

転移学習を応用した脳血流 SPECT 統計解析画像の識別

山本 泰司¹⁾ 上部 星雄²⁾ 矢田 伸広²⁾
 勝部 敬¹⁾ 吉廻 毅¹⁾ 北垣 一¹⁾

1) 島根大学医学部 放射線医学講座

2) 島根大学医学部附属病院 放射線部

Classification of the Statistical Analysis Images of Cerebral Blood Flow SPECT by Using the Transfer Learning.

Yasushi YAMAMOTO, Hoshio UWABE, Nobuhiro YADA,
Takashi KATSUBE, Takeshi YOSHIZAKO, Hajime KITAGAKI

1) Department of Radiology, Faculty of Medicine, Shimane University

2) Department of Radiology, Shimane University Hospital

(article received : Oct 20, 2020)

Key words : Artificial Intelligence, 3D-SSP, GoogLeNet, CNN, AlexNet,

1. 緒言 (はじめに)

我々は深層学習を応用し、核医学脳血流 Single photon emission computed tomography (SPECT) の統計解析画像からアルツハイマー型認知症 (Alzheimer's disease : AD), レビー小体型認知症 (Dementia with Lewy bodies : DLB), 認知症所見なし (Normal cognition : NC) の識別が可能であるかを本誌にて報告した^{1,2)}。Iizuka³⁾らは統計解析画像である Three-dimensional stereotactic surface projection (3D-SSP)⁴⁾からの脳表画像を利用して自ら構築した4層のネットワークを用い AD, DLB, NC の特異的血流低下領域の違いを、Gradient-weighted Class Activation Mapping (Grad-CAM)⁵⁾と呼ばれる畳み込みニューラルネットワーク (Convolutional neural network : CNN) の可視化技術を利用して、Artificial intelligence (AI) の妥当性を報告している。

我々による過去の GoogLeNet⁶⁾の転移学習の報告から²⁾ AD, NC については8割程度の正解率を得たが DLB については5割程度であり課題を残した。そこで今回は、GoogLeNet (G.N) よりシンプルなネットワーク構造を有する AlexNet⁷⁾(A.N)

の転移学習を利用し AD, DLB, NC の識別が可能であるかを G.N と同様に評価し、AlexNet と G.N の適合率、正解率について比較評価も行った。

2. 方 法

2-1. AlexNet による転移学習

A.N は2012年の画像認識コンテスト ILSVRC (ImageNet Large Scale Visual Recognition Challenge) において優勝したネットワークである。このモデルが初めて CNN の概念や活性化関数として ReLU 関数を採用し、少ない学習データを増やすため画像の反転や回転の手法を取り入れていた。現在は様々な形のモデルが登場しているがそれらの標準形となるモデルである。この研究で使用した A.N のネットワーク構造を Fig. 1 に示すが5層の畳み込み層と3つのプーリング層、3つの全結合層を有している。本来1000種類のカテゴリ分類ができるように構築されているため、AD, DLB, NC の3つのカテゴリ分類になるように、最後の全結合層のみを変更して使用した。それら変更処理や G.N 解析については Matlab 2019b (Mathworks 社) の Deep learning toolbox を用いた。

