

畳み込みニューラルネットワークと認知機能テストを用いた アルツハイマー病鑑別

遠藤 大樹^{1,a)} 伊藤 康一¹ 青木 孝文¹

概要

アルツハイマー病 (Alzheimer's disease: AD) は進行性の脳疾患であり, AD を発症すると記憶や思考能力がゆっくりと障害される. AD の診断では, MRI 画像を使用した脳萎縮の評価や認知機能テストによる評価に基づいて医師が総合的に判断している. 医師の診断を支援するために, 畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を用いて AD 鑑別を行う手法が検討されている. これらの手法では, MRI 画像から抽出される特徴のみが用いられているため, 脳の萎縮しか評価していない. 提案手法では, 脳 MRI 画像に加えて認知機能テストで得られたスコアを用いて CNN を学習する手法を提案する. 大規模データセットを使用した実験を通して提案手法の有効性を示す.

1. はじめに

アルツハイマー病 (Alzheimer's disease: AD) は, 進行性の脳疾患のひとつであり, 高齢者における認知症の最も一般的な要因である. AD を発症すると, 記憶障害が生じ, 進行にともなって見当識障害の症状も出現する. 現在までに AD を完治させることはできないが, 投薬治療により進行を抑制することが可能であるため, AD の早期発見および早期診断が極めて重要である. 高度高齢化社会の到来に伴って AD に関係する患者数が増加しても正確かつ迅速な診断のために医師を補助する診断支援システムが求められている. AD の診断は, 一般に, 医師が認知機能の評価したり, 脳画像を評価したりすることで行われている. 健常な場合に Cognitively Normal (CN), 記憶障害があるが日常生活は送れる場合に Mild Cognitive Impairment (MCI), 重度の障害がある場合に AD と診断されるため, 本論文でもデータに対して同様のラベルを付与する.

AD の診断では, 核磁気共鳴画像法 (Magnetic Resonance Imaging: MRI) で撮像された脳画像が用いられる. MRI は撮像条件を変えることで様々な画像を撮像可能であり, そ

れらの画像の中で, 脳の解剖学的な構造を捉えやすいことから T1 強調画像がよく用いられている. MRI 画像から AD 鑑別を行う手法として, ハンドクラフト特徴量を用いる手法 [11] や畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を用いる手法 [1], [5], [10] が提案されている. ハンドクラフト特徴量を用いる手法 [11] では, MRI 画像から得られる灰白質, 白質, 脳脊髄液の各脳組織を特徴量とし抽出し, Support Vector Machine (SVM) を用いて鑑別を行う. 脳組織を基本とする特徴量では, 正常加齢による萎縮と AD による萎縮とを明確に判別することができないため, 鑑別精度があまり高くない. CNN は, 画像認識などにおいて有効性が実証されているように, 脳画像解析においてもハンドクラフト特徴量よりも高精度である. 一方で, CNN を用いた従来手法では, 脳の萎縮度合のみを評価して AD 鑑別を行っているため, 認知機能テストのスコアも用いて総合的に判断を行っている医師の鑑別結果と差が生じる. また, 軽度認知障害 (Mild Cognitive Impairment: MCI) の診断は医師によって結果が異なることが多いため, MCI を含むデータセットではラベルにノイズが含まれている可能性が高い. そのようなデータセットを用いて CNN の学習を行うと分類精度が低下することが知られているため, ラベルノイズの影響を抑えて CNN を学習する必要がある.

本論文では, AD 鑑別の高精度化のために, 認知機能テストで得られた値を用いて CNN を学習する手法を提案する. まず, 過学習を抑制して CNN の学習を行うため, デノイジング自己符号化器 (Denoising Autoencoder: DAE) を用いて事前学習を行う. そして, 複数の認知機能テストからラベルを作成し, ラベルに存在するノイズの影響を抑えて CNN を学習することで, AD 鑑別の精度を改善する. 提案手法は, CNN を用いた既存手法に対する学習方法の改善であるため, どのような CNN にも用いることが可能である. 本論文では, 脳画像解析で一般的に用いられている VGG [6] を基本とした 3D CNN を用いる. Alzheimer's Disease Neuroimaging Initiative (ADNI)^{*1} データベースを用いた

