T1 強調 MRI による灰白質萎縮の Z スコア定量プラグイン（vis-brain-atrophy）の方法と評価

灰白質を中心とした脳萎縮の定量

テクニカルノート（GRAPHY-Next 用プラグイン、開発版 0.1.0）

Visionary Imaging Services, Inc.　2026 年 11 月 7 日　ライセンス：コードは MIT、基準 DB と IXI 由来のデータ・図は CC BY-SA 3.0（リポジトリの LICENSE・LICENSE-DATA・NOTICE）

本ソフトウェアは研究・教育目的であり、診断機器ではない。出力は健常者基準との差を示す参考値で、画像所見と臨床情報を合わせた医師の判断に置き換わらない。AD 例での感度は未検証である。

要旨

3 次元 T1 強調 MRI から、健常者基準との差を表す Z スコア（灰白質の萎縮が正）と、処理の経過・統計解析を図解したレポートを作る GRAPHY-Next 用プラグイン「vis-brain-atrophy」について解説する。

処理は HD-BET による脳抽出、N4 バイアス補正、Affine と SyN による標準脳へのレジストレーション、Atropos による組織分割、変調 GM/WM の作成（FWHM 8 mm）、ボクセルごとの回帰による Z スコアの算出の順に行われる。正常の基準コホートは IXI の健常者 373 名（20.0〜83.8 歳、分割 v2）で、共変量は全 DB 共通で年齢・年齢^2・性別・装置メーカー・磁場強度・log(detA) とした。品質管理（QC）は、標準脳位置合わせ時の拡大縮小の程度（detA）、位置合わせ後の相関（corr）、GM 体積（gm_ml）の 3 指標で行った。QC の合格基準は、detA と corr は基準 DB の ±3 標準偏差を超えないこと、gm_ml は上側（+3 標準偏差）を超えないこととし、下側（萎縮側）は警告にとどめた。海馬、内側側頭部（MTL）と背側脳幹白質の VOI 指標、全脳規格化後の Z も計算される。主な出力は、灰白質（GM）の萎縮（変調 GM の Z）である。海馬体積は HippoDeep からの抽出結果から求められる。GM の Z とは別系統である。

本ツールは、VSRAD advance 2（論文 Matsuda 2012 の SPM8＋DARTEL）の再実装ではない。処理（脳抽出・レジストレーション・組織分割・Z の作り方・VOI）も基準データベース（VSRAD は 54〜86 歳・1.5T の単一装置の 80 例、本ツールは IXI 373 名・20〜83.8 歳・3 施設）も異なり、数値は互換ではない（第 2 節）。IXI の T1 強調画像 は MPRAGE など VSRAD 用の T1 強調撮像条件で取得されていない。

評価は、正常の基準コホートからホールドアウトされた IXI 健常者（188 名）の評価コホートで行われた。正常の基準コホートと評価コホートは、施設×性別の層ごとに年齢順で 3 名ずつの組をつくり、各組から 1 名を無作為に評価コホートへ振り分けた（IXI では施設が装置メーカーと静磁場強度を決める）。全脳灰白質では、年齢・年齢^2・性別・磁場強度・log(detA)、装置メーカーを共変量として作成された基準 DB で合格を確認した（1.11%、下限との差は 0.01 ポイントで、余裕はほとんどない。4.3.2 節）。全脳白質は、基準 1・2・施設別に合格したが、基準 5（年齢を ±20 歳ずらす負例）に通らなかった。白質は加齢で増える範囲が長く、前提（一様な減少）が成り立たないためで、単調減少に頼らない 2 つの負例（模擬的な減少、減少側の年齢ずらし）には合格した。そこで全脳白質は警告つきの参考値として出す（4.3.3 節）。58 歳以

上に限った基準も試したが、減少側の負例に通らず併記しない。背側脳幹の灰白質は検査 D2（年齢の負例）に失敗したため、この指標はレポート上で記述しない。VSRAD advance 2 との突き合わせは、 IXI の健常者 5 例（正常の基準コホートに含まれていた）と、この 5 例を除いて作り直した正常の基準 DB（368 例） での測定で比べた。内側側頭部の萎縮度の順位相関は +0.70 で概ね対応したが（粗い目安）、背側脳幹の白質と比は対応しなかった。AD 例での感度、同一人物の別装置での再現性の評価は未達である。

キーワード：脳萎縮、Z スコア、T1 強調 MRI、基準データベース、IXI、HippoDeep、VSRAD、事前登録

1 はじめに

1.1 目的

アルツハイマー型認知症などの神経変性では、T1 強調 MRI で灰白質の萎縮が見られる。萎縮を健常者基準との差として数値化し、領域ごとの分布を図で示す処理パイプラインの例を示すことが、本プラグインの目的である。

対象の範囲：本ツールの主な出力は、灰白質（GM）の萎縮（変調 GM の Z）である。白質は、背側脳幹の白質 VOI の参考指標と、全脳白質の Z（年齢を使った事前登録の検査に通らないため、警告つきの参考値として出す。4.3.3 節）に使う。海馬は HippoDeep の体積と、その回帰 Z・百分位で、GM の Z とは別系統の指標である。

1.2 位置づけと既存手法との関係

日本では、SPM8 と DARTEL を用いた VSRAD advance が、内側側頭部の萎縮度（VOI 内で正の Z の平均）、VOI 内萎縮領域の割合（Z>2）、全脳萎縮領域の割合、VOI 間の比を出す（原論文：Matsuda 2012、AJNR）。本ツールは VSRAD の再実装でも互換品でもない。処理の流れと基準データベースの違いは第 2 節に表で示す。第 5 節で、公開データ IXI の健常者 5 例について、同じ T1 を VSRAD advance 2 に入れた値との対応を集計値を示す。

2 VSRAD との相違（処理パイプラインとデータセット）

VSRAD（VSRAD advance 2。論文 Matsuda 2012 AJNR が報告した SPM8＋DARTEL の方法）と本ツールは、処理パイプラインも基準データセットも異なる。下の表は、VSRAD 側を資料（処理の詳細条件・ログ、出力レポート、論文）に書かれている範囲だけで書き、書かれていない内部処理は「非公開／資料に記載なし」とした。

表 1　VSRAD advance 2 と本ツールの相違

区分	項目	VSRAD advance 2（資料の記載）	本ツール
処理	変換・等方化	DICOM→NIfTI、矢状→軸位、1.2 mm 等方化（trilinear）	本体の npz（LPS）を NIfTI へ。自前の等方化は行わない。N4・分割は元の格子（約 0.94×0.94×1.2 mm）、レジストレーション以降は 2 mm 等方の標準脳の格子（3.1 節）
処理	脳抽出	資料に記載なし（独立した工程の記載を確認していない）	HD-BET 2.0.1（CPU）

区分	項目	VSRAD advance 2（資料の記載）	本ツール
処理	バイアス補正	資料に記載なし（論文は「統合セグメンテーション」と記述。内部は非公開）	N4（ANTsPy、脳マスクつき）
処理	レジストレーション	MNI への線形変換（12 パラメータのアフィン、SPM2 テンプレート）、DARTEL（2 mm）。SPM の内部設定の詳細は資料に記載なし	ANTs の Affine→SyN（類似度 CC）。標準脳は ANTsPy 同梱の MNI（2 mm 格子へ再標本化）。Jacobian×detA
処理	組織分割	SPM（GM/WM/CSF、EuropeanBrains、ThoroughClean）	Atropos（k-means 3 クラス、MRF 重み 0.3、強度は 99.5 パーセンタイルで打ち切り）
処理	変調・平滑化	DARTEL の PreserveAmount（変調）、FWHM 8 mm	Jacobian×detA で変調、FWHM 8 mm
処理	Z の作り方	Z=(対照平均−個人値)/対照 SD。単一の健常者群との比較で、取得した本文の範囲では年齢・性別の項は無い。全脳規格化（target 50）あり、相対・絶対の 2 種	ボクセルごとの回帰（年齢・年齢^2・性別・メーカー・磁場強度・log(detA)）の残差 SD から Z。全脳規格化は既定ではなく、併記用の別の DB
処理	Z>2 の閾値	Z>2（論文は P<0.05 相当、ROC で表示閾値に採用と記述）	Z>2（片側正規分布の約 2.3% という統計的な意味で採用）
処理	VOI	内側側頭部（嗅内野・海馬・扁桃）。早期 AD 30 例と健常 40 例の群間比較で定義。座標・T 値（「T=7.0」の意味）は非公開／資料に記載なし。背側脳幹は灰白質と白質の両方を出力	MTL は CerebrA、背側脳幹は CerebrA と Harvard AAN v2.0 から自前で定義（VSRAD の VOI とは別）。背側脳幹は白質のみ出す
処理	VOI 間の比	出力の分子と分母に 0.01 を加える（出力レポートに明記。論文の式にこのオフセットは無い）	加えていない
データセット	基準 DB	「武蔵病院 DB for VSRAD advance」、GM・WM とも 80 例。DARTEL テンプレート用に若年 25 例。DB が公開されているかは未確認	IXI の健常者 373 名（公開データ、CC BY-SA 3.0）
データセット	年齢範囲	54〜86 歳	20〜83.8 歳
データセット	磁場・装置・施設	1.5T の Siemens Vision Plus 単一装置、単施設（Matsuda 2012）。3T は別の DB（65 例、Sone 2018）が報告され、出力レポートも 3T では萎縮比が低めに出ると注記	Philips 1.5T（Guys）、Philips 3T（HH）、GE 1.5T（IoP）の 3 施設・英国。磁場強度とメーカーを共変量にする
データセット	撮像条件（T1）	3D T1 MPRAGE（Matsuda 2012 の 80 例）	IXI の T1 は、MPRAGE など VSRAD 用の T1 強調画像の撮像条件で取得されていない。具体的なシーケンス名・パラメータは IXI の T1 強調画像に依存する。
データセット	集団・性別	日本人（総説の限界記述、二次情報）。男 37・女 43（Nakatsuka 2013 の記述）	英国の健常者。日本人への適用は未検証。train の性別内訳は本稿の資料に無く不明

