

INVITED REVIEW ARTICLE

肺癌の画像診断における AI の現状と可能性

木戸尚治¹

Current Status and Potential Utility of Artificial Intelligence in Diagnostic Imaging of Lung Cancer

Shoji Kido¹

¹Department of Artificial Intelligence Diagnostic Radiology, Osaka University Graduate School of Medicine, Japan.

ABSTRACT — Computer-aided diagnoses (CADs) of lung cancer have been studied for a long time. However, the development of a highly accurate CAD for lung cancer has been difficult due to the difficulty of determining appropriate features for various opacities. With the advent of deep learning, the core technology of the third artificial intelligence (AI) boom, it has now become possible to develop CAD systems with higher accuracy and a more general purpose, and expectations concerning the utility of CADs, such as lung cancer detection and differentiation, are increasing. In addition, research on radiomics and radiogenomics, which integrate non-imaging information with imaging information, is ramping up. Many companies are actively developing CAD systems; however, they are not yet fully practical. One issue is that AI cannot explain the diagnostic process, so research on AI that can explain the reason for the diagnosis is important. Five years ago, the possibility of AI replacing radiologists was widely discussed. However, at present, the shortage of radiologists has become a problem. In order for AI to be useful in lung cancer imaging in the future, not only the technical aspects of AI development but also the response of physicians who accept it are required.

(JLCC. 2021;61:282-288)

KEY WORDS — Artificial intelligence, Deep learning, Machine learning, Computer-aided diagnosis, Radiomics/Radiogenomics

Corresponding author: Shoji Kido.

要旨 — 肺癌の画像診断にコンピュータを用いる研究は以前から行われてきたが、肺癌の画像診断支援においては多様な陰影に対して、その特徴量を適切に設計することの困難さから精度の高いコンピュータ支援診断 (computer-aided diagnosis : CAD) システムを開発することは容易ではなかった。しかし、第3次 AI ブームの中核技術である深層学習の登場で、これまでより高精度で汎用的な CAD システムを開発することが可能となり、肺癌の検出や鑑別などの CAD に対する期待が高まっている。また、画像以外の情報と画像情報を統合する radiomics や radiogenomics などの研究も活発化している。

多くの企業による CAD システムの開発も活発であるが、まだ十分に実用的であるとはいえない。課題のひとつとして AI が診断に至る過程を説明できないという問題があり、説明可能な AI の研究が重要視されている。5 年前には放射線科医が AI に取って代わられる可能性が盛んに議論されていたが、現状はむしろ放射線科医の不足が問題化している。今後、肺癌の画像診断において AI が有用性を発揮するためには、AI 開発の技術面だけでなく、それを受け入れる医師の対応が求められている。

索引用語 — 人工知能 (AI), 深層学習, 機械学習, コンピュータ支援診断, Radiomics/Radiogenomics

¹大阪大学大学院医学系研究科人工知能画像診断学共同研究講座。

論文責任者: 木戸尚治。

AIの進歩

人工知能 (artificial intelligence : AI) とは、人間の持つ知的な能力 (知能) を機械的に作り出そうとする研究分野とされる。AIの歴史は、現在に至るまで3回のブームを経験しているが、2010年頃から現在に至るまで続いている第3次AIブームは機械学習 (machine learning) の時代といわれ、その核心的な技術は深層学習 (deep learning) である。機械学習はAIの一分野であり、機械学習の父といわれるSamuelによれば、「明示的にプログラムをしなくても学習する能力をコンピュータに与える研究分野」とされている。機械学習は迷惑メールフィルタなどに用いられているが、機械学習において重要なのは識別器とよばれる画像・音声などのデータに対して、意味を持つクラスに分類するアルゴリズムであり、サポートベクターマシン (support vector machine) やランダムフォレスト (random forest) などがその代表的な手法である。