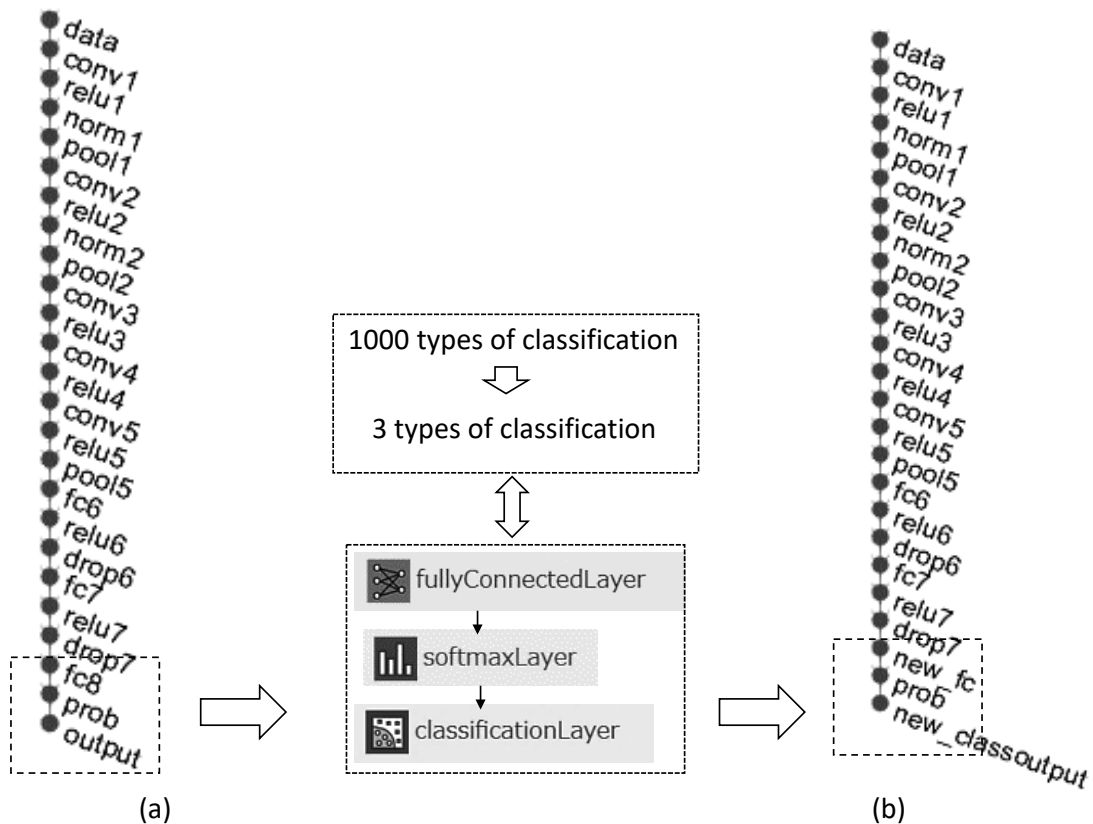


Fig. 1 AlexNet オリジナル1000のクラス分類を3クラス分類に変更したネットワーク図

(a) AlexNet 変更前ネットワーク図 (1000 クラス分類可能)

(b) AlexNet 変更後ネットワーク図 (3 クラス分類へ変更)

点線枠内の全結合層を1000 クラス分類から3クラス分類に変更した。

2-2. 解析データの作成

解析に使用したデータと使用装置の詳細については既に報告した G.N での転移学習の報告を参照いただきたい²⁾。

概要は、 ^{123}I -IMP 脳血流 SPECT 検査から得た 3D-SSP Z-score map (自施設ノーマルデータベース使用) の脳表 8 方向の画像から、診断では特に重要となる左右の外側面 (R. Lat, L. Lat) と内側面 (R. Med, L. Med) の 4 方向を用いて A.N の入力画像サイズである 247×247 ピクセルに作りなおした。その際、全データについて 4 つの画像が同じ座標上に配置されるようにした。CNN 解析には学習データ (Training), Training の精度評価用データ (Validation), 実際に識別するデータ (Test) の 3 群が必要となる。しかし、DLB のように画像症例の少ないものは Training 用に多くの数を準備することができない。そこで今回はデジ

タル的に作成した AD, DLB の画像を用いた (シミュレーション画像)。AD と DLB のシミュレーション画像は、健常者の SPECT データをベース画像に使用し、AD では左右の頭頂葉 (角回, 上頭頂小葉, 下頭頂小葉, 縁上回), 後部帯状回, 楔前部の 6 領域, DLB は、左右の後頭葉 (上後頭回, 中後頭回, 下後頭回), 一次視覚野 (楔部, 舌状回) の 5 領域に血流低下領域を作成することで AD と DLB の疾患例とした。具体的には、SPECT 画像を Statistical parametric mapping (SPM)⁸⁾ にて解剖学的標準化し Montreal neurological institute (MNI) 座標系に変換、全脳の神経学領域を MNI 座標の Volume of interest (VOI) 単位で解析可能な Voxel-based analysis-stereotactic extraction estimation (vbSEE)⁹⁾ を利用して必要な特異的血流低下領域に対して各ボクセルでの元のカウントから一律 20%~35% 低下させた画像

