

深層学習による脳血流SPECT統計解析画像の認知症分類

Dementia disease classification of the statistical analysis images of cerebral blood flow SPECT using deep learning

山本 泰司¹⁾, 白井 匡人²⁾, 勝部 敬³⁾, 吉廻 毅⁴⁾, 上部 星雄⁵⁾, 宮原 善徳⁵⁾, 北垣 一⁶⁾

1) 島根大学医学部 放射線医学講座

2) 島根大学 学術研究院理工学系 (助教)

3) 島根大学医学部 放射線医学講座 (助教)

4) 島根大学医学部 放射線医学講座 (准教授)

5) 島根大学医学部附属病院 放射線部

6) 島根大学医学部 放射線医学講座 (教授) (現 神戸市立西神戸医療センター)

Key words: GoogLeNet, AlexNet, Artificial intelligence, Grad-CAM, 3D-SSP

【Abstract】

Using the original deep learning model, we attempted to classify the following diseases from the cerebral blood flow SPECT 3D-SSP images: AD, DLB, and normal cognition. Then, it was compared with the transfer learning of GoogLeNet and AlexNet reported in the past. Furthermore, we tried to visualize accuracy of each analysis and feature image using the Grad-CAM technique. The accuracy was low for AlexNet in AD and for GoogLeNet in DLB. Stable results were obtained in the original model. The size of the feature area captured by each analysis method from the Grad-CAM image differed. The original model captured the area of reduced blood flow in a localized manner.

【要旨】

オリジナルモデルの深層学習を用いて, AD, DLB, normal cognitionにおける脳血流SPECT 3D-SSP画像の疾患分類を試みた。そして過去に報告したAlexNetやGoogLeNetの転移学習との比較も行った。さらにGrad-CAMにより各解析精度の視覚化を試みた。AlexNetではADで, GoogLeNetではDLBで精度が低く, オリジナルモデルでは安定した結果が得られた。Grad-CAMの画像から各モデルで捉えている特徴領域の位置や大きさに違いがあり, オリジナルモデルが血流低下領域を正確に捉えていた。

緒 言

近年, 放射線画像の分野でも人工知能 (Artificial intelligence: AI) の臨床応用が急速に進み, 論文での報告も多い。代表的なものでは頭部CTでの脳出血検出¹⁾ やMRAからの未破裂動脈瘤検出²⁾, Computed tomography (CT) やMagnetic resonance imaging

(MRI) でのノイズ低減技術³⁾ や再構成技術⁴⁾ への応用など多岐にわたる。2019年にはMagnetic resonance angiography (MRA) からの脳動脈瘤を検出するソフトが国内初の薬事承認を得る段階までできた。核医学分野に目を向けると, Positron emission tomography (PET) 画像からCT画像を生成し減弱補正に利用する試みが報告されている⁵⁾。Iizukaらは, 脳血流Single photon emission computed tomography (SPECT) からのThree-dimensional stereotactic surface projection (3D-SSP)⁶⁾ 画像を深層学習 (Deep Learning: DL) から認知症分類を行い, 特異的血流低下領域の視覚化から妥当性を報告している⁷⁾。著者らのグループも, 脳血流SPECT画像からの3D-SSP画像を用いてオリジナルの深層学習モデル (O.N) から認知症の分類を試みてきた⁸⁾。さらに上部らのシミュレーション学習データの考え⁹⁾ を応用し, 認知度の高いAlexNet (A.N)¹⁰⁾ やGoogLeNet (G.N)¹¹⁾ で認知症の分類も検証・報告している^{12,13)}。今回, O.NでもA.NやG.Nの報告同様のデータ群から識別精度を算出し比較評価した。さらに畳み込みニューラルネットワーク (Convolutional neural network: CNN) の可視化技術であるGradient-

YAMAMOTO Yasushi¹⁾, SHIRAI Masato²⁾, KATSUBE Takashi³⁾, YOSHIKAKO Takeshi⁴⁾, UWABE Hoshio⁵⁾, MIYAHARA Yoshinori⁵⁾, KITAGAKI Hajime⁶⁾

- 1) Department of Radiology, Faculty of Medicine, Shimane University
- 2) Institute of Science and Engineering, Academic Assembly, Shimane University
- 3) Department of Radiology, Faculty of Medicine, Shimane University
- 4) Department of Radiology, Faculty of Medicine, Shimane University
- 5) Department of Radiology, Shimane University Hospital
- 6) Department of Radiology, Faculty of Medicine, Shimane University

