

令和 6 年 5 月 17 日現在

機関番号：16301

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K08339

研究課題名（和文）ディープラーニングを用いた人工知能による皮膚病分類システムとデジタルバイオプシー

研究課題名（英文）Development of artificial-intelligence skin disease classifier and digital biopsy by using deep learning

研究代表者

藤澤 康弘 (Fujisawa, Yasuhiro)

愛媛大学・医学系研究科・教授

研究者番号：70550193

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：正答例と誤答例におけるどのような因子が診断に影響しているかをGradCamを使用し、検討したところ、当初はヒートマップにおいて腫瘍中心部に注目していない場合に正答率が下がると考えていたが、興味深いことに正答例でも誤答例でもヒートマップの分布にあまり違いが見られなかった。そこで画像の中央に着目するように設定して学習をさせてみたところ、これも興味深いことに全体を用いた場合と比べて正答率が低下することが分かった。これは先のGradCamによる結果と附合するものであり、腫瘍の判定において中央部の腫瘍部分だけでなくその周囲の情報も判定に用いられているということになる。

研究成果の学術的意義や社会的意義

GradCamの解析によるとヒートマップにおいて腫瘍中心部に注目していない場合に正答率が下がると考えていたが、興味深いことに正答例でも誤答例でもヒートマップの分布にあまり違いが見られなかった。逆の味方をする、ヒートマップで腫瘍部分に注目していなくとも正答してしまっている画像もかなり含まれていることを示している。また、そこで画像の中央に着目するように設定して学習をさせると逆に全体を用いた場合と比べて正答率が低下することが分かった。皮膚腫瘍の判定において中央部の腫瘍部分だけでなくその周囲の情報も判定に用いられているということになる。今後の機械学習におけるアノテーションの範囲にも検討が必要となる。

研究成果の概要（英文）：After examining the factors influencing diagnosis in both correct and incorrect examples using GradCam, it was initially thought that a decrease in accuracy would occur when not focusing on the tumor center in the heatmap. However, it was interesting to note that there was little difference in the distribution of heatmaps between correct and incorrect examples. Subsequently, when the model was trained to focus on the center of the images, it was observed that the accuracy decreased compared to using the entire image for training. This finding aligns with the earlier results obtained using GradCam, suggesting that not only the tumor center but also the surrounding information is utilized in tumor diagnosis.

研究分野：皮膚科

キーワード：人工知能

1. 研究開始当初の背景

画像認識技術の分野に革命的な進歩をもたらした深層学習 (Deep learning: DL) を用いた人工知能 (Artificial intelligence: AI) により、医療の分野においても AI の活用が大きな注目を集めている。我々は 2016 年から DL による AI 診断の研究を開始した。その結果、通常の臨床写真で皮膚腫瘍の良悪性を感度 95% 以上、特異度 90% 以上で判別できた。これは皮膚科専門医を上回る数値であり、この論文は掲載された Br J Dermatol の年間ダウンロード数トップ 10% に入るなど注目を集めている。

皮膚疾患診断は視診による皮疹形状判定が診断の決め手となることが多い。腫瘍の診断に頻用されるダーモスコピーは、対象物を拡大して詳細な特徴を観察することでその診断精度を上げるためのデバイスである。つまり、皮膚科医は臨床像 (画像) から「特徴を抽出」して診断という「判定」を行っている。この画像処理プロセスを AI による画像認識技術により再現する試みは以前から行われており、特にダーモスコピーのように規格が決まった画像が得られる分野では実用段階に近いレベルに達している。これまでの AI 診断は技術者が多大な労力を掛けて「特徴の抽出」と「判定」のアルゴリズムを設計しており、その経験やノウハウの蓄積により少しずつ制度を改善するという手法が取られていた (図 1)。この手法はある程度規格が決まった画像の認識は可能であったが、被写体と様々な角度や距離で撮影されてしまう臨床写真を用いた AI 診断システムは非常に困難と考えられた。

