

総説

医療と AI — 自らの研究で AI を学ぶ

¹⁾ 獨協医科大学 埼玉医療センター 放射線科

²⁾ 東京医科歯科大学 大学院医歯学総合研究科先端人工知能医用画像診断学講座

久保田一徳¹⁾, 藤岡 友之²⁾

要旨

医療分野での人工知能 (AI; artificial intelligence) は開発から利用のフェーズに移りつつある。最近の AI の発展の中心の技術となるのがディープラーニング (deep learning) という、複雑化させたニューラルネットワークを用いた機械学習である。結果が得られている画像をトレーニング・データとして多数用意しておけば、それをもとにコンピュータが自動で学習し、その後与えた画像に対して判別を行えるようになる。医療分野では画像領域の AI を中心に製品が登場しており、病変検出を補助する CADe (computer aided detection) や、欧米では診断を補助する CADx (computer aided diagnosis) も導入されつつある。画像診断への AI の活用においては、画像分類、物体検出、画像生成といったものがあり、我々は自験例をもとにこれらの AI の研究を行ってきた。ディープラーニングではトレーニング・データの質や内容によって AI の精度が決まり、単施設でも日常診療での精度の高い画像を用いれば診断能の高い AI を作る事が可能である。AI に触れることで理解し、今後の課題として医療にどのように実装していくかということを考えたい。

Key Words : 人工知能, 画像分類, 物体検出, 画像生成, 乳癌

1. はじめに

ほんの数年前まで玉石混交の状態だった医療分野での人工知能 (AI; artificial intelligence) だが、現在はその理解が進んで質や精度が高いもののみが残りつつある。すでに医療用の AI は研究や開発のフェーズから利用のフェーズに移行し、画像診断支援の領域を中心として製品が登場している。国内では内視鏡での検出支援、肺炎の検出、肺の結節検出や、脳動脈瘤検出、乳房超音波での病変検出などが認可されているが、海外と比べると使える製品は少ない。

2016年にAI研究者であるトロント大学のジェフリー・ヒントン博士は「AIが5年以内に放射線科医の能力を凌ぐので、もう放射線科医の育成は止めるべきである」と発言した。しかし、その後のAIの開発は読影の補助に役立ってきてはいるが、AIには総合的な判断を行うことはできず、放射線科医の必要性はさらに高まっている。

我々はAIの理解を深めるために自身のチームでの

AI研究を行ってきた。良質なデータを集めることでAIの研究を行うことが可能であり、その結果を分析することでどのように今後の医療に役立てるかを考えることにも繋がるのがわかってきた。

ここでは医療におけるAIについて知っておくべき技術的な側面を踏まえつつ、我々が行ってきた乳腺領域の画像診断の研究を中心に紹介し、今後の方向性について考えたい。

2. AI とディープラーニング

AIを利用する立場において、技術的な側面も少なからず知っておきたい¹⁾。

AIの定義は研究者によって分かれる。広い意味ではコンピュータさえ使っていれば何でも含まれるが、大まかには「機械が自分で物事を学び、認識し、判断する能力があること」がAIと考えられている。また、AIはいくつかの方法で分類される。

タスクにおける分類から見ると、現時点で「汎用AI」といえるような、何もかもを包括的に扱えるものはな

表1 AI (CAD) の利用形態による分類

セカンドリーダ型	CADなし読影→CAD結果を総合して判定
インタラクティブ型	CADなし読影→必要に応じてCADを参照
同時リーダ型	最初からCAD結果を見て読影
ファーストリーダ型	CADが見るべき画像のみを選定