Received July 7, 2021; accepted November 7, 2021

weighted Class Activation Mapping (Grad-CAM)¹⁴⁾を用いて、O.NがG.NやA.Nとどのように異なるかをGrad-CAMヒートマップ画像から比較した。

1. 方法

1-1 使用モデル (A.N, G.N, O.N)

O.Nは、4層のCNNと3層のPooling層で構成した。詳細な設定についてはFig.1に示す。解析にはMatlab 2019b (Mathworks社), Deep learning toolbox, Image Processing toolbox, Computer vision toolbox, Parallel computing toolboxを使用した。今回の比較対象であるG.NやA.Nの詳細については参考文献^{10,11)}で記述した。

1-2 画像データの作成

解析に使用した画像はN-Isopropy [¹²³I] -p-Iodoamphetamine (¹²³I-IMP) 167MBq, 脳血流検査から得た3D-SSP Z-score map, 収集装置 Discovery NM/CT 670pro, 画像再構成 3D-OSEM法 (subsets 10, iterations 6), 散乱線補正 Dual energy window 法, 減弱補正 CT based

Attenuation correction法を用いた。3D-SSP解析には、島根大学ノーマルデータベースを使用した。3D-SSP標準出力画像Fig.2 (C) は正面・後面・左右側面・左右内側面の脳表示8方向, 890×660ピクセルの画像となるが, G.NやA.Nとの比較評価の目的から画像サイズはA.Nと同じ224×224ピクセル (G.Nは227×227ピクセル) とした。その際, ターゲットとなる血流低下領域 (Z-score 描出領域) での画像縮小による検出精度の低下を避ける目的から, 8方向の脳表面画像から, Fig.2 (D) に示すようにAlzheimer's disease (AD) やDementia with Lewy bodies (DLB) での特異的血流低下領域の識別に重要と考えられる左右の外側面 (R.Lat・L.Lat), 左右の内側面 (R.Med・L.Med) の4方向を選択, 使用データ全てで同じ座標上に画像配置されるように再作成した。解析には学習用データ (Training), Trainingの精度評価用データ (Validation), クラス分類を行う臨床データ (Test) の3群が必要となる。Trainingについては, 上部らの報告同様⁸⁾にAD・DLBの特異的血流低下を模擬したSPECT画像のシミュレーションデータから作成した3D-SSP画像を用いることとした。シミュレーションデータのベースとなるSPECTデータは, 2013年にノーマルデータベース用に収集した¹²³I-IMP健常者データ28

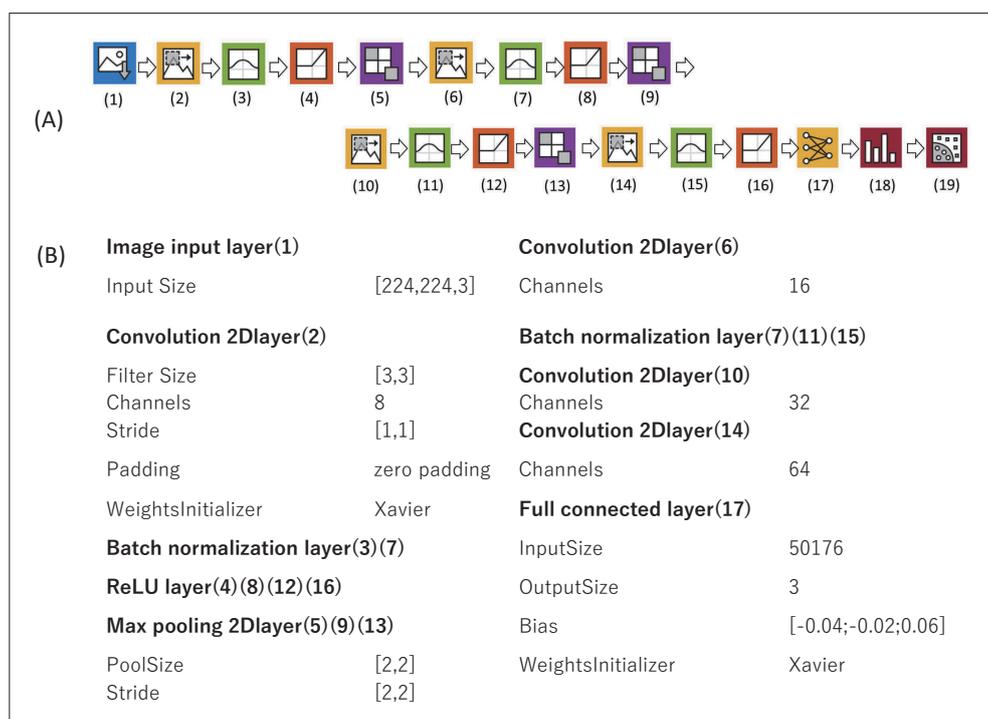


Fig.1 Configuration diagram and parameters of the original model.