そのような中、2012 年に機械が自ら教師データを読み込み学習しながら途中のパラメータを最適化していくディープラーニング (DL: 図 1) という技術が導入され、AI による画像認識にブレークスルーが起きた。その結果、2012 年の国際的な画像認識コンペティション (ILSVRC) でそれまでの 10 年分の改善がみられ、2015 年には人間の成績 (エラー率) を凌駕した (次ページ図 2)。現在様々なグループや企業が AI 診断システムの開発を行っており、欧米ではすでに眼底や放射線診断分野で実用化、本邦も内視鏡画像から腫瘍を見分ける AI 診断システムである EndoBRAIN が承認・発売されている。これらの分野は規格化された画像が得られるという特性から、DL との親和性が高い。一方、皮膚科は疾患の多様性のみならず、臨床写真の多様性 (規格化された画像が得られない) といった DL には不向きな分野とされてきた。

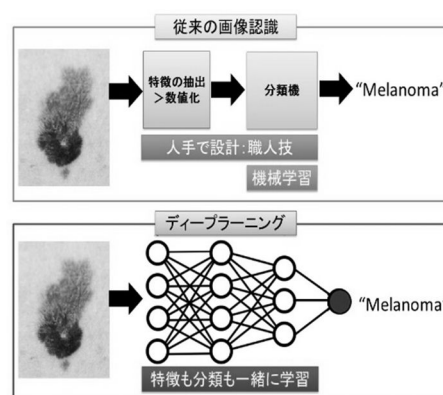


図 1 画像認識のパイプライン

2. 研究の目的

本研究の目的は、皮膚科専門医が臨床像から診断に至るまでの過程を AI に再現させること、そして病理所見を推定させることが出来る AI 診断補助システムを構築することである。

これまでの AI 診断装置はダーモスコピー写真や特殊な波長を用いた画像を使用することで悪性黒色腫の可能性を出力するもので、その診断精度は皮膚科医の診断と遜色が無いとする報告もある。しかし装置そのものが高価であることや悪性黒色腫に特化したシステムであることから余り普及していない。そこで我々は、そのような臨床写真を使用した皮膚腫瘍全般を診断できる DL による AI 診断システム開発・発表した (図 2) したが、その研究中に臨床写真とダーモスコピー写真を使用した DL による AI 診断システムが Nature に掲載された。コンセプトや手法はほぼ同じであったが我々がこの Nature の報告と異なるのは使用する写真の枚数が少ないこと (12 万枚対 6,000 枚)、そしてダーモスコピー写真は使用しないという点である。DL は写真の枚数が多いほど有利だが、我々は少ない枚数でも画像の前処理で精度が大きく改善できる独自の手法を確立しており、この手法は稀少疾患の多い皮膚疾患では有利となる。我々の構築したパイプラインは腫瘍以外の判定にも応用が可能となる点で発展性がある。

皮膚腫瘍鑑別 AI システムは世界的に見ても社会実装できたグループはなく、我々の研究成果をベースにさらに正診率を向上させ実用に耐えうる AI システムを開発できれば世界初の皮膚腫瘍を鑑別できる AI システムを社会実装できるチャンスがある。今後はこのシステムをベースとして、稀少疾患を含む炎症性皮膚疾患も鑑別できるシステムへと発展させていく。また、悪性腫瘍については臨床像と病理像とをリンクさせることで、生検しなくともその後の治療に大きく関わる腫瘍細胞の浸潤度合いを判別する AI バーチャルバイオプシーシステム構築にもチャレンジする。

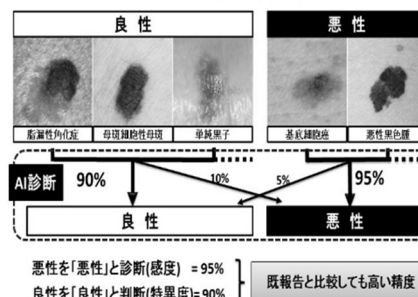


図 2 我がが発表した AI 診断の成績

3. 研究の方法

これまで我々が構築した皮膚腫瘍の臨床画像 AI 診断システムは次ページ図 3 に示すように 14 種類 21 カテゴリの良悪性含む皮膚腫瘍を高い感度と特異度で判別でき、スクリーニングツールとしては十分な成績である。図 4 の点線で囲われた良悪性判断を間違えた症例数は非常に少ないが、悪性を良性和間違える、偽陰性が 5%に生じていること、そして各腫瘍の正答率にはばらつきがあることが問題である。