¹ 東北大学

^{a)} daiki@aoki.ecei.tohoku.ac.jp

^{*1} <http://adni.loni.usc.edu/>

性能評価実験を通して、提案手法の有効性を実証する。

2. CNN を用いた AD 鑑別

本論文で提案する 3D CNN を用いた AD 鑑別について述べる。以下では、前処理、自己符号化器を用いた事前学習、損失関数、提案手法の概要について述べる。

2.1 前処理

頭の位置や角度などの違いや脳形状の個人差により、画像間で脳組織の位置が異なる。正確な解析のためには、T1 強調画像中の脳の位置や大きさなどを正規化する必要がある。提案手法では、Statistical Parametric Mapping 12 (SPM12)^{*2} を用いて個人の脳形態を標準的な脳形態に合わせる。標準脳と位置合わせを行う際の標準脳テンプレートとして、SPM12 に付属している International Consortium for Brain Mapping (ICBM) space templates を用いる。標準化後の T1 強調画像に対して、CNN に入力する領域を設定する。海馬周辺領域は、AD により顕著に萎縮するため、診断の際に関心領域 (Region of Interests: ROI) と設定される。CNN を用いた識別においても事前知識として ROI を設定することが有効であると考えられる。標準化された脳全体の T1 強調画像から、脳アトラスに基づいて ROI を抽出する。本論文では、脳アトラスとして、解剖学的意味をもつ 90 領域に分割された AAL-90 [8] を用いる。

2.2 自己符号化器を用いた事前学習

一般に、ニューラルネットワークの重みは、ランダムな値に初期化される [2], [3]。そのため、医用画像のように教師データが少ない場合は、十分な精度が得られなかったり、過学習が生じたりする。過学習を抑制する方法として自己符号化器で事前学習された CNN を利用することが考えられる。自己符号化器は、入力データのみを用いた教師なし学習により、データを表現する特徴を獲得することを目的とするニューラルネットワークである。自己符号化器の表現力が高すぎると、入力データ x に含まれるノイズも復元するような恒等写像を学習してしまう可能性がある。その場合、汎化性能が低下し、有用な特徴抽出器として機能しなくなる。この問題を解決するためにデノイジング自己符号化器 (Denoising Autoencoder: DAE) [9] が提案されている。DAE は、自己符号化器を拡張したものであり、データ x にランダムなノイズを付加した \tilde{x} を入力とすることで、自己符号化器が恒等写像を学習することを防ぐ。提案手法でも同様に DAE を用いて事前学習を行う。

2.3 認知機能テストからのラベル作成

認知機能テストは、認知症状の有無や程度を定量的に評

表 1 ADNI データベースにおける認知機能テストのスコアの統計量

Method	Class	Mean	Median	Std.	Range
CDR	CN	0.05	0.00	0.17	[0,1]
	LMCI	0.57	0.50	0.35	[0,1]
	AD	0.97	1.00	0.44	[0.5,2]
ADAS	CN	8.45	8.00	4.90	[0,32]
	LMCI	20.29	20.00	9.81	[3,78]
	AD	33.42	32.00	10.83	[11,78]
MoCA	CN	26.06	26.00	2.45	[15,30]
	LMCI	21.86	22.00	4.12	[0,30]
	AD	16.49	17.00	4.70	[0,27]

