

2022年8月30日

報道関係者各位

慶應義塾大学医学部

**日本初の乳がん超音波検診における
精密検査の要否判定を行う人工知能(AI)を開発**

— 専門医を凌駕する精度で画像診断を行い、乳がん検診の精度向上に貢献 —

慶應義塾大学医学部外科学教室（一般・消化器）の林田哲専任講師、北川雄光教授を中心とする多施設共同研究グループは、株式会社フィックスターズと共同で、ディープラーニング技術を用いた人工知能（AI）による画像診断システムを開発しました。このシステムは乳房超音波検査を対象としたもので、乳がん検診を受診した患者が、さらなる精密検査を受けるべきかどうかを高い精度で判定可能であることが明らかになりました。

ディープラーニング技術の一つである Convolutional Neural Network (CNN)（注1）と呼ばれる技術を利用して開発された AI は、乳房超音波画像診断の国際的基準である BI-RADS 判定基準（注2）に基づいた診断を行い、対象となる超音波画像に精密検査が必要な病変を含むかどうかを判定することを目的に開発されました。

この AI を用いて、教師データとは異なる 3,166 枚の乳房超音波画像（3,656 病変を含む）を対象として診断を行ったところ、感度 91.2%・特異度 90.7%（注3）の精度で診断が可能であることが明らかになりました。この結果は、日本乳がん検診精度管理中央機構が認定する「乳がん検診超音波検査実施・判定医師」の合格基準が感度 80%・特異度 80%であることを考慮すると、これらを大幅に上回る精度での診断結果でした。

さらに外科専門医 10 名を含む、20 名の医師による乳房超音波画像診断の結果と、AI による診断結果を比較したところ、AI が統計学的に有意に優れた精度で診断しました。

今後、乳がん検診や人間ドックなどにおける乳房超音波によるスクリーニング検査には、この AI システムが医師の診断補助として利用され、見逃しや過剰診断を防ぐことで、精度の向上や施設間格差の解消といった医療技術等の格差の是正に貢献することが期待されます。

本研究成果は、2022年8月3日（米国時間）に国際科学誌 *Cancer Science* オンライン版に掲載されました。

1. 研究の背景と概要

国内では乳がんの早期発見を目的とした乳がん検診が広く行われており、各自治体が提供する対策型検診や人間ドックなどの任意型検診では、マンモグラフィー検査と超音波検査が主要な検査方法として用いられています。

日本を含むアジア人女性は、乳腺の密度が高い「高濃度乳房」が多く、また乳がんの発症

年齢が欧米女性より若い 40 代～50 代に多いため、相対的に超音波検査の有用性が高いと考えられています。実際に日本で行われた大規模な臨床試験（J-START 試験）の結果から、乳がん検診において、超音波検査を併用することで、マンモグラフィー単独の検査に比べて早期乳がんの発見率が高まることが明らかになっています。

乳房超音波検査は、患者の身体を医師や技師が超音波装置を用いて直接調べる「生理機能検査」（注 4）に分類され、その精度は機材の良し悪しや、検査者の経験・疾患に対する知識・異常所見の見極め方などに左右されることが知られています。

そのため、乳がん検診における超音波検査の需要の増加に対応するには、検査技師や医師の育成および診断技術の向上が重要な課題と考えられていますが、十分な成果は得られていません。実際に乳腺診療を専門としない医師が検診やドックの最終的な判定を行う事例が数多く存在しています。

このような背景から研究グループは、近年目を見張る発展を遂げたディープラーニング技術を用いた人工知能（AI）を用いて、乳腺超音波検査の診断システムを構築し、診断精度の向上に貢献できるか検討を行いました。

2. 研究の成果と意義

この AI 診断システムの開発は、乳房超音波検査における病変の有無とその診断に、人間の能力を超えた客観性を持たせることを目指して開始されました。人の目では判断しきれない超音波画像の特徴を AI が読み取り、正確な診断を瞬時に提示することを目的としています。

これを実現するために、コンピュータの性能を最大限に引き出す高速化技術を社会に提供している株式会社フィクスターが開発した Convolutional Neural Network（CNN）をベースとしたディープラーニング技術を利用し、まず初めに慶應義塾大学が提供する約 1500 枚のアノテーション（注 5）済み乳房超音波画像を教師データとして学習させた AI 診断システムの構築を最初に行いました。その結果、AI が高い精度で検査画像中の腫瘍を認識して、良悪性の判定を一枚あたり 0.01 秒以下で行うことが可能でした。

この AI を臨床応用するために、さらなる精度向上を目指して、慶應義塾大学を含む以下の 8 施設と共同研究グループを構築し、7,194 枚の乳房超音波画像を収集し、その全てにアノテーション作業を行いました。これを教師データ（4028 枚）とテストデータ（3166 枚）に二分し、AI システムの構築と検証に使用しました。

