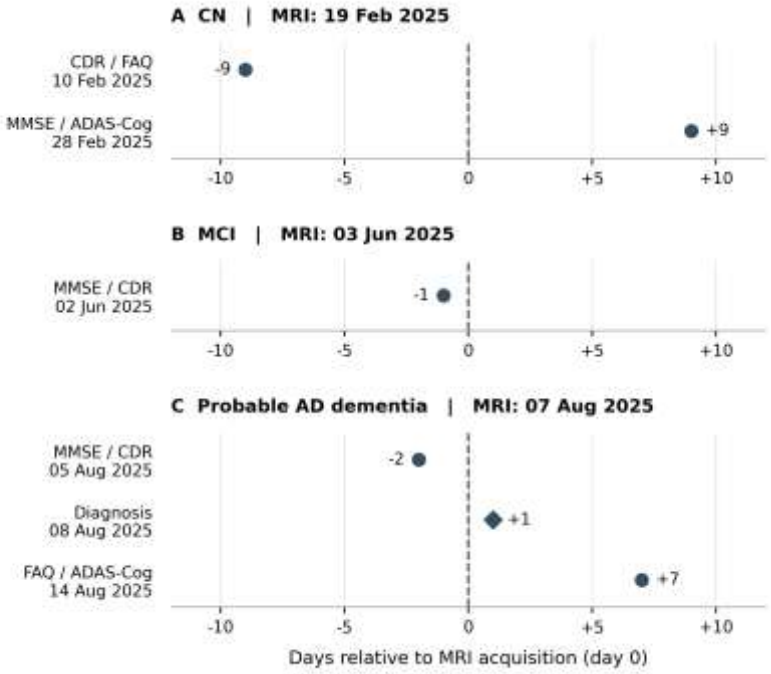


Figure 7. Clinical assessment timing in three illustrative ADNI4 cases. Each panel uses its MRI date as day 0 (dashed line). Circles mark cognitive or functional assessments; the diamond marks diagnosis. Offsets are calendar days. Case B's undated diagnostic entry is not plotted; matched ADNI4 FAQ and ADAS-Cog records were unavailable. These dates describe source records and do not establish predictor eligibility or diagnostic performance.

3.4 Interpretation of the descriptive findings

The available findings establish the scale and longitudinal structure of the supplied imaging inventory and provide traceable examples of clinical linkage. They do not establish an analysis-ready cohort or demonstrate that multimodal learning improves diagnosis. The difference between 7,263 records and 824 participants is consequential: record-level random splitting could place acquisitions from the same person in both training and testing. The proposed participant-level selection and nested validation directly address that risk, although their effectiveness can be judged only after implementation.

The illustrative ADNI4 cases show why apparently simple clinical tables require temporal and source-specific interpretation. A genotype may have been measured years before the displayed MRI, cognitive assessments may fall on different dates, and a diagnostic entry may lack a usable assessment date. Missingness must therefore be documented rather than silently resolved by nearest-record matching. The examples also belong to a different ADNI phase from most catalogue records and cannot demonstrate representativeness of the proposed analytical cohort.

This emphasis on traceability is consistent with the data-quality dimensions discussed by Uddandarao et al. [26], including incompleteness, inconsistency and staleness. Here, those concerns translate into specific checks of assessment dates, diagnostic conflicts, missing predictors and source versions. Their framework motivates these checks; it does not establish the quality or predictive value of the present ADNI inventory.

The primary reduced-proxy clinical set is intended to make the incremental contribution of MRI more interpretable. It does not remove every association between clinical variables and diagnosis, but it avoids directly including cognitive and functional scores that may contribute to diagnostic assignment. The expanded-set analysis will show how results depend on adding these measures. A higher score in that analysis would not, by itself, establish earlier detection or greater clinical usefulness.

The model comparison is designed to distinguish fusion benefit from model-selection bias. The unimodal comparator is selected only within development data and a lower-complexity clinical model is retained. Gated fusion must therefore be compared with alternatives on the same held-out participants. Calibration, class-wise sensitivity and specificity, and uncertainty remain necessary because a single average discrimination metric can conceal clinically relevant failure modes. Explanation stability similarly informs confidence in model behaviour without demonstrating causality.