深層学習は機械学習の一分野であるが、大阪大学のFukushimaにより1980年頃にネオコグニロンとして提唱された階層的・多層化された人工ニューラルネットワークが発想のもとになっている。¹ ネオコグニロンは現在の畳み込みニューラルネットワーク (convolutional neural network : CNN) とほぼ同じ構造を持ち、文字認識などにおいて高い識別精度を示した。ニューラルネットワークは、画像診断分野ではびまん性肺疾患の分類問題などに用いられたが、² 過学習や勾配消失問題がネックとなり使われなくなり、冬の時代を迎えることとなった。しかしながら、これらの問題に対して様々な技術的な解決策が示されたことや、インターネットの普及により大量のデータが利用できるようになったこと、深層学習の計算を高速に実行するための graphics processing unit (GPU) を手軽に利用できるようになったこと、見直されるようになった。

深層学習が再び注目されるきっかけとなったのは、2012年のImageNet Large Scale Visual Recognition Challenge (ILSVRC) とよばれる大規模な画像データベースを用いた画像分類コンテストとされるが、このコンテストでトロント大学のHintonらのチームが8層のCNNを用いて、2位以下のチームに10%以上の大差をつけて圧勝した。そのためにこれ以降はCNNが画像認識におけるデファクトスタンダードとなっている。さらに、2015年のILSVRCにおいてはMicrosoft Research Asiaのチームが152層のCNNを用いて誤識別率3.6%を達成したが、これは人間の誤識別率5.1%を下回るという結果となっている。このようにCNNは自然画像分類に関しては既に人間の認識能力を上回っているとされ、

医用画像診断における病変鑑別などにその高い画像識別能力が期待されている。さらに、画像分類だけでなく画像検出、画像領域抽出、画像生成などにも応用範囲が広がっている。

AIによる肺癌の画像診断支援

CTやMRI装置などの画像診断機器の進歩で高精細な胸部領域画像が極めて短時間で取得できるようになり、肺癌の画像診断に貢献している。しかしながら、このことが画像診断医の負担を増加させており、画像診断医の負担軽減や診断精度の向上を目的としてコンピュータを用いた画像診断支援に関する活発な研究が行われてきた。

胸部単純X線写真やCTなどの画像診断において病変を定量的に評価した結果を医師にセカンドオピニオンとして提示する役割を果たすものとして、コンピュータ支援診断 (computer-aided diagnosis : CAD) が提案されている。CADの特徴はコンピュータによる自動診断ではなく解析結果をセカンドオピニオンとして画像診断医に提供する支援診断である点にあり、1998年には米国のR2 Technology (現 Hologic) 社の乳癌検出CADが米国食品医薬品局 (FDA) の認可を得て、CADが画像診断医から大いに注目を集めるようになった。CADの研究はその後にも様々な疾患やモダリティを対象として行われているが、肺癌などの胸部疾患を対象として臨床で普及しているものはあまりないのが実情である。この理由としては、CADの設計においては病変や臓器の画像特徴設計が重要であるが、多様な画像陰影パターンを含む肺疾患に対しては、開発者側で適切にこれらの特徴量の設計をすることが困難な作業であることが指摘されている。

深層学習が従来の機械学習と決定的に違うのは、対象を識別するための特徴量の設計に関して、深層学習以前の機械学習においては人間側で行う必要があるのに対して、深層学習ではコンピュータ側で自動的に特徴量抽出が可能である。たとえば、肺結節において肺癌の悪性所見のひとつであるスピキュラの特徴量を設計しようとする場合には、その特徴をコンピュータにわかるように定量化する必要がある。また、同じような線状構造を持つ気管支や血管とはどのように違うのかも明確に数量化しなければならない。このような特徴量の設計は開発者にとりCADの性能を決定する重要な作業であるが、開発における最大のボトルネックであり、深層学習によりこれが解消されたことは画期的なブレイクスルーであった (Figure 1)。