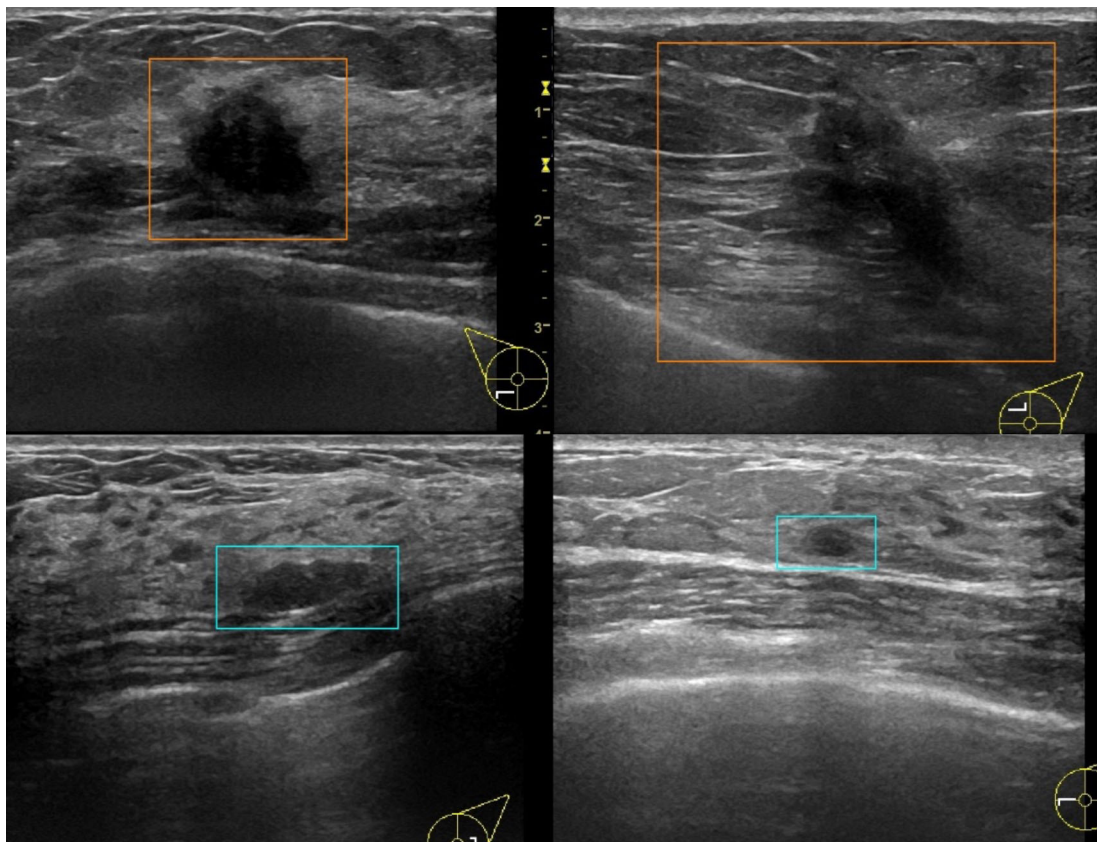
- ・ 慶應義塾大学医学部
- ・ 帝京大学医学部
- ・ 杏林大学医学部
- ・ 国立がんセンター中央病院
- ・ 国立がんセンター東病院
- ・ 埼玉医科大学国際医療センター
- ・ 北里研究所病院
- ・ 東京医療センター

乳房超音波検査において、病変の良悪性を鑑別することは重要な役割の一つですが、超音波検査のみで全ての病変を 100%の精度で診断することは不可能です。そのため、実際の臨床では検査で乳房に異常を発見した際に、「次にどのような行動を行うか」という判断が最も

大切なことと考えられています。すなわち、追加で画像検査を行うか、生検を行うか、経過観察でよいのかなど、発見した病変に対して適切な診療行為を行うことが必要です。

乳房超音波検査の国際的な判定基準である BI-RADS 基準では、病変の特徴をもとに 1 から 5 までのカテゴリに分類し、各カテゴリの病変が乳がんである頻度と、次に行うべき行動が規定されています。またこの BI-RADS 基準を検診や人間ドックなどのスクリーニング検査に適用する場合は、BI-RADS カテゴリ4 以上と判定された患者は組織生検を含む精密検査を行うことが強く推奨されます。

これらの背景から、本研究グループは AI の臨床応用を目指すにあたり、良悪性の判断を行うのではなく、乳房超音波検査画像に含まれる病変が、この BI-RADS 基準において乳がんの頻度が高まる BI-RADS カテゴリ4 以上なのか、乳がんの可能性がほとんどない BI-RADS3 以下なのかを判定する AI 診断システムを構築しました。(図 1)

