

令和 4 年 6 月 6 日現在

機関番号：17401

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K08155

研究課題名(和文)大規模肺がんCT検診データベースを用いた未病のコンピュータ支援診断

研究課題名(英文)Computer Assisted Diagnosis for Predicting Lung Cancer by Using Large Scale Low-Dose CT Database

研究代表者

白石 順二(Shiraishi, Junji)

熊本大学・大学院生命科学研究部(保)・教授

研究者番号：30551311

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：未病段階での肺がんの検出を目的とする本研究では、継続的に肺がんCT検診を受診している被検者のうち、最初のうちは正常と判定されながら、途中で異常所見が発見された被検者の、異常所見が発見される前の、正常と判定された最後の年を未病の状態と仮定し、世界で最初の未病の状態の症例の画像データベースを構築した。本研究の初期段階では、11年にわたり実施された被検者約2万3千人、検査数約4万5千件の肺がんCT検診全体の画像データベースを構築し、そこから世界初となる未病の状態の症例の画像データベースを抽出した。その後、未病の検出を前提として、喫煙者と非喫煙者の判別、若年者と高齢者の判別を、深層学習により試みた。

研究成果の学術的意義や社会的意義

「未病」とは、検査をしても結果に異常がない、もしくは、検査結果に異常があるけれども自覚症状が何もない状態を指し、未病の段階で、適切な治療を行うことで深刻な事態に進行することを防ぐことが可能になる。しかしながら、未病の段階で何らかの異常を見つけて病気を予防するための診断法に関しては未だ不明な点も多く、その解明には新しい視点からのアプローチが必要である。特に呼吸器領域においては、慢性閉塞性肺疾患や、超早期の肺がん、間質性肺疾患を対象とした未病の診断が考えられるが、専門医や放射線科医の読影に強く依存し、研究代表者らの知る限りでは、呼吸器領域における未病の検出の具現化を試みた研究はこれまでにない。

研究成果の概要(英文)：In this study, which aims to detect lung cancer in "Mibyo" stage, among the subjects who are continuously undergoing lung cancer CT screening, abnormal findings were discovered in the middle, although it was initially judged to be normal. We constructed the world's first image database of "Mibyo" cases, assuming that the subject's stage in the last year, before the discovery of abnormal findings, was normal. In the initial stage of this study, we constructed an image database of the entire lung cancer CT examination, which was conducted over 11 years with about 23,000 subjects and about 45,000 examinations, and was the first in the world to do so. An image database of cases in a diseased state was extracted. After that, on the premise of detecting lesions at "Mibyo" stage, we tried to distinguish between smokers and non-smokers, and young people and elderly people by deep learning.

研究分野：コンピュータ診断支援

キーワード：未病 人工知能 機械学習 肺がん検診 CT コンピュータ診断支援

1. 研究開始当初の背景

研究開始当初(2019年)の日本においては、65歳以上の高齢者が人口に占める割合が高くなるのに従って、医療費や介護費の高騰が問題となっていた。その一方で、少子化が進んでいるため、若年層に対する医療費の金銭的負担が増加し続けており、社会的な問題となっていた。そして、高齢化社会においては、寝たきりや要介護とならないように、普段から病気にならない努力が必要と言われていた。

「未病」とは、一般的に、軽い自覚症状はあるけれども発病していない(検査結果に異常がない)状態、もしくは、自覚症状は何もないけれども発病している(検査結果に異常がある)状態を指し、未病の段階で、その状態であることがわかり、適切な治療を行うことで深刻な事態に進行することを防ぐことが可能になる[1]。しかしながら、未病の段階で、何らかの異常を見つけて病気を予防するための診断法に関しては未だ不明な点も多く、その解明には新しい視点からのアプローチが必要である。現在では、高脂血症、糖尿病、高血圧なども未病の1つと考えられ、健康診断での血液検査や身体測定の結果から、未病の状態が確定できる場合もあるが、呼吸器領域においては、慢性閉塞性肺疾患(COPD)や、超早期の肺がん、間質性肺疾患といった未病の診断は専門医や放射線科医の読影に負うところが多く、研究代表者らの知る限りでは、呼吸器領域における未病の検出の具現化を試みた研究はこれまでにない。