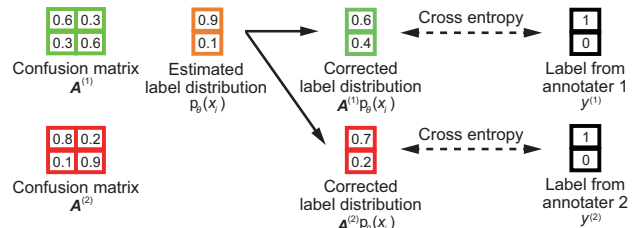


図 1 提案手法で使用する損失関数の概要

価するために、医療現場で用いられている。例えば、Clinical Dementia Rating (CDR) や、Alzheimer’s Disease Assessment Scale (ADAS), Montreal Cognitive Assessment (MoCA) などがある。CDR では、記憶、見当識、判断力と問題解決力、地域社会活動、家庭および趣味における活動、身の回りの世話の必要性などの 6 項目で重症度を評価する。ADAS は、重症度を 11 項目で評価する。MoCA は、軽度認知障害を評価することを主な目的としている。図 1 に、本論文で用いる ADNI データベースに含まれる CN, Late MCI (LMCI), AD 患者に対する認知機能テストの統計量を示す。解析に用いたデータの詳細は 3.3 節を参照されたい。提案手法では、認知機能テストで得られるスコアに閾値を設けて新たにラベルを作成する。例えば、CDR において、スコアが高いほど AD である可能性が高いため、閾値以上の患者を AD とし、閾値未満の患者を CN とラベルづける。

2.4 損失関数

Tanno らが提案した損失関数 [7] に基づいて複数ラベルに対する損失関数を設計する。図 1 に損失関数の概要を示す。この損失関数は、複数のアノテータが作成したラベルにノイズが含まれる場合に、ノイズの影響を抑制しながら CNN を学習させることが可能である。この損失関数は、(i) 交差エントロピー項と (ii) 正則化項に分けられる。(i) では、まず、それぞれのアノテーター r に対して混同行列 (Confusion Matrix) $A^{(r)}$ を定義する。そして、CNN が出力する推定確率分布 p_θ を混同行列を用いて補正し、アノテーター r が作成したラベル $y^{(r)}$ との交差エントロピーを計算する。入力データを x 、データ数を N 、アノテーター

^{*2} <https://www.fil.ion.ucl.ac.uk/spm/>

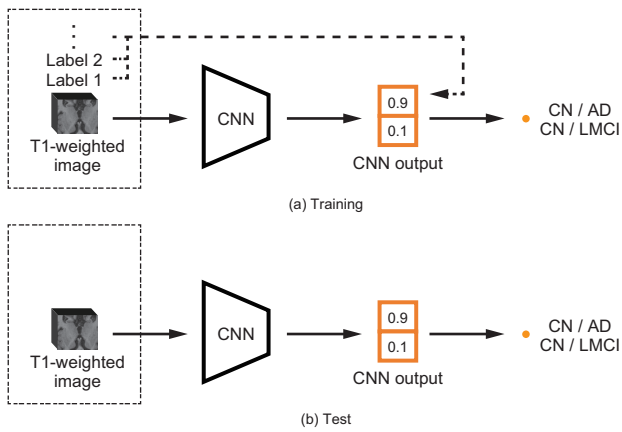


図 2 提案手法の概要

を R とすると、交差エントロピー項 (i) は次式で計算される。

$$\sum_{i=1}^N \sum_{r=1}^R CE(\mathbf{A}^{(r)} \mathbf{p}_{\theta}(x_i), y^{(r)}) \quad (1)$$

ここで、 $CE(\cdot)$ は交差エントロピーを計算する関数を表す。(ii) は、アノテータの信頼性を保つため、混同行列の対角和を最小化する項である。最終的な損失関数は、アノテータによるラベルが存在しないデータがあることを考慮して、次式となる。

$$\sum_{i=1}^N \sum_{r=1}^R \mathbf{1}(y_i^{(r)} \in S(x_i)) \cdot CE(\mathbf{A}^{(r)} \mathbf{p}_{\theta}(x_i), y^{(r)}) + \lambda \sum_{r=1}^R tr(\mathbf{A}^{(r)}) \quad (2)$$

ここで、 $S(x_i)$ はデータ x に対する全てのラベル、 λ は (ii) に対する重み、 $tr(\cdot)$ は対角和を計算する関数、 $\mathbf{1}(\cdot)$ は引数が真のときは 1、偽の時は 0 となる関数を示す。

2.5 提案手法の概要

図 2 に提案手法における学習と推論の概要を示す。学習では、図 2 (a) に示すように、事前学習を行った DAE の特徴抽出器を使用する。認知機能テストのスコアから作成されたラベルと T1 強調画像を CNN への入力とし、2.4 節で述べた損失関数を最小化するように学習を行う。推論では、T1 強調画像のみを学習済みの CNN へ入力することで、CN/AD または CN/LMCI の確率分布を出力する。