It comprises 4 convolution layers and 3 pooling layers. (A: Configuration diagram created with deep network designer in matlab.)

例（男性15人：67.7±6.2歳，女性13人：68.1±5.7歳）を使用し，Statistical parametric mapping 8 (SPM8)¹⁵⁾でMontreal neurological institute (MNI) 脳座標に変換後，Voxel-based analysis-stereotactic extraction estimation (vbSEE)¹⁶⁾ Level3の神経学的Volume of interest (VOI) 領域を利用し，ADでは左右の頭頂葉領域に位置する角回・上頭頂小葉・下頭頂小葉・縁上回に加えて，後部帯状回・楔前部の6領域，DLBは，後頭葉領域に位置する上後頭回・中後頭回・下後頭回に加え，一次視覚野に当たる楔部・舌状回の5領域を血流低下領域としてマスク画像を作成，各マスク内のカウントを35，30，25，20%と段階的に低下させたマスクデータをベース画像に掛け合わせることで，さまざまなパターンの認知症SPECTデータを作成した．それらの3D-SSP解析から得た各血流低下のZ-score mapより，Z-scoreの閾値設定 (Fig.2 (C) ②)，正規化部位設定 (全脳：GLB，視床：THL，橋：PNS，小脳：CBL③)を各画像で可変とすることでAD・DLB各341例のシミュレーションZ-score mapを作成した．(D)は①に当

たるADの特異的血流低下領域のカウントを35%減とし，4方向の抜粋で作成．(E)は特異的血流低下領域20%のカウント低下，(F)は(E)の参照部位をGLBからTHLに変更，(G)は(E)の閾値設定を変化させた各Trainingの一例を示す．以上の方法で健常者28症例からAD・DLBシミュレーションTrainingを341例に増やした．ただし，参照部位については認知症診断を得意とする経験20年以上の放射線科専門医がAD・DLBと判断できないものは除外している．詳細なデータ内訳については文献¹⁰⁾に記載した．Fig.2 (H)はDLB，(I)はN.CのTrainingの一例を示す．Testは，2020年3月から過去にさかのぼり¹²³I-IMP脳血流SPECTを施行した患者データのZ-score mapから，放射線科専門医によりAD・DLB・NCと判断された各20症例を採用，さらに正規化部位から最適と判断された2画像を採用することで40症例とした．患者背景は，ADでAge (y)：72.3±3.8，M/F：15/5，DLBで80.5±7.3，M/F：10/10，NCで67.8±15.4，M/F：6/14である．ValidationはAD・DLB・NCのTrainingから参照部位GLB，閾値下限設定1.0，上限