を多数作成し 3D-SSP 解析した。得られた Z-score map の閾値である Lower と Upper を可変とし、更には解析毎に出力される全脳 (GLB)、視床 (THL)、橋 (PNS)、小脳 (CBL) の 4 つの正規化部位 (参照部位) からできるだけ多くの画像を採用することで (参照部位として不適切なデータは除外)、AD、DLB 各 369 例のシミュレーション画像を作成した。うち 28 例は Validation に使用した。NC の Training については、自施設で収集した ^{123}I -IMP ノーマルデータベース 28 例から同様に閾値設定、複数の参照部位を採用することで 369 例を作成した。Test は、検査データから認知症の画像診断を得意とする経験年数 20 年以上の放射線専門医により AD、DLB、NC と判断された各 20 症例を採用した。次に、20 症例それぞれの 4 参照部位から医師が適切と判断した 2 参照部位を採用することで AD、DLB、NC 各 40 例を最終 Test とした。以上、使用データ群については既に報告した G.N の研究に揃えた²⁾。

2-3. 評価方法

A.N の解析を 20 回繰り返し、AD、DLB、NC について放射線科医の診断結果をリファレンスとして、3 カテゴリー全体での正解率: Accuracy (正解した数 / 全体の数)、陽性であると識別した総数の何割が当たっていたかを示す適合率: Precision ((真陽性 / (真陽性 + 偽陽性))、本来は陽性と判断すべきうちの、何割を陽性と判定できたかを示す再現率: Recall ((真陽性 / (真陽性 + 偽陰性))、適合率と再現率のバランスを評価する F 値: F-measure (($2 \times \text{適合率} \times \text{再現率} / (\text{適合率} + \text{再現率})$) の平均値と標準偏差を算出することで評価するが、Accuracy、Recall について G.N との比較評価も行った。また、CNN によりどの領域を特徴量として捉えたかを Grad-CAM から画像化し、G.N との違いについて調べた。

なお、本研究内容は ** 大学医の倫理委員会承認を得て行っている。

3. 結果

Table 1 から Training の Accuracy 0.96 に対し、Test は 0.68 と低値であった。Test の Recall は $\text{NC} > \text{DLB} > \text{AD}$ 、Precision は $\text{NC} > \text{AD} > \text{DLB}$ 、F-measure は $\text{NC} > \text{DLB} > \text{AD}$ 、Training と

Validation については Recall、Precision、F-measure すべて 0.9 以上であった。G.N と A.N の Recall の違いのグラフを **Fig. 2** に示す。また、**Table 1** の括弧内は G.N での解析結果を示す。G.N の Recall は $\text{AD} > \text{DLB}$ であったが、A.N では $\text{DLB} > \text{AD}$ であった。Accuracy については差がなかった。**Fig. 3(a)** は Test AD、**(b)** は Test DLB、**(c)** は Test NC で医師の診断と識別が一致した A.N の Grad-CAM ヒートマップ画像を示す (⇒ 脳内画像の低濃度領域がヒートマップ領域)。正しく識別された画像は特異的血流低下領域に一致してヒートマップされている。**(d)** は **(a)** と同一画像を G.N で解析し AD と判断された画像の Grad-CAM ヒートマップ、**(e)** は **(b)** と同一画像を G.N で解析し DLB と判断された画像の Grad-CAM ヒートマップ画像を示す。同一画像でも G.N と A.N のヒートマップ領域の違いを確認できる。**(f)** は A.N、Test AD の症例であるが、医師の診断と異なり DLB と判断された画像の Grad-CAM ヒートマップ画像を示す。**(b)** の DLB と異なるヒートマップ領域が存在する (➡)。A.N においても G.N 同様に識別がどの領域の特徴を捉えて判断されているか、画像化は可能であった。