肺結節の検出

胸部単純X線写真やCTを用いて肺結節の検出を行

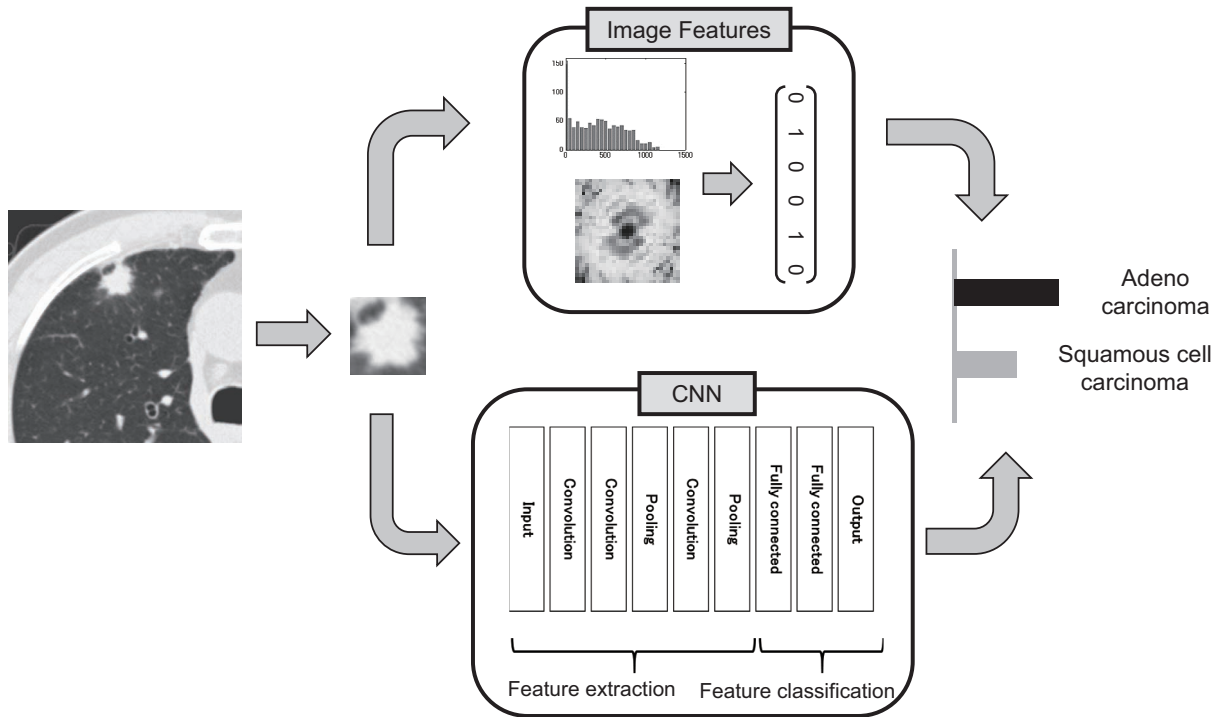


Figure 1. In previous CADs, the imaging feature design was important. However, in CADs using CNN, the imaging features design is not necessary, so a diagnosis based on images can be performed directly.

うことは、画像診断医にとって頻度の高い重要なタスクである。このため肺結節の検出を行うCADの研究・開発が盛んに行われてきた。画像における物体検出のテーマに関しては自動運転などで様々なアルゴリズムが提案されているが、最初に提案されたのは2013年に登場した regions with CNN features (R-CNN)³ である。R-CNN においては、画像中の濃淡情報やテクスチャ情報に基づき数多くの物体候補領域の設定が行われる。これらの物体候補領域は一定の大きさにリサイズされCNNにより特徴抽出が行われて、画像内の物体位置推定がなされる。従来手法では結節が胸壁に接したり、奇静脈陥凹にはまり込んでいたりする場合には、結節とそれに接する胸壁を分離するアルゴリズムを工夫する必要があった。また結節の性状は多彩であり、気管支透亮像を持つ結節やすりガラス結節などに対しても、それぞれ検出のためのアルゴリズムを工夫する必要があったが、R-CNNを用いた手法ではこのような工夫が不要であることにより、多彩な結節性病変に対するCADを统一的に設計することが可能である (Figure 2)。⁴

肺結節の良悪性・組織分類

肺結節の良悪性分類は日常臨床ではよく遭遇する問題である。局所特徴量として speeded up robust features (SURF) を用いて良悪性分類を行った従来手法の場合は

