

令和 4 年度
修士学位論文

深層学習による 3 次元 MRI 脳画像分類モデルの説明可能性に関する研究

A Study on Explainability of Deep Neural
Network-Based Classification Models for 3D Brain
MRI

1255115 藤 沢 元

指導教員 吉田 真一

2023 年 2 月 28 日

高知工科大学大学院 工学研究科 基盤工学専攻
情報学コース

要 旨

深層学習による 3 次元 MRI 脳画像分類モデルの説明可能性に 関する研究

藤 沢 元

CNN のブラックボックス性の問題によりどの特徴、情報を用いて結論づけを行っているかが不明であるという問題に対し、それらを解決するために判断の根拠が解釈可能な説明可能 AI が注目されている。医療分野での MRI 等の画像診断においても CNN による判別が研究されているが、説明性の問題を抱える。Grad-CAM による可視化で、画像のどの部分が判別結果に寄与したかがわかるが、まだ十分ではない。また、MRI 画像の判別のために多くの前処理が行われ、レジストレーションとセグメンテーションなどの前処理と判別精度の関連についても研究はあるが、十分な説明性を得ていない。そこで本研究では、前処理の各ステップの画像に対して Grad-CAM を適用し、可視化や部位ごとの精度差といった MRI 脳画像の説明性としてあげられる内容によって CNN モデルの識別領域の説明性を求める。そのため、レジストレーションとセグメンテーションを組み合わせ、9 種類の画像を作成し、精度の比較と Grad-CAM による可視化を行う。

識別結果は非線形レジストレーション後の脳領域抽出画像を除いて識別可能であり、精度の比較と Grad-CAM による可視化によって白質、特に小脳、視床周辺、脳梁などの部位が寄与していることがわかった。寄与部位として挙げられた部位は男女差があるという文献もあるが、ヒトの脳構造には違いはないとの報告もあり、少なくとも、本研究の結果からは、CNN での判別結果には上記の部位が寄与していることが分かり、それらの画像の何らかのパターンに違いがあることが示唆された。

キーワード 畳み込みニューラルネットワーク, 説明性, Grad-CAM, MRI 脳画像

Abstract

A Study on Explainability of Deep Neural Network-Based Classification Models for 3D Brain MRI

FUJISAWA, Moto

Explainable AI, in which the basis for decisions can be interpreted, is an important issue to solve the black box problem of convolutional neural-networks (CNNs). CNN-based discrimination has also been studied in medical imaging, such as MRI, and visualisation with Grad-CAM shows which parts of the image contributed to the discrimination result, however this is still not sufficient. Many pre-processing techniques are used to discriminate MRI images, and although there have been studies on the relationship between pre-processing, such as registration and segmentation, and discrimination accuracy, they have not been sufficiently explanatory. In this study, Grad-CAM is applied to the images at each pre-processing step, and the explanatory properties of the discriminative regions of the CNN model are determined by the explanatory properties of the MRI brain images, such as visualisation and accuracy differences between regions. To do this, registration and segmentation are combined to produce nine different images, which are then compared for accuracy and visualised using Grad-CAM.

The results were discriminative, except for the brain region images extracted after non-linear registration, and the comparison of accuracy and visualisation by Grad-CAM showed that regions such as white matter, in particular the cerebellum, peri-thalamus and corpus callosum, contributed to the discrimination.

Although it is thought that there are sex differences in the regions listed as con-

tributing regions, it has also been reported that there are no differences in human brain structure, suggesting that the regions that produce sex differences contribute to discrimination, but that other regions also contribute to discrimination, suggesting that at least some pattern differences are used for discrimination.

key words CNN, Explainability, Grad-CAM, MRI brain imaging

目次

第 1 章	序論	1
第 2 章	関連研究	3
2.1	男女識別タスクの精度と問題点	3
2.2	前処理画像ごとの男女識別精度の違いと未解決内容	3
2.3	可視化手法	4
第 3 章	実験	5
3.1	データセット	5
3.2	画像の前処理手順	5
3.2.1	元画像	6
3.2.2	有効視野の変更	6
3.2.3	トリミング	6
3.2.4	バイアス補正	7
3.2.5	線形レジストレーション	7
3.2.6	非線形レジストレーション	7
3.2.7	非線形レジストレーションの変位	8
3.2.8	線形レジストレーション後の脳領域抽出	8
3.2.9	非線形レジストレーション後の脳領域抽出	8
3.2.10	脳領域抽出	8
3.2.11	全脳セグメンテーション	8
3.2.12	白質セグメンテーション	9
3.2.13	灰白質セグメンテーション	9
3.2.14	脳髄液セグメンテーション	9

目次

3.3	機械学習タスク	10
3.3.1	モデル構成	10
3.3.2	実験設定	10
第 4 章	結果	12
4.1	Grad-CAM	12
第 5 章	考察	16
5.1	線形レジストレーション画像の考察	16
5.2	線形レジストレーション画像の脳領域抽出の差の考察	16
5.3	脳抽出画像ごとの考察	17
5.4	非線形レジストレーション画像の前処理の考察	17
5.5	考察まとめ	18
5.6	その他の考察と今後の展望	18
第 6 章	結論	20
	謝辞	21
	参考文献	22