4. 考察

Test で Accuracy が低いのは、AD の Recall 低値が起因している。DLB の Recall は高値であるが Precision で低値を示し、AD の多くが誤って DLB と識別され、DLB データの総数が増えたことが原因である。Training と Validation は Accuracy を含め各評価で高値を示した。これは、AD や DLB が典型例の画像群となり各カテゴリーの特徴が近似し、カテゴリー間の特徴の違いをはっきり識別できたことが予測できる。A.N と G.N との評価の差について、Validation では両解析の Recall や Accuracy は同じ傾向であったことから²⁾、両解析法ともにシミュレーションによる学習データを用いて解析する場合、学習データにあるような典型的なパターンの臨床データであれば妥当性はあると考える。両解析 Grad-CAM ヒートマップ画像について、**Fig. 3(a)** では AD の特異的血流低下領域である側頭・頭頂葉、及び後帯状回が、**(b)** の DLB についても両側側頭葉・後頭葉がヒート

Table 1 A.N 解析の評価結果を示す。Accuracy は AD, DLB, NC 全体での値を示す。Test データの Accuracy が低いのは AD の再現率 (Recall) 低値が一因。Training と Validation の各値が高いのは, AD, DLB 群をシミュレーションで作成したため画像の特徴が近似していることによる。

Data type	Evaluation method	AD	DLB	NC	Accuracy
Test	Recall	0.52±0.12 (0.82±0.12)	0.74±0.09 (0.42±0.12)	0.76±0.06 (0.78±0.08)	0.68±0.02 (0.68±0.01)
	Precision	0.70±0.08	0.53±0.08	0.93±0.04	
	F- measure	0.58±0.06	0.61±0.02	0.83±0.02	
Training	Recall	1.00	0.88±0.08	0.99±0.01	0.96±0.02
	Precision	0.91±0.05	0.99±0.01	0.98±0.03	
	F- measure	0.95±0.03	0.93±0.04	0.99±0.01	
Validation	Recall	0.96±0.03 (0.96±0.02)	0.94±0.04 (0.88±0.04)	0.98±0.01 (0.96±0.05)	0.96±0.01 (0.93±0.02)
	Precision	0.96±0.02	0.95±0.03	0.96±0.03	
	F- measure	0.96±0.01	0.94±0.01	0.97±0.01	

