医用画像を用いた深層学習 Data Augmentation

における SuperMix の有効性の検証

東北大学大学院医学系研究科医科学専攻

医用画像工学分野

稲森瑠星

1 Introduction

1-1 人工知能(AI)の概要

人工知能(Artificial Intelligence: AI)とは、一般に、推論、認識、判断な ど、人間の知的な処理能力を代替させた情報処理システムを指し、AI には機械 学習や深層学習などのアルゴリズムがある。

1-2 深層学習の概要と活用

深層学習とは、人間が機械に対して学ぶべき特徴量を定義する必要のあっ た機械学習とは異なり、機械が自ら入力データより特徴量を学習し、画像分類 や物体検出などの画像認識タスク、音声認識、自然言語処理などの系列データ に関するタスクを行う AI アルゴリズムである。特に画像認識タスクにおいては 畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) が近 年目覚ましい発展を遂げている。

CNN とは, 層構造を取るニューラルネットワークにおいて, 従来の全結合 層だけでなく, 新たに Convolution レイヤ (畳み込み層) と Pooling レイヤ (プーリング層)を用いることによって,入力データの形状を無視することな く学習を行うことができる. 畳み込みニューラルネットワークでは,入力デー タに対して,フィルタの畳み込み処理とプーリング処理を繰り返しながら学習 を進めていく. 畳み込み処理とは入力データに対して,フィルタを畳み込み演 算 (フィルターの要素と入力の対応する要素との積和演算)することで,画像 の特徴量を抽出する.プーリング処理では,特定の領域を1つの要素に集約す る処理を行うことで,空間サイズを小さくする効果がある.CNNでは,このよ うな畳み込み層とプーリング層を繰り返すことで,形態的情報を損なうことな く画像の特徴量を抽出,学習していくことができるという特徴がある.画像認 識タスクは,深層学習が最も得意とするタスクのひとつであり,医療分野にお ける研究も行われている.

1-3 現在までの深層学習の医療画像への応用について

近年,深層学習は医療画像の分野でも研究が行われており,様々な深層学習ベ ースの AI システムが急速に進化している.核磁気共鳴画像 (MRI) [1], コン ピュータ断層撮影 (CT) [2],マンモグラフィ (MMG) [3]など,さまざまな種 類の医療画像モダリティを用いて,分類[4],セグメンテーション[5],検出[6] など多種多様なタスクに利用されている.

1-3-1 脳腫瘍の概要と AI の応用

脳腫瘍は,他の部位に転移することはまれであるが,急速に成長し近くの部位 に拡散する際に脳を損傷する可能性があるため,危険である.良性腫瘍であっ ても,増殖により脳を圧迫し,重大な合併症を引き起こす可能性がある[7].脳 腫瘍は悪性新生物による死亡率全体の約 2.71%を占め,5 年生存率は膠芽腫で 約 5.6%と不良である[8].これらの要因から,腫瘍を予防する方法,より効率 的な治療法,より優れた診断検査,腫瘍を研究・分類するためのより良い方法 が模索されており,AIの研究・開発も取り組まれている.

1-3-2 COVID-19の概要とAIの応用

新型コロナウイルス(SARS-CoV-2)によって引き起こされる COVID-19 は, 2019 年 12 月に初めて検出され,世界中に急速に広がり,重大な影響をもたら

3

した. COVID-19 は感染力が高く,病原性が強く,潜伏期間が長いという特徴 がある.治療中に,症状が急速に変化するため,短期間に複数回の再検査が必 要となり,多くの検査が実施された.これらの問題により,医療現場は逼迫し, 医師にとって COVID-19 の正確かつ迅速な診断は非常に困難である.そこで, AI を用いた高速化かつ高精度な画像診断の研究・開発が行われている.いくつ かの研究では,AI モデルが放射線科医よりも感度,特異度が高く,放射線科医 は AI の支援により高速かつ高い診断性能を達成しており,AI の有用性が報告 されている[9].

1-3-3 乳がんの概要と AI の応用

乳がんは現在世界における女性の癌の中で,新規罹患数・死亡数ともに最 多の癌である[10](図1).日本においても.2021年の部位別予測がん罹患数 で,全体の22%を,死亡数では全体の10%を占めている[11].マンモグラフィ は乳がん早期発見のスクリーニング検査として広く用いられており,マンモグ ラフィ受診の継続によって,乳がん死亡率が低下することが医学的統計学的に 証明されている[12].現在の医師によるマンモグラフィの診断精度は感度が 86.9%,特異度が88.9%と報告されており[13],40歳~74歳を対象にしたマ ンモグラフィ単独での検査に関する5つのランダム化比較試験に関するメタア ナリシスでは,マンモグラフィ単独での検査において25%死亡率を低減すると いう報告がなされている[14].マンモグラフィによる乳がん画像診断へのAIの 応用例として,これまでに報告された放射線科医とAIの乳がんの診断能を比較 した研究では,放射線科医による診断のAUCは0.814,AIは0.840であり,AI



は一般的な放射線科医の平均よりも常に高い診断能を示した[15].