まとめ：本ツールは VSRAD の同一手法の再実装ではなく、数値は互換ではない。基準の年齢範囲（54〜86 歳対 20〜83.8 歳）、磁場・装置・施設、撮像条件も異なる。IXI 健常者 5 例での対応（第 5 節）は粗い目安にすぎず、差の原因は処理・基準データセット・撮像条件のどれが作っているか切り分けていない。

3 方法

3.1 処理パイプライン

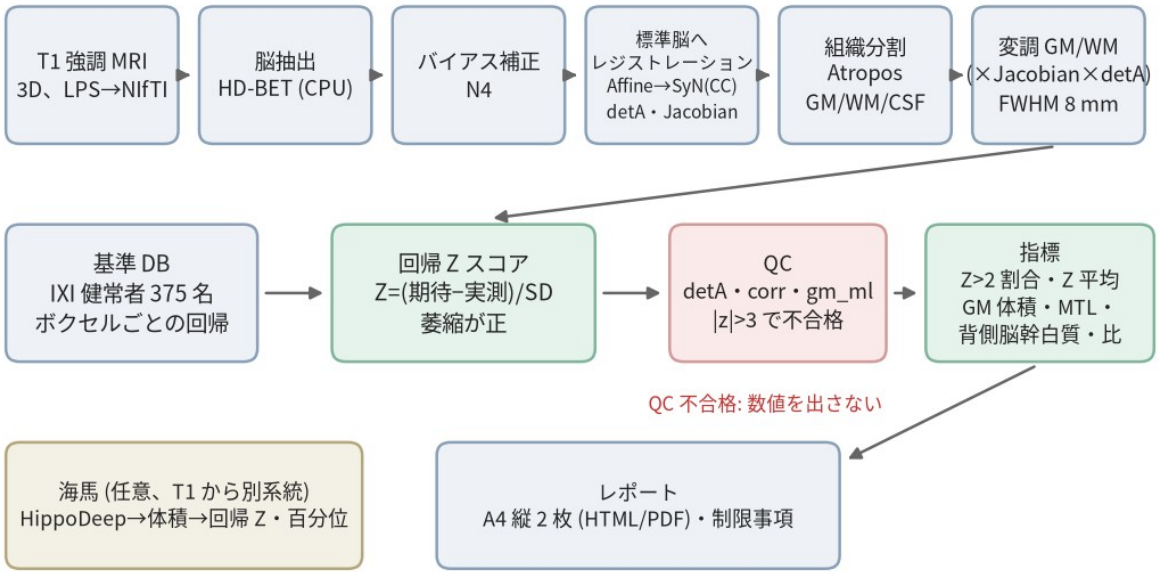


図 1　処理の流れ。海馬と VOI 指標は任意の追加解析で、全脳の処理とは独立に実行する。

各段の設定は実装（pipeline.py）のとおりで、次の値を使う。

段	方式と主な設定
入力	本体の npz（LPS）を NIfTI（RAS）へ変換する。幾何が欠けている入力・3 次元でない入力・間隔が 0 以下の入力は断る。
脳抽出	HD-BET 2.0.1。CPU、--disable_tta、脳マスクも保存する。
バイアス補正	ANTsPy の N4（脳マスクつき）。補正後の画像に脳マスクを掛ける。
レジストレーション	標準脳は ANTsPy 同梱の MNI（mni.nii.gz、182×218×182、1 mm）を 2 mm 格子へ再標本化したもの。Affine の後に SyN（類似度 CC、sampling 4）。乱数の種は 1。detA は Affine の 3×3 行列の行列式の絶対値（頭蓋内容積の代理。精度は未評価）。Jacobian は SyN の Jacobian に detA を掛ける。レジストレーション後の画像とテンプレートの相関（corr）を脳マスク内で計算する。
組織分割	Atropos（k-means 3 クラス、MRF 重み 0.3、反復 [5,0]、マスクは脳マスクを 1 画素収縮）。事前に、脳マスク内の強度を 99.5 パーセンタイルで打ち切る。乱数の種の引数を r=0 にして決定的にする。
変調と平滑化	GM・WM の確率画像を標準脳へ変形し、Jacobian を掛け（変調）、FWHM 8 mm で平滑化する。gm_ml は脳マスク内の変調 GM の合計に体素体積を掛けた値（mL）。
再現性の設定	ITK のスレッド数 1、乱数の種固定（ANTS_RANDOM_SEED=1）。この設定で同じ入力から同じ出力が得られることを、検査で確かめた（4.3 節）。

分割の変種（S0・S1・S2）の比較と S2 の採用の経緯は 4.2 節に記す。組織分割の方式について、Atropos が従来法より良いとする根拠は文献調査（research-E）では得られていない。自前の等方化は行わない（未実装）。N4 と組織分割は元の格子（IXI は約 0.94×0.94×1.2 mm）で動き、レジストレーション以降の定量（Jacobian による変調、FWHM 8 mm の平滑化、GM 体積、Z、VOI）はすべて標準脳 MNI の 2 mm 等方格子で行う。被検者画像の補間は変形の適用（線形補間）の 1 回である。HD-BET は内部で 1

mm 等方に再標本化して推論し、HippoDeep は内部で 1 mm 等方の MNI 切り出しで推論して、いずれも出力は元の格子に戻る（ツールのコードと設定の読み取りによる。HD-BET の設定名は重みの中身から読んだもので、実行時の確認は未実施）。Atropos の MRF 近傍と脳マスクの収縮はボクセル単位で、非等方の入力では mm では非等方になる（ANTs の慣例による。未検証）。入力を 1 mm 等方に再標本化して処理した診断では、Z>2 割合が高い 2 名で結果が大きく動いた（4.4.3 節）。

3.2 基準データベースと回帰 Z

基準データベース（DB）は、train の各被検者の変調 GM（標準脳 2 mm 格子）を、ボクセルごとに GM = Xβ + ε として回帰したものである。回帰するボクセルは、train の平均 GM マップの Otsu しきい値を超える場所（field の DB で 162,369 ボクセル）とする。残差 SD は自由度 n−p で補正する。被検者の Z は Z = (期待 GM − 実測 GM) ÷ SD で、萎縮を正にする。マスク外は NaN とし、0 で埋めない。

閾値 Z>2 は、正規分布の片側で 2.28% に当たる分布の裾として採用した。

表 2　基準 DB の仕様（共変量）と用途

仕様名	共変量	用途
field	age、sex、field_t（1.5 または 3 T）、log(detA)	全脳 GM の最初の仕様。新しい被検者の施設は未知なので、施設ダミーの代わりに磁場強度を使った。基準 1 に失敗（4.3 節）。
maker	age、sex、maker（Philips を基準、GE をダミー）、field_t、log(detA)	全脳 GM と海馬の採用仕様。IXI では GE ダミーが IOP ダミーと同値。Siemens とその他は学習に無く、Philips として計算し、補正なしの警告を出す。
makerq	maker に age^2 を加えたもの	採用仕様（分割 v2 で全 DB 共通）：全脳 GM・規格化版・海馬・DBS 白質。全脳白質と DBS 灰白質は再評価のみで不採用（白質の加齢変化が逆 U 字のため年齢 ^{2} を入れたが基準 5 に失敗）。
maker＋全脳規格化（gn）	v2 では makerq と同じ。被検者ごとにマスク内平均を train の平均へそろえてから回帰	全脳の萎縮を差し引いた、領域選択的な萎縮の Z。

detA は頭蓋内容積の代理である。IXI の施設と装置は、Guys（Philips 1.5T）、HH（Philips 3T）、IoP（GE 1.5T）で、IXI の公開記載による。

3.3 品質管理（QC）

QC は 3 指標、detA、corr（SyN 後の相関）、gm_ml で行う。各指標を DB の train の中央値と、MAD×1.4826 で作った頑健 SD から z = (値 − 中央値) ÷ 頑健 SD にし、|z|>3 を不合格とする（正規分布の 3 SD 外、約 0.27%）。固定の mL や相関の範囲は置かない。gm_ml は上側（z>+3）だけを不合格とし、下側（z<−3）は警告にとどめる。レジストレーションの失敗は detA と corr が拾い（脳マスク上側 25% を削った入力で確認）、gm_ml の下側は萎縮でも分割の失敗でも起こり、健常者の分布からは両者を区別できないためである（追加登録 12）。年齢が DB の範囲の外のときも警告を出す（外挿）。QC が不合格なら Z マップ・指標・海馬・VOI を出さず、QC のみのレポートを返す。年齢・性別・磁場強度の入力の誤りは、画像の指標ではないので QC では検出できない。

3.4 海馬の Z スコア（HippoDeep）

HippoDeep（Thyreau 2018）で左右の海馬体積を測る。入力は、製品の経路と同じ形（float32、qform と sform が同じ）に書き直した T1 である。体積（mL）を、基準の回帰（年齢、性別、磁場強度、log(detA)。採用の v2 はメーカーを加える）の期待値と比べ、Z = (期待 − 実測) ÷ 残差 SD（萎縮が正）と百分位（train の残差の経験分布での順位）を出す。Z≥1.645 は健常者でも約 5% が超える水準で、診断の閾値ではない。HippoDeep の体積は FreeSurfer や SynthSeg より系統的に小さい独自の尺度で（Sghirripa 2025）、基準は同じ手法で作り、他手法・他施設の体積とは比べていない。

3.5 VOI 指標（MTL・背側脳幹白質・比）

VOI は標準脳（2 mm 格子、91×109×91）上に自前で定義した。MTL は CerebrA の海馬・扁桃体・嗅内野（2,402 ボクセル、約 19.2 mL）。背側脳幹（DBS）は CerebrA の脳幹と、Harvard AAN v2.0 の DR・MnR・PAG・LC・PBC・PnO・mRt・PTg・LDTg（VTA を除く）の積を半径 3.4 mm（FWHM 8 mm に相当する σ）に膨張したもの（963 ボクセル、約 7.7 mL）。