55%というランダムな選択とほぼ変わらない識別率であったのに対し、CNNを用いた場合は86%という高い識別率が得られており、さらに良性和悪性をそれぞれ solid タイプと sub-solid タイプに分けた場合は、従来手法では25%でランダムな選択と変わらないのに対して、CNNを用いた場合は58%と、従来手法に比べて高い識別率が得られた。⁵

また肺結節は三次元的な構造を持つため、三次元の入力画像を用いたCNNによる画像分類が有用であると考えられる。The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) の症例を用いた肺結節の良悪性鑑別における3分割交差検証の結果は $79.4 \pm 3.5\%$ であり、放射線科医の評価とよく一致する結果が得られた。⁶ 三次元画像を用いた解析は結節全体を評価するために有用な手法であるといえるが、三次元画像を入力することは限られたGPUメモリ上では計算の負荷が大きいことや、転移学習を行うための事前学習モデルが利用できないなどのデメリットがある。このため、axial, coronal, sagittal などの多方向からの最大断面を入力画像として用いる手法などが工夫されている。⁷

肺結節の節領域抽出

2015年にDeconvNet⁸が、2016年にはV-Net⁹が提案

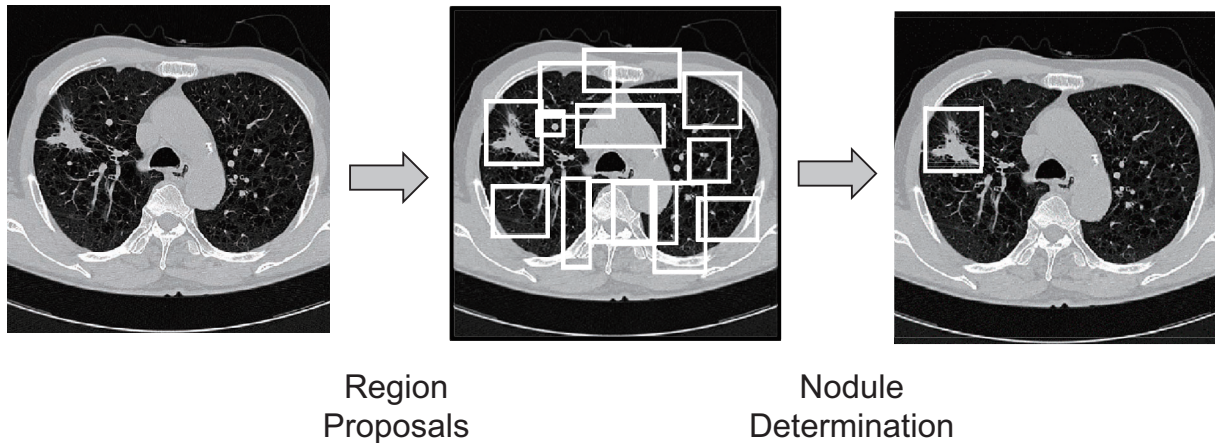


Figure 2. Lung nodule detection using R-CNN. The first step is the region proposals, which generates a large number of candidate regions based on shading and texture information. The next step is the nodule determination. Image features are extracted from a large number of nodule candidate regions using CNN, and the nodule is determined based on these features.