図1 2020年の癌の種類別罹患者数と死亡数([10]より引用)

1-4 医療画像を用いた深層学習における課題と解決策

深層学習では,一般的に大量の画像を用いて学習を行うことで,高い精度を実 現することが可能である.一方,医療画像は,個人情報保護などの観点から, 大量のデータセットを得ることが難しいという課題がある[16].

1-5 データ拡張 (Data Augmentation)

データ拡張(Data Augmentation: DA)とは、元画像に前処理を加えたり、 合成したりする事により新しい画像を生成し、データ数を増やす手法である. 深層学習において大量のデータが利用できない場合に活用されている。適切な DA を行う事により、画像の一般化を行い、過学習を回避し、モデルの汎化性 能を向上させる事が出来る[17].また、PyTorch などの深層学習フレームワー クでは、学習用データセットを物理的に拡張するオフライン拡張ではなく、学 習中にその場でデータ補強を行うことが可能なオンライン拡張という方法もあ る.一般的な DA 手法としては、ランダムな回転、トリミング、反転、ノイ ズ付加などがある[18].しかし、医療画像のような複雑な高次元のデータを扱 う場合には、これらの単純な操作では十分ではない可能性がある.

1-6 医療における Data Augmentation の期待

医療分野により適した DA 手法の提案には,新しく生成されたデータを実デー タと区別することが不可能,あるいは少なくとも困難となるように,実データ にできるだけ近いデータ分布を再現することが求められる.そこで,全く新し い画像を生成する GAN (Generative Adversarial Network) や複数の画像を 合成する事によるデータ拡張である mixup や CutMix などの DA 手法の医療分 野への応用が期待され,研究されている[19][20][21].また,近年では拡散モデ ル (Diffusion Model: DM) と呼ばれる新しい深層学習生成モデルが出現し, 現実的で高品質な画像を生成出来るとされている. 医療分野でも DM を用いた 画像生成について研究が進み、その応用が期待されている[22][23].

1-7 様々な Data Augmentation 手法

1-7-1 Rand Augment

Rand Augment は, Google が提案したデータ拡張の自動最適化手法の1つ である. 14 個 (Identity, autoContrast, equalize, rotate, solarize, color, posterize, contrast, brightness, sharpness, shear-x, shear-y, translate-x, translate-y)のデータ拡張手法の中からN個の手法,強度(10段階)の中から 強度 M を選択し,グリッドサーチで最適化するという単純な手法である[24] (図 2). これまで Auto Augment という最適化手法もあったが,パラメータ の探索コストが 10³² と非常に大きく実用的ではないとされていた. Rand Augment は 10²の探索コストで行えるため,低コストに最適な手法を探索可能 であるため有用である.



図2 RandAugmentの例([24]より引用)

1-7-2 mixup

mixup は, 2つの訓練サンプルのペアを混合して新たな画像を作成する DA 手法の 1 つである[25] (図 3).具体的には,データとラベルのペアから,下 記の式により新たな画像を生成する.

$$\begin{split} \tilde{x} &= \lambda x_i + (1 - \lambda) x_j \\ \tilde{y} &= \lambda y_i + (1 - \lambda) y_j \\ \lambda &\in [0, 1], \quad \lambda \sim Beta(\alpha, \alpha) \quad \alpha \in (0, \infty) \end{split}$$

x:画像, y:ラベル, λ:Beta 分布によって取得される混合割合







[non-COVID , COVID] [non-COVID , COVID]



[0.3, 0.7] [non-COVID , COVID]

図3 mixupの概念図

1-7-3 SuperMix

SuperMix は、画像上の重要な領域(例:乳癌)を認識し、重要な領域に重 みをつけた上で複数の画像の合成を行う DA 手法である[26]. 画像上の重要な 領域を認識することで、視覚的な画像の特徴を保持したまま合成することが可 能である. この特徴により, CutMix や MixUp の課題であった画像の一部の領 域が別の画像に置き換えられることや画像全体が平均化されることにより、情 報が失われてしまうという問題を解決できると報告されている[26] (図 4). 図 4 で示されている青い領域は、背景などあまり正解ラベルに意味を持たない 領域であり、赤い領域は画像の中で正解ラベルとして意味を持つような領域で ある.