指標の定義は Matsuda 2019 に従う。Z スコア指標は VOI 内の正の Z の平均（負は含めない）、Extent は VOI 内で Z>2 の割合、比は DBS の指標 ÷ MTL の指標である。VOI 内に正の Z が無いときは 0、比の分母が 0 のときは算出不可とし、0 で埋めない。MTL は全脳の灰白質の Z を使う。DBS は白質のみとし、背側脳幹専用の基準（マスク＝DBS VOI、仕様 makerq）の Z を使う。背側脳幹の灰白質は、事前登録の検査に通らなかったので出さない（4.3 節）。全脳の白質は警告つきの参考値として別に出す（4.3.3 節）。論文の簡易基準・決定木基準は両方の比を使うため、本ツールでは当否を判定できない。VSRAD の比に加わる +0.01 は、本ツールでは加えていない。

3.6 全脳規格化後の Z とレポート

全脳規格化後の Z は、gn の DB で計算し、現行の Z と並べて出す（現行の Z は変えない）。脳全体の萎縮を差し引くので、脳全体の萎縮が強い例では領域の Z が小さく出る。

レポートは A4 縦 2 枚（QC 不合格は 1 枚）の HTML と PDF で、1 枚目に見出し・処理の経過・結果の表・Z マップ・最大値投影、2 枚目に統計の図・海馬・VOI・説明文・所見欄を置く。注記は本文に並べず、末尾の「制限事項」節に集約する。患者名・ID は表紙にだけ載る。表示の向きは放射線科の慣例（患者の右が画像の左、前が上）で、向きは既知の位置の点を 3 種の向き行列で置いて確かめた。

3.7 評価の設計

- **事前登録：**判定基準と閾値は、被検者の処理や DB の作成の前に記録し、合格・失敗・未達を分けて数える。
- **コホート：IXI の T1。ファイル 581（Guys 322、HH 185、IOP 74）のうち、年齢と性別が確定する 561 名を使う。採用する分割 v2 は、施設×性別の層ごとに年齢順で 3 名ずつの組をつくり、各組から 1 名を検証用に無作為に選ぶ（numpy.random.default_rng(20261105)。train 373、held-out 188。held-out は Guys 105、HH 60、IOP 23）。傾向スコア・標準化平均差・年齢の KS で均衡を検査した（4.3.2 節）。**

- **除外：**20 名。18 名は IXI.xls に年齢か性別が無い。IXI.xls に同じ ID の行が複数ある ID が 26 あり、年齢か性別が食い違う ID のファイルは本人を決められないので除外した。
- **閾値の由来：**基準 1 の「中央値 1.1〜4.6%」（理論 2.28% の 0.5〜2 倍）は、導出ではなく事前の目安である。n が小さい段階で置いたもので、後の分析で、被検者ごとの割合が空間的に相関する前提に合わない可能性が示された（6.1 節）。海馬と施設別の Z 平均の閾値は、n から導いた（H1：|平均| ≤3/√n、H2：1±3/√(2n)、H3：片側 5% の二項分布の 99% 区間）。
- **負例：**年齢を −20 歳ずらすと Z>2 割合が上がり、+20 歳ずらすと下がること（ずらした年齢が train の範囲内の被検者だけ）。当初の登録は向きを誤っていたため、結果を見る前に訂正した。

4 評価

4.1 判定の数え方

下の表の「判定」は、合格・失敗・未達・変更の 4 つに分ける。未達は前提（データ・環境・実機での確認）がそろわず検査できていないもので、合格に数えない。変更は、登録した条件と異なる結果になったもので、合格に数えない。1 つの検査に自動検査で確かめられた部分と確かめられていない部分があるときは、行を分けて数える。

4.2 候補実装の経緯（要約）

- **候補 (a)（ANTsPy のみ、57 名）：**不合格。数値が壊れ（GM 体積が 200〜400 mL の被検者が多数、施設間差が大きい）、プロセスが完走しても失敗を検出できなかった。アフィン後の相関は −0.02〜0.12 で整列していなかった。
- **脳抽出の前置き (a')：**テンプレートは頭蓋除去済み、被検者は頭蓋を含むのが整列しない主因という切り分けが、6 名と画像で支持された。HD-BET を前置きした 57 名では、サイト別の GM 体積の中央値が Guys 432・HH 413・IOP 416 mL にそろった。
- **決定性：**原因は Atropos の乱数の種で、r=0 で固定できた（3 名を HD-BET からやり直して、S0・S1・S2 の GM/WM の 9 組すべてで最大絶対差 0）。
- **分割 S2 の採用：**S0（元の分割）、S1（99.5 パーセンタイルで打ち切り）、S2（S1＋MRF 0.3）の中で、事前の規則（GM 体積の外れ ≤3/57、規格化ありの HH の Z>2 割合 ≤4.6%、基準 1・2 が規格化あり・なし両方で合格）を満たすのは S2 だけだった。ただし S2 は GM 体積の水準が約 9% 高く（中央値 454→495 mL）、train の頑健 SD も 33→63 mL に広がるため、外れが 10→0 になったのは SD の拡大の効果を含む。基準 DB は必ず同じ処理で作る。
- **事前登録の不備：**追試 1 の「元の Guys の分布（653 mL）から大きく外れない」という目安は、整列していない元の実装の値に基づいており、基準として不適切だった（記録済み）。
- **SynthSeg 系：**動かせず未達（重みの手動取得と古い Keras 向けのコード）。CAT12 系も未測定。

4.3 事前登録の判定と結果

以下は旧分割 v1（held-out 186 名、規格化なし）での結果で、閾値は登録のとおりである。採用 DB（分割 v2）での再評価は 4.3.2 節に示す。AD 例・別装置・OS・配布物に関する項目は未達としてまとめて示す。

注（判定の読み方）：表の「合格」「失敗」は、結果を見る前に決めた許容範囲に実測値が入ったかどうかを示すものであり、統計的検定（p 値や信頼区間）による「有意な外れ」の判定ではない。失敗は、実測値が許容範囲の外に出たことだけを意味し、差の大きさは別である（例：全脳白質の基準 1 は下限を 0.02〜0.04 ポイント下回っただけで、基準 5 は変化の向きが逆だった）。合格も、範囲に入ったことを示すだけで、性能の高さの証明ではない（例：メーカーを加えた基準 DB の基準 1 は、下限との差が 0.04 ポイントだった）。閾値は結果を見てから動かしていない。

許容範囲の設定の根拠：評価は、基準 DB の作成に使わない保持（held-out）の被検者に対して行った。基準 1（Z>2 割合の中央値 1.1〜4.6%）は、正規分布の片側 Z>2 が 2.28% であることを基に、その 0.5〜2 倍を許容範囲とした。自由度や予測分散の影響で理論値は少し上がるが、これを厳密に反映せず、n が小さいことも考慮した事前の目安であり、統計的に導出した値ではない。基準 2（Z 平均の中央値の絶対値 <0.25）は、被検者ごとの Z 平均の理論的な標準誤差の目安として置いた事前の目安である。基準 3（決定性）は、スレッド数と乱数の種を固定した同一入力の再実行で出力が完全に一致すること（最大絶対差 0）とした。基準 5（年齢の負例）は、保持の被検者の年齢を人為的にずらしたときに Z>2 割合が期待の向きに変わることで、回帰が年齢を捉えているかを確かめるものである。基準 6 は処理の失敗を 0 件とし、失敗は除外せず数えた。基準 7・8 は検査できていないため未達とし、合格に数えない。施設別の Z 平均の限界（追加登録 10 の M1）は、施設内平均の絶対値が $3/\sqrt{n}$（n は施設の被検者数）以下とした。なお、閾値の幅は導出ではなく目安であるため、基準 1 のように下限や上限の近くに実測値が来た場合、合否だけでは判断できず、余裕（下限との差）を併記して読む必要がある。

表 3　全脳灰白質（旧分割 v1、held-out 186 名）

ID	検査	合格の基準（事前登録）	結果	判定
基準 1（field）	Z>2 割合の中央値	1.1〜4.6%	0.99%（最大 18.69%）。下限をわずかに下回る。検出が少ない側の失敗で、閾値は動かしていない	**失敗**
基準 2（field）	|Z 平均の中央値|	<0.25	−0.054	**合格**
基準 5（field）	年齢 −20 歳で Z>2 割合が上がる	上がる	0.79%→3.27%（対象 120 名）	**合格**
基準 5（field）	年齢 +20 歳で下がる	下がる	0.97%→0.28%（対象 157 名）	**合格**
M2 基準 1（maker）	Z>2 割合の中央値	1.1〜4.6%	1.14%（対照の field は 0.99%）。下限との差 0.04 ポイントで余裕は小さい	**合格**
M2 基準 2（maker）	|Z 平均の中央値|	<0.25	−0.034	**合格**
M2 基準 5（maker）	年齢 −20 歳で上がり、+20 歳で下がる	両方向	−20：0.97→4.13%（120 名）、+20：1.13→0.30%（157 名）	**合格**
M1（maker）	施設別の Z 平均	|平均|≤$3/\sqrt{n_site}$（Guys 0.30、HH 0.39、IOP 0.63）	maker：−0.028・+0.014・−0.004。対照 field：−0.091・+0.015・+0.302（IOP の偏りが +0.30→0.00）	**合格**
N1（gn）	Z>2 割合の中央値	1.1〜4.6%	1.30%（最大 9.47%）	**合格**
N2（gn）	|Z 平均の中央値|	<0.25	0.005	**合格**
N3（gn）	施設別の Z 平均	M1 と同じ限界	Guys +0.002・HH +0.003・IOP +0.017	**合格**
N4（gn）	決定性	DB の SHA-256 一致、Z の差 0	2 回の構築で SHA-256 一致、3 名の Z の再計算で差 0.0	**合格**
基準 3	決定性（パイプライン）	ビット一致（最大差 0）	3 名を HD-BET からやり直し、GM/WM の 9 組すべて最大絶対差 0（57 名の追加登録 2）	**合格**

