

1. はじめに

脳の皮質厚や皮質下領域の体積は、加齢や精神・神経疾患と関連していることが知られている。Magnetic Resonance Imaging (MRI) を用いた脳画像解析の従来の研究では、脳年齢予測や疾患分類等のタスクに特化した個別モデルで実現されている。そのため、複数の下流タスクに適用可能な基盤モデルの実現が期待されている。画像や自然言語処理の分野では、自己教師あり学習による基盤モデルの構築が有効とされている。そこで、本研究では、脳 MRI を対象とした基盤モデルの構築を目的とする。また、下流タスクにおいて基盤モデルがどの程度有用な表現を獲得できているかを検証する。

2. 従来研究

Siegel らは、MRI からの脳年齢予測タスクにおいて、CNN と Transformer の有用性を評価している [1]。結果として、いずれのモデルにおいても高精度な脳年齢予測が可能であることが示唆された。しかし、従来手法は教師あり学習を前提としており、大規模なラベル付きデータの確保や他タスクへの展開という課題がある。これに対し、自己教師あり学習を用いることで、ラベルなしデータからさまざまなタスクに共通する汎用的な特徴を抽出でき、基盤モデルの作成が可能と考える。これにより、脳年齢予測や疾患分類といった複数のタスクに特化したモデルを構築するためのデータ収集やアノテーションコストを低減できる。

3. 提案手法

本研究では、教師あり学習による脳 MRI 解析手法で必要となる大規模なラベル付きデータを用いず、自己教師あり学習により基盤モデルを構築する。

3.1 データセットと前処理

基盤モデルの構築に向けて、8つのデータベースおよび公開サイトから 54,521 件のデータを収集した。データの内訳は、事前学習用データ 47,699 件、年齢予測タスク用データ 8,824 件、3 クラス分類の疾患データが 1,421 件、2 クラス分類の疾患データが 571 件である。なお、分類タスクの健常は年齢予測タスクと共有されている。各データベースでは撮影条件や保存形式が異なり、学習に悪影響を及ぼす可能性が高いため、3つの前処理を実施した。図1に本研究で用いた前処理を示す。1つ目に RAS 座標系の統一、2つ目に 1mm ボクセルへのリサンプリング、3つ目に Z スコアによる正規化を行う。これらの処理により、複数のデータバンク間に存在する差異を低減した統一的な MRI データを得ることができる。

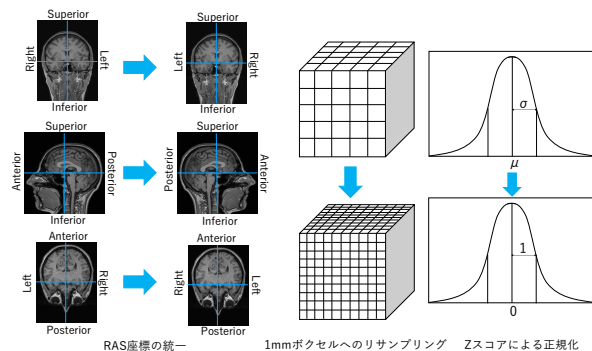


図 1：本研究における MRI データの前処理

3.2 基盤モデルの構築

事前学習では、自己教師あり学習に基づく手法として 3D Masked Autoencoder (3D MAE) を採用する。入力となる 3 次元 MRI を 3D パッチに分割し、ランダムに選択したパッチをマスクする。マスクされていないパッチのみを 3D Vision Transformer (3D ViT) の Encoder に入力し、得られた潜在表現から Decoder によりマスクされたパッチの復元を行う。このとき、Decoder で出力した復元パッチ

と元画像中の対応するパッチとの差を平均二乗誤差として計算し、その誤差を最小化するように学習を行う。この事前学習によって得られた Encoder を基盤モデルとして用いる。

4. 実験概要

事前学習は 47,699 件のデータを使用し、マスク率 75% で学習する。基盤モデルの汎用性を検証するために下流タスクとして脳年齢予測タスクと疾患分類タスクを実施し、教師あり学習で学習を行った ViT-Large と比較検証する。

4.1 脳年齢予測タスク

MRI から被験者の実年齢を予測する脳年齢予測タスクを行う。脳年齢予測タスクには、8,824 件のデータを使用している。脳年齢予測タスクにおける教師あり学習と自己教師あり学習の精度比較を表 1 に示す。表 1 より、教師あり学習モデルに対し、自己教師あり学習で事前学習を行ったモデルでは MAE が 1.59pt 低下し、決定係数 R^2 は 0.16pt 向上した。

表 1：脳年齢予測タスク

学習方法	MAE (\downarrow)	R^2 (\uparrow)
教師あり	6.11 \pm 0.37	0.48 \pm 0.06
自己教師あり	4.52 \pm 0.12	0.64 \pm 0.02

4.2 アルツハイマー分類タスク

疾患分類として、健常者と認知機能障害、アルツハイマーの 3 クラス分類を行う。3 クラス分類では、2,037 件の MRI データを 8:2 に分割した交差検証を用いている。表 2 に、アルツハイマーの 3 クラス分類タスクの結果を示す。表 2 より、自己教師あり学習で事前学習を行ったモデルの方が、教師あり学習で学習を行ったモデルと比較して精度が向上した。

表 2：3 クラス分類の精度比較

学習方法	Accuracy (%)	Macro F1 (%)
教師あり	44.31 \pm 2.24	33.01 \pm 1.30
自己教師あり	65.43 \pm 1.58	57.36 \pm 1.72

4.3 自閉スペクトラム症分類タスク

疾患分類として、健常者と自閉スペクトラム症の 2 クラス分類を行う。2 クラス分類では、1,099 件の MRI データを 8:2 に分割した交差検証を用いている。表 3 に、自閉スペクトラム症の 2 クラス分類タスクの結果を示す。表 3 より、自己教師あり学習で事前学習を行ったモデルの方が、教師あり学習で学習を行ったモデルと比較して精度が向上した。

表 3：2 クラス分類の精度比較

学習方法	Accuracy (%)	Macro F1 (%)
教師あり	55.16 \pm 4.09	54.74 \pm 4.14
自己教師あり	58.90 \pm 2.04	58.71 \pm 2.99

これらの結果から、大規模な事前学習を行うことでそのタスクに特化した学習と比較して、より良い精度を得られることが確認できた。

5. おわりに

本研究では、MRI を用いた脳画像解析のための基盤モデル構築に向け、3D Vision Transformer による大規模データを用いた自己教師あり学習の有効性を検証した。MRI データを用いて自己教師あり学習を行うことで、従来研究である教師あり学習を用いた手法と比較して精度の向上を確認した。このことから、特定のタスクに特化させるための教師あり学習よりも、自己教師あり学習による汎用的な特徴獲得の有効性が示された。今後は、拡散モデルによるデータ拡張や、他の事前学習手法の検討を行う。

参考文献

- [1] N. T. Siegel, *et al.* “Do Transformers and CNNs Learn Different Concepts of Brain Age?”, Wiley, 2025.