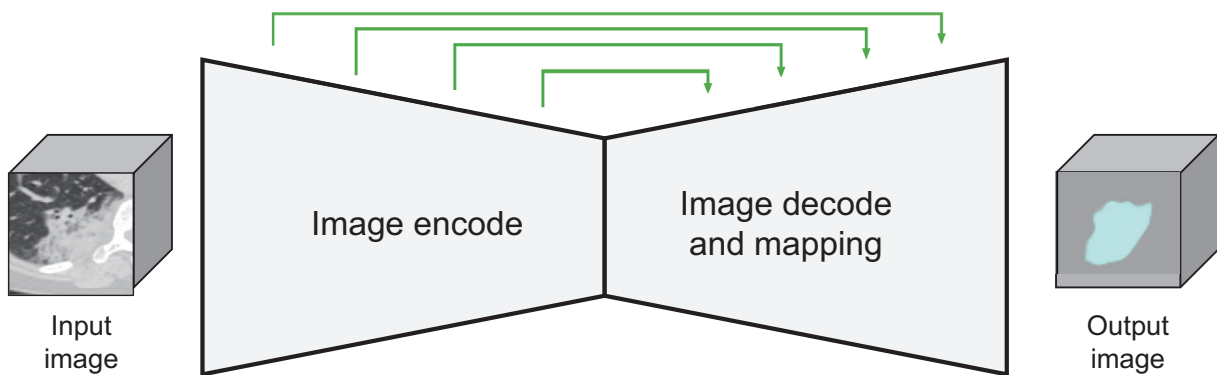


Figure 3. Lung nodule segmentation. In the segmentation model, the image encode step extracts the features of the object, while the image decode and mapping step uses the features to create a region map. Since the corresponding feature maps of the first and second steps are connected, location information can be included when recovering the image.

されているが、いずれも画像の領域抽出を目的としている。DeconvNetは画像分類で用いたCNNと同じく前半部分はconvolution networkであり対象の特徴量の抽出を行うが、後半部分のdeconvolution networkではその特徴量を用いて領域マップの作成が行われる。またV-Netは、U-Net¹⁰と同様に医用画像の領域抽出を目的として提案された手法である。学習用画像として放射線科医の指導のもとにアノテーションされた画像を用い、肺結節を中心として128×128×64画素に切り出されたものを入力して、DeconvNetとV-Netを用いて肺結節の領域抽出を行った結果は、放射線科医によるアノテーションとよく一致し、dice係数ではそれぞれ0.740±0.012, 0.810±0.016と良好な結果が示された (Figure 3).^{11,12}

肺結節の radiomics/radiogenomics

2015年に当時のオバマ米国大統領の一般教書演説において“Precision Medicine Initiative”が発表され、¹³患者個人に最適な治療方法を選択するというオーダーメイド治療が注目を集めている。Personalized medicineは遺伝子情報や様々な臨床情報を統合して個々の患者に最適な治療法を選択を行うことをめざすが、近年、放射線医学領域においてもこのような考えに基づいたradiomicsやradiogenomicsに関する研究発表が増加している。Radiomicsとは、放射線医学に研究分野を意味する接尾辞の「omics」をつけた造語で、放射線画像情報と非画像情報を統合してより高い診断精度をめざしている。肺癌

の radiomics においては、CT 画像から肺癌領域を抽出し、濃度、形状、テクスチャ、ウェーブレットなどの画像情報の抽出を行い、さらに遺伝子情報や臨床情報などから特徴量を作成してより高い質的診断をめざそうとしている。¹⁴ 肺腺癌の CT 画像と遺伝子変異の関係に対する radiomics 解析を行った結果では、放射線科医と比較してより高い診断能を示すことができた。¹⁵ 一般的に radiomics 解析においては、特徴量抽出のために対象となる肺結節の正確な領域抽出が重要であるが、肺結節の領域抽出を日常臨床で行うことは画像診断医にとり負担が大きい作業であり、前述のような CNN を用いた肺結節領域の自動抽出は radiomics の前処理技術として必須となる。

肺癌診断支援の動向と課題

Ardila らは、Google の CNN モデルを用いた肺結節の検出と鑑別を行うシステムの性能評価を行った。教師データとして放射線科医により診断がなされた National Lung Cancer Screening Trial (NLST) の低線量 CT 画像を用いて、モデルのトレーニングとテストを行った。前回 CT が利用できない場合は、このモデルは 6 人の放射線科医の検出能をすべて上回っており、擬陽性を 11%、偽陰性を 5% 減少させることができたが、前回 CT が利用できる場合はモデルの性能は放射線科医と同等であったと報告された。¹⁶ Google の CNN モデルの性能は確かに優れたものであるが、無条件で放射線科医を上回っているわけではないことには注意が必要である。