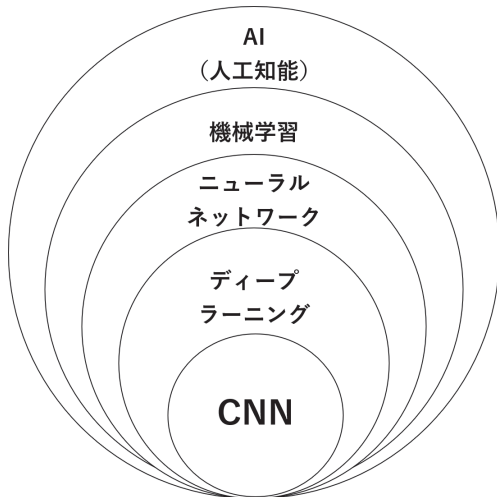


図1 AIのベン図

い。従って、基本的には特定の用途のみに対応した「特化型AI」が用いられている。

以前から一定のアルゴリズムをベースとしたプログラムで動くAIがあり、これらも様々な業務に用いられてきた。近年のAIの発展の多くは事前に用意されたアルゴリズムを用いるものではなく、機械自身が学習する「機械学習」と呼ばれる手法によるものである。その中でも、「教師あり学習」と呼ばれる手法では、多数の画像とその答えを教師データとして用意すると、コンピュータがそれらを自動的に学習して特徴量を見つけ判別を行うことができるようになる。

ディープラーニング (Deep learning ; 深層学習) は複雑化させた neural network を用いた機械学習の一つである (図1)。ディープラーニングの普及が現在のAIの発達をさせてきたといっても過言ではない。その中でもとくに convolutional neural network (CNN) という画像認識に適した手法が用いられている。Convolution (畳み込み) とは、コンピュータが画像から特徴量を抽出する作業に相当し、人間による特徴量の決定が不要である。つまり、学習用の画像データセットを用意さえおけば、コンピュータが自ら特徴量を抽出し、予測モデルが構築される。例えば良性、悪性に分類した画像データセットを用意すれば、コンピュータが自動的に良性か悪性かを判断できるAIを作成してくれるわけである。ディープラーニングでは学習データの量や質によって

AIの精度は大きく変わってしまう。また、過学習によって別のデータに対しては低性能になるというような問題もある。例えば、単施設で精度高く行われた検査画像データの範囲であれば、数百症例程度のトレーニング・データでも人よりも高い診断能のAIを作ることが可能であるが、そのAIを他施設で使えるとは限らないわけである。良性・悪性の比率をどの程度にするかによってもAIの精度は異なると考えられ、やみくもにビッグデータを集めればどうにかなるというわけでもない。

病変検出を行う製品としての分類では大きく CADe (computer aided detection), CADx (computer aided diagnosis) に分けられる。CADeは病変(異常)の検出を補助するものであり、例えばマンモグラフィ読影においてはアルゴリズムベースで腫瘍や石灰化などの病変の検出を補助してくれるものが従来から存在している。CADxは病変(異常)の質的な評価(良悪性など)をおこなうものであり、病変の診断を補助するものである。医師の最終的な診断にも直結する可能性が高く、その後の診療方針に大きく影響することも考えなければならない。Non-CADeとって、臓器の判別など、病変検出を行わないものも分類に含めることがある。

AI (CAD) を利用形態として分類することもある (表1)。最もAI (CAD) に依存したものがファーストリーダ型であり、例えば検診においてCADが異常なしと判定したものは医師が読影を行わないことになるため、効率是非常に高いがそれだけAI (CAD) を信用してよいか、その際の責任の所在はどこにあるのか、といった課題があり得る。

オランダのアンケート調査によると、AIのみでのマンモグラフィ読影を78%の女性が望んでおらず、2次読影のみAIに行わせることも25%の女性は望んでいないと報告されている²⁾。AI導入によって診断が疎かになることを心配されるのかも知れない。受診者の意向も踏まえて導入を考えていく必要もあると考えられる。

現在、医療現場に製品として導入されるAIはすでにディープラーニングなどで学習済みのものであり、そこから自律的な成長はしない。しかし、医療において常にデータは増え続けていき、我々医師が経験に基づいて成長していくようにAIも成長させる必要はある。このよ