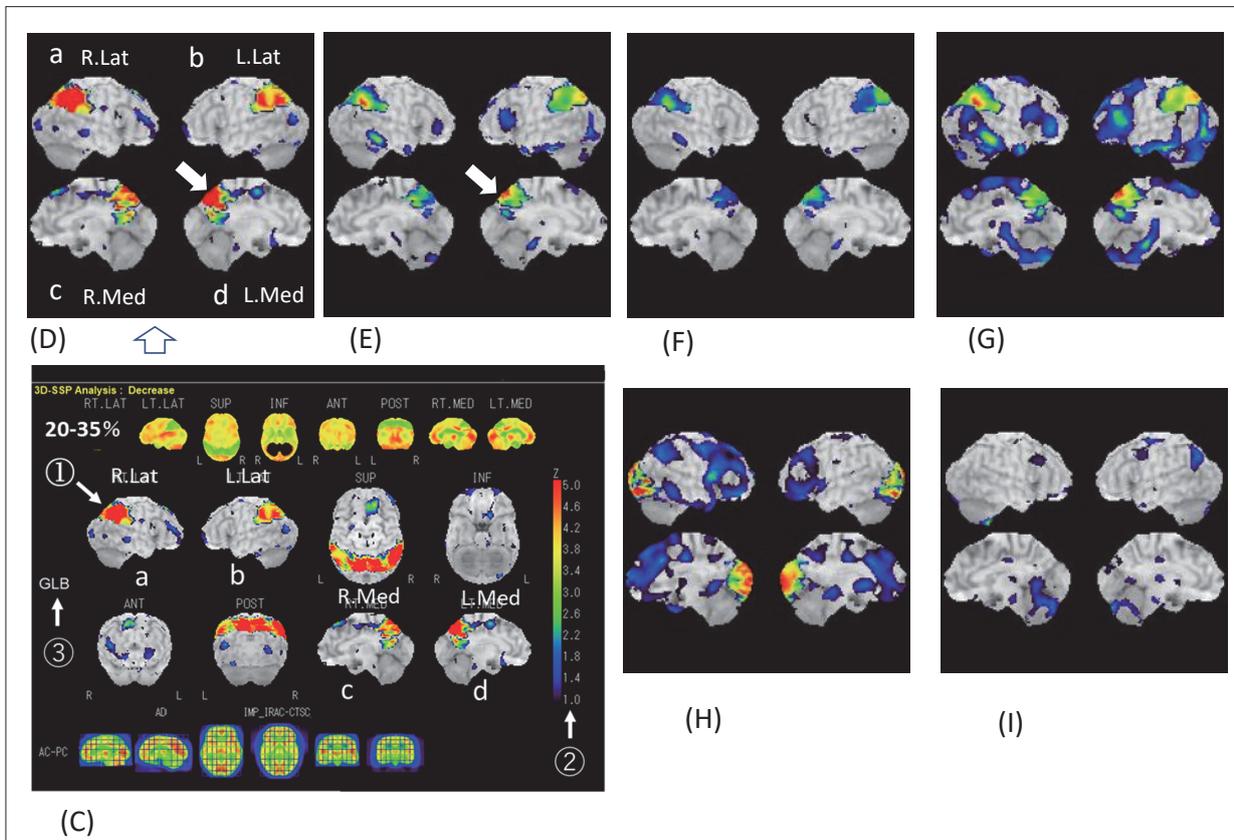


Fig.2 AD and DLB simulation images and the NC image used for the training.

(C) Original image of 3D-SSP, (D) AD simulation data created from C (⇔：-35%)，(E) Simulation image of AD (⇔：-20%)，(F) Simulation image of AD (Changed reference region from GLB to THL)，(G) Simulation image of AD (Changed Z-score threshold)，(H) Simulation image of DLB，(I) NC image.

設定5.0の28例を固定解析で使用した。

1-3 評価方法

AD・DLB・NCの判断について、放射線科専門医の診断結果を正解として、正解データ数の割合である正解率 (Accuracy), 全陽性のうち正解した割合 (精度) を示す適合率 (Precision), 感度・検出力を示す再現率 (Recall), 総合的な指標値となるF値 (F-measure) で評価した。真陽性 True positive : TP, 偽陽性 False positive : FP, 偽陰性 False negative : FN, 真陰性 True negative : TN で示すと、 $Accuracy = TP + TN / \text{総数}$, $Precision = TP / TP + FP$, $Recall = TP / TP + FN$, $F\text{-measure} = 2 \times (Precision \times Recall) / (Precision + Recall)$ となり、各指標値は20回の平均値と標準偏差で比較評価した。またどの血流低下領域からクラス分類が決定されたのかは、Grad-CAMによるヒートマップの画像から視覚的に評価した。なお、本研究内容は島根大学医学部医の倫理委員会の承認を得て行っている。

2. 結果

Fig.3に、過去の報告から引用^{10,11)}したG.NとA.Nでの病型別RecallにO.Nの結果を加えたグラフを示す。Fig.3とTable 1からADのRecallはO.N (0.83 ± 0.09) > G.N (0.82 ± 0.12) > A.N (0.52 ± 0.12), DLBではO.N (0.85 ± 0.05) > A.N (0.74 ± 0.09) > G.N (0.42 ± 0.12), NCではO.N (0.86 ± 0.06) > G.N (0.78 ± 0.08) > A.N (0.76 ± 0.06) であり、ADのG.N, O.Nで有意差はなかったが、他は有

意差をもってO.Nが高値を示し、O.NのTraining, Validationでは多くの指標値が1.0であった。Fig.4は、参考文献¹¹⁾から引用したA.N・G.NのGrad-CAMヒートマップ画像に、O.Nの結果を加えた。解析モデルによりヒートマップされている領域の大きさと分布形状は異なるが、(J, K, L)の各モデルのADでは両側後部帯状回・楔前部・側頭・頭頂葉、(M, N, O)のDLBでは両側側頭・後頭葉にヒートマップされてい