図目次

3.1	年齢分布	6
3.2	元画像	7
3.3	FOV 変更	7
3.4	トリミング	7
3.5	バイアス補正	7
3.6	線形	9
3.7	非線形	9
3.8	変位	9
3.9	マスク	9
3.10	線形抽出	9
3.11	非線形抽出	9
3.12	脳領域	9
3.13	全脳	9
3.14	白質	9
3.15	灰白質	9
3.16	脳髄液	9
3.17	モデル構成図	11
4.1	画像 A スライス 41	13
4.2	画像 A スライス 30	13
4.3	画像 B スライス 45	13
4.4	画像 A 小脳周辺	13
4.5	画像 A 視神経	13
4.6	画像 B 小脳周辺	13

図目次

4.7 画像 C スライス 70	14
4.8 画像 D スライス 67	14
4.9 画像 E スライス 72	14
4.10 画像 C 小脳周辺	14
4.11 画像 D 小脳周辺	14
4.12 画像 E 全体	14
4.13 画像 F スライス 68	14
4.14 画像 F スライス 52	14
4.15 画像 F 視床下部周辺	14
4.16 画像 F 上部	14
4.17 画像 G スライス 47	15
4.18 画像 G スライス 50	15
4.19 画像 I スライス 45	15
4.20 画像 G 小脳周辺	15
4.21 画像 H 後頭部	15
4.22 画像 I 小脳周辺	15

表目次

3.1 3DCNN モデルの構成表	10
4.1 3DCNN モデルの各画像での識別結果	12

第 1 章

序論

ヒトの脳について、身体他の部分と同じように性的二型性がある、すなわち男性と女性とで脳の構造的特徴が異なるか否かについては、様々な議論が行われてきている。文献 [1] や文献 [2] では、前視床下部間質核の性的二型性について述べられており、一般的に男性の方が女性よりも脳の大きさは大きい、脳梁の横断面積は平均的に男性が大きい、脳梁膨大部は選択的に女性が大きいなどの報告があげられているが、一方でヒトの脳構造に性差はあまりないとの報告もある。

近年、医療や神経科学分野において MRI 脳画像に対する畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) の適用も進んでいるが、MRI 脳構造画像に対する男女識別モデル [3] も報告されており、入力画像の前処理方法を変えた CNN の判別結果の違いについての報告 [4] では、様々な前処理入力画像においても CNN によって MRI 脳画像から男女を識別することが可能である。しかしながら、CNN の判別結果からでは寄与部位はどこであるかは未知であった。これはニューラルネットのブラックボックス性の問題により、どのような特徴や情報を用いて結論づけているかが不明であるためであり、この問題に対して判断根拠が解釈可能な説明可能型 AI が注目されている。医療分野における説明性は、寄与部位の可視化や画像の前処理ごとの可視化の比較、識別精度の比較があげられ、これらの説明性により男女識別可能モデルの説明性を求めることが本研究の目的となる。個人差がある脳の大きさを揃えたうえで、脳領域以外のその他の領域 (特に頭蓋) を除去した画像を作成し、大きさや脳領域以外の影響を排除した識別を行い、Gradient-weighted Class Activation Mapping (Grad-CAM) [5] によって可視化を行うことで、寄与部位の特定や、画像の前処理ごとの寄与部位の違いについて示し、男女識別可能モデルの説明性と

する.

第 2 章

関連研究

本章では，関連研究と本研究で用いる技術について記述する．

2.1 男女識別タスクの精度と問題点

文献 [3][6] では，男女識別タスクの精度が 90%前後と高精度である．使用した画像は，頭蓋あり画像，頭蓋あり線形レジストレーション画像，頭蓋なし全脳画像であり，頭蓋ありの場合においてはその他の前処理画像と比べて高精度であることから，脳領域以外の特に頭蓋が識別へ影響を与えていると考えられている．また，頭蓋なし全脳画像はレジストレーションを行わない場合，男女差があるとされている脳の大きさの違いが影響していると考えられている．そのため，脳領域以外のその他の領域を取り除き，レジストレーションが行われた画像を使用した男女識別を行うことが提案される．

頭蓋の影響は新たに前処理を行うことで排除できるが，実際に影響を与えていたことを確認するため，頭蓋あり画像を識別可能なモデルについて可視化によって寄与部位を確認することで CNN モデルがどの特徴に着目しているかを探る．

2.2 前処理画像ごとの男女識別精度の違いと未解決内容

文献 [4] において，前処理画像ごとの男女識別の精度は，頭蓋あり (レジストレーション画像を含む) \geq セグメンテーション \geq マスク \geq 変位 $>$ チャンスレベルの順であることが報告されている．識別精度の上位には，脳領域以外の情報や脳の大きさから識別していると考えられる画像が含まれ，変位画像においても識別可能であることがわかる．前処理画像の中

2.3 可視化手法

でも、変位画像が識別可能であることは、脳の形をしていない画像ではなく脳の局所的な形状情報があるとされているため解析対象となる。この変位画像は、非線形レジストレーションを行う際に足し合わせる微小変位であり、3.2.7 節にて詳細を記述する。