うな枠組みをどうしていくかについても、今後の検討が必要である。

3. マンモグラフィでの AI 利用

マンモグラフィは検診における死亡率低減効果が示されていることから世界的に検診で用いられており、アルゴリズムベースで病変検出を補助する CADe 製品が従来から存在する。米国では医師がダブルチェックを行う代わりに CAD を用いた読影が行われることが多く、報告によると 2016 年までに 92% の施設で CAD が導入されているとされている。腫瘍や石灰化を検出してくれることで見逃しを防ぐ効果は期待できるものの、従来の CAD では偽陽性が多い、つまり病変でないものを病変の候補として拾い上げすぎることが多いとされている³⁾。日本でもマンモグラフィ読影用の CADe 製品が承認されているが、さほど普及が進んではいない。

近年はディープラーニングを用いたマンモグラフィの AI 開発が大規模に行われている。2017 年には Dream Challenge という名の下にマンモグラフィの AI コンテスト、総額 100 万ドル。AI のコードを作成し、online でアップロード。最高成績：感度 0.8 での特異度 0.81, AUC=0.87 といった成績が示されている。

2020 年には Google 傘下の deep mind 社の開発された AI が Nature 誌に論文発表されている⁴⁾。AI により、UK 症例では偽陽性を 1.18%、偽陰性を 2.70% 減少させ、USA 症例では偽陽性を 5.70%、偽陰性を 9.40% 減少させた。AI は読影医師 2 名とほぼ同等以上の能力を有していたという。この様に開発が進むだけでなく、実際に最近では諸外国において良悪性判別など病変の質的評価も行う CADx 製品が商用化され、導入され始めている。今のところ日本人女性を含むアジア人女性のデータを含むものは少なく、日本での導入においてはデータの追加や検証作業が必要となるか、あらたに国内での開発にも期待される。

マンモグラフィ読影における問題点の一つに、高濃度乳房 (dense breast) がある。密度が高い乳腺に病変が紛れてしまうことや、乳腺の密度が高ければ乳癌発症リスクも高くなると考えられている。どの程度の乳腺があるかを乳房の構成として分類するが、目視での評価には評価者間の不一致もあるため、AI を用いた乳房の構成の自動判定に期待される。すでに欧米では商用ソフトウェアが広く用いられ、自動判定が良好に BI-RADS 分類と一致することが示されている。日本人を含めたアジア人女性での商用ソフトを用いた大規模なデータに乏しく、検診における有用性についてはさらなる検証が必要と考えられている⁵⁾。

4. ディープラーニングの研究を行ってみる

2017 年頃から医療分野でのディープラーニングを用いた研究が増え始めており、必要な機器、ソフトウェアなどのプラットフォームの情報も容易に得られるようになった。手順通りに行えば自身の PC においても、犬や猫の画像セットを用意しておけば画像判別ができることから、これを医療画像に置き換えれば医療画像 AI ができるとわかってきた。適切なハードをどの様に用意するか、プログラミング技術をどうするかといった問題は残るが、マウス走査だけで本格的な AI 開発が可能となる GUI 深層学習ツール Deep analyzer (GHELIA 社、東京) がインストールされ、AI 開発に必要な GPU を搭載した PC である Deepstation (UEI 社、東京) が通常の PC ハードウェア代金と同程度で購入可能であり、これを用いることにした。

4-1. AI による画像分類に挑戦

最初に我々が取り組んだのが、CNN を使って乳房超音波画像の良悪性鑑別診断を試みることである。犬や猫の画像の代わりに、良性和悪性のそれぞれの超音波画像を用意することが必要になるため、まずはこれらの画像のデータセット作成を行った。後ろ向き研究として乳房超音波検査にて描出された腫瘍 (嚢胞を含む) をランダムに抽出し、腫瘍が入るように皮膚から胸壁の画像をトリミングして jpeg 形式で良性・悪性それぞれのフォルダに保存した。その中から 937 枚 (良性 480, 悪性 457) をトレーニング・データ、120 枚 (良性 48, 悪性 72) とテストデータに分けて利用することとした。