図4 SuperMix の概念図([26]より引用)

1-8 SuperMix に関する関連研究

SuperMix による DA が深層学習モデルの分類精度に与える影響が報告され ている. Ali Dabouei らは CIFAR-100 と ImageNet の 2 つの自然画像データセ ットを用いて,画像分類と知識蒸留の 2 つのタスクに対する SuperMix の性能 を評価している. 5 つの実験のうち 4 つで SuperMix は高い精度を発揮し,さ らに SuperMix と RandAugment を組み合わせることで全ての実験において分 類の精度がさらに向上する事を示した[26]. しかし,自然画像に対しての分類 精度しか調査されておらず,医療画像を用いた調査はされていない. また,他 の研究でも医療画像に対する Super Mix の有効性を調査した研究の前例はない.

1-9 仮説と目的

自然画像を用いた画像分類において有効性が示されている SuperMix を医療画像へ応用することは、データ取得が困難な医療画像 AI の DA 手法としても 有用であると仮説を立てた.本研究の目的は、複数の医療画像データセットに 対し、DA を用いない深層学習と 3 つの DA 手法、1) Rand Augment、2) mixup, 3) SuperMix を用いた深層学習の比較を行うことで, SuperMix の有 効性を検証することである.

2 Materials & Methods

2-1 使用したデータセットについて

本研究には, SARS-CoV-2 CT-scan dataset [27], BraTS2019 [28], DatasetBUSI [29], CMMD [30]の4つの異なるモダリティのオープンデータセ ットを使用した (表1).

表1使用したデータセットの概要

データセット	モダリティ	撮影部位	対象疾患
SARS-CoV-2 CT-scan dataset	СТ	肺	COVID 関連肺炎
BraTS2019	MRI	脳	Glioma(神経膠腫)
DatasetBUSI	US	乳腺	乳癌
CMMD	MG	乳腺	乳癌

SARS-CoV-2 CT-scan dataset

SARS-CoV-2CT-scan dataset は, ブラジルのサンパウロ公立病院(HSPM) で入院している 120 名の患者から得た CT 画像で構成されている. SARS-CoV-2 に感染した患者(60 名)の 1,252 枚と, SARS-CoV-2 以外の肺疾患を有す る患者(60 名)の 1,230 枚の画像からなる[27].

BraTS2019

BraTS2019 データセットは,複数施設から得た 3T の MRI 画像であり, 259 件の High grade glioma (HGG) および 76 件の low grade glioma (LGG) MRI 画像で構成されている. すべての画像の正解ラベルは, 放射線科医によって付けられた. また, T1, T2, FLAIR, T1CE の 4 つのシーケンスのデータが 含まれており, 1 件あたり 155 枚の画像からなる [28].

DatasetBUSI

DatasetBUSI データセットは,2018 年にバヘヤ病院で25 歳-75 歳の女性 患者 600 名の患者から得た乳房超音波画像で構成されている.LOGIQ E9 超音 波装と LOGIQ E9 Agile 超音波装置を用いて撮影され,使用したリニアプロー ブの周波数は 1~5MHz である。内訳は,良性腫瘤 487 枚,悪性腫瘤 210 枚, 正常乳腺 133 枚の合計 780 枚である [29].

CMMD

CMMD データセットは、2012 年 7 月から 2016 年 1 月の間にマンモグラ フィ検査を受けた中国人患者 1775 名の患者から得たマンモグラフィ画像で構 成されている. 放射線科医 1 名による画像診断を行い、2601 乳房のうち、cv ポートが挿入されている 8 乳房、葉状腫瘍と診断された 6 乳房、神経繊維腫症 と診断された 2 乳房、リンパ浮腫と診断された 1 乳房、異物が写り込んでいる 5 乳房、画像に不備があり診断が困難である 3 乳房、乳癌患者で病変が同定不 能な乳房 140 乳房、合計 165 乳房を除外した. 最終的に、合計 2436 乳房を対 象とし、放射線科医の診断をもとに乳癌 1167 乳房、良性病変 215 乳房、正常 乳腺 1054 乳房と定義した. 乳癌 1167 乳房には、腫瘤が 941 乳房、石灰化が 508 乳房、Focal asymmetric density: FAD が 15 乳房、Distortion が 53 乳房 含まれていた. 一部の乳房は、複数病変が含まれている.

2-2 データセット

本研究では、DA 手法の比較のために、DatasetBUSI 以外の3つのデータセットにおいて、4 つのデータセットの大きさを等しくするために、全画像データうち 30-40%程度の画像を抽出し、小規模なデータセットを作成した(図5).

SARS-CoV-2 CT-scan dataset

本研究では, kaggle より得られた, SARS-CoV-2 の画像 1,252 枚と, SARS-CoV-2 以外の肺疾患の画像 1,230 枚の CT 画像から一部抽出し, 最終的 に SARS -CoV-2 の画像 250 枚と, SARS-CoV-2 以外の肺疾患の画像 245 枚を 使用した(図 5).