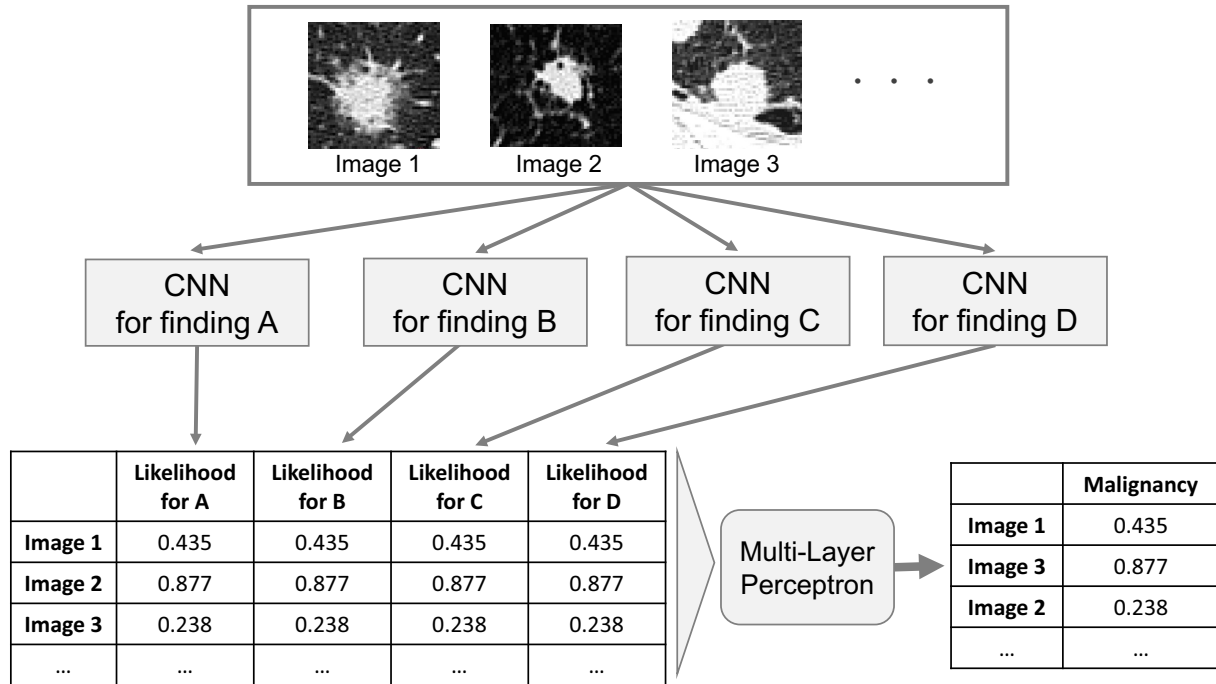
肺癌の診断支援を行うための商用システムで日常臨床において使用できるものはまだ多くはないが、エルピクセル株式会社からリリースされた EIRL X-Ray Lung nodule は胸部単純 X 線写真から肺結節を検出するシステムであり、これを利用した場合は放射線科専門医で 9.95%、放射線科非専門医で 13.1% の感度の向上がみられたと報告されている。米国 Riverain 社の ClearRead シリーズは、読影を行う上で障害となる胸部単純 X 線写真の骨組織の透過 (ClearRead BS) や胸部 CT の肺血管の透過 (ClearRead CT-VS) により、肺結節などの異常陰影の視認性を向上させている。これらのシステムは日本でも薬機法が取得されており利用可能であるが、肺結節を検出する ClearRead Xray | Detect や ClearRead CT | Detect に関しては FDA の承認は受けているが、日本では未承認である。富士フイルム株式会社は胸部 CT 画像から肺結節の検出を行う肺検出機能と、検出した肺結節の性状を分析して医師の所見文作成を支援する肺結節性状分析機能を開発した。後者の肺結節性状分析機能では、肺結節の大きさや辺縁部や内部構造の性状分析を行い、その結果に基づいて所見文の作成を行う。画像診断医の