まずはトレーニング・データを Deep Analyzer にアップロードするが、これは良性 (benign)、悪性 (malignant) のそれぞれのフォルダに画像を入れて zip 圧縮したものを読み込むだけで完了する。次に、ディープラーニングのプログラム作業となる。マウス操作で Model として GoogleNet (Inception v2) を、Framework に Chainer を選択し、行う作業を画像分類 (Image Classification) とした。これらのマウスクリックだけで自動でディープラーニングのモデルのプログラムが完成し、Epoch 数 (繰り返し学習) を 50 回として学習を開始した (図 2)。しばらく放置しておくと、診断モデルの構築が完了した。構築されたモデルに画像を読み込ませると、良性・悪性のどちらかを AI が判別した結果が表示される。テストデータに対しての AI の診断能は、感度 95.8%、特異度 87.5%、正診率 92.5%、ROC 解析では AUC 91.3% であった。同一のテストデータに対して 4 年目、9 年目、20 年目の医師による読影試験を行った

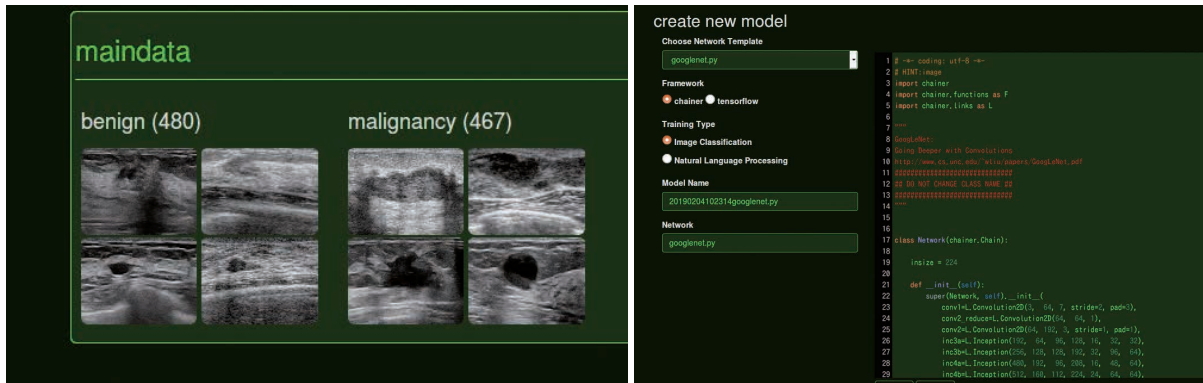


図2 GUIベースのAIソフトウェアの利用画面

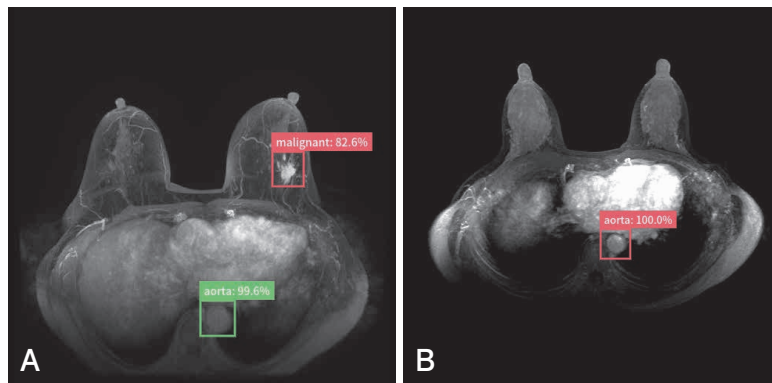


図3 AIによる造影乳房MRI (MIP像) での物体検出
A: 左乳癌症例. B: 正常例.

が, AIの診断は4年目放射線科医より優れ(有意差あり), 9年目および20年目の放射線科医と同等以上(有意差はなかった)という結果となった⁶⁾.

この研究を応用し, 乳房超音波で硬さを評価するエラストグラフィによる良悪性の判別⁷⁾や, 同様の手法で乳房造影MRIのMIP(maximum intensity projection)画像での良悪性の判別にも応用した⁸⁾. どちらも, AIが放射線科医と同等以上の診断能を示していた.