ID	検査	合格の基準（事前登録）	結果	判定
基準 6	処理の失敗	0 件	561/561 完了、失敗 0。QC 不合格 2/561（train 2、held-out 0、いずれも IOP の corr）	合格
基準 7	AD 例での分離	—	検査していない（AD 例のデータ未取得。ADNI は利用条件の確認が先）	未達
基準 8	同一人物の別装置での再現性	—	IXI に該当データ無し	未達

基準 4（実行時間・メモリ）は合否を置かず、参考値のみである：HD-BET を含む 1 例が約 4 分 40 秒・最大約 4.1 GB（5 並列の負荷下、ITK 1 スレッド。単独の計測ではない）、DB 構築では 1 例 約 6〜7 分（5 並列）。基準 5 は gn には適用していない（規格化版では意味が変わり、追加登録 4 の規格化版は +20 側が失敗した）。

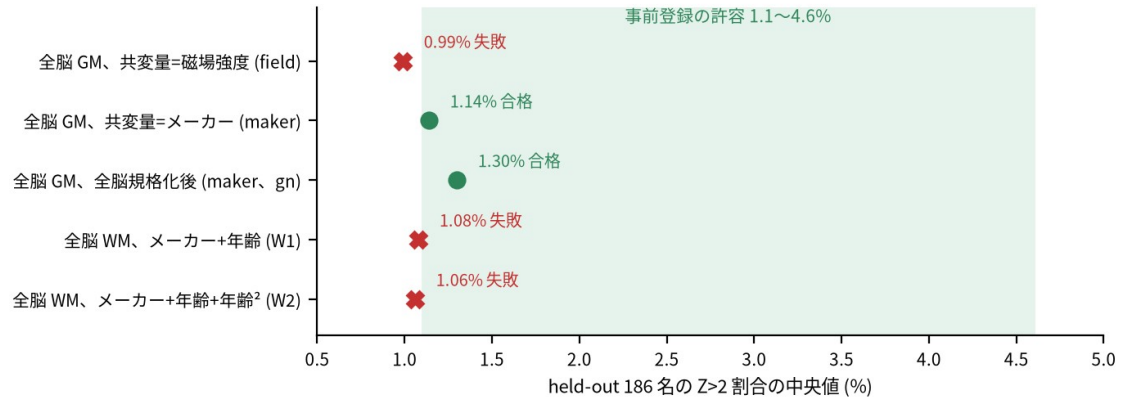


図 2　旧分割 v1 の held-out 186 名の Z>2 割合の中央値と、事前登録の許容範囲（緑帯）。全脳灰白質の 3 つの基準と、全脳白質の 2 つ（失敗）。マーカーの形と文字で判定を示す。

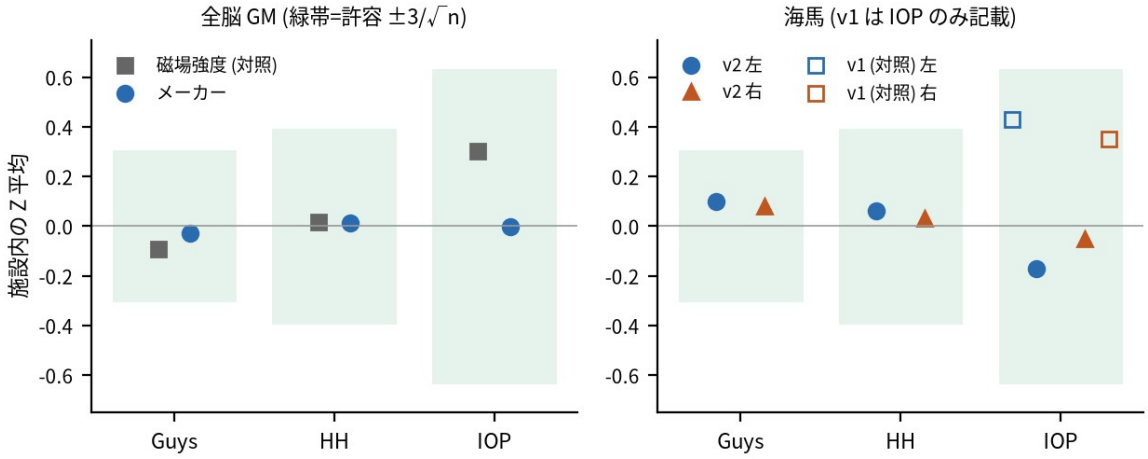


図 3　施設別の Z 平均（held-out、被検者ごとの Z 平均の施設内平均）。緑帯は各施設の許容 ±3/√n_site。左：全脳 GM、右：海馬（v2 はメーカー補正あり、v1 は対照として IOP のみ）。IXI では GE ダミーが IOP ダミーと同値であり、新しい GE 装置への一般化は検証できていない。

表 4　海馬の Z スコア（旧分割 v1、held-out 186 名、左／右）

ID	検査	合格の基準（事前登録）	結果	判定
H1（v1）	Z の平均	|平均|≤0.22	+0.059／+0.050	合格
H2（v1）	Z の SD	0.844〜1.156	1.032／1.039	合格
H3（v1）	Z>1.645 の割合	0.9〜9.1%	6.45%／6.45%	合格

ID	検査	合格の基準（事前登録）	結果	判定
H4（v1）	年齢 −20 歳で上がり +20 歳で下がる	両方向	−20：6.45→10.75／12.37%、+20：→3.23／2.15%	合格
H5（v1）	3 名を 2 回実行	マスク・体積が完全一致	完全一致	合格
M1（v2、メーカー）	施設別の Z 平均	|平均|≤3/√n_site	Guys +0.10／+0.08、HH +0.06／+0.03、IOP −0.17／−0.05。対照 v1 の IOP は +0.43／+0.35	合格
M2 H1〜H3（v2）	平均・SD・Z>1.645 の割合	H1〜H3 と同じ	平均 +0.056／+0.048、SD 1.047／1.042、割合 7.0%／6.5%	合格
M2 H4（v2）	年齢の負例	両方向	−20：13.4／12.9%、+20：3.2／2.2%	合格
（海馬）基準 7 相当	AD 例での感度	—	未検証	未達

海馬の診断（合否なし）：体積は左 2.07〜4.33 mL、右 2.23〜4.51 mL（held-out）。24 名の目視（Z の上位 6・下位 6・体積の外れ 6・無作為 6、判定者 1 名、低解像度の montage）で、海馬を外す・他の構造を含む明確な失敗は 0/24、形がやや不整に見えたのは 2 名。HippoDeep は 1 例 3.6〜4.3 秒・約 0.7 GB（CPU）。スペーシング 1e-6・回転 1e-6 rad の摂動 18 回で体積は mm^3 まで同じ（1e-6 だけの確認）。

表 5　全脳白質と背側脳幹（DBS）の専用基準（旧分割 v1、held-out 186 名）

ID	検査	合格の基準（事前登録）	結果	判定
W1 基準 1（maker）	全脳白質の Z>2 割合の中央値	1.1〜4.6%	1.08%（下限に 0.02 ポイント足りない）	失敗
W1 基準 2	|Z 平均の中央値|	<0.25	−0.024	合格
W1 基準 5	年齢の負例	−20 で上がり +20 で下がる	−20：1.00→0.85%（上がらない）、+20：0.91→1.59%（下がらない）	失敗
W2 基準 1（makerq）	全脳白質の Z>2 割合の中央値	1.1〜4.6%	1.06%（下限に 0.04 ポイント足りない）	失敗
W2 基準 2	|Z 平均の中央値|	<0.25	−0.024	合格
W2 基準 5	年齢の負例	両方向	−20：1.03→1.06%、+20：0.99→1.82%（下がらない）	失敗
W2 施設別	施設別の Z 平均	|平均|≤3/√n_site	Guys −0.043・HH −0.008・IOP +0.076	合格
D1 DBS 灰白質	VOI 内 Z 平均（全体、施設別）	全体 ≤0.22、施設別 ≤3/√n_site	全体 −0.137、Guys −0.050・HH −0.107・IOP −0.607（限界 0.63）	合格
D2 DBS 灰白質	年齢の負例	両方向	−20：−0.154→−0.177、+20：−0.140→−0.117（向きが逆）	失敗
D1 DBS 白質（makerq）	VOI 内 Z 平均	同上	全体 −0.117、Guys −0.171・HH −0.054・IOP −0.035	合格
D2 DBS 白質	年齢の負例	両方向	−20：−0.043→+0.029、+20：−0.119→−0.336	合格
D3	決定性	SHA-256 一致、Z の差 0	最大絶対差 0.0	合格
（VOI 指標）	Matsuda 2019 の簡易・決定木基準の当否	—	灰白質の比が無いため判定不能	未達

取り扱い（登録どおり。のち 4.3.3 節で、全脳白質は警告つきの参考値として出すことに改めた）：全脳白質の Z は出さない。DBS の灰白質指標（DBS_Z_gray と灰白質の比）は出さない。DBS の白質指標と、DBS_Z_white÷MTL_Z は出す。診断（合否なし）：train の白質は年齢と逆 U 字（20〜30 歳 0.501、30〜40 歳 0.518、70 歳〜0.489。線形の年齢項の R^2=0.005）で、基準 5 の仮定が白質に合っていない可能性が高い。DBS の灰白質は VOI 内の変調 GM が薄く（平均 0.075）、年齢の信号が基準に入っていない。