2.3 可視化手法

ニューラルネットには内部構造が不明であるブラックボックス性の問題のため、どの特徴や情報に基づいて結論付けを行なっているかが不明である。そのため判定の根拠として Class Activation Map (CAM) [7] が提案された。CAM は Global Average Pooling (GAP) 層が必要な技術であるが、学習済みモデルの特徴マップと、特徴マップのチャンネル毎の重みを求め、重み付けを行なったチャンネルを足し合わせ元画像のサイズにリサイズし顕著性マップとする。

Grad-CAM[5] は GAP の制限を受けず、特徴マップごとに勾配の平均を取り重みとし、特徴マップと重みを足し合わせ ReLU を通すことで顕著性マップとする。GAP の制限を受けないことで、CNN モデルがどのような構成でも Grad-CAM の実装が可能とされる。

第 3 章

実験

本章では，使用するデータセットとデータセットの画像に対する前処理方法，そして構築する 3DCNN モデルの構成と機械学習を行うタスク，設定を記述する．

3.1 データセット

使用するデータセットは，ロンドンの異なる 3 箇所の病院で収集された 581 人の T1 強調画像が含まれる IXI-Dataset[8] である．MRI 装置は，Philips 社の 3T，Philips 社の 1.5T，GE 社の 1.5T をそれぞれの病院が使用しているため機種と磁場強度によって違いは生まれるが，本研究では区別をせずに使用する．全被験者の年齢分布は図 3.1 の通りである．

3.2 画像の前処理手順

1 人の被験者を用いて，使用する画像を前処理する手順を示す．全ての画像はバイアス補正までの手順は順番に行われるが次の手順はレジストレーションとセグメンテーションに分歧し，それぞれ前処理が行われる．

レジストレーションを行う際の標準脳は，カナダの Montreal Neurological Institute (MNI) が作成した MNI152 標準脳を用いる．

FMRIB Software Library (FSL) [9][10][11][12][13] を用いて前処理を行う．

本研究で使用する前処理画像は，図 3.6，図 3.7，図 3.8，図 3.10，図 3.11，図 3.13，図 3.16，図 3.15，図 3.14 である．

3.2 画像の前処理手順

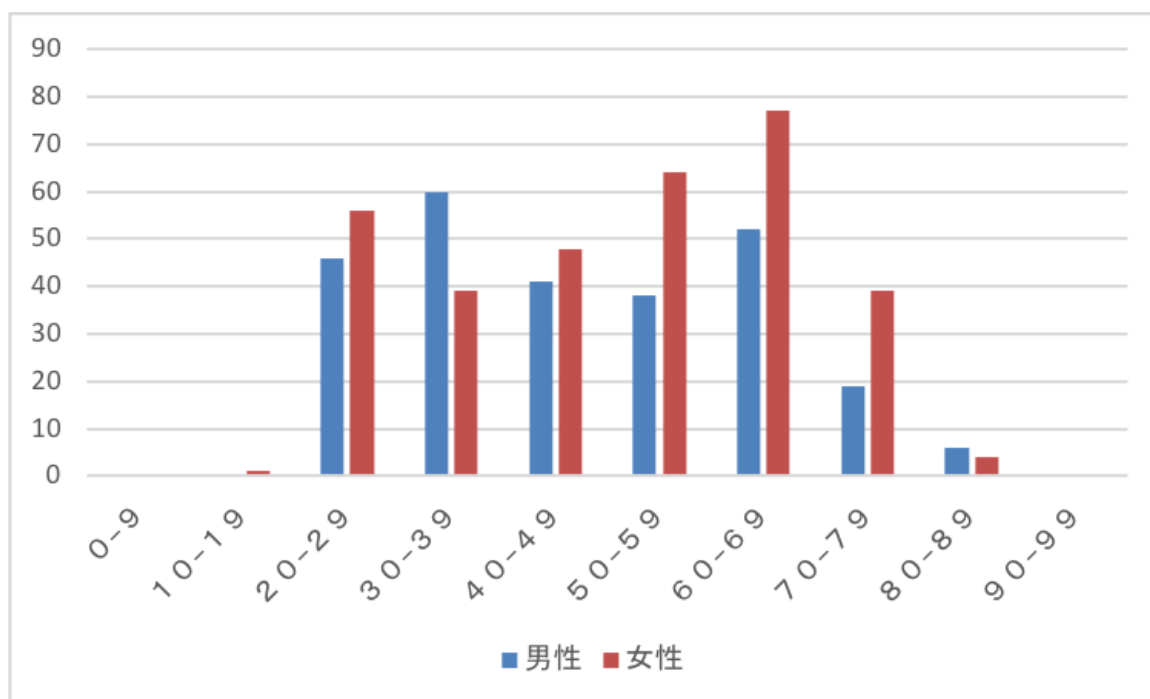


図 3.1 年齢分布

3.2.1 元画像

図 3.2 は、撮像時の元画像である。

3.2.2 有効視野の変更

図 3.2 に対して、有効視野（Field of view : FOV）の変更を行った画像が図 3.3 である。広範囲に撮像後、見たい範囲が部分的である場合に FOV を狭くすることが多いが、今回の前処理は変更前後の FOV が同じであるため変更は行われない。