AIが好成績であることはデータセットが単一施設のもので, 質的に担保された画像ばかりであることが大きな理由の一つと考えられるが, 大規模な多施設でのデータ収集を行えば汎用的な乳房超音波診断AIの作成が可能と推測される. AIの診断能は高いが完璧というわけではなく, 読影医のみ正解するものもあるので, AIの特徴を知り診療に活かす必要がある.

4-2. AIによる物体検出

先の研究は, AIによる診断(良性・悪性の判別)を行うものであったが, 医療画像においては物体検出, つまり画像の中のどこに病変があるのかを知る必要がある. Deep analyzerには物体検出(Object detection)のメ

ニューもあり, これを利用した, MRI画像での物体検出においては, 正常例は何も検出せず(動いていることがわかるように正常構造として大動脈(aorta)をマークさせた), 異常(つまり病変)がある際にはそれを検出してマークさせることができ, らに検出した病変の良性・悪性の可能性について示すことが可能であった(図3)⁹⁾.

4-3. AIはどこを見て判断しているのか?

AIの診断根拠はブラックボックス化されていてわからない, と言われることが多い. 医師と同等以上の圧倒的な診断能であっても, 診断根拠がわからなければ医師が診断の責任をとるには何かしら担保できるものが欲しいという考えもあるのであろう. AIによる画像分類の可視化する手法の一つにGrad-CAMがある. 画像のどの部分が予測に最も影響を与えたかをヒートマップとして出力することが可能となる. これを用いた研究も行って見たところ, 造影乳房MRIにおいて正常例では両側の乳房がヒートマップで色がつき, 悪性例では病変のある乳房にヒートマップで色がつく結果となった(図4)⁹⁾.

病変をどのようにAIが良悪性判別したかまでの根拠は得られていないものの, どこを注視したかがわかれば

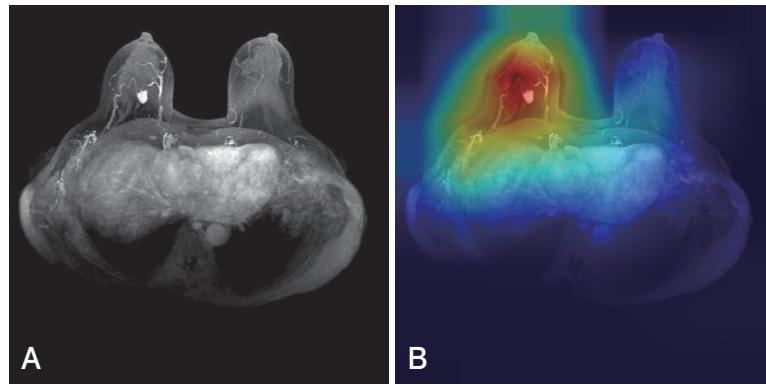


図4 Grad-CAMでAIがどこをみて、右乳癌と判定したのか
A：造影乳房MRI (MIP 像). B：Grad-CAMによるヒートマップ.