BraTS2019

本研究では, kaggle より得られた, 脳腫瘍がある画像 1,503 枚と脳腫瘍が ない画像 1,503 枚の MRI 画像から一部抽出し, 最終的に脳腫瘍がある画像 300 枚と, 脳腫瘍がない画像 299 枚を使用した(図 5).

DatasetBUSI

本研究では,乳房超音波画像全データ 780 枚のうち,正常乳腺 133 枚を除く,良性腫瘤 487 枚,悪性腫瘤 210 枚のみを使用した(図 5).

CMMD

乳癌 1167 乳房と良性病変 215 乳房の中には,複数の病変がある乳房が含まれている。本研究では,乳癌 1167 乳房,良性病変 215 乳房から,病変部分を中心に 512 × 512 に切り取りを行い,乳癌病変 1515 枚,良性病変 255 枚の

画像を抽出した. 正常乳腺 1054 乳房からは、ランダムに 512 × 512 に切り取 りを行い, 1054 枚の画像を抽出した. さらに乳癌病変 1515 枚を乳癌, 良性病 変 255 枚と正常乳腺 1024 枚の画像を合わせた 1279 枚を非乳癌と定義した. その後, 乳癌の画像 1515 枚と, 非乳癌の画像 1279 枚のマンモグラフィ画像 から一部抽出し, 最終的に乳癌の画像 274 枚, 非乳癌の画像 276 枚を使用した (図 5).



図5 データセット抽出の概要

各 デ ー タ セ ッ ト は , 学 習 用 デ ー タ (Train) , 検 証 用 デ ー タ (Validation) , テスト用データ (Test) にそれぞれ 6 : 2 : 2 の割合でランダ ムに分割した. 各データセットの学習用データ,検証用データ,テスト用デー タの内訳は表 1 の通りである. 表1各データセットの内訳

データセット	Train	Validation	Test
SARS-CoV-2 CT-scan dataset	297	99	99
(COVID/Non-COVID)	(150/147)	(50/49)	(50/49)
BraTS2019	360	119	120
(Tumor/Non-Tumor)	(180/180)	(60/59)	(60/60)
DatasetBUSI	389	129	129
(Benign/Malignant)	(263/126)	(87/42)	(87/42)
CMMD	325	114	111
(Cancer/Non-Cancer)	(165/160)	(53/61)	(56/55)

2-3 実験方法

本研究では、4つの異なるモダリティのデータセットに対し,DAを用いな い深層学習と3つのDA手法,1)RandAugment,2)mixup,3)SuperMix を用いた深層学習,合計4つの実験を行い,比較検証した(表2). また,各実験について深層学習モデルの学習およびテストを1セットとして各 実験に対して10セットずつ繰り返した。

表 2 実験方法の概要

データセット	実験手法	
SARS-CoV-2 CT-scan dataset	DAなし	
	RandAugment	
	mixup	
	SuperMix	

BraTS2019	DAなし
	RandAugment
	mixup
	SuperMix
DatasetBUSI	DAなし
	RandAugment
	mixup
	SuperMix
CMMD	DAなし
	RandAugment
	mixup
	SuperMix

2-4 深層学習モデル

本研究の深層学習モデルには ResNet50 [31]を使用した. ハイパーパラメ ータは以下の通りである. バッチサイズ:32, エポック:100, 最適化関数: Adam (learning-rate=0.0001, weight-decay=1.0×10⁻⁶), 損失関数: BCEWithLogitsLoss. また,実装に用いたコンピュータの環境は以下の通りで ある. Intel Corei7-7800X (6core), NVIDIARTX8000 (48 Giga Bytes), Ubuntu18.04.5 long term support (LTS): Xenial Xerus. すべての解析は, Python Language, version 3.7.3 (Python Software Foundation at http://www.python.org)を使用し,深層学習フレームワークとして, PyTorch 1.13.1を使用した.

2-5 評価方法

本研究では,深層学習モデルの診断精度を評価するために、Accuracy, Sensitivity, Specificity, PPV (Positive Predictive Value), NPV (Negative Predictive Value), AUC (area under receiver operating characteristic curve)を評価指標として用いた.各実験について深層学習モデ ルの学習およびテストを1セットとして各実験に対して行った 10 セットそれ ぞれの評価指標の平均値をAverage Accuracy, Average Sensitivity, Average Specificity, Average PPV, Average NPV, Average AUC と定義した.また、 各平均評価指標の 95%信頼区間を求めた.DA 手法の違いによる有意差を調査 するため,SuperMixを用いた深層学習の平均 AUC の値を参照値として定義し, p<0.05 を統計的に有意として他の手法における深層学習モデルの有意差を評価 した.統計的有意差検定にはDelong 検定を用いた.