3. 性能評価実験

本節では、提案手法の性能評価実験について述べる。DAE を用いた事前学習に IXI^{*3} を使用し、ADNI^{*1} データベースを使用して AD 鑑別の精度を評価する。

3.1 ネットワークアーキテクチャ

提案手法は、複数のラベルに基づいた学習により CNN

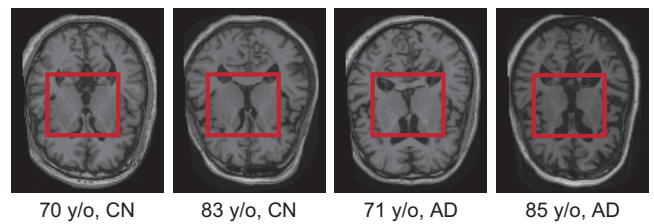


図 3 実験に用いた T1 強調画像の例 (赤枠は ROI を示す.)

表 2 実験に用いた ADNI データベースに含まれる T1 強調画像の詳細

Class	# of Subjects	# of Scans	Age range
CN	500	1,589	[55, 94]
AD	207	578	[55, 93]
LMCI	163	894	[55, 94]

の識別精度を向上させるため、どのような CNN にも適用できる。本論文では、脳画像解析で用いられる一般的な CNN である VGG [6] を基本とした 3D CNN を使用する。3 次元畳み込み層 (カーネルサイズ: $3 \times 3 \times 3$, スライド: 1), バッチ正規化 (Batch Normalization: BN) 層, 活性化関数, Max Pooling 層で構成される畳み込みブロックを 4 回繰り返す。畳み込みブロックの活性化関数には Rectified Linear Unit (ReLU) を用いる。畳み込みブロックのチャンネル数は 16, 32, 64, 128 である。識別器は、全結合 (Fully-Connected: FC) 層と活性化関数で構成し、過学習を抑制するために、50% の確率でドロップアウトを行う。FC 層の活性化関数には ReLU と Softmax を用いる。事前学習で用いる DAE の Encoder には、上記の畳み込みブロックを用いる。Decoder には、3 次元転置畳み込み層 (カーネルサイズ: $2 \times 2 \times 2$, スライド: 2), バッチ正規化層, 活性化関数で構成される畳み込みブロックを 4 層繰り返す。Decoder の活性化関数には、はじめの 3 ブロックに ReLU を用い、4 ブロック目に Tanh を用いる。

3.2 事前学習

本実験では、IXI^{*2} が公開しているデータベースを用いて事前学習を行う。19 歳から 86 歳の健常者のみで構成される 563 枚のデータを使用し、全データの 90% を学習に、10% を評価に用いる。入力データに対してノイズとして平均 0, 標準偏差 0.1 のガウシアンノイズを付加する。学習に使用できる画像枚数が少なく、過学習を引き起こす可能性があるため、データ拡張を行う。本実験では、水平反転, スケーリング, 画像シフトの 3 種類のデータ拡張を用いる。学習時の損失関数として平均二乗誤差 (Mean Square Error: MSE) を用い、最適化手法として Adam [4] を用いる。学習率の初期値を 0.0005, 重み減衰を 0.0001 とし、400 エポックの学習を行う。

*3 <https://brain-development.org/ixi-dataset/>

表 3 AD 鑑別の実験結果

	DAE	Multi-Label Learning				CN vs LMCI		CN vs AD	
		Physician	CDR	ADAS	MoCA	Acc.	AUC	Acc.	AUC
Baseline						0.652	0.736	0.877	0.950
Proposed	✓					0.715	0.753	0.898	0.950
Proposed	✓	✓	✓	✓		0.732	0.774	0.888	0.946
Proposed	✓	✓	✓	✓	✓	0.747	0.769	0.891	0.946