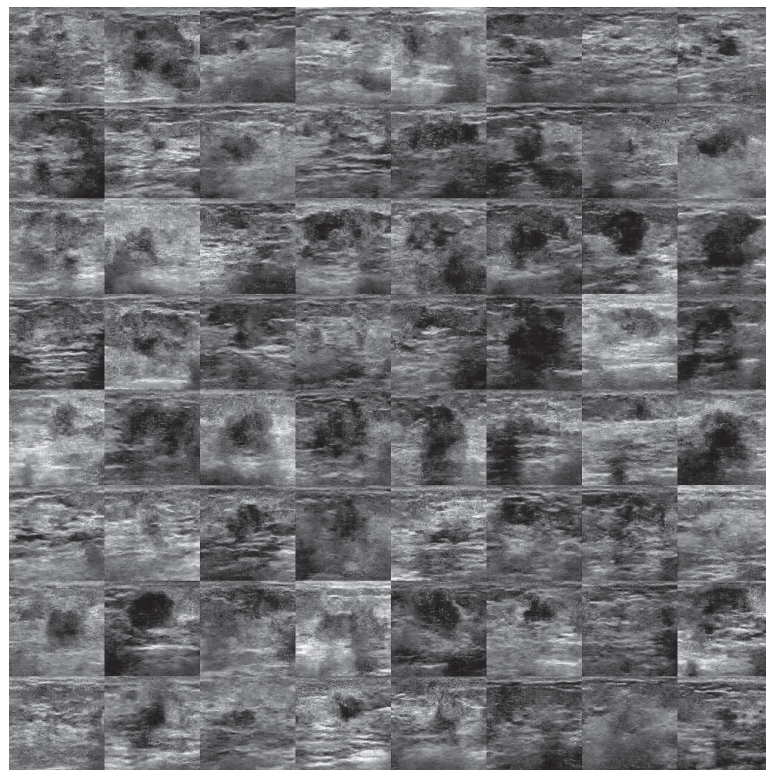


図5 GANで作った乳癌の超音波仮想画像
これらはすべて、AIが作り出した偽物の乳癌の画像である。

少なくとも本当に病変なのかアーチファクトでよいから落として良いかなどの判別はできるし、AIが注視していない部分を人間が見落とさないようにすることもできるかもしれない。

4-4. 画像生成

AIによる画像生成は既に、CTやMRIでのノイズ除去を中心とした deep learning reconstruction は日常臨床において既に多くで用いられている。これらは、高画質の画像を教師画像とすることにより、低線量や低画質

の画像の改善を行うことを可能としている。このように、既にCTやMRIでは画像生成においての実用化が進んでいる。われわれも今後の画像生成の可能性を検討するためにも、いくつかの研究を行ってみた。

Generative Adversarial Network : GAN 敵対的生成ネットワークという画像生成モデルが2014年に発表され、それ以降様々な業界でとても注目を集めている¹⁰⁾。GANにおいては、generator network (画像を生成するもの)と、Discriminator network (本物かどうか識別するもの)の2つのネットワークが敵対的 (Adversarial)