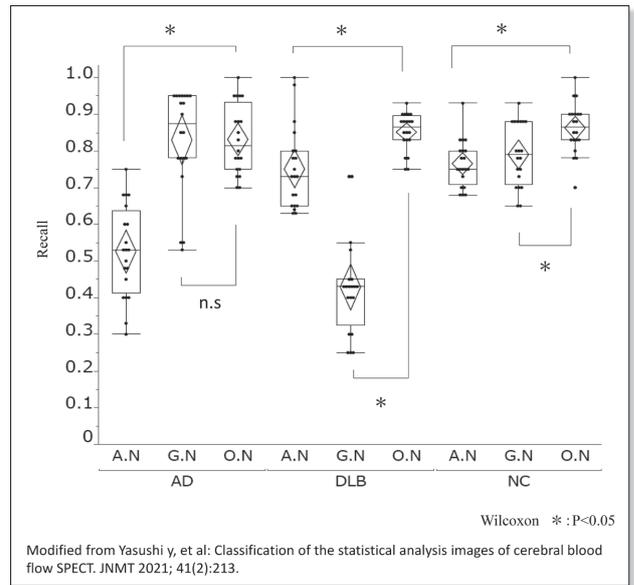


Fig.3 The graph shows the recall in each model by disease type.

In the AD group, G.N and O.N showed high values, and there was no significant difference in them. In the DLB group, O.N showed the highest value, which was significantly different from those of G.N and A.N. Additionally, O.N was high in NC, and it differed from A.N and G.N.

Table 1 Recall, Precision, and F-measure in O.N analysis method were calculated separately for AD, DLB, and NC.

Data type	Evaluation method	AD	DLB	NC	Accuracy
Test	Recall	0.83±0.09	0.85±0.05	0.86±0.06	0.84±0.02
	Precision	0.84±0.05	0.83±0.06	0.87±0.06	
	F-measure	0.83±0.04	0.83±0.04	0.86±0.03	
Training	Recall	1.00	1.00	1.00	1.00
	Precision	1.00	1.00	1.00	
	F-measure	1.00	1.00	1.00	
Validation	Recall	1.00	1.00	0.98±0.02	0.99±0.01
	Precision	0.98±0.02	1.00	1.00	
	F-measure	0.99±0.01	1.00	1.00	

る。(P, Q, R)のNCでは血流低下のない領域をヒートマップした。

3. 考 察

最初に本研究におけるO.Nのシミュレーション学習データについて、Table 1でのTrainingのAccuracyは全病型1.0であり、AD・DLBは異なった画像パターン(異なった特徴量画像)で構成されていることになり、Validationの結果からも典型的な血流低下パターンの3D-SSP画像であれば正確に分類可能であり、本研究での学習データ群として問題ないと判断した。Fig.4 (J, K, L)のADヒートマップからA.NやG.NではZ-scoreの高い領域は正しく捉えているが、ヒートマップが広い領域に及び、(J) R.MedのようにDLBでの特異的低下領域である後頭葉にまで広がっているケースや、(K) L.Latのように頭頂葉でのヒートマップが弱いケースがあることから、Z-score描出領域がADやDLBの特異的領域に限られ、かつ高値のZ-scoreであることが正確に分類できる条件

であることが予測できる。Fig.4 (M, N, O)のDLBでの後頭葉について注目するとO.N・A.N・G.N全てでR.Medで強くヒートマップされ、L.Medは弱いことから、(K) L.Latと同様にZ-scoreの「高い」「低い」を特徴の一つとして認識されていることが示唆される。またG.Nの(N)では左右の後頭葉が1つの大きな領域として捉えられていることから、ヒートマップの中心が画像のない部分に存在している。他にも同じ傾向の症例が存在し、解析モデルそのままの設定条件の転移学習ではZ-scoreの領域に局限した特徴量検出はできなかった。Fig.4 (P, Q, R)のNCについては特異的血流低下領域が存在しないことから、画像全体がヒートマップされ特徴領域を絞り切れない結果と判断できる。Fig.5 (S, U, X)は各モデルで放射線科専門医の診断と異なるクラス分類となった画像を示す、(S)はA.NでADをDLBと分類した例であるが、ヒートマップがDLBの血流低下領域である後頭葉(⇒)まで広がったことで識別を誤った。そこでA.NのAD Training画像を調べると、(M)に示すように作成した血流低下領域より下方のZ-scoreにヒート