3.3 AD 鑑別による精度評価

AD 鑑別を行うネットワークの学習には、ADNI データベースを用いる。図 3 に画像の例および ROI ($54 \times 47 \times 35$ ボクセル) を示し、データセットの詳細を表 2 に示す。モデルの汎化性能を評価するために、全データを 5 分割して交差検証を行う。ADNI データベースには、1 人の被験者に対して、撮像時期が異なる複数の T1 強調画像が含まれている。それらのデータが学習用と評価用の両者に存在しないように、全データの 80% を学習に、20% を評価に用いるように分割する。また、学習データの 12.5% をバリデーションデータとして分離し、過学習を確認するために使用する。DAE の事前学習と同様にデータ拡張を適用する。2.2 節で述べた DAE による事前学習を行い、2.4 節で述べた損失関数を用いて学習を行う。認知機能テストで得られたスコアに基づいてラベルを作成する際に、学習データにおいて 2 つのクラスが最も分離されるスコアを閾値とする。医師によるラベルと、CDR, ADAS, MoCA から作成したラベルを学習に用いる。最適化手法として momentum SGD を使い、momentum を 0.9, 重み減衰を 0.0001, 学習率の初期値を 0.0005 とする。バリデーションデータで最も精度が高いモデルを用いて評価を行う。精度評価の指標として、予測に対する正答率を示す Accuracy と、真陽性率と真陰性率で表される Receiver Operating Characteristic (ROC) 曲線の Area Under the Curve (AUC) を用いる。

実験結果を表 3 に示す。提案手法の有効性を示すために、VGG [6] (Baseline) に対して同条件で実験を行った。DAE による事前学習を行うと精度が向上した。事前学習により過学習が抑制されたと考えられる。複数のラベルを用いた学習では、CN と LMCI の鑑別で精度が向上した。ラベルに含まれるノイズの影響を抑えて学習が行われたと考えられる。CN と AD の鑑別の場合、複数のラベルを用いずに学習した場合と同程度の精度であった。以上より、提案手法を用いることで医師でも判断が難しい CN と LMCI の鑑別において精度が改善されている。

4. まとめ

本論文では、脳画像と認知機能テストを用いた 3D CNN による AD 鑑別手法を提案した。DAE を用いた事前学習により過学習を抑制するとともに、認知機能テストのスコアを考慮することで高精度化した。ADNI データベースを

用いた実験を通して、提案手法の有効性を実証した。

参考文献

- [1] Bäckström, K., Nazari, M., Gu, Yu-Hua, I. and Jakola, S.: An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images, *IEEE Int'l Symp. Biomed. Imaging*, pp. 149–153 (2018).
- [2] Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, *Proc. Int'l Conf. Artificial Intelligence and Statistics*, Vol. 9, pp. 249–256 (2010).
- [3] He, K., Zhang, X., Ren, S. and Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification, *Proc. IEEE Int'l. Conf. Computer Vision*, pp. 1026–1034 (2015).
- [4] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, *Proc. Int'l Conf. Learning Representations*, Vol. abs/1412.6980, pp. 1–15 (2015).
- [5] Nigri, E., Ziviani, N., Cappabianco, F., Antunes, A. and Veloso, A.: Explainable deep CNNs for MRI-based diagnosis of Alzheimer's disease, *Int'l Joint Conf. Neural Networks*, pp. 1–8 (2020).
- [6] Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, *CoRR*, Vol. abs/1409.1556, pp. 1–14 (2014).
- [7] Tanno, R., Saeedi, A., Sankaranarayanan, S., Alexander, D. C. and Silberman, N.: Learning from noisy labels by regularized estimation of annotator confusions, *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, pp. 11244–11253 (2019).
- [8] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B. and Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI single-subject brain, *NeuroImage*, Vol. 15, No. 1, pp. 273–289 (2002).
- [9] Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.-A.: Extracting and composing robust features with denoising autoencoders, *Proc. Int'l Conf. Machine Learning*, pp. 1096–1103 (2008).
- [10] Wen, J., Thibaut-Sutre, E., Diaz-Melo, M., Samper-Gonzalez, J., Routier, A., Bottani, S., Dormont, D., Durleman, S., Burgos, N. and Colliot, O.: Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation, *Medical Image Analysis*, Vol. 63, pp. 1–19 (2020).
- [11] Yang, W., Lui, R. M. L., Gao, J.-H., Chan, T. F., Yau, S.-T., Sperling, R. A. and Huang, X.: Independent component analysis-based classification of Alzheimer's disease MRI data, *J. Alzheimer's Disease*, Vol. 24, pp. 775–783 (2011).