各腫瘍の正答率を上げるにあたり鍵となる誤回答が興味深く(図 4), 例えばボーエン病はその誤回答のほぼ全てが有棘細胞癌か日光角化症と診断されている。他方、脂漏性角化症の誤回答は母斑細胞性母斑が多いが、悪性腫瘍である有棘細胞癌、悪性黒色腫、基底細胞癌などの間違いも多くなっており、皮膚科医が通常鑑別に挙げる診断が並ぶ。このような誤回答の症例を集積し、どのような要因が診断に影響しているかを解析することで精度向上を図り、またこの解析から見えて来る違いが組織学的な違いと相関するのかを含めた解析も行う。この解析と並行し、皮膚病ナショナルデータベースで集積した 30 万枚の臨床画像と 3 万枚の病理バーチャルスライドを用いて臨床像から病理像を推定するシステムの構築を行う。

機械学習の基となる教師付データについても事前の処理がその後の正答率に影響を与えることから、様々な事前の前処理を試み(図 4), 成績の良い組み合わせが明らかとなりつつある。他方でまだ多数の組み合わせや試す価値のある前処理法もあり、今後これらの検討を進めることでさらなる精度の向上が見込める。

4. 研究成果

画像一枚からの判定は画像のほんの少しの違いにより判定結果が揺らぐことがあり、90 度回転ただけで、「母斑細胞性母斑」が「メラノーマ」に判定結果が変わるなどの揺らぎが問題になっている。これはすでに報告されている現象であり、その克服には学習に用いるが増数を増やしたり、文字情報を加えたりするなどの工夫が報告されている。そこで判定 AI を複数用いることで判定結果が向上することがわかった(図 5)。

また、動画を用いた判定システム(図 6)は AI そのものではなく、AI の運用方法による精度改善であり、これはこれまでに報告の無い新しいアプローチである。動画を入力すると 1 秒間に数回判定を繰り返すことが可能で、判定結果を積算して最終判定結果を出すように設計している。この方法は個人情報のやりとりがなく、情報漏洩のリスクは極めて低いことが特徴である。一方で判定システムはオンラインでのアップデートが可能のため最新のシステムに更新も出来る。現在、社会実装のために動画判定の優位性を検討する臨床研究、インターフェースの熟成、版の実地テストを検討している。

様々な皮膚疾患の分類 AI を検討しており、その中で AI のクラス間類似度と医学的分類クラスと整合性があることもわかり、AI と医師とで同じような特徴を捉えて分類していることがわかった。このような結果から、AI 学習は医師が無意識に行っている分類(診断)を模倣するような判断基準のような法則を学習の中から得ている可能性が示唆された。

AIの回答

	日光角化症	Bowen 病	有棘細胞癌	基底細胞癌	悪性黒色腫	青色母斑	先天性母斑	母斑細胞性母斑	スピッツ母斑	汗孔腫	脂漏母斑	単純黒子	扁平母斑	脂漏性角化症
日光角化症	33	1	5	13				1						1
Bowen 病	1	51	5	2	1					1				
有棘細胞癌	7	5	156	15	5									1
基底細胞癌	7	12	16	2	4			6		1				3
悪性黒色腫			11	5	69	1		4		2		1		2
青色母斑						7		1						
先天性母斑		1		2			38	5	1					3
母斑細胞性母斑			2	9	3		8	189	2					4
スピッツ母斑				1				1	3	3				
汗孔腫		1	2	1				4	2	1				1
脂漏母斑				3				1					17	
単純黒子					1									4
扁平母斑														31
脂漏性角化症	2	4	7	5	6	1		17	1	4				66