3.5 Strengths and limitations

The report's current strengths are explicit separation of catalogue facts from proposed analyses, traceable example timing and a defined prospective evaluation specification. Final eligibility counts, diagnostic-group distributions, missingness proportions and quality-control exclusions remain unknown. The procedural minimum class count is not a sample-size justification for a complex model. The authors must report learning curves, confidence intervals, and training behaviour, and conclusions must remain cautious if precision is poor. No claim of MCI predominance or adequate class-specific sample size is supported at this stage. No completed formal sample-size or precision calculation is available. Before implementation, the investigators must justify model complexity against the eligible class counts and the intended precision of performance estimates; the threshold of 20 per class is insufficient for that purpose.

ADNI is a research cohort, and internal participant-level validation will not establish transportability to routine clinical records, other scanner settings or underrepresented populations. MRI–diagnosis matching within 90 days also permits temporal mismatch, addressed only partly by the narrower-window sensitivity analysis. The three examples are selected illustrations. The planned models classify existing clinical categories and do not establish Alzheimer pathology, prospective progression, treatment benefit or a clinical decision-support system.

3.6 Clinical relevance and further work

If the internal study is completed, subsequent evaluation should use an independent cohort with clearly specified intended users, disease prevalence and acquisition conditions. Prospective work would need to test whether predictions or explanations change referral decisions and whether benefits outweigh false-positive and false-negative consequences. Community-health relevance should be assessed through accessibility, calibration across settings and patient-centred outcomes, rather than inferred from model complexity.

4. CONCLUSION

The supplied ADNI inventory reports 824 participants and 7,263 original T1-weighted MRI records. Three separate ADNI4 examples illustrate MRI–clinical linkage and its timing limitations. This report proposes an explicit participant-level protocol for comparing gated fusion, concatenation and unimodal models. Full-cohort curation and completed validation are required before any diagnostic-performance, superiority or deployment claim can be made.

The descriptive inventory also shows why record-level counts must be interpreted cautiously. The 7,263 MRI records represent repeated and longitudinal observations from 824 participants, with follow-up records forming the largest

component. The phase and visit graphs therefore describe the structure of the available resource rather than the prevalence of cognitively normal, mild cognitive impairment or Alzheimer disease participants. The three scan examples further show that clinical assessments and imaging acquisitions are not uniformly timed.

The proposed analysis addresses these issues through one eligible MRI per participant, explicit clinical linkage windows, image-quality review, participant-level nested cross-validation and comparison with clinical-only models. This design is intended to estimate the incremental value of MRI under controlled conditions. It also separates the primary reduced-proxy clinical predictor set from an expanded set containing cognitive measures that may overlap with diagnostic assignment.

The next phase must establish the final eligible cohort, diagnostic-group counts, missingness patterns, quality-control exclusions and a reproducible preprocessing record before model fitting. Only after these steps can the study report held-out macro-F1, class-specific performance, calibration, uncertainty and explanation stability. Independent validation in a separate cohort will be required before any claim of generalisability, clinical usefulness or decision-support readiness. Within its current scope, the manuscript provides a transparent ADNI imaging inventory and a reproducible protocol for completing that evaluation.

Acknowledgements

ADNI collection and data sharing are supported by the US National Institute on Aging through NIH grant U19 AG024904, with the Northern California Institute for Research and Education as grantee. Earlier support included the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research and private contributions administered through the Foundation for the National Institutes of Health.

ADNI's acknowledged private contributors include AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica; Biogen; Bristol-Myers Squibb; CereSpir; Cogstate; Eisai; Elan Pharmaceuticals; Eli Lilly; EuroImmun; F. Hoffmann-La Roche and Genentech; Fujirebio; GE Healthcare; IXICO; Janssen Alzheimer Immunotherapy Research & Development; Johnson & Johnson Pharmaceutical Research & Development; Lumosity; Lundbeck; Merck; Meso Scale Diagnostics; NeuroRx Research; Neurotrack Technologies; Novartis; Pfizer; Piramal Imaging; Servier; Takeda; and Transition Therapeutics. We thank ADNI participants, investigators, coordinators and imaging personnel. The maintained investigator roster is linked in the author note, and the official acknowledgement guidance accompanies the ADNI documentation [2].