3.2.3 トリミング

図 3.3 に対して、水平面（Horizontal）をトリミングした画像が図 3.4 である。

3.2 画像の前処理手順

3.2.4 バイアス補正

図 3.4 に対して、バイアス補正をかけた画像が図 3.5 である。

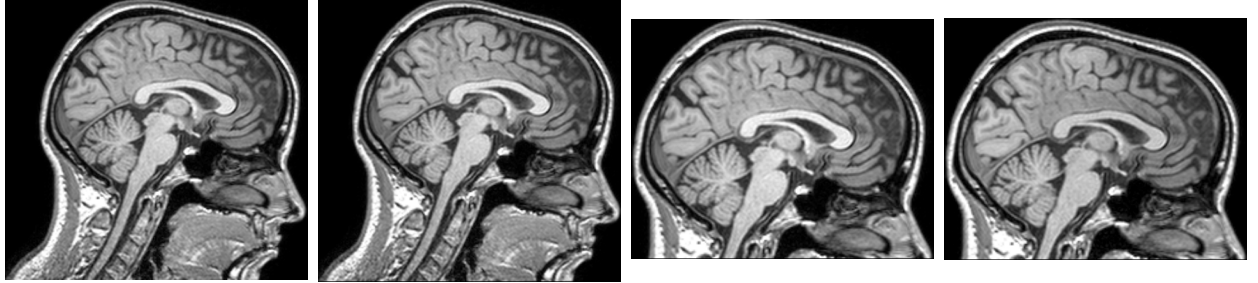


図 3.2 元画像

図 3.3 FOV 変更

図 3.4 トリミング

図 3.5 バイアス補正

3.2.5 線形レジストレーション

図 3.5 に対して、MNI152 標準脳へ線形レジストレーションを行なった画像が図 3.6 である。

3.2.6 非線形レジストレーション

図 3.5 に対して、MNI152 標準脳へ非線形レジストレーションを行なった画像が図 3.7 である。式 (3.1) の左辺 $\begin{bmatrix} x' & y' & z' & 1 \end{bmatrix}^T$ が非線形レジストレーションによって得られる画像である。レジストレーション前の図 3.5 である、式 (3.1) の右辺 $\begin{bmatrix} x & y & z & 1 \end{bmatrix}^T$ に対し、アフィン変換行列 \mathbf{A} (平行移動, 回転, 拡大縮小, 角度変更の線形変換) を乗算し, 非線形変換の位置 (x, y, z) における微小変位 $d(x, y, z)$ を加算することで画像を作成する。

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} + \begin{bmatrix} d_x(x, y, z) \\ d_y(x, y, z) \\ d_z(x, y, z) \\ 0 \end{bmatrix} \quad (3.1)$$

3.2 画像の前処理手順

3.2.7 非線形レジストレーションの変位

図 3.5 に対して，MNI152 標準脳へ非線形レジストレーションを行ない，変位を画像にしたものが図 3.8 である．変位とは，式 (3.1) の $\begin{bmatrix} d_x(x, y, z) & d_y(x, y, z) & d_z(x, y, z) & 0 \end{bmatrix}^T$ である．

3.2.8 線形レジストレーション後の脳領域抽出

図 3.6 に対して，図 3.9 を用いて，脳領域の抽出を行なった画像が図 3.10 である．用いるマスク画像は MNI152 標準脳にレジストレーションしているため全被験者共通である．

また，脳領域抽出後にレジストレーションを行うことはできるが，この方法では被験者ごとに大きな差ができ，使用できる画像であるかの選別が必要になり，使用できる枚数が減るため不適切である．

3.2.9 非線形レジストレーション後の脳領域抽出

図 3.7 に対して，図 3.9 を用いて，脳領域の抽出を行なった画像が図 3.11 である．

3.2.10 脳領域抽出

図 3.5 に対して，脳領域抽出を行なった画像が図 3.12 である．

3.2.11 全脳セグメンテーション

図 3.12 に対して，脳髄液，灰白質，白質の 3 部位にセグメンテーションを行うが，部位ごとに画像を分割せずに 1 つの画像として作成した画像が図 3.13 である．図 3.12 は使用せず，図 3.13 を脳領域抽出全脳画像として使用する．セグメンテーション前後で比べて，寄与部位の可視化時にわかりやすいと考え使用する．

