

平成 30 年 7 月 12 日

報道関係者各位

国立大学法人 筑波大学
京セラコミュニケーションシステム株式会社

皮膚腫瘍の良悪性を判定する人工知能診断補助システムを開発 ～専門医よりも正確な診断が可能に～

研究成果のポイント

1. 臨床画像から、14 種類の皮膚腫瘍を識別する人工知能 (AI) 診断補助システムを開発しました。
2. 従来、AI の学習に必要とされていた画像数の半数以下の画像でも、90%以上という高い識別率を達成しました。
3. 腫瘍の良悪性の識別に関して、皮膚科専門医よりも AI 診断補助システムの方が正確に行えることがわかりました。

国立大学法人筑波大学医学医療系の藤本学教授、藤澤康弘准教授と京セラコミュニケーションシステム株式会社の共同研究グループは、筑波大学が所蔵している臨床写真を用いて、90%以上という非常に高い診断精度を有する皮膚腫瘍人工知能診断補助システムを開発しました。

通常、AIのディープラーニングによる画像の識別には1つのカテゴリごとに最低1,000枚の画像を用いた学習が必要であるとされており、今回、本研究グループが設定した14種類の皮膚腫瘍を識別するシステムの構築には14,000枚以上の画像が必要となります。しかし本研究グループは、その半分以下である約6,000枚の臨床写真を用いたAI学習でも、皮膚腫瘍の良悪性を判断できる割合（識別率）が90%を超えるシステムが構築可能であることを示しました。このような高い識別率を達成した要因には、教師画像として使用した臨床写真の大半が、病理組織診断で診断が確定している質の高いデータであるという点と、学習前の画像処理の段階で京セラコミュニケーションシステムがこれまでに蓄積してきた画像解析ノウハウを活かし、さまざまな工夫を加えた点が挙げられます。

今回開発した診断システムを評価するため、同じ画像セットを診断するテストを用いて、AI診断補助システムと日本皮膚科学会認定皮膚科専門医13名とを比較したところ、皮膚科専門医による皮膚腫瘍の良悪性の識別率が85.3%±3.7%であったのに対して、AI診断補助システムの良悪性の識別率は92.4%±2.1%と有意に高いことが明らかとなりました ($P<0.0001$, Welch's t-test)。良悪性の識別より難しい14種類の詳細な診断の正答率でも、皮膚科専門医が59.7%±7.1%であったのに対して、AI診断補助システムの正答率は74.5%±4.6%であり、こちらも有意に優れていました ($P<0.0001$, Welch's t-test)。

皮膚腫瘍の良悪性が写真で判定できるようになれば、皮膚科専門医が不足している地域において皮膚がんの早期発見が可能となり、必要な治療が手遅れになる前に受けられるようになります。今後は診断精度の向上だけでなく、皮膚がん以外の皮膚疾患も診断出来るよう、さらに研究を進めていきます。

* 本研究成果は、6月28日付「British Journal of Dermatology」でオンライン公開されました。

研究の背景

皮膚疾患の診断で最も大きなウェートを占めるのは「視診」です。しかしながら、例えば悪性黒色腫（メラノーマ）とよばれる皮膚がんは、発見が遅れると死に至ることもあります。いわゆる「ほくろ」と見た目が非常に似ています。そのような背景から、古くから画像診断技術によるほくろとメラノーマの鑑別に関する研究が行われてきました。しかし、一般的な臨床写真から得られる情報だけではその判断は難しく、特殊な10倍の拡大鏡（ダーモスコープ）を用いて撮影された画像によるAI診断補助システムの開発が進められており、皮膚科専門医と変わらない診断精度が報告されています（参考文献1）。ダーモスコープによる画像は常に一定の距離で撮影されることから、肺のエックス線画像のようにある程度規格化された画像となり、AIに非常に適していることがその成功の背景にあると考えられます。