Funding

No direct funding for this secondary study is declared. Funding for the ADNI resource is acknowledged separately above.

Author Contributions

G. Gifta Jerith: conceptualisation, methodology and supervision. R. Poornima: data curation and clinical-variable organisation. G. Naveen Sundar and D. Narmadha: methodological review and analysis planning. S. Joyson Abraham: academic supervision and manuscript review.

Conflict of Interest

The authors declare no competing interests.

Ethics Statement

ADNI participant recruitment and data collection were conducted under participating-site ethics oversight and informed-consent procedures [19]. This proposed secondary study uses de-identified data obtained through authorised access and involves no new participant contact. Documentation of the institutional ethics determination for this secondary analysis is pending. No local approval number or exemption is claimed. The applicable determination will be documented before the proposed analysis begins.

Data Availability Statement

The underlying data are available to approved investigators through the LONI Image and Data Archive under ADNI access conditions. The manuscript contains selected illustrative images and clinical summaries, not a downloadable participant-level

dataset. No public code repository or trained model is currently claimed. Any future sharing of analysis code, model outputs or weights will follow ADNI conditions and institutional requirements. The manuscript will undergo the required ADNI publication review before journal submission [2].

REFERENCES

[1] Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., Kim, B., 2018. Sanity Checks for Saliency Maps. Advances in Neural Information Processing Systems 31. Curran Associates, Inc. https://doi.org/10.48550/arXiv.1810.03292

[2] Alzheimer's Disease Neuroimaging Initiative, n.d. ADNI Documentation: Data Use Agreement and manuscript citation guidance. https://adni.loni.usc.edu/help-faqs/adni-documentation/. Accessed 29 September 2026.

[3] Castellano, G., Esposito, A., Lella, E., Montanaro, G., Vessio, G., 2024. Automated detection of Alzheimer's disease: a multi-modal approach with 3D MRI and amyloid PET. Scientific Reports 14(1), 5210. https://doi.org/10.1038/s41598-024-56001-9

[4] Chen, Z., Liu, Y., Zhang, Y., Zhu, J., Li, Q., Wu, X., 2025. Enhanced Multimodal Low-Rank Embedding-Based Feature Selection Model for Multimodal Alzheimer's Disease Diagnosis. IEEE Transactions on Medical Imaging 44(2), 815-827. https://doi.org/10.1109/tmi.2024.3464861

[5] Collins, G.S., Moons, K.G.M., Dhiman, P., Riley, R.D., Beam, A.L., Van Calster, B., et al., 2024. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 385, e078378. https://doi.org/10.1136/bmj-2023-078378

[6] Debnath, M., Sharma, P., Prashanth, J., Arunkumar, R.S., Kumar, S., Sharma, V., Mukherjee, C., 2026. NeuroFusion: A Multimodal Transformer Framework for Early and Explainable Detection of Alzheimer's Disease. Natural Resources for Human Health 6(5), 189–201. https://doi.org/10.53365/nrfhh.1044

[7] Frisoni, G.B., Fox, N.C., Jack, C.R. Jr, Scheltens, P., Thompson, P.M., 2010. The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology 6(2), 67-77. https://doi.org/10.1038/nrneurol.2009.215

[8] Jack, C.R. Jr, Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., et al., 2018. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia 14(4), 535-562. https://doi.org/10.1016/j.jalz.2018.02.018

[9] Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Medical Image Analysis 5(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

[10] Jerith, G.G., Kumar, P.N., 2020. Recognition of Glaucoma by means of Gray Wolf Optimized Neural Network. Multimedia Tools and Applications 79(15-16), 10341-10361. https://doi.org/10.1007/s11042-019-7224-1