役割として所見文の作成は重要であり、正しく結節の性状を記述するように支援することは、特に経験の浅い読影医にとっては有用である。¹⁷

AI が正しく認識している元画像にノイズを付加することにより、AI に意図的に画像を誤認識させることができる敵対的サンプル (adversarial example) が知られている。有名な例として GoogleNet に対する敵対的サンプルがある。¹⁸ これはバンダの画像にわずかなノイズデータを付加すると、AI にテナガザルの画像と誤認識させることが可能である。これらの画像は人間ではその違いを認識することができないので、誤認識の原因を明らかにしようとする、AI に推論過程を説明させる必要があるが、深層学習を用いた AI では推論過程を説明させることができない。この問題を解決するために、説明可能な人工知能 (explainable AI : XAI)¹⁹ の技術が注目を浴びている。XAI は、AI の予測結果がどのような推論過程を経て得られたのか説明できないというブラックボックス問題をホワイトボックス化することをめざしている。特に医療の場合は、AI の診断や治療の根拠を医師や患者に説明できなければ、実際に臨床で用いることは困難である。AI の判断根拠を示す手法のひとつに注目部位の可視化があるが、その代表的な手法として gradient-weighted class activation mapping (Grad-CAM)²⁰ がある。この手法では CNN が分類のために注視している範囲をマッピングすることにより、AI の判断根拠を推定することが可能である。またそれを拡張した guided Grad-CAM では、より詳細にどのような特徴をとらえて分類したかが可視化可能である。肺結節の良悪性鑑別においては、Grad-CAM や guided Grad-CAM の結果から、悪性結節では結節内部のすりガラス領域や結節の辺縁領域が着目されているのに対して、良性結節では CNN の注目領域がほとんどないことが示された。²¹ また、肺結節の形状、明瞭性、気管支透亮像などの画像所見に基づいた良悪性鑑別モデルも提案されている。この報告では結節の画像上の各所見に基づく CNN モデルを作成し、これらの所見に基づく尤度を多層パーセプトロンで分類することにより良悪性の鑑別を行った。この結果は 79% 程度と良好であったが、画像から直接良悪性の鑑別を行った場合と有意差はなかった。しかしながら、この方法では画像所見の重みから診断根拠を明確化することが可能である利点がある (Figure 4)。²²

おわりに

第 2 次 AI ブームにおいて伝染性の血液疾患の診断と抗生物質の選択のコンサルテーションを行うようにデザインされたエキスパートシステムの MYCIN の診断精度は 65% 程度で、80% 程度の専門医には及ばないものの一



A: clarity of edges, B: shape, C: notch, D: air-bronchogram

Figure 4. Benign and malignant differentiation model based on imaging findings. A CNN model was created based on each finding of lung nodule CT images, and the likelihood based on these findings was classified by a multi-layer perceptron to differentiate between benign and malignant nodules. Courtesy of Hirano Y, Yamaguchi Univ.

般医よりは優れていたが、臨床の現場で日常的に使われることはなかった。この理由としては MYCIN では対話形式で疾患を絞り込むために、日常診療に忙しい医師にとっては時間がかかりすぎることや、仮に医療事故となった場合の責任の所在はどうするのかという問題が指摘されている。米国においてはマンモグラフィ CAD の普及率が高いが、わずか 2% の放射線科医しか CAD を信用しておらず、半数の放射線科医は CAD を全く信用していないという報告がある。

メリーランド大学の Siegel は CAD の利用と放射線科医の CAD に対する信頼性にはミスマッチがあるとして、次世代の AI が CAD を改善するための 7 つの提言をしている。この提言の中には、AI を放射線科医のワークステーションで利用可能とすることや、読影のワークフローをさまたげてはならないといったこと、AI が放射線科医に対して「whether」ではなく「why」を答えるべきであるとか、放射線科医に質の高いフィードバックをするべきであるというようなことが含まれており、今後の AI 開発において非常に重要な助言であると思われる。²³

現在、世界中で大流行している COVID-19 に対しては、非常に多くの医療画像 AI 関連の論文や製品が発表されているが、医療現場への導入では十分な成果をあげ

ることができていない。画像診断医は、このような AI の技術面と社会面での実装のギャップを埋めるために積極的な役割を果たす必要があると考えられ、そのために AI のリテラシーを身につけて、画像を含む様々な医療データ分析能力を高めていくことが期待される。