表 6　QC・実装・画面（コードの検査と本体アプリ上の確認を含む）

ID	検査	合格の基準（事前登録）	結果	判定
Q1	QC 不合格数（新規則）	現行の規則より増えない	held-out 186：現行 0・新 0。全 561：現行 2・新 2。下側の警告 0	合格
Q2	QC の負例	頭蓋含み・雑音は不合格。脳マスク上側 25% を削った入力は不合格か警告	頭蓋含み：不合格（detA +5.7・corr −15.8・gm_ml +10.5）。雑音：不合格。上側 25% を削った入力：不合格（corr z=−4.0）、gm_ml 下側は警告	合格
QC 負例（IXI419）	強度 0.1 倍	通る（k-means は強度の尺度に依らない）	通過	合格
Q4	pytest	全件	102 合格・4 スキップ（当時）	合格
E1・E2	撮像条件の DICOM 抽出（正例・負例）	メーカー・磁場の正例、欠落・範囲外・同名の別タグ等の負例	正例・負例とも期待どおり	合格
E3	本体の REST 経由の自動入力	画面の値が DICOM の値と一致	IXI175（HH 3T）の DICOM で磁場 3 T・Philips が一致	合格
E3	基準に無い装置の DICOM を本体へ取り込んでの確認	—	未実施（抽出規則は単体検査のみ）	未達
E4	画面の挙動	値の入力・上書き禁止・失敗時の理由	期待どおり。磁場が未選択なら実行を止める	合格
P1・P2・P3	進捗の解釈・状態遷移・解析の不変性	進捗の有無で GM・WM・detA・corr・gm_ml がビット一致	IXI419 でビット一致、イベントは N4→レジストレーション→分割→変調の順	合格
P5	本体での進捗・サムネイル・「画像を見る」	期待どおり	全プロセスの遷移とサムネイルを観測	合格
P4	印刷ボタン（iframe の print の呼び出し）	1 回呼ばれる	本体で 1 回呼ばれた	合格
P4	OS の印刷ダイアログ	—	自動検査できず未確認	未達
R4	キャンセル	止まって再実行できる	HD-BET の実行中 4.0 秒、レジストレーションの実行中 6.2 秒で停止。取り消し直後に関連プロセス 0 件	合格
R5・R6	判定 UI・画面構成	期待どおり	本体で位置・表示を測定、目視	合格
R3	既定のファイル名（保存ダイアログ）	脳萎縮解析_<患者 ID>_<日付>	title と保存ダイアログの既定名が一致	合格
R3	Print to file の保存名	—	OS の印刷ダイアログ（ユーザーの実機で確認する）	未達
R1	レポートの頁数（通常）	2 頁	PDF 2 頁・HTML 2 頁（Chromium 印刷も 2 頁）	合格
R1	レポートの頁数（QC 不合格）	ちょうど 2 頁（登録）	1 頁（内容が少なく、2 頁にすると白い頁になるため）	変更
R2	余白（最長の白い帯）	白い空間を作らない	PDF 1.5〜6.2%、HTML 印刷 最大 14.0%、QC 不合格 6.8%。検出器の負例で 20%・51% となり落ちる	合格
（本体 spike）	2 枚化後の本体での再確認	—	未実施	未達
（本体 spike）	旧基準：本体経路と直接実行の Z>2 割合の差	1 ポイント以内	20.52% 対 18.84%（差 1.68）。旧基準の失敗は残す	失敗
配布	Windows・macOS での動作、パッケージの配布・署名・リリース	—	行っていない	未達

旧基準の失敗の後に、期待値の NIfTI を本体と同じ DICOM 経路（傾きを標準軸に直したもの）から作り直して比べたところ、Z>2 割合は直接 21.16%・本体 21.46%、MTL_Z は 2.666・2.68、GM 体積は 436.6・436.3 mL で、以前の差は期待値が傾きを残した別の NIfTI だったことによる入力の違いだった。旧基準の行は変えていない。

4.3.1 判定の集計

表 7　判定の件数（表 3〜6 の行を数えた値）

表	合格	失敗	未達	変更	行数
全脳灰白質	13	1	2	0	16
海馬	8	0	1	0	9
白質・DBS	7	5	1	0	13
QC・実装・画面	15	1	5	1	22
合計	43	7	9	1	60

合格が多いのは、事前登録の基準を満たすことが分かった検査と、コードの動作の検査を含むためである。合格の件数は、臨床での有用性や AD 例での感度を表さない。なお表 7 の件数は旧分割 v1 の行を数えたもので、4.3.2 節の再評価は含まない。

4.3.2 分割 v2・共変量統一での再評価（採用 DB）

採用 DB は、均衡を検査した分割 v2（train 373、held-out 188）と、全部位で共通の共変量（年齢・年齢^2・性別・装置メーカー・磁場強度・log(detA)、仕様 makerq）で作り直した（追加登録 17）。閾値は結果を見て動かしていない。画像は再処理しておらず、保存した変調 GM・WM と海馬の体積から再フィットした。

分割 v2 は、施設×性別の層ごとに年齢順で 3 名ずつの組をつくり、各組から 1 名を検証用に無作為に選ぶ（乱数の種 20261105。端数の組は組の大きさに比例した確率）。均衡は、傾向スコア（検証用に入る確率を年齢・年齢^2・性別・メーカー・磁場強度から推定）、各共変量の標準化平均差（SMD）、年齢の KS 統計量で検査し、すべて 0.1 未満を合格とした（最初の種で合格）。IXI では施設がメーカーと磁場強度を決めるため、施設×性別で割るとこれらも同時に均衡する。年齢は train 48.5±16.4 歳、held-out 48.8±16.7 歳である。最大 |SMD| は 0.036、PS の SMD は +0.042、KS は 0.027 だった。

表 7a　分割 v2 の内訳（施設×性別、人数）

施設・性別	全体	基準 DB 用（train）	検証用（held-out）
Guys・女	174	116	58
Guys・男	139	92	47
HH・女	93	62	31
HH・男	87	58	29
IOP・女	44	29	15
IOP・男	24	16	8
合計	561	373	188

表 7b　均衡の指標（標準化平均差。基準は |値| < 0.1）

指標	分割 v2（採用）	v2 の判定
年齢	+0.020	合格
年齢 ^{2}	+0.036	合格
性別（女）	−0.004	合格
メーカー（GE）	+0.005	合格
磁場（3 T）	−0.005	合格
傾向スコア	+0.042	合格
年齢の KS 統計量	0.027	合格

表 7c　採用 DB の判定（分割 v2、held-out 188 名: Guys 105・HH 60・IOP 23）

DB	判定	主な数値	対照・備考
全脳 GM	合格（基準 1・2・5・施設別）	Z>2 割合の中央値 1.11%（許容 1.1〜4.6%。下限との差 0.01 ポイント）、Z 平均の中央値 −0.009、年齢 −20 で 0.92→4.08%・+20 で 1.05→0.34%、施設別 Z 平均 +0.002／−0.009／−0.009	—
全脳 GM 規格化	合格（N1〜N3）	中央値 1.31%、Z 平均の中央値 +0.001、施設別 −0.001／−0.003／+0.010	—
海馬	合格（H1〜H4）	左 Z 平均 +0.037・SD 0.917・Z>1.645 が 3.2%、右 +0.031・0.929・3.2%。年齢 −20 で 11.2%／9.6%、+20 で 1.1%／1.1%	施設別（登録外の参考値）: IOP 左 −0.628 は目安 ±0.626 をわずかに超える
DBS 白質	合格（D1・D2）	VOI 内 Z 平均 +0.119（限界 ±0.219）、年齢 −20 で 0.136→0.238・+20 で 0.133→−0.061	—
全脳白質（再評価）	失敗（基準 5）	基準 1 は 1.37%、基準 2 は +0.057。年齢 −20 で 1.33→1.29%（上がらない）	不採用のまま
DBS 灰白質（再評価）	失敗（D2）	D1 は合格（−0.049）。年齢 −20 で −0.045→−0.091、+20 で −0.054→+0.001（向きが逆）	不採用のまま
決定性	合格	6 DB すべて 2 回の構築で内容の SHA-256 が一致。3 名の Z の再計算は最大絶対差 0	—

基準 7（AD 例での感度）と基準 8（同一人物の別装置での再現性）は未達のままである。採用 DB を v2 に切り替えても、全脳 GM の基準 1 の余裕はほとんどない（0.01 ポイント）。海馬の施設別 Z 平均は IOP（GE、23 名）で期待より大きい側に偏りが残り、左は目安をわずかに超える（登録外の参考値）。

4.3.3 全脳白質の再検討：警告つきの参考値として出す（追加登録 18・20・23）

全脳白質（分割 v2・makerq の DB、held-out 188 名）は、基準 1（Z>2 割合の中央値 1.37%）・基準 2（Z 平均の中央値 +0.057）・施設別には合格し、基準 5（年齢 −20 歳で Z>2 割合が上がり +20 歳で下がる）だけが失敗した。診断（追加登録 18）では、失敗は年齢の信号が無いためではなく、年齢曲線が逆 U 字（マスク内平均の頂点 57.9 歳、20 歳から 80 歳までの変化は +0.93 SD で、最大は 60 歳付近）で、基準 5 の前提（一様な減少）が成り立たないためだった。年齢で有意なボクセルは 70%、向きが逆のボクセルは −20 歳で 60.5%・+20 歳で 56.6% で、3 施設すべてで同じ向きだった（施設の交絡ではない）。

単調減少に頼らない負例（追加登録 20）では合格した。(i) 模擬的な減少（白質を 5・10・20% 小さくする）では、Z の増分が解析の予言と一致し（相対誤差 7e-13）、Z>2 割合の中央値が 1.4→3.1→6.9→23.7% と単調に増え、半球だけの減少は減らした側にだけ出た。全体を 10% 減らすと 56%、20% で 97% の被検者が健常者の 95 パーセンタイルを超えた。(ii) 減少側（頂点より高齢、n=43）で年齢を ±5 歳ずらすと、マスク内 Z 平均の差の 95% 区間が期待どおりの向きで 0 を含まなかった（効果は 0.03〜0.05 SD と小さい）。