3.2 画像の前処理手順



図 3.6 線形

図 3.7 非線形

図 3.8 変位

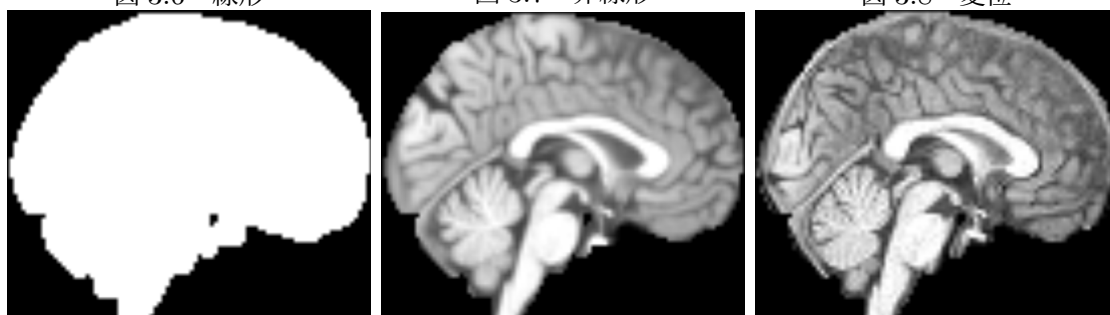


図 3.9 マスク

図 3.10 線形抽出

図 3.11 非線形抽出

3.2.12 白質セグメンテーション

図 3.12 に対して、白質のセグメンテーションを行なった画像が図 3.14 である。

3.2.13 灰白質セグメンテーション

図 3.12 に対して、灰白質のセグメンテーションを行なった画像が図 3.15 である。

3.2.14 脳髄液セグメンテーション

図 3.12 に対して、脳髄液のセグメンテーションを行なった画像が図 3.16 である。



図 3.12 脳領域

図 3.13 全脳

図 3.14 白質

図 3.15 灰白質

図 3.16 脳髄液

3.3 機械学習タスク

3DCNN モデルを用いた男女識別タスクを行う実験環境について記述する。

3.3.1 モデル構成

図 3.17 に使用する 3DCNN モデルの構成を示す。

表 3.1 に Conv3D (フィルタ数, フィルタサイズ), MaxPooling (プーリングサイズ), Dense(クラス数) としてモデル詳細を示す。フィルタサイズやプーリングサイズで書かれる数字は $N \times N \times N$ のことを指す。

3.3.2 実験設定

3.2 節で作成する画像を入力とし, 男性と女性の 2 クラス分類を行う。最適化手法は Adam, エポック数を 300 とし, 学習率は $10^{-2} \sim 10^{-6}$ の範囲で検証精度の増減に応じて 1/2 倍ずつ減衰させる。性能評価は Accuracy を用い, 300 エポックを 1 セットとして評価し 3 セットの内最高精度のモデルを Grad-CAM に用いる。チャンスレベルは男性 53 人, 女性 66 人の 55%である。

表 3.1 3DCNN モデルの構成表

No.	Details
1	Conv3D(8, 4) → BatchNormalization → ReLU → MaxPooling3D(2) →
2	Conv3D(16, 4) → BatchNormalization → ReLU → MaxPooling3D(2) →
3	Conv3D(32, 4) → BatchNormalization → ReLU →
4	GlobalAveragePooling3D → Dense(2, softmax)

3.3 機械学習タスク

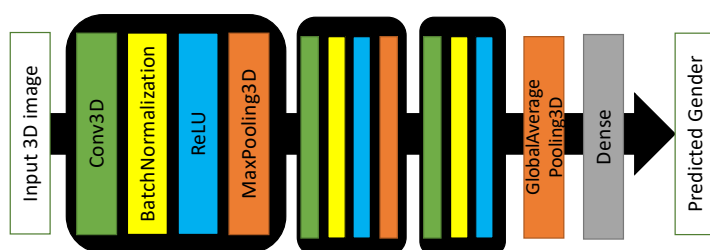


図 3.17 モデル構成図

第 4 章

結果

3.2 節で作成し，使用する 9 種類の画像を A. 線形レジストレーション（図 3.6），B. 線形レジストレーション後の脳領域抽出（図 3.10），C. 全脳セグメンテーション（図 3.13），D. 白質セグメンテーション（図 3.14），E. 灰白質セグメンテーション（図 3.15），F. 脳髄液セグメンテーション（図 3.16），G. 非線形レジストレーション（図 3.7），H. 非線形レジストレーションの変位（図 3.8），I. 非線形レジストレーション後の脳領域抽出（図 3.11）とする．構築した 3DCNN に各 A～I の画像を入力し，男女識別を行なった各精度を表 4.1 に示す．Acc1 から Acc3 に 1 セット目から 3 セット目の精度を示し，MaxAcc に各セットの最高精度を示す．

表 4.1 3DCNN モデルの各画像での識別結果

Image	A	B	C	D	E	F	G	H	I
Acc1	87	71	76	80	75	63	83	73	55
Acc2	69	67	74	82	71	66	80	72	55
Acc3	85	68	76	75	69	70	62	73	55
MaxAcc	87	71	76	82	75	70	83	73	55

4.1 Grad-CAM

各入力画像を Grad-CAM によって可視化を行う．重ね合わせは，被験者個人ごとの Grad-CAM による可視化ではなく，全被験者の画像の平均と全被験者の Grad-CAM の結果の平均の重ね合わせを行った．これは，個人ごとの差を目視で判断することは難しく，

4.1 Grad-CAM

Grad-CAM の結果を平均して確認する方を良しとしたためであり，それによって入力画像も平均をとり重ね合わせた。

3次元画像であるため2次元表示する際には，全スライスの中で特に寄与部位がわかりやすいスライスを抽出し，各画像と Grad-CAM によるヒートマップを重ね合わせた画像を並べて表示する．また，図 4.18 は変位画像ではなく，非線形レジストレーションが行われている部位の比較が行いやすい非線形レジストレーション画像を表示する．