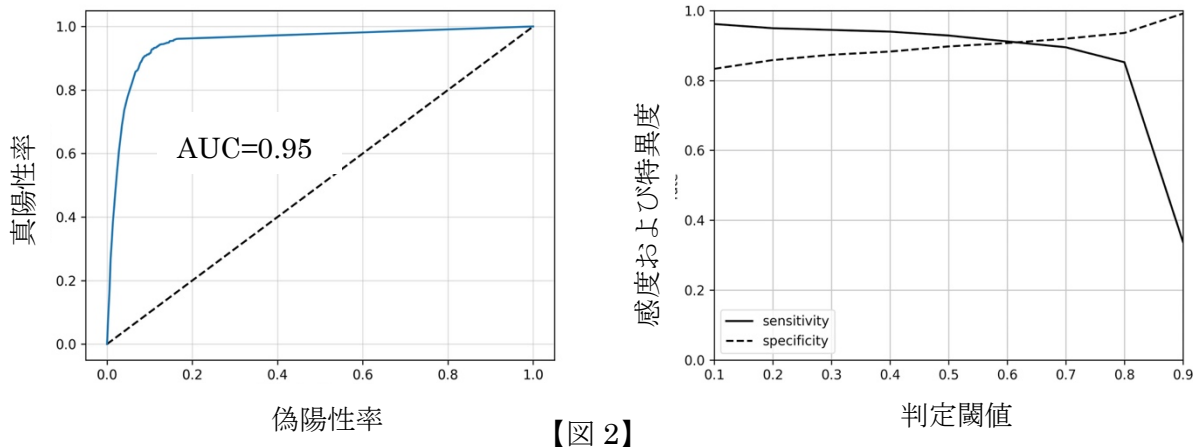


【図 1】

AI が病変を認識し、BI-RADS4 以上（精密検査が必要）ならばオレンジ、BI-RADS3 以下（精密検査は必要なし）ならば青枠で表示

この AI を用いて、前述の教師データとは異なるテストデータ（3,166 枚の乳房超音波静止画像）を対象として診断を行ったところ、感度 91.2%・特異度 90.7%の精度で診断が可能であることが明らかになりました。

判定の閾値を変化させて描いた ROC 曲線（注 6）における AUC（注 7）の値は 1 に近づくほど診断システムの精度がよいことを表しますが、この AI 診断システムの AUC は 0.95 であり、非常に精度が高いことが示されました（図 2）。



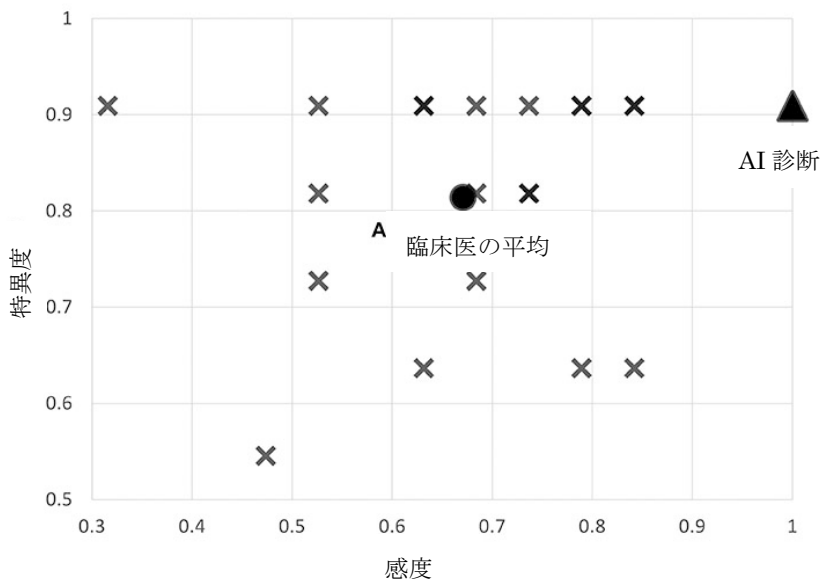
【図 2】

診断の閾値を変化させた際の ROC カーブ（左）と感度・特異度の変化（右）

日本乳がん検診精度管理中央機構が認定する「乳がん検診超音波検査実施・判定医師」は、乳房超音波検査を行う医師および技師が十分な水準の技量と診断能力をもつことを証明するものです。

この認定試験では、静止画のみではなく動画の判定も評価基準に含まれるため、今回の結果と単純な比較はできないものの、合格基準が感度 80%・特異度 80%であることを考慮すると、AI はこれらを凌駕する精度での診断結果を示し、最高クラスの診断能力を持つ専門医・放射線技師と同等以上の能力を持つことが期待されます。