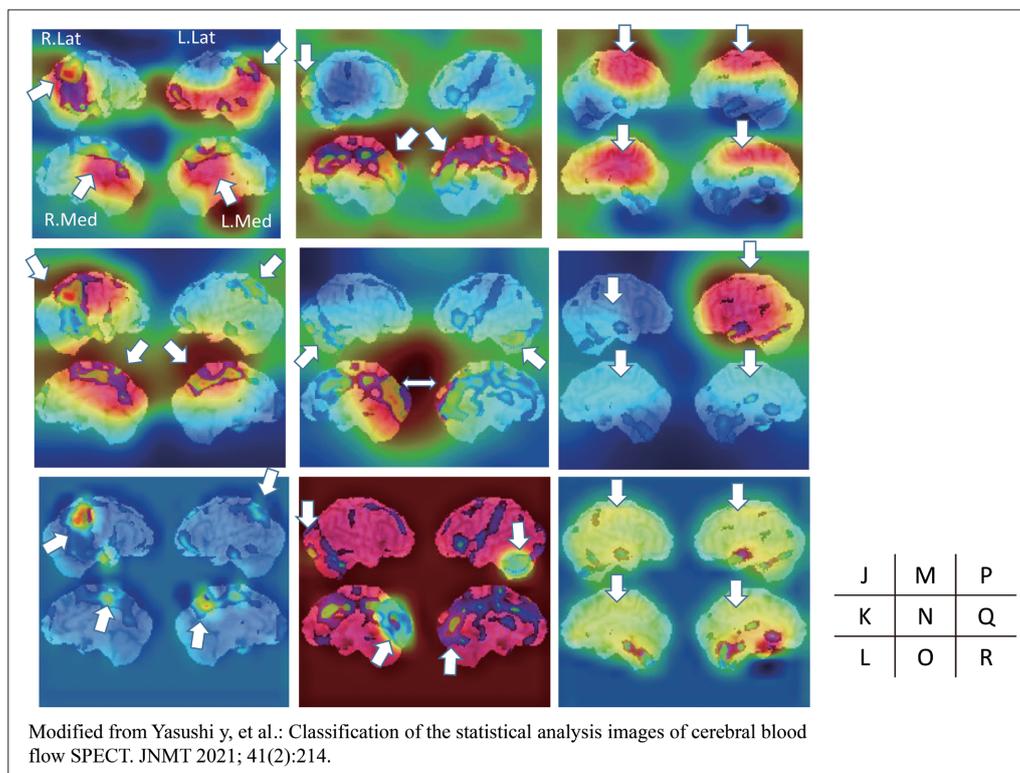


Fig.4 Heat map of A.N, G.N, and O.N analysis images classified as AD, DLB, and NC by matching with the radiologist's reading results.

Heat map of images classified as (J) AD, (M) DLB, and (P) NC by A.N. Heat map of images in which the same image is classified as (K) AD, (N) DLB, and (Q) NC by G.N. Heat map of images in which the same image is classified as (L) AD, (O) DLB, and (R) NC by O.N. (⇒) indicates a heat map region that correctly recognizes the specific blood flow reduction region in AD and DLB, and NC heat map region without blood flow reduction.