図 3 AI 診断の結果詳細

前処理種類

	Rot	Phase	Brit	Sat	Blur	Crop	診断名	良悪性
A	○	0.00	0.0	0.1	5	0	70.1	92.5
B	○	0.02	0.1	0.1	5	0	73.2	94.8
C	○	0.02	0.0	0.1	5	0	71.5	92.3
D	○	0.00	0.1	0.1	5	0	71.6	93.8
E	○	0.00	0.1	0.1	5	224	72.7	94.2
F	○	0.00	0.1	0.1	5	0	72.2	93.6
G	x	0.00	0.0	0.1	0	0	69.7	93.0
H	x	0.00	0.0	0.0	5	0	70.4	92.6
I	x	0.00	0.0	0.0	0	0	68.2	92.8
J	x	0.00	0.0	0.0	0	0	69.5	92.0
K	x	0.00	0.0	0.1	0	0	69.8	92.5
L	○	0.02	0.0	0.1	5	0	70.8	91.9
M	○	0.00	0.0	0.1	5	0	72.2	94.8
N	○	0.00	0.0	0.1	5	0	73.6	93.4
O	○	0.00	0.0	0.2	5	0	71.1	92.1
P	○	0.00	0.0	0.0	0	0	73.4	93.2
Q	○	0.00	0.1	0.1	5	0	74.0	94.7
R	○	0.04	0.1	0.2	5	0	73.4	93.2

図 4 前処理による精度の改善効果

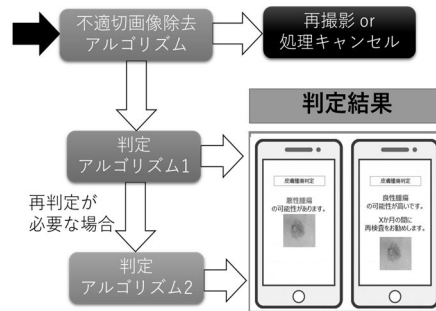


図 5 複数の判定 AI 導入による精度向上

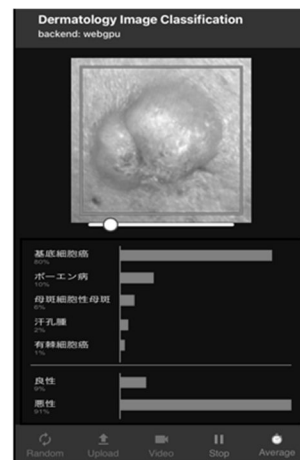


図 6 動画による判定システム

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 1件 / うちオープンアクセス 0件）

1. 著者名 Fujimoto Atsushi, Iwai Yuki, Ishikawa Takashi, Shinkuma Satoru, Shido Kosuke, Yamasaki Kenshi, Fujisawa Yasuhiro, Fujimoto Manabu, Muramatsu Shogo, Abe Riichiro	4. 巻 10
2. 論文標題 Deep Neural Network for Early Image Diagnosis of Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis	5. 発行年 2022年
3. 雑誌名 The Journal of Allergy and Clinical Immunology: In Practice	6. 最初と最後の頁 277 ~ 283
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jaip.2021.09.014	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Masaya Tanaka, Atsushi Saito, Kosuke Shido, Yasuhiro Fujisawa, Kenshi Yamasaki, Manabu Fujimoto, Kohei Murao, Youichirou Ninomiya, Shin'ichi Satoh, Akinobu Shimizu	4. 巻 16(11)
2. 論文標題 Classification of large-scale image database of various skin diseases using deep learning	5. 発行年 2021年
3. 雑誌名 International Journal of Computer Assisted Radiology and Surgery	6. 最初と最後の頁 1875-1887
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s11548-021-02440-y	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 藤澤康弘
2. 発表標題 皮膚科領域におけるAI開発について
3. 学会等名 愛媛県医師会
4. 発表年 2022年

1. 発表者名 藤澤康弘
2. 発表標題 人工知能(AI)と医療への応用について
3. 学会等名 皮膚アレルギー学会
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------