一方で、一般の臨床画像は写真を撮影するときの条件がバラバラであることから、ダーモスコープと比べて画質を均一化することが難しいと考えられます。さらにメラノーマの場合、そもそも日本ではアメリカと比べて非常に症例数が少なく（年間1,500症例対90,000症例程度）、ほくろとメラノーマの判別だけ出来たとしても活用出来るシチュエーションはあまり多くありません。そのため本研究グループでは、メラノーマ以外の良悪性を含む14種類の皮膚腫瘍を診断するための、ディープラーニングを用いたAI診断補助システムを2017年から開発してきました。昨今のAIブームの火付け役となったディープラーニングは、生体の脳の情報処理を人工的に模倣したニューラルネットワークがベースになっています。ニューラルネットワークを多層にすることで、画像等データに含まれる特徴をより詳細に学習することを可能にする技術です。これを導入することで、画像の判別システムの精度は飛躍的に向上しており、すでに一般的な物体の認識では人間と遜色ないレベルに達しています。

しかしながら、このディープラーニングは、学習のために多数の画像をシステムに投入する必要があります（分類1つにつき各1,000枚以上、例えば分類が10種類あれば合計10,000枚以上）、希少な疾患が多い皮膚疾患の診断システムの構築には不向きです。実際、2017年1月にはStanford大学のグループが、皮膚がんの識別精度が皮膚科医と変わらないAI診断補助システムを開発したとの報告（参考文献2）がありますが、その論文では、ダーモスコープを用いて撮影されたものを含む129,450枚の画像が使用されています。

そこで本研究グループは、一桁少ない画像数で高い精度を達成できる技術開発を行うことを目標としました。

研究内容と成果

筑波大学でこれまで蓄積してきた臨床写真の中で、病理検査を行い皮膚腫瘍の診断が確定したものを中心に収集し、診断ごとに仕分けた上で間違った写真が含まれていないかチェックを行い、診断違いの写真を除いた結果、14種類の皮膚腫瘍に関する約6,000枚のデータセットが得られました。このデータセットからテスト用の1,200枚を取り分け、残りの4,800枚をディープラーニングの学習データとして使用しました。ディープラーニングは、120万枚の一般的な物体画像により事前学習されているGoogLeNet^{注1)}をベースとして使用し、学習データは皮膚腫瘍が画面の中心に来るようにトリミングした後に、1,000x1,000ピクセルにリサイズ、その後15度ずつ傾けながら別画像として保存することで、1枚の写真から24枚の学習画像を作成しました。学習時には画像にわざとぼかしを加えたり、明るさを±10%の範囲で変化させたりすることで、撮影時の状況のばらつきを含めて学習できるように工夫しました。

その結果、皮膚悪性腫瘍の判定感度96.5%、特異度が89.5%という高い診断精度を達成しました。実際にこの診断精度の有効性を示すため、同じ画像セットを診断するテストにより、AI診断補助システムと日本皮膚科学会認定皮膚科専門医13名とを比較したところ、皮膚科専門医による良悪性の識別率が85.3%±3.7%であったのに対して、AI診断補助システムの良悪性の識別率は92.4%±2.1%と有意に高いこ