医用画像の読影に関しては、研究開始当初にはAIを用いて診断支援を試みる研究が盛んに行われるようになっていた。この背景には、2010年代になり飛躍的に研究が発展し、その性能が大幅に改善された「ディープラーニング」(深層学習: deep learning)の存在がある。深層学習は多層の「人工ニューラルネットワーク」(ANN: Artificial Neural Network)による機械学習であるが、多層ではなく、3層構造のANNを応用し、医用画像からコンピュータで自動的に抽出した画像の特徴量をANNの入力として、病変部の検出や鑑別診断に必要な情報を得る手法の開発が、それ以前から進められている。このような手法でコンピュータから得られた情報を医師が「第2の意見」として診断の参考にするのを「コンピュータ支援診断」(CAD: Computer aided diagnosis)と呼び、1980年代にシカゴ大学から提案されて以降、現在でも研究が進められ、米国では、乳がんの画像診断の82%にCADが利用される程、実用化が進んでいる(2015年のデータ)。AIによる診断支援は、従来型のCADで必要とされた画像特徴量の抽出を深層学習によって置き換えたもので、ヒトの視知覚では判断できない高次のレベルにおける画像の特徴が深層学習によって導き出され、診断支援に用いられる。研究代表者らはCAD研究が始まった当初からCADの技術開発およびCADによる診断能向上の定量的評価に関する研究を進めており、2000年には、CAD技術の開発支援と、開発された技術の公平な評価を目的として、パブリックに使用可能な世界で最初の胸部単純X線像の画像データベースを開発し、国内外の研究者に公開した[2]。この画像データベースは現在でも国内外の多くの研究施設で利用されており、その文献引用数は443編(2022年5月現在)になっている。特にAI診断支援研究が盛んになった現在においては、2020年以降で140編以上の引用がある。さらに、CADによる診断能向上の評価に関する研究では、コンピュータが与える情報の真偽に関係なく、医師の判断基準がコンピュータ出力によって大きく影響を受けることを、世界で最初に観察者実験において証明した[3]。また、診断能の評価を行う手法として国内外で広く用いられている観察者受動特性(ROC: Receiver Operating Characteristic)解析を用いた観察者実験を含む295編の論文について、その手法や実験方法を解析し、診断能評価の正しい手法について総評により提案した[4]。

2. 研究の目的

現在、AI診断支援の研究は、国内外で盛んに行われているが、それらの対象は、肺がんや大腸がん、乳がんといった、従来型のCADが対象としていた疾患で、未病の検出を試みた報告はわれわれの知る限りでは存在しない。その理由は、従来型のCADでは早期の疾患でさえ検出が困難な場合が多く、その前の状態の未病の検出は不可能と考えられていたことと、未病の状態を定義し、その未病の状態の症例の画像データベースを構築する方法が存在しなかったためと考えられる。そこで、本研究では、長期にわたり継続的に検診を受診している被検者のうち、最初のうちは正常と判定されながら、途中で異常所見が発見された被検者の、異常所見が発見される前の、正常と判定された最後の年の段階を未病と仮定し(図1)、世界で最初の未病の状態の症例の画像データベースを構築する。その上で、これまでに蓄えたCADの技術開発の経験と知識を生かし、呼吸器領域における未病の検出を具現化し、そのためのCADを開発することを目的とする。