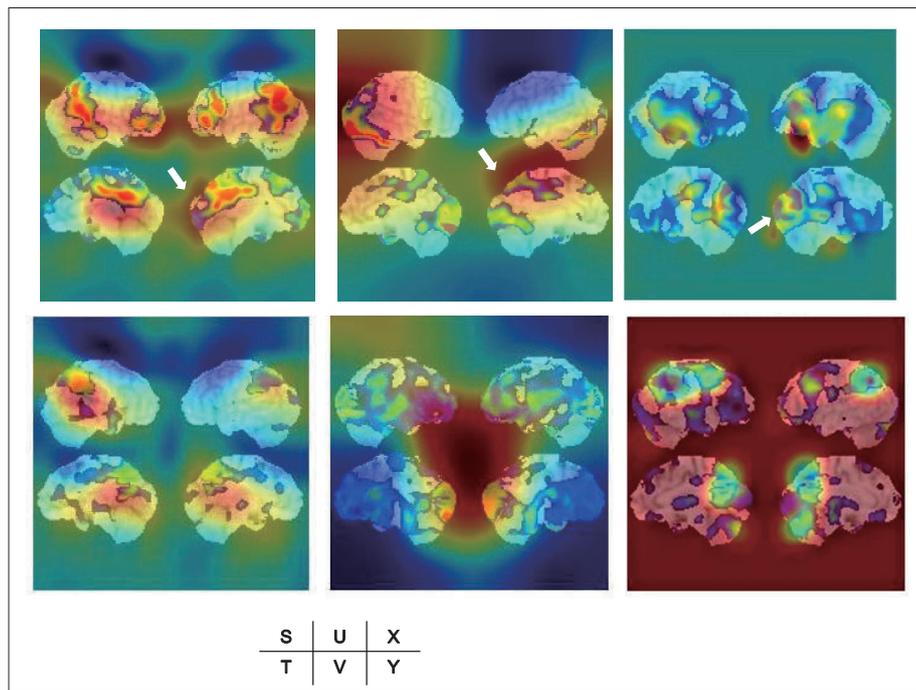


Fig.5 Heat map of images in which the radiologist's reading results and classification results did not match. (S) In A.N, AD was erroneously classified as DLB. (U) In G.N, DLB was erroneously classified as AD. (X) In O.N, DLB was erroneously classified as AD. (⇒) indicates the heat map area that caused the erroneous classification of AD and DLB. (T)AD in A.N (V) DLB in G.N (Y) AD in O.N, heat map area of each train data.

マップされた画像も存在し、DLBとの区別化が可能となるようにADの特徴を持ちつつ、後頭葉にも低いZ-scoreが存在する学習データの追加も再検討する必要がある。(U)はG.NでDLBをADと分類した例であるが、後頭葉のヒートマップが弱く、帯状回領域を強く捉えている。(V)はG.NのDLB Trainingの一例を示す。Fig.4 (N)でも同様にR.Lat, L.Latの後頭葉を1つの特徴として広くヒートマップされたことから、左右対称ではなく片側のZ-scoreが小さいデータもTrainingに加える必要がある。(X)はO.NでADをDLBと分類した例であるが、帯状回の血流低下が強く、放射線科専門医は高Z-scoreの影響と判断できたがO.Nでは判断できなかった例である。(Y)はO.NのAD Trainingの一例を示すが、O.NではわずかなZ-scoreや位置の違いも識別はできることから、(X)のような学習データの追加で対応が可能と推測する。

一般的に、G.Nのように層が深くなるモデルでは学習対象のパラメーターが多いことから、学習データの数を増やして検証する必要がある。A.NについてはO.Nに比較して畳み込み層が1層多いのみであるが、それぞれの畳み込み層が持つチャンネル数が多いことから、G.N同様に多くの学習データが必要である。また正規化処理でLocal Response Normalizationを

使用しているが、O.NではBatch Normalizationにすることで内部の変数分布の変化を小さく抑えることが可能になり、学習時の結果をより正確に推定結果に反映できる。この研究のように、学習データの特徴のバラツキが小さい、クラス分類数が少ない、画像全体としての形状差がない、異なるのはZ-scoreの分布形状とZ-scoreのみというケースの場合、O.Nのようなシンプルなモデルでも分類が可能であった。ただし、学習データを充実させ、各モデルに適したパラメーターを使用すれば、結果は異なることも考えられる。

4. 結 語

深層学習を用いて、脳血流SPECT・3D-SSP画像の認知症の分類を試みた。学習データはシミュレーションから作成した341データ、症例はAD・DLB・異常なしの各40データ、解析モデルは4層のCNN、3層のPooling層を持つO.Nを使用。放射線科専門医の診断に対し8割以上の精度で分類可能であった。われわれが過去に報告したA.N・G.Nの転移学習より良い結果であった。Grad-CAMによる特徴量領域の視覚化と比較すると、O.NはA.NやG.Nより限局した特異的血流低下領域を捉えていた。今回のように、同じ種類

の画像, 分類数が少ないケースでは層の少ないD.Lでも対応可能であった. しかし, さらに多くの学習データを準備して各モデルを比較することも必要である.

利益相反

筆頭著者および共著者が開示すべき利益相反はない.

表の説明

Table 1 O.NでのRecall・Precision・F-measureの値をAD・DLB・NCに分けて算出.