[11] Lin, W., Gao, Q., Du, M., Chen, W., Tong, T., 2021. Multiclass diagnosis of stages of Alzheimer's disease using linear discriminant analysis scoring for multimodal data. Computers in Biology and Medicine 134, 104478. https://doi.org/10.1016/j.compbiomed.2021.104478

[12] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., et al., 2017. A survey on deep learning in medical image analysis. Medical Image Analysis 42, 60-88. https://doi.org/10.1016/j.media.2017.07.005

[13] Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik 29(2), 102–127. https://doi.org/10.1016/j.zemedi.2018.11.002

[14] Manivannan, D., Jerith, G.G., Sekhar, G.C., Jagadeesh, S., Samsudeen Shaffi, S., Babu, S.A., 2026. Assessing the Comparative Effectiveness of Ensemble CatBoost Versus XGBoost Models in Predicting Heart Disease Outcomes. Journal of Computer Science 22(4), 1484–1493. https://doi.org/10.3844/jcssp.2026.1484.1493

[15] McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R. Jr, Kawas, C.H., et al., 2011. 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 & Dementia 7(3), 263-269. https://doi.org/10.1016/j.jalz.2011.03.005

[16] Moons, K.G.M., Damen, J.A.A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., et al., 2025. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ 388, e082505. https://doi.org/10.1136/bmj-2024-082505

[17] Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., et al., 2005. The Alzheimer's Disease Neuroimaging Initiative. Neuroimaging Clinics of North America 15(4), 869-877. https://doi.org/10.1016/j.nic.2005.09.008

[18] Petersen, R.C., 2011. Mild Cognitive Impairment. New England Journal of Medicine 364(23), 2227-2234. https://doi.org/10.1056/nejmcp0910237

[19] Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., et al., 2010. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology 74(3), 201–209. https://doi.org/10.1212/WNL.0b013e3181cb3e25

[20] Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 1135–1144. https://doi.org/10.1145/2939672.2939778

[21] Selvaraj, J., Jerith, G.G., Karthikeyan, R., Senthil, K., 2025. Assessment of CatBoost for Diabetes Prevention in Comparison to XGBoost: AI model capable of predicting the onset of diabetes. EAI Endorsed Transactions on Internet of Things 11. https://doi.org/10.4108/eetiot.5880

[22] Smith, S.M., 2002. Fast robust automated brain extraction. Human Brain Mapping 17(3), 143–155. https://doi.org/10.1002/hbm.10062

[23] Suk, H.I., Lee, S.W., Shen, D., Alzheimer's Disease Neuroimaging Initiative, 2014. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage 101, 569–582. https://doi.org/10.1016/j.neuroimage.2014.06.077

[24] Tu, Y., Lin, S., Qiao, J., Zhuang, Y., Zhang, P., 2022. Alzheimer's disease diagnosis via multimodal feature fusion. Computers in Biology and Medicine 148, 105901. https://doi.org/10.1016/j.compbiomed.2022.105901

[25] Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., et al., 2010. N4ITK: improved N3 bias correction. IEEE Transactions on Medical Imaging 29(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

[26] Uddandarao, D.P., Nagpal, R., Challa, A., Pushpam, A., Vats, S., Gupta, N., 2026. Benchmarking the Impact of Data Quality Dimensions on Predictive Analytics Systems Across Healthcare and Life Science Workflows. Natural Resources for Human Health 6(5), 121–135. https://doi.org/10.53365/nrfhh.778

[27] Van Calster, B., McLernon, D.J., van Smeden, M., Wynants, L., Steyerberg, E.W., STRATOS initiative, 2019. Calibration: the Achilles heel of predictive analytics. BMC Medicine 17(1), 230. https://doi.org/10.1186/s12916-019-1466-7

[28] von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P., et al., 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Medicine 4(10), e296. https://doi.org/10.1371/journal.pmed.0040296

[29] Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., et al., 2020. Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 111–119. https://doi.org/10.1109/CVPRW50498.2020.00020

[30] Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., et al., 2017. The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. Alzheimer's & Dementia 13(5), 561-571. https://doi.org/10.1016/j.jalz.2016.10.006