3 Result

3-1 SARS-Cov-2 CT-scan dataset における結果

各実験の平均の評価指標は, non-DA/RA/mixup/SuperMix の順に Accuracy: 0.803/0.860/0.759/0.835, Sensitivity: 0.740/0.778/0.574/0.804, Specificity: 0.867/0.943/0.947/0.867, PPV: 0.854/0.933/0.922/0.864, NPV: 0.767/0.807/0.691/0.816, AUC: 0.872/0.929/0.886/0.924 であった (表 3).

Statistic (95%CI)	non-DA	RA	mixup	SuperMix
	0.803	0.860	0.759	0.835
Accuracy	(0.777-0.829)	(0.847-0.873)	(0.719–0.799)	(0.823-0.848)
	0.740	0.778	0.574	0.804
Sensitivity	(0.700-0.780)	(0.758–0.798)	(0.480-0.668)	(0.763-0.845)
Specificity	0.867	0.943	0.947	0.867
	(0.823–0.912)	(0.928-0.958)	(0.927–0.967)	(0.832–0.903)
	0.854	0.933	0.922	0.864
PPV	(0.815–0.893)	(0.917–0.950)	(0.902–0.942)	(0.836–0.892)
NPV	0.767	0.807	0.691	0.816
	(0.739–0.795)	(0.793–0.821)	(0.649–0.733)	(0.789–0.842)
AUC	0.872	0.929	0.886	0.924
	(0.852–0.891)	(0.916–0.941)	(0.869–0.902)	(0.912–0.937)

表 3 SARS-CoV-2 CT-scan dataset における平均の評価指標



図 6 SARS-CoV-2 CT-scan dataset における ROC 曲線

*p-value<0.05 SuperMix に有意差のある手法

3-2 BraTS2019 における結果

各実験の平均の評価指標は, non-DA/RA/mixup/SuperMix の順に Accuracy:0.798/0.826/0.773/0.84, Sensitivity:0.737/0.765/0.617/0.773, Specificity:0.858/0.887/0.928/0.913, PPV:0.845/0.870/0.899/0.901, NPV: 0.779/0.795/0.710/0.803, AUC: 0.887/0.917/0.905/0.936 であった (表4).

Statistic	non-DA	RA	mixun	SuperMix
(95%CI)		101	шлар	Buperivita
Acouroou	0.798	0.826	0.773	0.843
Accuracy	(0.757–0.838)	(0.788–0.864)	(0.747 - 0.798)	(0.820-0.867)
Soncitivity	0.737	0.765	0.617	0.773
Sensitivity	(0.634–0.840)	(0.699–0.831)	(0.566 - 0.668)	(0.733–0.814)
Specificity	0.858	0.887	0.928	0.913
Specificity	(0.804–0.912)	(0.861–0.912)	(0.897 - 0.960)	(0.883 - 0.944)
DDV	0.845	0.870	0.899	0.901
F F V	(0.810-0.880)	(0.840-0.901)	(0.862–0.936)	(0.870-0.933)
NDV	0.779	0.795	0.710	0.803
INI V	(0.715–0.842)	(0.746 - 0.844)	(0.684–0.735)	(0.775–0.830)
	0.887	0.917	0.905	0.936
AUC	(0.858–0.917)	(0.893–0.942)	(0.889–0.921)	(0.918–0.950)

表 4 BraTS2019 における平均の評価指標

各実験の ROC 曲線および Delong 検定の結果を以下に示す(図 7).



図 7 BraTS2019 における ROC 曲線

*p-value<0.05 SuperMix に有意差のある手法

3-3 DatasetBUSI における結果

各実験の平均の評価指標は, non-DA/RA/mixup/SuperMix の順に Accuracy:0.678/0.703/0.731/0.66, Sensitivity:0.719/0.800/0.405/0.845, Specificity: 0.658/0.656/0.889/0.581, PPV: 0.506/0.534/0.649/0.503, NPV: 0.832/0.873/0.758/0.892, AUC: 0.761/0.799/0.764/0.809 であった (表 5).