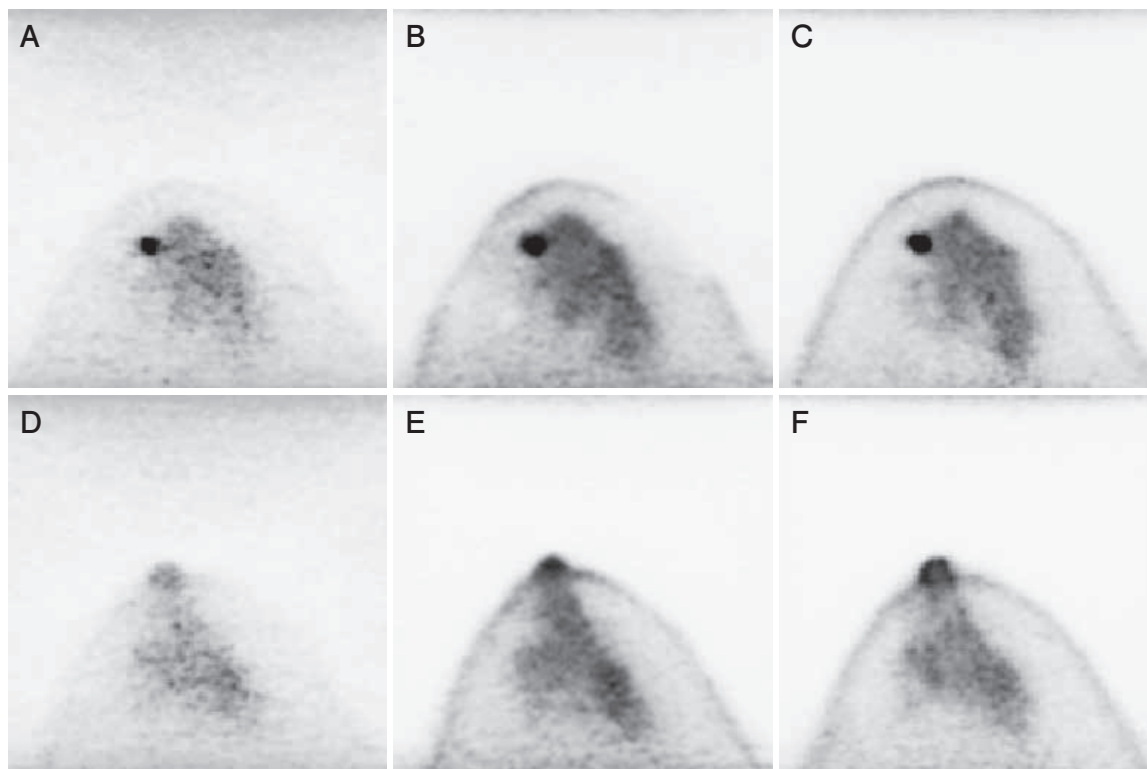


図6 Pix2Pixで作った仮想画像 低画質→高画質への変換

トレーニング済みのAIにinput画像(26秒)としてA(乳癌)およびD(正常)を入力すると、ノイズの少ない画像(BおよびE)が出力され、420秒のスキャン画像(CおよびF)と遜色ない。

に競い合うように学習することで、より実際に近い生成画像を作成する。

GANに対して畳み込みニューラルネットワーク(CNN)を適用し高精度の生成画像を行うDCGAN(Deep Convolutional GAN)というモデルを用い、元画像(本物)から、生成画像(偽物)を作成することを行った¹¹⁾。図は乳房超音波画像を用い、457枚の乳癌の画像に対して500繰り返し学習を行って生成した、いずれも偽物の乳癌画像である(図5)。同様に、良性病変の偽物を作ることもでき、また作成したものから中間的な画像を生成できる。良性あるいは正常から乳癌までの中間画像を作ることで、境界明瞭なものから不明瞭なものを段階的に作ることもできる。GANによって作成した偽画像を用いて医師の教育を行うことや、AIじたいの教育を行うことも可能であろう。

GANの一種でPix2Pixというものもしばしば用いられている。ペア画像の関係を学習し、1枚の画像からその関係を考慮し仮想の画像を作成するものである。この手法により、CT画像からMRI画像を作る、と言ったことも発表されている¹²⁾。我々も、MRIで脂肪抑制を行った画像と脂肪抑制を行っていない画像について3000枚のペア学習をすることで、脂肪抑制なしの画像

から脂肪抑制ありの画像に変換することを試行してみた¹³⁾。このような手法が、画質向上や画像比較などの上でも役立つことに期待される。

また、Pix2Pixで作った仮想画像により、低画質の画像を高画質画像に変換することも十分にできる。PETの撮像においては時間をかけることでカウントが増え、短時間のみでは病変からの集積を十分にカウントできずにノイズが多くなる。49乳房についてのフルカウント(420秒)とローカウント(26秒)の乳房専用PET画像のペアを用いることで、ローカウントの画像だけでもノイズの少ない合成画像を作成することに成功した(図6)¹⁴⁾。このような手法を用いることで、撮像時間の短時間化や投与薬剤が減少できる方向性が見いだせる。

最近では自身のPCを用いずとも、Webサービスを用いたAI研究も可能となってきた。情報管理の上で、外部Webサービスでの研究をどうするかについては問題となる可能性はあるが、自身でAIを作って試してみることは一度やってみる価値があると思う。

5. おわりに

医療AIの現状と、乳腺領域を中心としたAI研究の

実際について画像分類, 物体検出, 画像生成を中心に, 我々のチームで行ってきた実際の研究を紹介しつつ解説した. 臨床において画像診断に AI (とくにディープラーニング) を利用することは, 時間の節約, 術者や読影者の疲労軽減, 経験やスキルの不足を補う, といった用途から非常に重要である. どのように AI を用いていくのかを十分に理解するためには, 各自が AI 研究を少なからず実践してみるのも良いと考える. また, 自施設での診療結果を AI で分析することにより, 施設内でのトレーニングにも役立てることができる. AI は完璧ではなく, 汎用型の医療 AI が登場するまではまだ時間がかかるだろうが, 現時点でも使えるツールをそれぞれのシーンでの診療に活かせるようにしていきたい.