図の説明

- Fig.1 オリジナルモデルの構成 (A) とパラメーター (B). 4層の畳み込み層と3層のプーリング層で構成. (A: Matlabのディープネットワークデザイナーで作製した図).
- Fig.2 学習に使用したADとDLBのシミュレーション画像およびNC画像の1例を示す. (C) 3D-SSPのオリジナル画像, (D) Cから作成したADのシミュレーションデータ (⇒: -35%), (E) ADのシミュレーションデータ (⇒: -20%), (F) ADのシミュレーションデータ (参照領域をGLNからTHLへ変更), (G) ADのシミュレーションデータ (Z-score閾値変更), (H) DLBのシミュレーションデータ, (I) NCのデータ.
- Fig.3 各モデルでの適合率 (Recall) を病型別に示す. (山本泰司, 他: 転移学習を応用した脳血流SPECT統計解析画像の識別. 核医学技術, 41 (2): 213, 2021. より引用一部改変). AD群ではG.NとO.Nで高値を示し有意差がなかった. DLB群ではO.Nが最も高値を示し, G.NやA.Nと有意差があった. NCでもO.Nが高値であり, A.NやG.Nに対して有意差があった.
- Fig.4 放射線科専門医の読影結果と一致してAD・DLB・NCと分類されたA.N・G.N・O.N解析画像のヒートマップ (山本泰司, 他: 転移学習を応用した脳血流SPECT統計解析画像の識別. 核医学技術, 41 (2): 214, 2021. より引用一部改変). A.Nで (J) AD, (M) DLB, (P) NCと分類された画像のヒートマップ (⇒). 同一画像をGNで (K) AD, (N) DLB, (Q) NCと分類された画像のヒートマップ (⇒). 同一画像をO.Nで (L) AD, (O) DLB, (R) NCと分類された画像のヒートマップ (⇒).
- Fig.5 放射線科専門医の読影結果と分類結果が一致しなかった画像のヒートマップを示す. (S) A.NでADを誤ってDLBと分類, (U) G.NでDLBをADと分類, (X) O.NでADをDLBと分類. (⇒)はAD・DLBの識別を誤った原因と予測できるヒートマップ領域. (T)はA.NのAD, (V)はG.NのDLB, (Y)はO.NのADの各Trainingデータ.

参考文献

- 1) Chilamkurthy S, Ghosh R, Tanamala S, et al.: Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*, 392 (10162): 2388-2396, 2018.
- 2) Ueda D, Yamamoto A, Nishimori M, et al.: Deep Learning for MR Angiography: Automated Detection of Cerebral Aneurysms. *Radiology*, 290(1): 187-194, 2019.
- 3) Tatsugami F, Higaki T, Nakamura Y, et al.: Deep learning-based image restoration algorithm for coronary CT angiography. *Eur Radiol*, 29(10): 5322-5329.
- 4) Ryu K, Nam Y, Gho SM, et al.: Data-driven synthetic MRI FLAIR artifact correction via deep neural network. *J Magn Reson Imaging*, 50(5): 1413-1423, 2019.
- 5) 福井亮平, 藤井進, 二宮宏樹, 他: 深層学習を用いたPET画像の減弱補正を目的とした疑似CT画像の作成. *日放技学誌*, 76(11): 1152-1162, 2020.
- 6) Minoshima S, Robert A. Koeppe M. A, et al.: Anatomic standardization; Linear scaling and nonlinear warping of functional brain images. *J Nucl Med*, 35(9): 1528-1537, 1994.
- 7) Iizuka T, Fukasawa M, Kameyama M: Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. *Sci Rep*, 9(1): 8944, 2019.
- 8) 山本泰司, 上部星雄, 矢田伸広, 他: 人工知能を利用した脳血流SPECT統計解析画像の識別. *核医学技術*, 40 (4): 407-412, 2020.
- 9) 上部星雄, 山本泰司, 矢田伸広, 他: シミュレーションデータを用いたディープラーニングによる脳血流SPECT統計解析画像の血流低下パターンの分類精度. *日放技学誌*, 77 (6): 581-588, 2021.
- 10) Krizhevsky A, Sutskever I, Hinton G: ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the Acm*, 60(6): 84-90, 2017.
- 11) Szegedy C, Liu W, Jia Y, et al.: Going Deeper with Convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7-12, 2015.
- 12) 山本泰司, 上部星雄, 矢田伸広, 他: GoogLeNetを利用した脳血流SPECT統計解析画像の識別. *核医学技術*, 41 (2): 204-209, 2021.
- 13) 山本泰司, 上部星雄, 矢田伸広, 他: 転移学習を応用した脳血流SPECT統計解析画像の識別. *核医学技術*, 41 (2): 210-215, 2021.
- 14) Selvaraju R, Cogswell M, Das A, et al.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, 128(2): 336-359, 2020.
- 15) 織田圭一: 核医学における臨床解析ソフトウェアの基礎と応用 1-3 SPM. *放射線医療技術学叢書*(28), 16-22, 日本放射線技術学会, 2011.
- 16) Uruma G, Hashimoto K, Abo M.: A new method for evaluation of mild traumatic brain injury with neuropsychological impairment using statistical imaging analysis for Tc-ECD SPECT. *Annals of Nuclear Medicine*, 27(3): 187-202, 2013.