Statistic	non-DA	RA	mixun	SuperMix
(95%CI)			mingb	Supermin
Acourcou	0.678	0.703	0.731	0.667
Accuracy	(0.656 - 0.699)	(0.662 - 0.744)	(0.703–0.759)	(0.627–0.707)
Sensitivity	0.719	0.800	0.405	0.845
Sensitivity	(0.664–0.774)	(0.741–0.859)	(0.316-0.494)	(0.779–0.912)
Specificity	0.658	0.656	0.889	0.581
specificity	(0.614–0.702)	(0.600-0.713)	(0.849-0.928)	(0.495–0.666)
PPV	0.506	0.534	0.649	0.503
I I V	(0.481–0.531)	(0.492–0.575)	(0.571-0.726)	(0.460 - 0.547)
NPV	0.832	0.873	0.758	0.892
INI V	(0.807–0.856)	(0.838–0.907)	(0.732–0.783)	(0.866 - 0.918)
AUC	0.761	0.799	0.764	0.809
AUC	(0.734–0.788)	(0.753–0.845)	(0.726–0.803)	(0.792–0.826)

表 5 DatasetBUSI における平均の評価指標

各実験の ROC 曲線および Delong 検定の結果を以下に示す(図 8).



図 8 DatasetBUSI における ROC 曲線

*p-value<0.05 SuperMix に有意差のある手法

3-4 CMMD における結果

各実験の平均の評価指標は, non-DA/RA/mixup/SuperMix の順に Accuracy:0.672/0.690/0.691/0.635, Sensitivity:0.493/0.493/0.505/0.350, Specificity: 0.855/0.891/0.880/0.926, PPV: 0.778/0.829/0.843/0.831, NPV: 0.629/0.637/0.649/0.588, AUC: 0.772/0.811/0.812/0.824 であった (表 6).

Statistic (95%CI)	non-DA	RA	mixup	SuperMix
	0.672	0.690	0.691	0.635
Accuracy	(0.637–0.707)	(0.660–0.720)	(0.639–0.743)	(0.596 - 0.674)
	0.493	0.493	0.505	0.350
Sensitivity	(0.396–0.589)	(0.407–0.578)	(0.359–0.652)	(0.249–0.451)
Specificity	0.855	0.891	0.880	0.926
	(0.816–0.893)	(0.859–0.923)	(0.826–0.934)	(0.896–0.955)
	0.778	0.829	0.843	0.831
PPV	(0.738–0.818)	(0.799–0.860)	(0.776-0.909)	(0.796–0.865)
	0.629	0.637	0.649	0.588
NPV	(0.594–0.664)	(0.607–0.668)	(0.596–0.701)	(0.555–0.621)
	0.772	0.811	0.812	0.824
AUC	(0.753–0.791)	(0.800-0.823)	(0.791–0.832)	(0.809–0.840)

表6 CMMD における平均の評価指標

各実験の ROC 曲線および Delong 検定の結果を以下に示す(図 9).



図 9 CMMD における ROC 曲線

*p-value<0.05 SuperMix に有意差のある手法

4 Discussion

4-1 各種実験に関する考察

深層学習における DA の有効性や自然画像における SuperMix を用いた DA の有効性については既に研究が報告されている[17][26].しかし,医療画像に 対する Super Mix の有効性については十分に研究されていない.本研究では, 複数の医療画像データセットに対し, DA を用いない深層学習と 3 つの DA 手 法, RandAugment, mixup, SuperMix を用いた深層学習の比較を行うことで, SuperMix の有効性を検証した.

CMMD を除く3つのデータセットにおいて,SuperMix を用いた DA は, DA を用いない場合や mixup を用いた DA と比較して AUC が有意に高かった. また,有用な DA 手法とされている RandAugment と同等またはそれ以上の AUC の結果が得られた.これは先行研究の結果と同様の傾向を示した[26].こ の事から,自然画像とは画像特性の異なるドメインの医療画像分類タスクにお いても SuperMix を DA に用いることの有効性が示唆された.

医療画像のようなグレースケール画像が主体の医療画像への適応について は,MRI や CT など,正常組織や病変部などコントラストが大きい画像データ に対して,SuperMix による DA の効果がより高かったと考えられる.CMMD は病変を中心に 512 × 512 に切り取りを行っているため,画像の大半の領域を 病変が占めており,正常乳腺などの背景となる部分の割合が少なく,画像内の コントラストの差がが少ないため,SuperMix による DA の効果が低かったと 考えられる. SuperMix は先行研究で、従来の勾配降下法よりも効率的な計算手法を開発し ており、より高速に計算できるとされている[26].効率よく RandAugment と 同等な性能が発揮できる SuperMix を用いることで、将来的に様々な疾患に対 する高性能な AI の開発を全体的に底上げすることへの可能性が期待できる.今 後多くの検証が必要であるが、多様な疾患やタスクにおいて AI の精度の底上げ を出来れば、将来的に画像診断や治療の方針の決定など医療全体の効率化に繋 がると考えられる.