文 献

- 1) 寺本篤司, 篠原範充, 久保田一徳: 乳がん診療に活かすやさしい AI 入門. 中外医学社, 東京, 2022.
- 2) Ongena YP, Yakar D, Haan M, et al : Artificial Intelligence in Screening Mammography : A Population Survey of Women's Preferences. *J Am Coll Radiol* **18** : 79-86, 2021.
- 3) Geras KJ, Mann RM, Moy L. Artificial Intelligence for Mammography and Digital Breast Tomosynthesis : Current Concepts and Future Perspectives. *Radiology* **293** : 246-259, 2019.
- 4) McKinney SM, Sieniek M, Godbole V, et al : International evaluation of an AI system for breast cancer screening. *Nature* **577** : 89-94, 2020.
- 5) 日本乳癌学会編: 乳癌診療ガイドライン 2022 年版② 疫学・診断編. 金原出版, 東京, 2022.
- 6) Fujioka T, Kubota K, Mori M, et al : Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network. *Jpn J Radiol* **37** : 466-472, 2019.
- 7) Fujioka T, Katsuta L, Kubota K, et al : Classification of Breast Masses on Ultrasound Shear Wave Elastography using Convolutional Neural Networks. *Ultrason Imaging* **42** : 213-220, 2020.
- 8) Adachi M, Fujioka T, Mori M, et al : Detection and Diagnosis of Breast Cancer Using Artificial Intelligence Based assessment of Maximum Intensity Projection Dynamic Contrast-Enhanced Magnetic Resonance Images. *Diagnostics (Basel)* **10** : 330, 2020.
- 9) Fujioka T, Yashima Y, Oyama J, et al : Deep-learning approach with convolutional neural network for classification of maximum intensity projections of dynamic contrast-enhanced breast magnetic resonance imaging. *Magn Reson Imaging* **75** : 1-8, 2021.
- 10) I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al : Generative adversarial nets. *Proc. NIPS 2014*, pp.2672-2680, 2014.
- 11) Fujioka T, Mori M, Kubota K, et al : Breast Ultrasound Image Synthesis using Deep Convolutional Generative Adversarial Networks. *Diagnostics (Basel)* **9** : 176, 2019.
- 12) Jin C, Kim H, Jung W, et al : Deep CT to MR Synthesis using Paired and Unpaired Data, <https://arxiv.org/abs/1805.10790>, arXiv preprint, 2018.
- 13) Mori M, Fujioka T, Katsuta L, et al : Feasibility of new fat suppression for breast MRI using pix2pix. *Jpn J Radiol* **38** : 1075-1081, 2020.
- 14) Fujioka T, Satoh Y, Imokawa T, et al : Proposal to Improve the Image Quality of Short-Acquisition Time-Dedicated Breast Positron Emission Tomography Using the Pix2pix Generative Adversarial Network. *Diagnostics (Basel)* **12** : 3114, 2022.

Medicine and AI - Learning about AI Through Own Research

Kazunori Kubota¹⁾, Tomoyuki Fujioka²⁾

¹⁾ *Department of Radiology, Dokkyo Medical University Saitama Medical Center, Saitama, Japan.*

²⁾ *Department of Diagnostic Radiology, Tokyo Medical and Dental University, Tokyo, Japan.*

Artificial intelligence (AI) in the medical field is moving from the development phase to the utilization phase. The central technology in the recent development of AI is deep learning, which is machine learning using complex neural networks. By preparing a large number of images that have already yielded results as training data, a computer can automatically learn from them and then discriminate the images that are given to it. In the medical field, AI products have emerged mainly in the imaging area. In Japan, CAdE (computer aided detection) has been introduced, and CAdx (computer aided diagnosis) is being

introduced in other countries. AI applications in diagnostic imaging include image classification, object detection, and image generation. We have been studying these AI applications based on cases at our institution. In deep learning, the accuracy of AI is determined by the quality and content of training data, and it is possible to create AI with high diagnostic capability even at a single institution by using highly accurate images from daily practice. We would like to understand AI by conducting some AI research independently and consider how to implement AI in medical care as a future challenge.