図 4.1 画像 A スライス 41

図 4.2 画像 A スライス 30

図 4.3 画像 B スライス 45

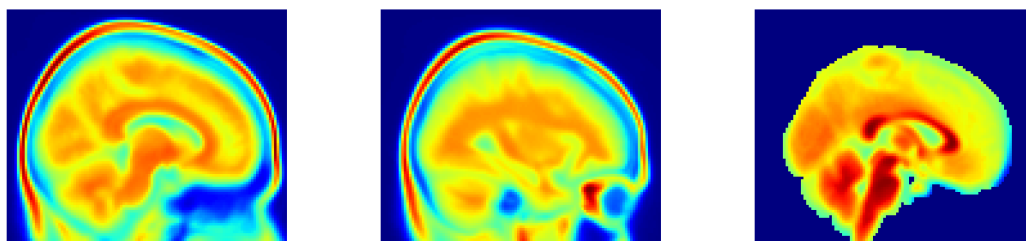


図 4.4 画像 A 小脳周辺

図 4.5 画像 A 視神経

図 4.6 画像 B 小脳周辺

4.1 Grad-CAM



図 4.7 画像 C スライス 70



図 4.8 画像 D スライス 67



図 4.9 画像 E スライス 72

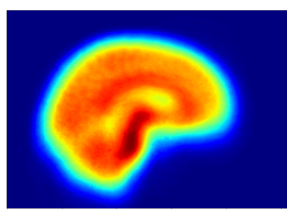


図 4.10 画像 C 小脳周辺

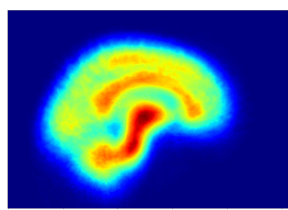


図 4.11 画像 D 小脳周辺

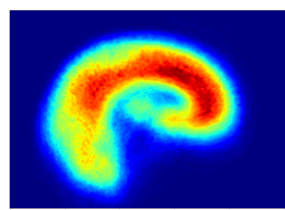


図 4.12 画像 E 全体



図 4.13 画像 F スライス 68



図 4.14 画像 F スライス 52

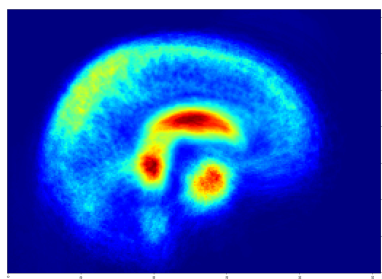


図 4.15 画像 F 視床下部周辺

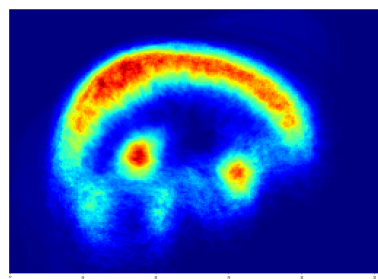


図 4.16 画像 F 上部

4.1 Grad-CAM

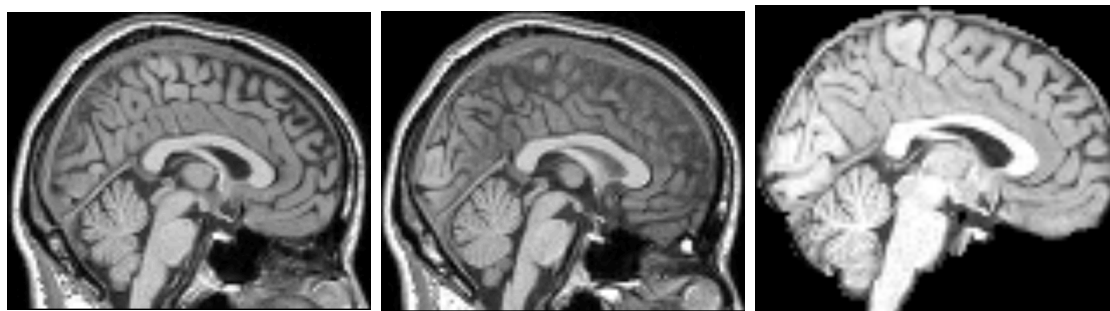


図 4.17 画像 G スライス 47 図 4.18 画像 G スライス 50 図 4.19 画像 I スライス 45

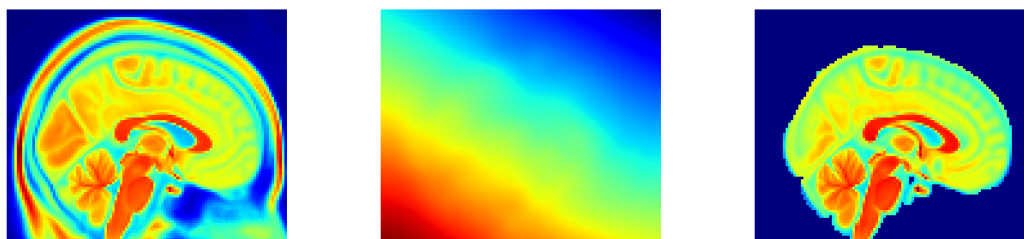


図 4.20 画像 G 小脳周辺

図 4.21 画像 H 後頭部

図 4.22 画像 I 小脳周辺

第 5 章

考察

各精度と Grad-CAM によるヒートマップにおける各画像に対する考察を記述する.