実際に、この AI 診断システムと 10 名の外科専門医含む計 20 名の臨床医による、30 枚の乳房超音波検査画像に対する診断精度比較を行ったところ、感度・特異度ともに統計学的に有意な差をもって AI が優れた精度での診断を行うことが示されました。（図 3）



【図 3】

20 人の臨床医と AI による診断の感度・特異度

3. 今後の展望

これらの研究結果から、開発された乳房超音波検査の AI 診断システムは、乳がんの可能性が否定できない精密検査が必要な患者を適切に診断可能であることが明らかとなりました。

今後、本研究グループはこの AI 診断システムを社会実装するために薬事承認へのプロセスを進め、乳がん検診や人間ドックなどでスクリーニングとして実施される乳房超音波検査において、医師が診断の補助として利用することを目指しています。

AI の補助により診断精度の向上がなされれば、乳がんの見逃しを防ぐことはもちろん、乳がんの可能性がないのに要精密検査と判定してしまう過剰診療を防ぐことで、患者の身体的・精神的負担のみならず不要な医療費の増加を抑制することが期待されます。また、診断能力の差による施設間の格差も是正されることで、日本全国どこでも同じレベルの検査が可能となるといった医療技術等の格差の是正が図られると考えられます。

4. 特記事項

本研究で用いられた AI アルゴリズムは、株式会社フィクスターズより提供されました。また、同社から提供された研究資金が AI 診断システム構築のために使用されました。さらに、本研究の遂行には JSPS 科研費 JP20K08993 による支援を受けて行われました。

5. 論文

英文タイトル : Establishment of a deep-learning system to diagnose BI-RADS4a or higher using breast ultrasound for clinical application

タイトル和訳 : 臨床応用に向けた乳腺超音波診断のためのディープラーニングシステムの構築
著者名 : 林田哲、小谷依里奈、菊池雅之、永山愛子、関朋子、高橋麻衣子、二木紀行、松本暁子、村田健、綿貫瑠璃奈、横江隆道、前田日菜子、大西達也、麻賀創太、北條隆、神野浩光、五月女恵一、松井哲、首藤昭彦、井本滋、北川雄光

DOI : 10.1111/cas.15511.

掲載誌 : *Cancer Science* オンライン版

【用語解説】

- (注 1) Convolutional Neural Network (CNN) : AI が画像認識を行う際に使用されるディープラーニング技術の代表的手法。
- (注 2) BI-RADS : breast imaging reporting and data system の略。米国放射線専門医会が中心となって作成された、マンモグラフィ・超音波・MRI の読影用語・所見に基づいたカテゴリー分類と、報告書の記載方法の標準化を図るガイドライン。
- (注 3) 感度・特異度 : 病気を診断する際の正確さを表す尺度。感度は疾患のある人のうち検査結果が陽性となる割合。特異度は疾患のない人のうち検査結果が陰性となる割合。
- (注 4) 生理機能検査 : 検査者が直接患者の身体に触れて検査を行う生体検査
- (注 5) アノテーション : AI に覚え込ませる教師データを作成するために、テキストや音声、画像などあらゆる形態のデータにタグ (注釈) を付ける作業のこと。本研究では、画像上の腫瘍の位置と、その腫瘍に精密検査が必要かどうかのタグを作成した。
- (注 6) ROC 曲線 (Receiver Operating Characteristic curve) : 診断検査の有用性を検討する手法。縦軸を感度、横軸を偽陽性率 (1-特異度) として診断の閾値を変化させた際のそれぞれの値をプロットしたグラフのこと。

(注7) AUC (Area Under Curve) : ROC 曲線の下部分の面積で、0.5~1.0 の値をとる。この AUC が高ければ高いほど診断能力が高いと判断でき、一般的には AUC > 0.9 であれば高精度と考えられている。

※ご取材の際には、事前に下記までご一報くださいますようお願い申し上げます。

※本リリースは文部科学記者会、科学記者会、厚生労働記者会、厚生日比谷クラブ、各社科学部等に送信しております。

【本発表資料のお問い合わせ先】

慶應義塾大学医学部 外科学教室 (一般・消化器外科)

専任講師 林田 哲 (はやしだ てつ)

秘書 小野 祐子 (おの ゆうこ)

TEL : 03-5363-3802 FAX : 03-3355-4707 E-mail : yukorascal@keio.jp

【本リリースの配信元】

慶應義塾大学信濃町キャンパス総務課 : 山崎・飯塚・奈良

〒160-8582 東京都新宿区信濃町 35

TEL : 03-5363-3611 FAX : 03-5363-3612 E-mail : med-koho@adst.keio.ac.jp

<https://www.med.keio.ac.jp>

※本リリースのカラー版をご希望の方は【本リリースの配信元】までご連絡ください。