58 歳以上に限った基準（追加登録 23、train の 58 歳以上 134 名、仕様 maker、held-out 58 歳以上 71 名）も試した。基準 1（1.66%）・基準 2（+0.032）・施設別・模擬的な減少・決定性は合格したが、減少側の年齢の負例に通らなかった（−5 歳で Z 平均の差 −0.0069、向きが逆のボクセル 52.5%）。58 歳以上でもマスク内平均の年齢変化は +0.16 SD/10 歳（加齢で増える）で、白質の総量は 70 歳前後まで緩やかに増える。「58 歳以降は減少」という前提で範囲を絞る方法は、この負例には通らなかった。登録した表示規則どおり、58 歳以上の基準の値は併記しない。

取り扱い（ユーザー判断）: 全脳白質の Z は、警告つきの参考値として出す（レポートは主表に Z>2 の割合・Z 平均を橙で示し、制限事項に警告、画面は結果欄に警告）。警告の内容は、年齢を ±20 歳ずらす検査に通っていないこと、とくに 58 歳未満では Z が高いことを萎縮と読めないこと、AD 例での感度が未検証であること。Z マップ・最大値投影は出さない。基準 5 の失敗の記録は書き換えていない。

表 7d　全脳白質の再検討（held-out 188 名、58 歳以上の基準は held-out 58 歳以上 71 名）

検査	判定	主な数値	備考
基準 1・2・施設別（全年齢の DB）	合格	Z>2 割合の中央値 1.37%、Z 平均の中央値 +0.057、施設別 +0.077／+0.006／−0.049（限界 ±0.29／0.39／0.63）	—
基準 5（年齢 ±20 歳。登録どおり）	失敗	−20 歳で 1.33→1.29%（上がらない）、+20 歳で 1.27→1.95%（向きが逆）	前提（一様な減少）が成り立たない
W-sim（模擬的な減少）	合格	解析との一致 7e-13、1.4→3.1→6.9→23.7%（単調）、半球だけの減少は減らした側にだけ出る	追加登録 20
W-limb（減少側、±5 歳、n=43）	合格	Z 平均の差 +0.033［+0.030,+0.037］（−5 歳）、−0.048［−0.052,−0.044］（+5 歳）。向きが逆のボクセル 39%／35%	効果は小さい
58 歳以上の基準 基準 1・2・施設別・W-sim・決定性	合格	1.66%、+0.032、Guys +0.067／HH +0.040（IOP 6 名は報告のみ）、1.7→3.6→7.9→25.6%、SHA-256 一致	追加登録 23
58 歳以上の基準 減少側の負例（63〜78.8 歳、n=42）	失敗	−5 歳で Z 平均の差 −0.0069（向きが逆）、向きが逆のボクセル 52.5%。58 歳以上でも年齢変化は +0.16 SD/10 歳	併記しない

4.4 診断（合否なし）：Z>2 割合の分布と測定の揺れ

4.4.1 基準 1 の失敗の診断（追加登録 5、field DB）

- プールした Z>2 は 1.99%（正規 2.275%）、Z<−2 は 2.92%。Z>2 の側はほぼ正規で、負の側（期待より GM が多い側）に裾がやや重い（歪度 −0.20、超過尖度 +0.63）。

- 被検者ごとの Z>2 割合は右に歪み（10/25/50/75/90/95% 点：0.12/0.44/0.99/2.14/5.1/7.8%）、被検者ごとの Z 平均の SD は 0.265。全脳で同じ向きにずれる成分があり、中央値が 2.3% より下になる。これは登録外の追加集計である。
- 途中画像 24 名（上位 6・下位 6・外れ 6・無作為 6、判定者 1 名）の目視では、明確な失敗（マスクの欠け・過剰、分割の取り違え）は 0/24。高齢で脳溝・脳室が大きい 6〜8 名で、レジストレーション後の皮質外縁に数 mm の局所的な残差があった（数値化していない）。

4.4.2 ヘッダの符号化に対する測定の揺れ（追加登録 7・8）

同じ画素・同じ向きの入力でも、ヘッダの 1 ulp（スペーシングの相対 6.4e-8）の違いで Z>2 割合が動く。原点のずれは効かず、回転・スペーシング・符号化の違いが効いた。水準（1e-7〜1e-4）に依らず同じ大きさで動く（離散的な分岐）。2 回の実行の差の 95% 区間（$1.96 \cdot \sqrt{2} \cdot SD$、ポイント）は、被検者の Z>2 割合に比例して増える。

被検者（IXI）	Z>2 割合 %	SD（ポイント）	95% 区間（ポイント）
IXI075	0.05	0.014	0.04
IXI354	1.2	0.108	0.30
IXI052	3.0	0.207	0.57
IXI015	7.3	0.527	1.46
IXI175	19.4	0.614	1.70
IXI331	16.2	1.213	3.36

N4 が原因の段であることを 2 被検者の 1 対で確かめた（N4 だけで 83〜95% を再現）が、別の対（1 ulp 違い）ではレジストレーション側・N4 側の両方が同じ大きさの変化を起こし、相互作用は未特定。N4 が常に原因とは言えない。検査は 2 被検者・2 対だけである。基準 1 の中心（Z>2 が 1% 前後）では SD が約 0.1〜0.2 ポイントで影響は小さい。レポートは、この揺れの幅より小さい差を読まないよう注記する。

4.4.3 追加の記録（合否なし）

- 模擬萎縮（held-out の GM を内側側頭部 VOI 内で 30% 減らす）での VOI Extent の中央値：現行 0.5%→52.7%、規格化 0.7%→51.0%。健常の 95 パーセンタイル（13.2%）を超える割合はどちらも 99%（領域の萎縮の検出力は規格化で落ちない）。これは模擬であり、実際の萎縮での感度ではない。
- 年齢帯別の Z>2 割合の中央値（field DB）：〜40 歳 1.44%（66 名）、40〜60 歳 0.72%（64）、60 歳〜0.83%（56）。20 歳代・80 歳代の端だけの値は未取得。
- サイト別（field DB）：Guys 0.80%（103）、HH 0.99%（60）、IOP 5.76%（23）。メーカー共変量（maker DB）で IOP の偏りは解消した（図 3）。

入力の格子の影響（登録 17 の診断。合否なし）：入力 T1 を 1 mm 等方（線形補間）に再標本化して HD-BET から処理し直し、元の格子の結果（DB は v2）と比べた（IXI175・IXI075・IXI331、各 1 回）。detA は 0.1〜0.6% しか変わらなかった一方、Z>2 割合は IXI175 で 17.5%→10.1%、IXI331 で 7.3%→10.0%、IXI075 で 0.06%→0.20%、GM 体積は IXI175 で 437.5→490.5 mL（+53 mL）、IXI331 で 582.2→564.6 mL、IXI075 で 529.7→525.6 mL だった。Z>2 割合が高い 2 名では、4.4.2 節の保存形式・ヘッダの揺れ（約 16〜19% の被検者で ±1.7〜3.4 ポイント）より大きく動いた。原因（HD-BET のマスク・N4・Atropos のどの段か）は未特定で、3 名・各 1 回のため一般化できず、どちらの格子が正しい

かも決められない。基準 DB は元の格子の入力で作られているため、同じ条件で処理するのが現状の整合した使い方である。

4.5 未達・未検証の一覧

表 8　未達・未検証の項目（合格に数えない）

項目	状況
AD 例での感度（基準 7、海馬を含む）	未検証。ADNI の利用条件の確認後。
同一人物の別装置での再現性（基準 8）	IXI に該当データ無し。
新しい GE 装置・Siemens など基準に無い装置への一般化	IOP 1 施設しか無く検証不能。Siemens とその他は補正なし（警告）。
日本人・IXI（20〜86 歳の英国の健常者）以外の集団	未検証。
Windows・macOS での動作、配布物（wheel）、署名・リリース	行っていない。Windows では、初期の 6 名の追試のみを実行し、Linux で再現した（それ以外は未検証）。
本体の最小版数	暫定 >=0.5.1（未確認）。
QC 不合格の画面の実機確認	実入力が無く未確認。
OS の印刷ダイアログ、Print to file の保存名、PDF の埋め込み表示	未確認（PDF の埋め込みは load は起きるが描画を確認できず不採用）。
等方化、被検者脳への重畳（SyN の逆写像）、脳表 8 方向、VSRAD 互換の 4 指標	未実装（VOI のアトラスは自前定義で、VSRAD とは別）。
分割方式の優位性	Atropos が従来法より良い根拠は無い（文献調査）。SynthSeg・CAT12 は未測定。
全脳白質の追加検証（単調減少を仮定しない負例）	未実施（白質は逆 U 字）。

5 VSRAD advance 2 との突き合わせ（IXI 健常者 5 例、集計値のみ）

同じ T1 を VSRAD advance 2 に入れた出力 PDF の値と、本ツールの値を比べた。データは公開データ IXI の健常者で、個別症例の値は載せない。例数が少ないので、換算も閾値の変更もせず、一致・不一致を数えるだけである。この突き合わせは採用 DB（分割 v2）で更新した。5 例は分割 v2 の train に含まれるため、一次の値は 5 例を train から除いて作り直した DB（368 名）で計算した（採用 DB そのものでも計算し、集計はほぼ同じだった）。症例と IXI の対応表は手元に無かったので、症例の年齢・性別・磁場から候補を作り、記録済みの集計値がすべて再現できることで確かめた。VSRAD の出力 PDF との直接の突き合わせではない。

5.1 IXI 健常者 5 例の差と順位相関

同じ T1（矢状断の DICOM、右から左）を入れた。年齢は 36〜72 歳（VSRAD の基準は 54〜86 歳）、1.5T 4 例・3T 1 例、Philips 4 例（3T の 1 例を含む）・GE 1 例。本ツールは灰白質 maker の DB と DBS 白質の DB を使った。