5.1 線形レジストレーション画像の考察

図 4.4 はスライス 41 の位置であり, この位置からは頭蓋, 小脳, 視床周辺, 脳梁が寄与部位としてあげられる. また, 図 4.5 はスライス 30 の位置であり, この位置からは視神経周辺も寄与部位としてあげられる.

文献 [3][6][4] において, 頭蓋あり画像は複数ある前処理方法の中でも 1 番精度が高いため, 男女識別を行う際に寄与していると考えられていた. そのため, Grad-CAM によって得られるヒートマップによって頭蓋の影響が説明された.

人間の顔面を用いた男女識別においては, 肌や頭髮の他に頭部の大きさが識別要因の一つに挙げられる. そのため, 頭蓋や顔の形の情報が残る画像においてはそれらが寄与すると考えられる.

5.2 線形レジストレーション画像の脳領域抽出の差の考察

図 3.6 と図 3.10 の差は, 頭蓋と脳領域以外のその他の領域が取り除かれたことである. 特に頭蓋の影響はこれまで考えられており, 5.1 節での考察から取り除くことで脳領域のみの識別精度となる. その中でも識別可能であるということ, Grad-CAM によるヒートマップにおいて図 4.6 ではスライス 45 の位置より, 小脳や視床周辺, 脳梁, 脳幹が寄与していることがわかった. これらの部位は, 5.1 節で挙げられた部位と同じような部位である.

5.3 脳抽出画像ごとの考察

図 3.13, 図 3.14, 図 3.15, 図 3.16 は全て脳領域のみの画像であり, 図 4.10 ではスライス 70 の位置, 図 4.11 ではスライス 86 の位置, 図 4.15 ではスライス 68 の位置からこれまでと同様な部位である小脳や視床周辺, 脳梁, 脳幹が寄与していることがわかった。

図 4.9 ではスライス 72 の位置にはこれまでの寄与部位が写っていないため, その他の部位が全体的に寄与している。

図 4.16 ではスライス 52 の位置には上部のみが写っているため寄与していると考ええる。

また, 可視化ではない説明性として脳領域のみの画像ごとの精度の違いによる比較は, 白質 > 全脳 \geq 灰白質 \geq 線形脳領域抽出 \geq 脳髄液 > 非線形脳領域抽出であり, 白質を含む画像が上位に多くあることから白質が男女識別において寄与していることが考えられる。

5.4 非線形レジストレーション画像の前処理の考察

図 3.7, 図 3.8, 図 3.11 は非線形レジストレーションによって脳の大きさにおける男女差を 1 番軽減している画像タイプである。そのため, 男女識別が行える場合の寄与部位は特に男女差がある部位となる可能性が高い。図 4.20 ではスライス 47 の位置より, これまでと同様な部位である頭蓋, 小脳, 視床周辺, 脳梁が寄与していることがわかった。

しかし, 図 4.21 ではスライス 50 より後頭部の脳領域以外の部分が寄与部位であるとわかった。図 3.8 は, その他の前処理画像と比べて特にどのような情報によって男女識別が行われているか未知であるが, Grad-CAM によって首の脂肪などの部位が寄与し, 脳領域による識別は行われていない可能性が高い。

図 4.19 ではスライス 45 の位置より, これまでと同様な部位である小脳や視床周辺, 脳梁, 脳幹に着目して識別を行おうとしていることがわかる。識別精度である 55% はチャンスレベルであり識別不能であるが, 脳領域抽出前後で注目領域は変わらず, 精度に差が出る理由として頭蓋やその他の領域が識別へ大きく寄与していると考えられる。そのため, 図 3.7 と図 3.8 において識別可能であり図 3.11 では識別不能であることから, 脳の大きさの差が

5.5 考察まとめ

少なく、脳領域のみの画像、図 3.11 には識別可能な情報がないと考えられる。

5.5 考察まとめ

文献 [1] において、前視床下部間質核 (interstitial nuclei of the anterior hypothalamus : INAH) の中で INAH-1 や INAH-3 は性的二型性について議論されている。また、脳梁の水平面積は、一般的に男性の脳が大きいとされるが、平均的に女性よりも大きい。さらに、脳梁の後端である脳梁膨大部においては選択的に女性の方が大きいと報告されている。議論がされている内容であるが、Grad-CAM によって頭蓋、視神経周辺、小脳、視床周辺、脳梁、脳幹、後頭部周辺の脂肪などの脳領域以外のその他の領域が寄与していることがわかり、男女識別可能であるモデルにおいてその部位に着目して識別を行なっているため、特に視床周辺や脳梁は議論されている部位と同じ部位が示された。

しかし近年、脳構造に性的二型性はないと結論付ける報告もあり、文献 [1] においてもヒトの脳構造の性的二型性の最も信頼できる結論は差はあまりないということと記述されている。これらより、CNN での判別結果には上記の部位が寄与していることが分かるが、CNN はそれらの画像の何らかのパターンに違いがあることが考えられる。