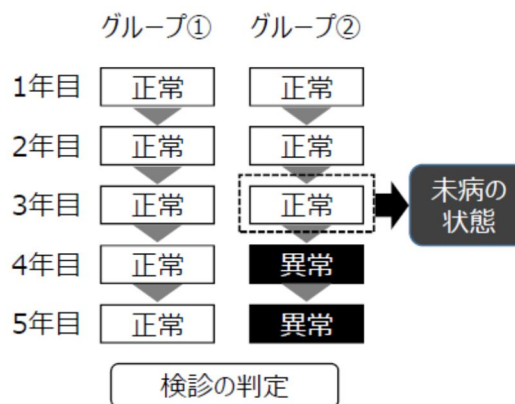


図1 母集団のグループ分類と未病の状態

3. 研究の方法

本研究では、初年度の 2019 年度に未病の画像データベース（以後、未病 DB）を構築するためのベースとなる肺がん CT 検診（3mm スライス厚）の画像データベース（以後、石川 DB）を構築した。石川 DB は石川県予防協会において 2006 年から 2016 年の 11 年間に撮影された肺がん検診の低線量 CT 画像で構築され、構築にあたっては当該医療機関の倫理審査委員会の承認を得た上で、画像や被検者情報のデータ収集および匿名化の作業を開始した。被検者情報に含まれるのは、性別、年齢、喫煙歴、読影レポート、病変がある場合はその位置情報、確定診断等の付帯情報とした。研究分担者らの初年度の作業は、画像データベースに収録する画像や被検者情報の匿名化と匿名化後に個人を同定するための対応表の作成、および、異常ありの判定があった場合の異常所見の領域のマークとその程度の判定である。2020 年度に石川 DB の構築が完了した後は、検診データの結果から、被検者の母集団を、経年的に受けた検診の回数と、最初の検診から最終の検診までの判定結果から、初回の検診から最終まで「異常なし」と判定されたグループ、初回の検診では「異常なし」と判定されたが、その後の経年的な検診で「異常あり」と判定されたグループ、最初の検診から「異常あり」と判定されたグループに分類した。そして、未病の検出の初期検討として、このグループについて、[A]非喫煙者と喫煙者のサブグループ、[B]被検者の実年齢を用いて分割した高齢者と若年者のサブグループ、の 2 つのサブグループについて、CT 画像情報だけで、両者の判別が可能かどうかを検討した。判別には深層学習による畳み込みニューラルネットワーク(CNN)を利用した。また、最終年度に当たる 2021 年度には、前述のこのグループから、最後に正常と判定された段階での画像データを評価対象として、未病の状態の症例の未病 DB を構築し、グループの正常な被検者と未病の状態の判別が CNN を用いることで可能となるかどうかを検討した。

4. 研究成果

本研究で構築した石川 DB は、最終的に検査数 45,337 件、被検者数 23,065 名で構築される非常に大規模な画像データベースとなった（図 2）。

喫煙者・非喫煙者の分類においては、感度は 66.2%、特異度は 82.6% で全体の分類精度は 67.2% となった。肺年齢推定においては、実年齢と推定肺年齢の間の相関係数は、男性で 0.56、女性で 0.65 となり、それぞれ、平均パーセント誤差は 11.20%、8.38%、推定年齢の平均誤差は 5.91 歳、4.33 歳となった。

本研究で用いた未病の定義「最終的に肺がんと診断された被検者で、ある年の検査で肺がんの疑いを指摘された被検者が、その検査の前年、もしくは直近年に受けた検査で、正常と判断されていた時点の状態」に準じて構築した未病の画像データベースは、この定義に従って石川 DB をベースとして、未病症例を 17 症例抽出し、同じ年齢性別の正常な 17 症例も加え、34 例とした（図 3）。

CNN を用いて試行的に行った正常症例と未病症例の分類において、感度は 81.2%、特異度は 53.3% で全体の分類精度は 67.2% となった。

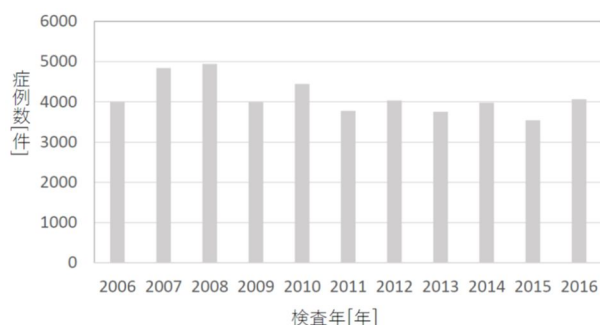


図 2 石川 DB に収録された 11 年間の CT 検査数

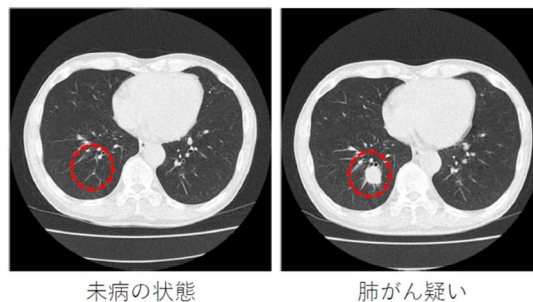


図 3 未病 DB に収録された未病データの 1 例

< 引用文献 >

- 1) 野口 久. 呼吸器専門医の立場から. 日本未病システム学会雑誌 12(2): 293-294, 2006
- 2) Shiraishi J, Katsuragawa S, Ikezoe J, et.al. Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. AJR 174:71-74, 2000.
- 3) Shiraishi J, Abe H, Engelmann R, et.al. Computer-aided diagnosis for distinction between benign and malignant solitary pulmonary nodules in chest radiographs: ROC analysis of radiologists' performance. Radiology 227: 469-474, 2003.
- 4) Shiraishi J, Pesce L, Metz CE, Doi K: Experimental design and data analysis in receiver operating characteristic studies: Lessons learned from reports in Radiology from 1997 to 2006, Radiology, 253:822-830, 2009.