() : GoogLeNet

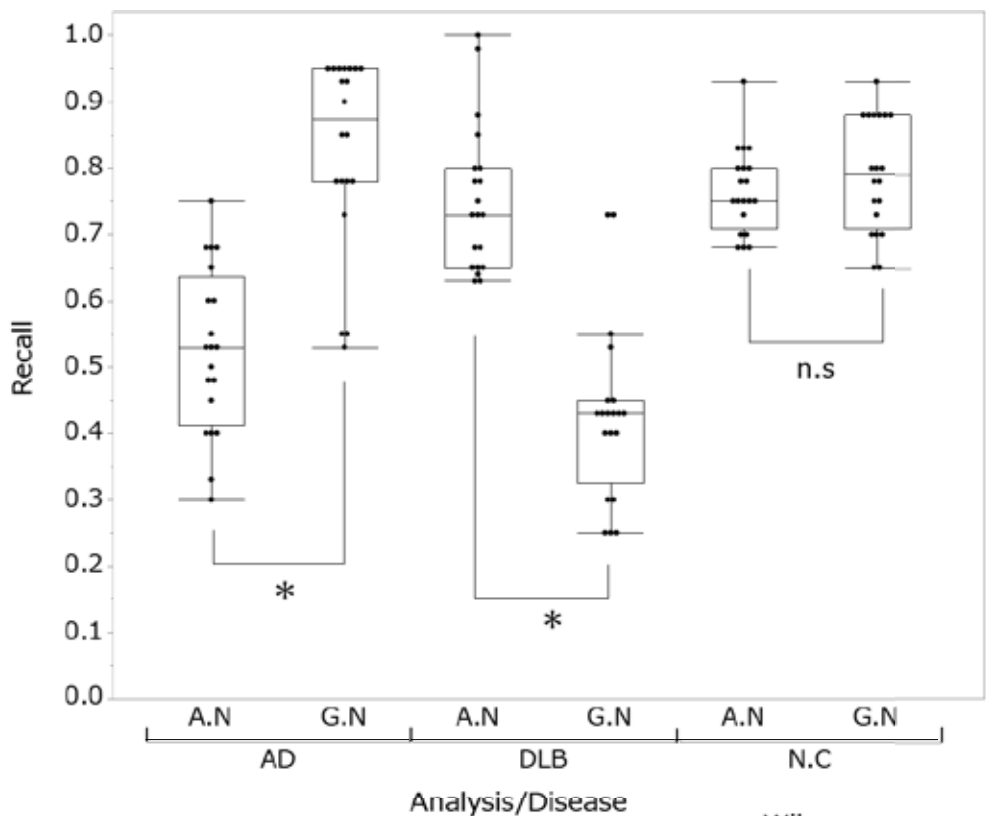


Fig. 2 AlexNet (A.N) と GoogLeNet (G.N) での AD, DLB, NC の適合率 (Recall) を示す。A.N では DLB が AD より高い値を示し, G.N では DLB より AD が高い値を示し有意差があった。NC では A.N と G.N で有意差がなかった。

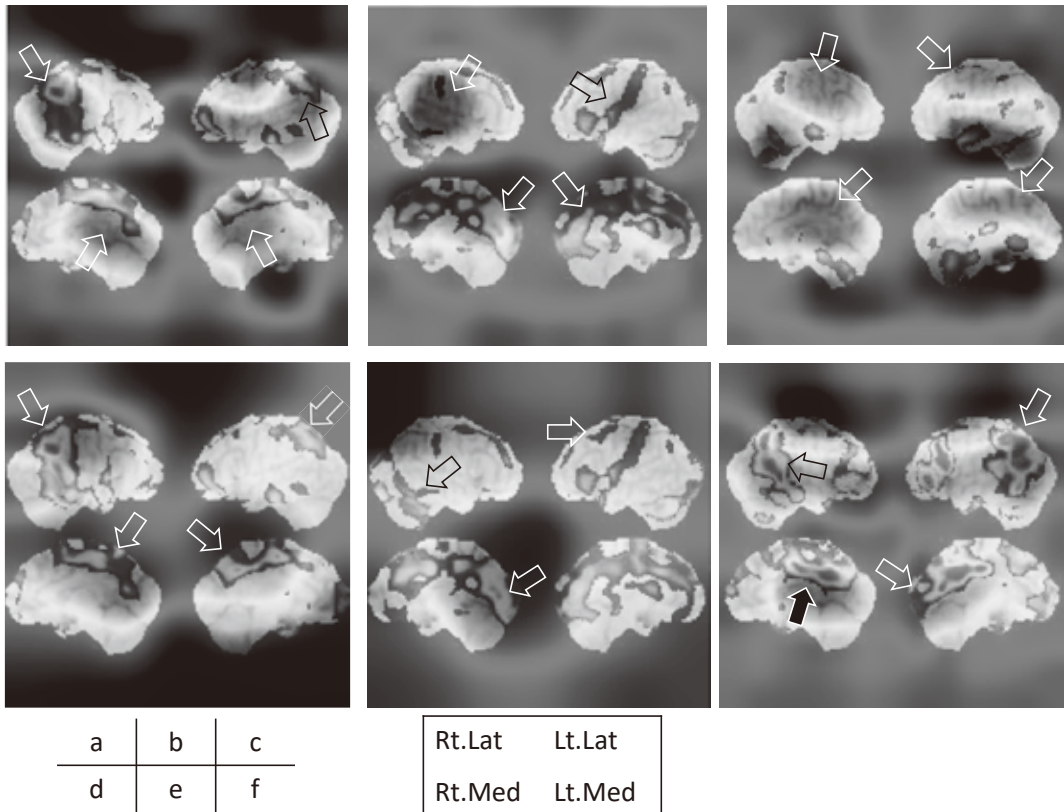


Fig. 3 Grad-CAN を利用したヒートマップ画像を示す (⇒: 脳内画像の低濃度がヒートマップ領域)。
4つの画像は Rt.Lat: 右側面脳表, Lt.Lat: 左側面脳表, Rt.Med: 右内側脳表, Lt.Med: 左内側脳表画像を示す。

- a: A.N, Test AD にて医師の診断と同じく AD と判断された画像の Grad-CAM ヒートマップ
- b: A.N, Test DLB にて医師の診断と同じく DLB と判断された画像の Grad-CAM ヒートマップ
- c: A.N, Test NC にて医師の診断と同じく NC と判断された画像の Grad-CAM ヒートマップ
- d: a と同一画像を G.N で解析し AD と判断された画像の Grad-CAM ヒートマップ
- e: b と同一画像を G.N で解析し DLB と判断された画像の Grad-CAM ヒートマップ
- f: A.N, Test AD にて医師の診断と異なり DLB と判断された画像の Grad-CAM ヒートマップ