表 9　IXI 健常者 5 例：差（本ツール−VSRAD）の中央値と範囲、順位相関（採用 DB〔5 例を除いて作り直したもの〕）

指標	差の中央値（範囲）。採用 DB（5 例除外）	順位相関
内側側頭部の萎縮度	−0.07（−0.68〜+0.73）	+0.70
全脳の Z>2 割合（%）	−1.85（−4.66〜+12.02）	+0.50
内側側頭部の Extent（%）	−1.29（−18.36〜+5.71）	+0.30
内側側頭部の VOI/全脳 Ratio	−0.99（−4.33〜+1.91）	+0.30
背側脳幹（白質）の萎縮度	+0.59（−0.84〜+0.84）	+0.20
白質の比（オフセットなし／+0.01）	+0.31／+0.32（−1.02〜+1.47）	−0.20

判定の一致は、白質の比≥0.2 が 3/5（オフセットの有無で同じ）、内側側頭部<2.0 が 4/5。VSRAD の値で論文の簡易基準（MTL<2.0 かつ両比≥0.2）を満たした健常者は 3/5 だった。内側側頭部の萎縮度は、年齢が VSRAD の基準の範囲内の 3 例で差が −0.07〜+0.73 と概ね対応した。背側脳幹の白質の萎縮度と比は対応しなかった（健常者 188 名で被検者間 SD 0.82 と大きく、雑音が主と考えられるが、検証していない）。

5.2 対応の読み方

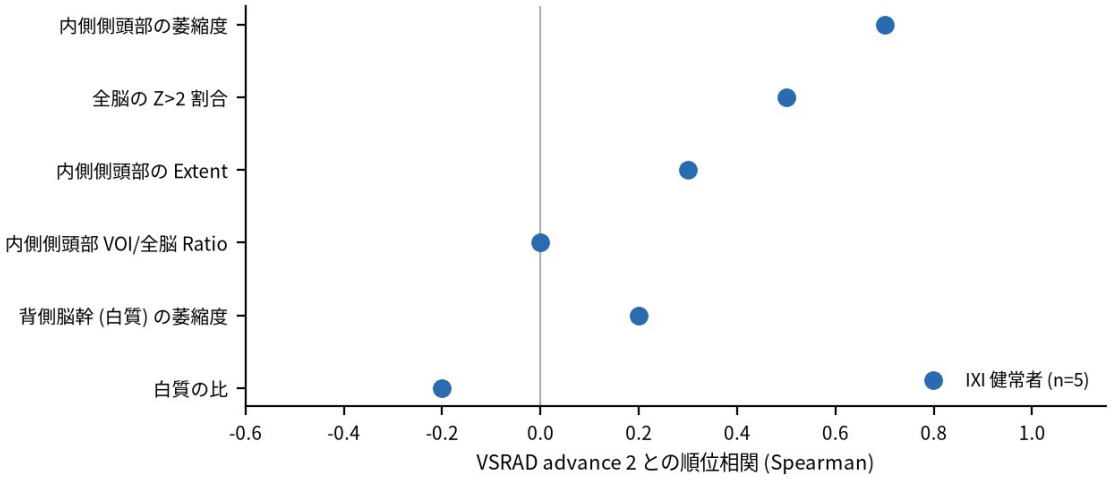


図 4　VSRAD advance 2 との順位相関（IXI 健常者 5 例）。n が 5 なので、相関は粗い目安にとどまり、一致度の推定として読まない。

撮像条件が異なる（IXI の T1 は MPRAGE など VSRAD 用の T1 強調の撮像条件で取得されていない）ため、差の原因は切り分けていない。差の要因（いずれも検証していない仮説）：年齢が VSRAD の基準の範囲外の 2 例（36・53 歳）で VSRAD が高い、VOI の定義が別、基準 DB の違い（80 例 1.5T Siemens MPRAGE 対 IXI 373 例）、IXI の T1 のコントラストが MPRAGE と異なる、組織分割・レジストレーションの違い（SPM+DARTEL 対 HD-BET+ANTs+Atropos）、全脳の分母の違い（VSRAD は全灰白質内、本ツールは基準マスク内）、3T の 1 例について VSRAD 自身が比を 3 割強低く出すと注記していること。いずれも第 2 節の相違に当たる。

結論：本ツールは VSRAD の同一手法の再実装ではなく、数値は互換ではない。IXI 健常者 5 例での対応は粗い目安にすぎない。AD 例での一致は検討していない。

6 考察

6.1 基準 1 と DB の仕様

年齢・性別・磁場強度・log(detA) を共変量とする DB は基準 1 に失敗した（0.99%）。これに装置メーカーを加えた DB は合格した（1.14%）。ただし余裕は 0.04 ポイントで、結果を見る前に置いた閾値の由来も「事前の目安」であり導出ではない。診断（4.4.1）は、被検者ごとの Z>2 割合が全脳で同じ向きにずれる成分を持ち、独立な二項分布を前提にした登録の範囲が当てはまりにくいことを示唆する。これは推測であり、基準 1 の置き方の見直しはユーザーの判断事項として残している。メーカー補正は IXI では施設ダミーと同値なので、IOP の偏り（Z 平均 +0.30→0.00、海馬 +0.43/+0.35→−0.17/−0.05）が消えたことは、装置差を吸収したことの証拠ではあるが、新しい装置への一般化の証拠ではない。分割 v2・共変量統一の再評価（4.3.2 節）でも全脳灰白質は合格したが、1.11% は下限 1.1% との差が 0.01 ポイントで、境界に近い。

6.2 指標の範囲を絞った判断

背側脳幹の灰白質は、事前登録の検査（年齢の負例）に失敗したため出さないことにした。全脳白質も基準 5 に通らなかったが、白質は逆 U 字の加齢変化で、基準 5 の前提が成り立たないことを診断で確かめ（4.3.3 節）、登録は変えず、事実を記録した上で、警告つきの参考値として出すことにした（ユーザー判断）。背側脳幹の白質だけが D1〜D3 を通ったので出すが、VSRAD との突き合わせでは健常者で対応せず（順位相関 +0.20、比 −0.20）。したがって、背側脳幹の指標と比は参考であり、論文の閾値を当てる材料にならない。

6.3 VSRAD との対応

内側側頭部の萎縮度は、IXI 健常者 5 例で VSRAD と概ね対応した（順位相関 +0.70）が、粗い目安にとどまる。処理（レジストレーション・分割）、基準データベース（80 例 1.5T 単一装置対 IXI 373 例の 3 施設）、VOI、全脳の分母、撮像条件（IXI の T1 は MPRAGE など VSRAD 用の T1 強調の撮像条件で取得されていない）がいずれも異なるため、どの要因が差を作っているかは切り分けていない。例数は 5 例で、一致度を推定する規模ではない。

6.4 測定の揺れと読み方

ヘッダの符号化の違い（1 ulp 級）で Z>2 割合が 1〜3 ポイント動くことは、保存形式を変えた同一画像の解析で再現する。割合が大きい（約 16〜19%）被検者ほど、同じ入力の 2 回の差が大きい。小さい差や小数点以下の桁を読まないことが、レポートの説明に入っている。海馬（HippoDeep）は 1e-6 の摂動に対して体積が変わらなかったが、確認した水準は 1e-6 だけである。

6.5 萎縮と処理の失敗の区別

QC は、GM 体積が基準より小さい場合に、萎縮と脳抽出・分割の失敗を区別できない。萎縮例を弾かないため、下側は警告にとどめ、処理の確認（段ごとの画像）での目視を促す設計にした。レジストレーション

の失敗は detA と corr が拾うこと（脳マスクの上側 25% を削った入力で corr の z=−4.0）を確認したが、他の種類の失敗が QC を通る可能性は否定できない。

7 制限事項

- 診断機器ではない。研究・教育目的の参考値で、医師の判断に置き換わらない。
- 基準は IXI の健常者 373 名（20.0〜83.8 歳）で、英国の 3 施設（Philips 1.5T・3T、GE 1.5T）のデータである。日本人、AD 例、他の装置・集団での感度・再現性は未検証である。
- 装置メーカーの補正は IOP（GE）1 施設から推定した。他の GE 装置や Siemens などの基準に無いメーカーでは補正されず、偏る可能性がある。磁場強度は線形の 1 項で、IXI の 3T は 1 施設だけなので、磁場と施設の影響は分けられていない。
- Z マップの背景は標準脳で、被検者の脳ではない。位置はレジストレーションの精度に依存し、皮質外縁で数 mm の残差がありうる（被検者脳への重畳は未実装）。
- 基準 1 は余裕 0.04 ポイントで通っている（maker）。閾値の由来は事前の目安であり導出ではない。
- 測定の揺れ：同じ画素でもヘッダの符号化で Z>2 割合が約 1〜3 ポイント（割合が大きい被検者で）動く。小さい差を読まない。
- GM 体積が小さいとき、萎縮と脳抽出・分割の失敗は QC で区別できない。年齢・性別・磁場強度の入力の誤りも QC では検出できない（使った値は結果に再掲する）。
- 海馬の輪郭は自動抽出で、誤りがありうる。体積は HippoDeep 独自の尺度で、FreeSurfer・SynthSeg・他施設の体積とは比べられない。AD 例での感度は未検証。HippoDeep は LICENSE ファイルが無く、同梱しない。
- VOI は本ツール独自の定義（MTL＝CerebrA、背側脳幹＝CerebrA と Harvard AAN v2.0 の共通部分）で、VSRAD の VOI とは別である。論文の閾値（0.2 など）は VSRAD の Z に対して作られたもので、本ツールの Z にそのまま当てはまらない。灰白質の比が無いため、論文の簡易・決定木基準は判定できない。比は DLB と AD の鑑別（認知症が前提）の文脈の値で、健常者でも満たしうるため、スクリーニングには使えない。
- 背側脳幹の灰白質の Z は出さない（検証に通らなかったため）。全脳白質の Z は警告つきの参考値で、年齢を ±20 歳ずらす検査に通っておらず、とくに 58 歳未満では Z が高いことを萎縮と読めない。全脳規格化後の Z は、脳全体の萎縮が強い例で領域の Z が小さく出る。
- ライセンスは、コードを MIT、IXI 由来の基準 DB・測定値・図を CC BY-SA 3.0 とし、DB はコードとは別の成果物として配布する。ShareAlike が配布物全体に及ぶ範囲の法的判断は未了で、公開前に法務の確認が要る。パイロット DB は製品に載せない。CerebrA のデータ自体の明示のライセンスは、TemplateFlow の LICENSE（MNI の許諾文）以外未確認である。
- Windows・macOS、配布物（wheel）、署名・リリースは未達。評価の主体は Linux・CPU（16 コア）での解析である。
- VSRAD との比較は IXI 健常者 5 例で、一致度を推定する規模ではない。VSRAD は 54〜86 歳・1.5T 単一装置の 80 例を基準とする別の手法で、数値は互換ではない（第 2 節）。IXI の T1 は MPRAGE など VSRAD 用の T1 強調の撮像条件で取得されておらず（具体的なシーケンスとパラメータは資料に無く不明）、差の原因は撮像条件を含めて切り分けていない。