5. 主な発表論文等

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

1. 著者名 Tanaka Rie、Nozaki Shiho、Goshima Futa、Shiraishi Junji	4. 巻 9
2. 論文標題 Deep learning versus the human visual system for detecting motion blur in radiography	5. 発行年 2022年
3. 雑誌名 Journal of Medical Imaging	6. 最初と最後の頁 1,9
掲載論文のDOI（デジタルオブジェクト識別子） 10.1117/1.JMI.9.1.015501	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9件（うち招待講演 0件／うち国際学会 1件）

1. 発表者名 森寛人, 田中利恵, 小林聡, 南哲弥, 松永哲夫, 白石順二
2. 発表標題 低線量CTを対象に深層学習を用いて非喫煙・喫煙肺分類を行う試み
3. 学会等名 第77回日本放射線技術学会総会学術大会
4. 発表年 2020年～2021年

1. 発表者名 北野瑞稀, 松永哲夫, 小林 聡, 南 哲弥, 三輪洸亮, 田中利恵, 白石順二
2. 発表標題 機械学習による低線量胸部CT画像の肺年齢分類の試み -CT画像による実年齢グループ分類の可能性の検証-
3. 学会等名 第77回日本放射線技術学会総会学術大会
4. 発表年 2020年～2021年

1. 発表者名 Hiroto Mori, Rie Tanaka, Satoshi Kobayashi, Tetsuya Minami, Tetsuo Matsunaga, Junji Shiraishi
2. 発表標題 Deep learning-based lung classification of low-dose chest CT images: Smoking or non-smoking?
3. 学会等名 Korean Societly of Radiological Science Conference 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 三輪洸亮, 田中利恵, 松永哲夫, 小林聡, 南哲弥, 白石順二
2. 発表標題 低線量胸部CT画像を対象としたディープラーニングによる喫煙指数別分類の試み
3. 学会等名 第47回日本放射線技術学会秋季学術大会
4. 発表年 2019年

1. 発表者名 北野瑞稀, 松永哲夫, 小林聡, 南哲弥, 田中利恵, 白石順二
2. 発表標題 肺がんを対象とした「未病」の低線量CT画像データベース構築の試み
3. 学会等名 第47回日本放射線技術学会秋季学術大会
4. 発表年 2019年

1. 発表者名 北野瑞稀, 松永哲夫, 小林聡, 南哲弥, 田中利恵, 白石順二
2. 発表標題 11年間分の肺がんCT検診画像データベース構築
3. 学会等名 第14回九州放射線医療技術学術大会
4. 発表年 2019年

1. 発表者名 北野瑞稀, 松永哲夫, 小林聡, 南哲弥, 田中利恵, 白石順二
2. 発表標題 機械学習による低線量胸部CT画像の肺年齢分類の試み CT画像による実年齢グループ分類の可能性の検証
3. 学会等名 第77回日本放射線技術学会総会学術大会
4. 発表年 2021年

1. 発表者名 北野瑞稀, 松永哲夫, 小林聡, 南哲弥, 田中利恵, 白石順二
2. 発表標題 機械学習による低線量胸部CT画像の 肺年齢推定の試み
3. 学会等名 第49回日本放射線技術学会秋季学術大会
4. 発表年 2021年

1. 発表者名 北野瑞稀, 松永哲夫, 小林聡, 南哲弥, 田中利恵, 白石順二
2. 発表標題 機械学習による低線量胸部CT画像の 未病の肺がん検出の試み
3. 学会等名 第16回九州放射線医療技術学術大会
4. 発表年 2021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
研究分担者	小林 聡 (Kobayashi Satoshi) (30313638)	金沢大学・保健学系・教授 (13301)	
研究分担者	田中 利恵 (Tanaka Rie) (40361985)	金沢大学・保健学系・准教授 (13301)	
研究分担者	南 哲弥 (Minami Tetsuya) (60436813)	金沢医科大学・医学部・教授 (33303)	

6. 研究組織（つづき）

	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
研究協力者	松永 哲夫 (Matsunaga Tetsuo)		
研究協力者	北野 瑞稀 (Kitano Mizuki)		

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