4-2 本研究における限界点

本研究にはいくつかの制限がある.第1に、今回の実験は分類タスクのみ の検討であることである。今後セグメンテーションや物体検出などの異なるタ スクに応用することで、より汎用的な SuperMix の有効性を示すことが出来る と考えられる.第2に、本研究では単一のモデル(ResNet50)のみの検討しか 行っていないことである。今後、VGG16 や Xception などの他の CNN モデル や VisionTransformer などの Transformer ベースのモデルを用いた検討を行う ことで、手法の有効性を示すことが出来ると考えられる。第3に、ResNet50 モデルおよび学習に使用したハイパーパラメータについて、全ての可能な組み 合わせを実装することはできなかった。この研究で選択したハイパーパラメー タの設定は一般的なものだが、さらに適正な設定が可能である可能性がある。 第4に、DA 手法の組み合わせによる精度向上の可能性を検討していないこと である。先行研究でも SuperMix と RandAugment を組み合わせた DA 手法が 最も精度が高くなると報告されており[26]、精度向上に最適な DA 手法の組み 合わせを検討することにより、将来的に更なる精度向上が期待されると考えら れる.

5 Conclusion

複数の医療画像を用いた分類タスクにおいて,SuperMix による DA は従来の DA 手法と比較して,分類精度の向上に有効であることが示唆された.

References

- Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. doi: https://doi.org/10.1016/j.zemedi.2018.11.002
 Zeitschrift für Medizinische Physik 2019, 29, 102–127.
- [2] Song, Y.; Zheng, S.; Li, L.; Zhang, X.; Zhang, X.; Huang, Z.; Chen, J.; Wang, R.;
 Zhao, H.; Chong, Y.; et al. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. IEEE/ACM Trans. Comput. Biol. Bioinform. 2021,18, 2775–2780.
- [3] Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Sci Rep. 2019 Aug 29;9(1):12495. doi: 10.1038/s41598-019-48995-4. PMID: 31467326; PMCID: PMC6715802.
- [4] Shimokawa D, Takahashi K, Oba K, Takaya E, Usuzaki T, Kadowaki M, Kawaguchi K, Adachi M, Kaneno T, Fukuda T, Yagishita K, Tsunoda H, Ueda T. Deep learning model for predicting the presence of stromal invasion of breast cancer on digital breast tomosynthesis. Radiol Phys Technol. 2023 Jul 19. doi: 10.1007/s12194-023-00731-4. Epub ahead of print. PMID: 37466807.
- [5] Nemoto T, Futakami N, Yagi M, Kunieda E, Akiba T, Takeda A, Shigematsu N. Simple low-cost approaches to semantic segmentation in radiation therapy planning for prostate cancer using deep learning with non-contrast planning CT images. Phys Med. 2020 Oct; 78:93-100. doi: 10.1016/j.ejmp.2020.09.004. Epub 2020 Sep 17. PMID: 32950833.
- [6] Shou Y, Meng T, Ai W, Xie C, Liu H, Wang Y. Object Detection in Medical Images Based on Hierarchical Transformer and Mask Mechanism. Comput Intell

Neurosci. 2022 Aug 4; 2022:5863782. doi: 10.1155/2022/5863782. PMID: 35965770; PMCID: PMC9371842.

- [7] Anaya-Isaza A, Mera-Jiménez L, Verdugo-Alejo L, Sarasti L. Optimizing MRIbased brain tumor classification and detection using AI: A comparative analysis of neural networks, transfer learning, data augmentation, and the crosstransformer network. Eur J Radiol Open. 2023 Mar 14; 10:100484. doi: 10.1016/j.ejro.2023.100484. PMID: 36950474; PMCID: PMC10027502.
- [8] Reynoso-Noverón, N., Mohar-Betancourt, A., Ortiz-Rafael, J. (2021). Epidemiology of Brain Tumors. In: Monroy-Sosa, A., Chakravarthi, S.S., de la Garza-Salazar, J.G., Meneses Garcia, A., Kassam, A.B. (eds) Principles of Neuro-Oncology. Springer, Cham. https://doi.org/10.1007/978-3-030-54879-7_2
- Chang Z, Zhan Z, Zhao Z, You Z, Liu Y, Yan Z, Fu Y, Liang W, Zhao L.
 Application of artificial intelligence in COVID-19 medical area: a systematic review. J Thorac Dis. 2021 Dec;13(12):7034-7053. doi: 10.21037/jtd-21-747.
 PMID: 35070385; PMCID: PMC8743418.
- [10] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray,
 Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA. Cancer J. Clin. 71 (2021) 209-249. https://doi.org/10.3322/caac.21660.
- [11] が ん の 統 計 2022 https://ganjoho.jp/public/qa_links/report/statistics/2022_jp.html
- [12] M. Marmot, Review The benefi ts and harms of breast cancer screening:an independent review Independent UK Panel on Breast Cancer Screening*, Lancet.

380 (2012) 1778-1786.https://doi.org/10.1016/S0140-6736(12)61611-0.