8 再現性

8.1 ソフトウェアと環境

表 10　解析に使ったソフトウェアの版と環境

項目	値
gn_brain_atrophy	0.1.0（開発版、リリース前）
Python	3.11.17（解析・pytest の環境）
antspyx（ANTsPy）	0.6.3
hd-bet	2.0.1（CPU、--disable_tta）
torch	2.14.1+cpu
numpy／scipy／nibabel／matplotlib	2.3.5／1.15.3／5.4.2／3.11.2
HippoDeep	commit 310305fe5a928b1da3e98e7c1384490881690593。実行に使う 4 ファイルの内容の識別子 fd639702...（先頭 8 桁）
標準脳	ANTsPy 同梱の mni.nii.gz（182×218×182、1 mm、SHA-256 先頭 fc3b241d）。2 mm へ再標本化
ITK の設定	スレッド数 1、乱数の種 1、Atropos r=0
実行環境	Linux、16 コア、メモリ 93 GB、CPU のみ（GPU なし）。CPU の型番は記録なし（不明）
テスト	pytest 112 合格・4 スキップ、ui の node --test 50 合格（本稿の根拠とした最終時点）
本体（GRAPHY-Next）	engines.graphy >=0.5.1（暫定、最小版は未確認）。本体の実機確認は Linux のみ

8.2 基準 DB（SHA-256 の先頭 8 桁）

SHA-256 は DB の内容（メタデータと配列）から計算した値で、ファイルのハッシュではない。読み込み時に検証する。

表 11　基準 DB

DB	仕様	SHA-256（先頭 8 桁）	状態
DBS 灰白質 ixi_dbs_gm_makerq_v2	makerq、マスク＝DBS VOI	7910574c	不採用（D2 失敗）
全脳白質 ixi_wm_makerq_v2	makerq	6d9f23a6	不採用（基準 5 失敗）
DBS 白質 ixi_dbs_wm_makerq_v2	makerq、マスク＝DBS VOI	dfdf4e39	採用
海馬 ixi_hippo_v3	age・age²・sex・maker・field_t・log(detA)	401e8abe	採用
全脳 GM 規格化 ixi_train_makerq_gn_v2	makerq、全脳規格化	4554da84	採用（併記用）
全脳 GM ixi_train_makerq_v2	makerq（373 名）	af102685	採用
全脳 GM（train 375 名、field）	field	ccfda1c8	判定用（基準 1 失敗）
全脳白質 ixi_wm_maker_v1／ixi_wm_makerq_v1	maker／makerq	7f392dee／c341cfec	不採用（W1・W2 失敗）
DBS 灰白質 ixi_dbs_gm_maker_v1	maker、マスク＝DBS VOI	55256b18	不採用（D2 失敗）

train は 373 名（分割 v2）、年齢範囲 20.0〜83.8 歳。回帰マスクは全脳 GM の DB で 161,835 ボクセル。train の性別・施設別の内訳は 4.3.2 節（表 7a）に示す。DB の構築スクリプトは db-build/（split_ixi.py＝分割、eval_v2.py＝再評価）、分割は fw/split_v2.csv（IXI の公開 ID と属性。乱数の種 20261105）である。DB は IXI（CC BY-SA 3.0）の派生物で、コードとは別に CC BY-SA 3.0 で扱い、パイロット DB は製品に載せない。

8.3 実行と再現の手順

- 解析環境へ python -m pip install -e python[test]（antspyx 0.6.3、hd-bet 2.0.1、matplotlib、nibabel、CPU 版 torch）。環境はプラグインが書き換えない。
- python -m gn_brain_atrophy precheck --db <db.npz> で不足と版違いを確認する（実行はしない。使われる Python を名指しした導入コマンドを返す）。
- python -m gn_brain_atrophy run --t1 <T1.nii.gz> --db <db.npz> --out <dir> --age <年齢> --sex <M|F> --field <1.5|3> --maker <Philips|GE|Siemens|other> --report。海馬は --hippo-db、DBS 白質は --dbs-wm-db、全脳規格化は --gn-db を付ける。終了コードは 0（成功）、2（入力・環境）、3（QC 不合格）。
- 評価は db-build/ のスクリプト（eval_full.py・eval_maker.py・eval_hippo.py・eval_voi.py・eval_gn.py など）で、結果と閾値の記録は fw/benchmark-protocol.md・fw/benchmark-result-a.md。
- 決定性：同じ入力から同じ出力が得られること（ITK スレッド 1、種固定、Atropos r=0）を、DB の SHA-256 の一致と Z の再計算の差 0 で確かめた。

参考文献・出典・ライセンス

以下は、本ツールの設計・評価の資料（fw/ の調査ファイル）に書誌・ライセンスの記載があるものに限る。記載の無い項目は「未取得」とした。

参考文献

[1] Matsuda H, Mizumura S, Nemoto K, Yamashita F, Imabayashi E, Sato N, Asada T. Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated Lie algebra improves the diagnosis of probable Alzheimer disease. AJNR Am J Neuroradiol 2012;33(6):1109-1114. DOI 10.3174/ajnr.A2935, PMID 22300935.

[2] Matsuda H, Yokoyama K, Sato N, Ito K, Nemoto K, Oba H, et al. Differentiation between dementia with Lewy bodies and Alzheimer's disease using voxel-based morphometry of structural MRI: a multicenter study. Neuropsychiatr Dis Treat 2019;15:2715-2722. DOI 10.2147/NDT.S222966, PMID 31571887.

[3] Hirata Y, Matsuda H, Nemoto K, Ohnishi T, Hirao K, Yamashita F, et al. Voxel-based morphometry to discriminate early Alzheimer's disease from controls. Neurosci Lett 2005;382(3):269-274. DOI 10.1016/j.neulet.2005.03.038, PMID 15925102.

[4] Thyreau B, Sato K, Fukuda H, Taki Y. Segmentation of the hippocampus by transferring algorithmic knowledge for large cohort processing. Med Image Anal 2018:214-228. DOI 10.1016/j.media.2017.11.004, PMID 29156419（HippoDeep）。巻は未取得。

[5] Sghirripa S, Bhalerao G, Griffanti L, Gillis G, Mackay C, et al. Evaluating traditional, deep learning and subfield methods for automatically segmenting the hippocampus from MRI. Hum Brain Mapp 2025:e70200. DOI 10.1002/hbm.70200, PMID 40143669.

[6] Isensee ら. HD-BET の原論文. Hum Brain Mapp 2019;40(17):4952-4964. DOI 10.1002/hbm.24750, PMID 31403237。題名・全著者は本稿の資料に無く未取得（DOI・PMID で確認されたい）。

[7] Avants BB, Tustison NJ, Wu J, Cook PA, Gee JC. An open source multivariate framework for n-tissue segmentation with evaluation on public data. Neuroinformatics 2011:381-400. DOI 10.1007/s12021-011-9109-y, PMID 21373993（Atropos）。巻は未取得。

[8] Manera ら. CerebrA（MNI-ICBM152 2009c 対称 1 mm のアトラス）. Sci Data 2020。題名・巻・頁・DOI は本稿の資料に無く未取得。

データ・ソフトウェアの出典とライセンス（資料に記載のもの）

対象	出典	ライセンス・条件
IXI データセット（T1、IXI.xls）	brain-development.org/ixi-dataset	CC BY-SA 3.0。「acknowledge the source of the IXI data」。Hammersmith（Philips 3T）、Guy's（Philips 1.5T）、IoP（GE 1.5T）。年齢範囲 20〜86 歳・約 581 名は二次情報＋本件のヘッダ走査（581 ファイル）。DB は派生物で、ShareAlike の扱いは配布前に確認が要る。
CerebrA（アトラス）	Manera 2020（TemplateFlow の tpl-MNI152NLin2009cSym）	TemplateFlow の LICENSE は MNI の許諾文（目的を問わず無償で使用・複製・改変・配布でき、著作権表示を付ける）。データ自体の明示のライセンスは未確認。
Harvard Ascending Arousal Network（AAN）v2.0	Dryad、doi:10.5061/dryad.zw3r228d2	CC0 1.0（Dryad の README の記載）。著者は資料に記載なし。
HippoDeep	github.com/bthyreau/hippodeep_pytorch（commit 310305f）	README に MIT と記載、LICENSE ファイルは無い（MRA モデルは MIT+Attribution）。同梱しない。
HD-BET	github.com/MIC-DKFZ/HD-BET（重みは Zenodo 2540695）	Apache-2.0（LICENSE 取得）。重み単独のライセンスは README に独立の記載なし。
ANTs／ANTsPy（antspyx 0.6.3）	github.com/ANTsX/ANTs、PyPI antspyx	Apache-2.0。MNI テンプレート mni.nii.gz は ANTsPy 同梱（来歴の詳細は未取得）。
nibabel	framework.md の調査	MIT。
matplotlib・numpy・scipy・torch・scikit-learn	PyPI のパッケージ情報（2026-11-05 に確認）	matplotlib は PSF 系（matplotlib ライセンス）、numpy・scipy は BSD、torch は Apache-2.0 ほか BSD 系、scikit-learn は BSD-3-Clause。同梱せず利用者が入れる。詳細はリポジトリの NOTICE。