本論文内容に関連する著者の利益相反：木戸尚治 [企業などが提供する寄附講座] 富士フィルム株式会社

謝辞：本研究は JSPS 科研費 JP21H03840 の助成を受けた。

REFERENCES

1. Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern.* 1980;36:193-202.
2. Ashizawa K, Ishida T, MacMahon H, Vyborny CJ, Katsuragawa S, Doi K. Artificial neural networks in chest radiography: Application to the differential diagnosis of interstitial lung disease. *Acad Radiol.* 1999;6:2-9.
3. Girshick R, Donahue J, Darrell T, Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014:580-587.

4. Kido S, Hirano Y, Hashimoto N. Detection and Classification of Lung Abnormalities by use of Convolutional Neural Network (CNN) and Regions with CNN Features (RCNN). International Workshop on Advanced Image Technology. 2018.
5. Kido S, Hirano Y, Hashimoto N. Computer-aided Classification of Pulmonary Diseases: Feature Extraction Based Method versus Non-feature Extraction Based Method. International Workshop on Advanced Image Technology. 2017.
6. Ito T, Hirano Y, Suzuki K, Kido S. First-reader computerized system for distinction between malignant and benign nodules on thoracic CT images by means of end to end deep learning: convolutional neural network (CNN) and neural network convolution (NNC) approaches. RSNA. 2017.
7. Setio AA, Ciampi F, Litjens G, Gerke P, Jacobs C, van Riel SJ, et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. *IEEE Trans Med Imaging*. 2016;35:1160-1169.
8. Noh H, Hong S, Han B. Learning of deconvolution network for semantic segmentation. Proceedings of the IEEE International Conference on Computer Vision. 2015:1520-1528.
9. Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *IEEE*. 2016:565-571.
10. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015 Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015:234-241.
11. Kidera S, Kido S, Hirano Y, Mabu S, Tanaka N. Segmentation of lung nodules on MDCT images by using 3D Conv-DeconvNet, Computer Assisted Radiology and Surgery. 32th International Congress and Exhibition (CARS2018). 2018.
12. 木寺俊介, 木戸尚治, 平野 靖, 田中伸幸. DeconvNet と V-Net を用いた 3 次元 CT 画像中の肺結節領域抽出. 信学技報. 2019;118:103-106.
13. Precision Medicine Initiative. <https://obamawhitehouse.archives.gov/node/333101>
14. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
15. Odo T, Kido S, Hashimoto N, Hirano Y, Tanaka N. Radiomic analysis for lung adenocarcinoma. Proceedings of the EMBC2017. 2017.
16. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019;25:954-961.
17. Wataya T, Yanagawa M, Yamagata K, Suzuki Y, Kido S, Tomiyama N. Can Deep learning-based Reading Support System Aid Lower Experienced Radiologists To Recognize Accurately The Features Of Pulmonary Nodules/masses? RSNA2021. (submitted)
18. Goodfellow IJ, Shlens J, Szegedy C. Explaining and Harnessing Adversarial Examples. arXiv. 2015:6572.
19. Core M, Lane H, Lent M, Gomboc D, Solomon S, Rosenberg M. Building Explainable Artificial Intelligence Systems. *AAAI*. 2006:1766-1773.
20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Proceedings of the IEEE International Conference on Computer Vision. 2017:618-626.
21. 平島 翔, 平野 靖, 木戸尚治, 岩野信吾, 本田 健, 関順彦, 他. 胸部 CT 像中の肺結節の良悪性鑑別における自動抽出された画像特徴の可視化. 第 38 回日本医用画像工学会大会. 2019.
22. 福嶋一茂, 平野 靖, 木戸尚治, 岩野信吾. ディープラーニングを用いた胸部 CT 像中の肺結節の画像所見の検出および良悪性鑑別. *Medical Imaging Technology*. 2019; 37:244-254.
23. Casey B. 7 steps to better AI algorithms. 2017. <http://www.auntminnie.com/index.aspx?sec=ser&sub=def&page=dis&ItemID=117553>