マップされ A.N も正しく低下領域を捉えている。(c) は NC であるが血流低下のない領域を特徴としてヒートマップされたと理解する。(f) AD の Rt.Med では (a) のヒートマップ領域と同じ領域を捉えていることから正しく AD の特徴を捉えている, しかし Lt.Med では後部帯状回近位の強い血流低下領域が後頭葉近位まで及び広範囲にヒートマップされたことから DLB と識別したと考える。最終的に DLB と識別された原因については本研究のみでは考察できない。(d) は (a) と同じ画像の G.N ヒートマップを示すが, Rt.Med と Lt.Med の後部帯状回領域では A.N が G.N より脳内側を大きくヒートマップしている。(e) は

(b) と同じ画像の G.N ヒートマップを示すが後頭葉領域に注目すると A.N では R.Med と L.Med を別領域としてヒートマップされているが, G.N では左右後頭葉は一つの大きな領域としてヒートマップされている。この傾向は全画像のヒートマップで確認され A.N が G.N より細かな領域の特徴を捉えることが可能であったことが, DLB の Recall が G.N より高値である理由の一つとも考える。Grad-Cam ヒートマップ画像から A.N と G.N は共に医師の診断のように限局した小さな領域の特徴も含めて総合的に判断するのではなく, 比較的広範囲な領域の特徴と位置情報から学習データに存在する近似した画像を見つけて分類

を決定する。G.N, A.Nで安定した精度が得られなかったモデル上の原因として、G.Nでは層の数が多く、そのぶん多くの学習パラメータを必要していることから Training データを増やしての検討が必要、A.Nでは層の数は少ないものの畳み込み層のチャンネル数が多く、やはり多くの学習パラメータを必要とすることが考えられ、Training データを増やして再検討する必要がある。

以上から、A.NでAD, DLB, NCの識別を行うにはADの学習データの再検討に加えてA.Nのアルゴリズムをそのまま3D-SSP画像に用いて良いのかを研究する必要がある。A.NとG.Nで評価が異なったのはネットワーク構造の違い、Validationデータの適正など様々な因子が想定される。今後の研究では、学習データの追加も含めてこれらも問題点に取り組んでいく。

5. 結 論

AlexNetの転移学習を使用して、AD, DLB, 認知症所見なしの3カテゴリについて、統計画像(3D-SSP Z-score map)の識別を放射線科医の診断と比較評価した。但し、学習データはシミュレーションから作成した。また、Grad-CAMのヒートマップ画像から各カテゴリの特徴をどのように捉えているか画像での確認を試みた。更にはGoogLeNetでの解析と比較した。AlexNetでは、DLBと認知症所見なしについては7割以上の正解率であったがADについては半数程度がDLBと識別され、この結果はDLBよりADの正解率が良かったGoogLeNetと異なる結果であった。Grad-CAMから医師の診断と識別が一致したAD, DLBでは特的血流低下領域の特徴を正しく捉えていることが確認できた。また、同じ画像でもGoogLeNetとヒートマップされる領域が異なった。今後は、学習データの再構築、他のネッ

トワークの検討、Validationデータの選択も検討していく必要がある。

引用文献 (references)

- 1) 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 2020; **40**(4): 407-412
- 2) 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 2021; **41**(2): 204-209
- 3) Iizuka T, Fukasawa M, Kameyama M. Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. Sci Rep 2019; **9**(1): 8944
- 4) Minoshima S, Robert A, Koeppe M. A, et al. Anatomical standardization; Linear scaling and nonlinear warping of functional brain images. J Nucl Med 1993; **23**(11): 1296-1309
- 5) Selvaraju R, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision 2020; **128**(2): 336-59
- 6) Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions. 2015 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr). 2015: 1-9
- 7) Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the Acm 2017; **60**(6): 84-90
- 8) 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). 日本放射線技術学会, 2011: 16-22
- 9) 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 2013; **27**(3): 187-202