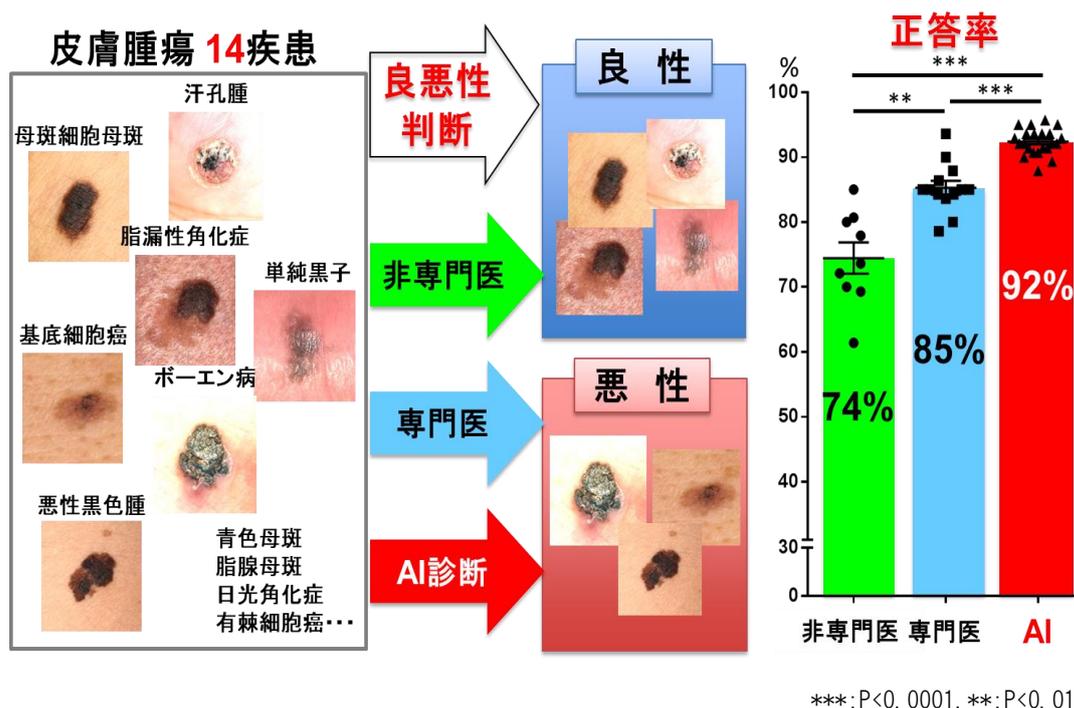
とが明らかとなりました ($P < 0.0001$, Welch' s t-test、参考図)。良悪性の識別より難しい14種類の詳細な診断の正答率についても、皮膚科専門医が59.7%±7.1%であったのに対して、AI診断補助システムの正答率は74.5%±4.6%であり、こちらもAI診断補助システムの方が有意に優れていました ($P < 0.0001$, Welch' s t-test)。

今後の展開

今回開発した皮膚腫瘍 AI 診断補助システムは、十分な性能評価を行った上で、数年以内に実際の臨床の現場で使用できるようになることが目標です。皮膚腫瘍の良悪性が写真で判定できるようになれば、皮膚科専門医が不足している地域においても皮膚がんの早期発見が可能となり、手遅れになる前に必要な治療が受けられるようになります。

また、現状のシステムには非常にまれな皮膚腫瘍が含まれていないため、今後は対応する診断名を増やすことでさらに実用的なものに発展させていくことが必要です。さらに、皮膚腫瘍だけでなくその他の皮膚疾患も診断が出来るシステムに拡張し、あらゆる皮膚疾患に対応できるようにバージョンアップを進めていきます。

参考図



本研究で開発した AI 診断補助システムと日本皮膚科学会認定皮膚科専門医の皮膚腫瘍の良悪性識別率比較

用語解説

注 1) GoogLeNet

ニューラルネットワークの一種であり、2014年に画像認識コンペティションで優勝したモデルです。従来の画像認識技術は、画像の特徴をどのように数値化するかを人間が長い時間をかけて複雑に設計する手法が主流でしたが、2012年の画像認識コンペティションにおいてニューラルネットワークを用いた AlexNet が、従来の方法に大差をつけて優勝したことで一躍注目されるようになりました。

現在では画像認識の領域では標準的な技術となっており、音声や翻訳などの他分野でも活用されています。

参考文献

- 1) Ferris LK et al, Computer-aided classification of melanocytic lesions using dermoscopic images, J Am Acad Dermatol, 2015
- 2) Esteva A et al, Dermatologist-level classification of skin cancer with deep neural networks, Nature, 2017

掲載論文

【題名】 Deep learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumor diagnosis

(ディープラーニングを用いたコンピューター診断補助装置は、教師画像の数が少なくても皮膚科専門医より正確に皮膚腫瘍を診断する)

【著者名】 Y. Fujisawa Y. Otomo Y. Ogata Y. Nakamura R. Fujita Y. Ishitsuka R. Watanabe N. Okiyama K. Ohara M. Fujimoto

【掲載誌】 British Journal of Dermatology (doi: 10.1111/bjd.16924)

*掲載されている情報は、発表日現在の情報です。