- [13] C.D. Lehman, R.F. Arao, B.L. Sprague, J.M. Lee, D.S.M. Buist, K. Kerlikowske,
 L.M. Henderson, T. Onega, A.N.A. Tosteson, G.H. Rauscher, D.L. Miglioretti,
 National performance benchmarks for modern screening digital mammography:
 Update from the Breast Cancer Surveillance Consortium, Radiology. 283 (2017)
 49-58. https://doi.org/10.1148/radiol.2016161174.
- [14] C. Hamashima, K. Ohta, Y. Kasahara, T. Katayama, T. Nakayama, S. Honjo, K.
 Ohnuki, A meta-analysis of mammographic screening with and without clinical breast examination, Cancer Sci. 106 (2015) 812-818.
 https://doi.org/10.1111/cas.12693.
- [15] A. Rodriguez-Ruiz, K. Lång, A. Gubern-Merida, M. Broeders, G. Gennaro, P. Clauser, T.H. Helbich, M. Chevalier, T. Tan, T. Mertelmeier, M.G. Wallis, I. Andersson, S. Zackrisson, R.M. Mann, I. Sechopoulos, Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists, J. Natl. Cancer Inst. 111 (2019) 916–922. https://doi.org/10.1093/jnci/djy222
- [16] Price, W.N., Cohen, I.G. Privacy in the age of medical big data. Nat Med 25,37–43

(2019). https://doi.org/10.1038/s41591-018-0272-7

- [17] Y. Miao and W. Luo, "Improve Generalization Ability of CNN by Data Augmentation and SE Block in Landmark Classification," 2022 14th International Conference on Computer Research and Development (ICCRD), Shenzhen, China, 2022, pp. 250-255, doi: 10.1109/ICCRD54409.2022.9730256.
- [18] Goceri, E. Medical image data augmentation: techniques, comparisons and

interpretations. Artif Intell Rev (2023). https://doi.org/10.1007/s10462-023-10453-z

- [19] C. Han, K. Murao, T. Noguchi, Y. Kawata, F. Uchiyama, L. Rundo, H. Nakayama, S. Satoh, Learning More with Less: Conditional PGGAN-based Data Augmentation for Brain Metastases Detection Using Highly-Rough Annotation on MR Images, In ACM International Conference on Information and Knowledge Management (CIKM, acceptance rate: ~19%), Beijing, China, November 2019.
- [20] Snider EJ, Hernandez-Torres SI, Hennessey R. Using Ultrasound Image Augmentation and Ensemble Predictions to Prevent Machine-Learning Model Overfitting. Diagnostics (Basel). 2023 Jan 23;13(3):417. doi: 10.3390/diagnostics13030417. PMID: 36766522; PMCID: PMC9914871.
- [21] Zhu, Q., Wang, Y., Yin, L., Yang, J., Liao, F., Li, S. (2022). SelfMix: A Self-adaptive Data Augmentation Method for Lesion Segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds) Medical Image Computing and Computer Assisted Intervention MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science, vol 13434. Springer, Cham. https://doi.org/10.1007/978-3-031-16440-8_65
- [22] Pierre Chambon, Christian Bluethgen, Curtis P. Langlotz, Akshay Chaudhari. Adapting Pretrained Vision-Language Foundational Models to Medical Imaging Domains. 2022 Oct. arXiv:2210.04133
- [23] Walter H. L. Pinaya, Petru-Daniel Tudosiu, Jessica Dafflon, Pedro F da Costa,Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, M. Jorge Cardoso.Brain Imaging Generation with Latent Diffusion Models. 2022 Sep.

arXiv:2209.07162

- [24] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, Quoc V. Le, RandAugment: Practical automated data augmentation with a reduced search space, 2019 Nov. arXiv:1909.13719
- [25] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz, mixup:Beyond Empirical Risk Minimization, 2018 Apr. arXiv:1710.09412
- [26] Ali Dabouei, Sobhan Soleymani, Fariborz Taherkhani, Nasser M. Nasrabadi, SuperMix: Supervising the Mixing Data Augmentation, 2021 Dec. arXiv:2003.05034.
- [27] Soares, Eduardo, Angelov, Plamen, Biaso, Sarah, Higa Froes, Michele, and Kanda Abe, Daniel. "SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification." medRxiv (2020). doi: https://doi.org/10.1101/2020.04.24.20078584.
- [28] https://www.med.upenn.edu/cbica/brats2019/data.html
- [29] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, Aly Fahmy, Dataset of breast ultrasound images, https://doi.org/10.1016/j.dib.2019.104863.
- [30] Wang, J., Yang, X., Cai, H. et al. Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning. Sci Rep 6, 27327
 (2016). https://doi.org/10.1038/srep27327
- [31] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.