5.6 その他の考察と今後の展望

医療分野の観点からでは、今回作成された各前処理画像の 3D Grad-CAM について専門家と話し合うことがあまりなかったため、この結果に対する意見をもらい、寄与部位の考察が行われるとより良いと考える。その場合、脳地図を用いたより詳細な部位の分割を行うことで、特定の部位に関する寄与度について議論することが可能であると考え。

また、機械学習の観点では、本研究で用いた 3DCNN モデルは各画像ごとに入力サイズを合わせ、9 つの学習済みモデルを作成するように設定したが、全ての画像に対して同じスライスにリサイズを行うことができるため、一つのモデルに対して複数の前処理画像を入力して識別を行うことで一般的な機械学習による画像分類モデルを作成することができる。こ

5.6 その他の考察と今後の展望

れにより、より複雑なタスク解決となり、それでも識別可能である場合の情報はより価値が高いものとなる可能性がある。本研究の 3DCNN モデルは近年の一般的な深層モデルと比較すると層数は少ないため、複雑なタスクは解決不能である可能性はあるが、アルゴリズムや実験環境の向上により、より多層なモデルによる学習は行えるため、新たに実験を試すべきである。

さらに、本研究では可視化手法として Grad-CAM を実装したが、様々な CAM 手法が近年提案されており、より良い CAM 手法による可視化が行われると寄与部位の特定はわかりやすくなり、より良い説明性となると考える。しかし、CAM 手法ごとに寄与部位に差が生まれることやどの CAM 手法を用いてもあまり差がみられないこともあるため、単なる可視化以外の方法による説明性を求めることが重要であると考え。特に、Adversarial Attack と呼ばれる AI に対する攻撃の研究が進んでおり、可視化だけを判断根拠とすることが難しくなっていく可能性があるため様々な手法による説明性が重要となっていくことが考えられる。

第 6 章

結論

本研究は 9 種類の前処理画像の男女識別を行い、精度の違いや Grad-CAM による可視化によって男女識別における説明性を求めた。3DCNN を用いた精度によってセグメンテーションごとの結果の違いから白質が寄与部位として考えられる。また、Grad-CAM による可視化によって頭蓋、小脳、視床周辺、脳梁といった部位が主に寄与することがわかった。文献 [1] においても INAH-1 や INAH-3 や脳梁水平面積、脳梁膨大部など男女差や性的二型性が考えられる部位であるため、同じ部位が可視化によって寄与していることがわかった。しかし、ヒトの脳構造には男女に差はないと結論づけられることもあり、CNN での判別結果には上記の部位が寄与していることが分かり、それらの画像の何らかのパターンに違いがあることが考えられる。

謝辞

本研究を進めるにあたり，高知工科大学情報学群 吉田真一教授には研究のご指導いただき心から感謝いたします。高知工科大学情報学群 繁榊博昭教授ならびに，高知工科大学情報学群 中原潔教授には副査ををしていただき，深く感謝いたします。

参考文献

- [1] Mark Bear, Barry Connors, Michael A. Paradiso : Neuroscience:Exploring the Brain, Enhanced Edition, Jones & Bartlett Learning, (2016)
- [2] Laura S. Allen, Roger A. Gorski : Sexual dimorphism of the anterior commissure and massa intermedia of the human brain, JOURNAL OF COMPARATIVE NEUROLOGY, 312(1), 97/104 (1991)
- [3] Takenobu Sasatani, Kaechang Park, Shinichi Yoshida, Yuki Shinomiya : Prediction of human characteristics from brain structural mri using cnn, International Symposium on Frontier Technology(ISFT-2019), 529/533 (2019)
- [4] 西森涉, “MRI 脳画像の前処理方法の違いによる 男女識別精度の比較,” 高知工科大学 学士学位論文, 2020.
- [5] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra : Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2921/2929 (2016)
- [6] 藤沢元, “MRI 脳画像からの男女識別 CNN モデルの構成の検討,” 高知工科大学 学士学位論文, 2020.
- [7] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba : Learning Deep Features for Discriminative Localization, IEEE International Conference on Computer Vision (ICCV), 618/626 (2017)
- [8] Biomedical Image Analysis Group, Ixi-dataset.
- [9] Mark W. Woolrich, Saad Jbabdi, Brian Patenaude, Michael Chappell Salima Makni, Timothy Behrens, Christian Beckmann, Mark Jenkinson, Stephen M. Smith : Bayesian analysis of neuroimaging data in FSL, NeuroImage, 45:S173-

86 (2009)

- [10] Mark Jenkinson, Stephen M. Smith : A global optimisation method for robust affine registration of brain images, *Medical Image Analysis*, 5(2), 143/156 (2001)
- [11] Mark Jenkinson, Peter R. Bannister, Michael Brady, Stephen M. Smith : Improved optimisation for the robust and accurate linear registration and motion correction of brain images, *NeuroImage*, 17(2), 825/841 (2002)
- [12] Jesper L. R. Andersson, Mark Jenkinson, Stephen Smith : Non-linear registration, aka spatial normalisation, FMRIB technical report TR07JA2 (2007)
- [13] Yongyue Zhang, Michael Brady, Stephen Smith : Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm, *IEEE Trans Med Imag*, 20(1), 45/57 (2001)