



Computer-aided diagnosis of chest X-ray for COVID-19 diagnosis in external validation study by radiologists with and without deep learning system

宮崎, 亜樹

(Degree)

博士 (医学)

(Date of Degree)

2024-03-25

(Resource Type)

doctoral thesis

(Report Number)

甲第8782号

(URL)

<https://hdl.handle.net/20.500.14094/0100490007>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



(課程博士関係)

学 位 論 文 の 内 容 要 旨

Computer-aided diagnosis of chest X-ray for COVID-19 diagnosis in external validation study by radiologists with and without deep learning system

COVID-19 の胸部 X 線診断における、
深層学習の人工知能を使用した場合と使用しない場合の
放射線科医による診断能についての検討

宮崎 亜樹, 池島 健吾, 西尾 瑞穂, 藪田 実, 松尾 秀俊, 尾上 宏治,
松永 卓明, 西岡 瑛子, 河野 淳, 山田 大輔, 大庭 建, 石藏 礼一,
村上 卓道

神戸大学大学院医学研究科医科学専攻
放射線医学分野
(指導教員： 村上 卓道 教授)

宮崎 亜樹

【目的】

新型コロナウイルスは 2019 年に中国で初めて発見され、今日まで世界で膨大な感染者や死者を出している新しい感染症である。新型コロナウイルス感染症の診断方法として多くの医療機関で用いられているのが RT-PCR(Real-time polymerase chain reaction)検査だが、感度は 70-80%程度であり、CT の感度の方が高かったとする報告もある。さらに、RT-PCR には結果が出るまで一定の時間がかかるのに対して、CT やレントゲン検査は数分で結果が出るため、より迅速に新型コロナ感染症を診断する手段となり得る。

レントゲンによる COVID 肺炎の検出に対する、深層学習による人工知能 (AI) の臨床応用が注目されている。CT 検査は小規模のクリニックには設置されていないことが多いのに対して、レントゲン検査は CT よりも検出力は劣るものの簡便でかつコストも低い。AI とレントゲンの組み合わせによる新型コロナの診断が可能となれば、医者のない施設でも新型コロナ肺炎のスクリーニングが可能となるかもしれない。

AI による新型コロナ肺炎の CT/レントゲン診断についてはすでに沢山の研究が行われており、その多くが良好な成績をおさめている。ただ実際に AI を臨床応用する場合、医師は自身の診断と AI の診断を比較して検討することになる。しかし、放射線診断医が AI を併用して読影実験を行った研究はまだ少ない。

この研究の目的は、AI 診断が臨床的に有用であることを確認するために、放射線診断医が AI の結果を踏まえて実際に診断を変更するか、またその診断能が変化するかについて調べることである。AI と放射線診断医は、用意されたレントゲン写真をそれぞれ正常・肺炎・コロナ肺炎に診断する。AI ソフトは、既存の 2 つの公開データセットと我々の収集したデータセットを用いて作成したものを使用する。さらに新規の医療機関から読影実験用のデータセットを収集し、外的妥当性(external validation)を考慮した上で計 8 人の放射線診断医による読影実験を行った。

【方法】

今回の研究に利用する AI のレントゲン写真のデータセットには三つの疾患のレントゲン写真が含まれており、正常・COVID-19 でない肺炎、COVID-19 肺炎の三つである。今回は二つの公開されたデータセットと以前に我々が用意したデータセットに一部追加データを合わせたものを使用し、合計 26,213 例のレントゲン写真を開発セット(development set)として使用した。また、今回新規に development set に追加されたレントゲンのうち、一つの医療機関から収集した 168 例(正常 80・肺炎 37・COVID51 例)を AI の内的妥当性 (internal validation)チェックおよび放射線診断医の読影練習用として採用した。

今回の研究は、AI の診断能をより正しく評価するために、前述のデータセットとは

異なる医療機関から読影実験用のデータ 180 例を収集し、外的妥当性の検討を行った。このデータセットにおいて COVID は男性 46 例、女性 17 で、年齢 53.4 ± 38.6 歳であった。診断時期 は 2020 年 2 月 29 日から 2020 年 5 月 2 日の間だった。

今回の研究では、AI の結果は AI 単独の成績の計算および読影実験の際に放射線診断医が参照する情報として主に使用した。今回は 3 分類の検討になるため、AI にはそれぞれのレントゲン写真について正常・肺炎・COVID である確率を合計 100%となるように算出させた。また、読影実験用・練習用のデータについては、Grad-CAM および Grad-CAM++を用いて AI の診断の根拠となった部分を可視化した画像を作成した。AI の Grad-CAM および Grad-CAM++の結果では、正常例においては肺野に比較的左右対称な関心領域が見られることが多かった。肺炎は肺炎像と一致して（多くの場合）片側の肺野に関心領域が、COVID19 では両肺および縦隔に関心領域がみられる傾向があった。こうした Grad-CAM、Grad-CAM++の解釈法についても、放射線診断医には読影実験の前に教示した。

読影実験は、合計 8 人の放射線診断医（経験年数：5~14 年）が一か月以上の期間をあけて合計二回の読影実験を行った。そのうち一回はレントゲン写真のみを参照し（“reading session without AI”）、もう一回はレントゲン写真に加えて AI の読影結果を参照しながら（“reading session with AI”）行った。

放射線診断医はレントゲン写真を参照し、正常・肺炎・COVID らしさを 100 点満点で採点した。放射線診断医は“reading session without AI”の場合、AI の算出した正常・肺炎・COVID の確率の数値を必ず参照し、AI の診断に疑問があった場合には、Grad-CAM・Grad-CAM++の結果を参照可能とした。なお、練習用の 168 例についてもそれぞれ Grad-CAM および Grad-CAM++で画像を作成し、こちらも放射線診断医に提示した。なお、読影実験および練習にはとくに時間制限はないものとした。読影実験ではテストセットに含まれる正常・肺炎・COVID の割合のみを提示し、レントゲン写真以外の臨床情報は提示しなかった。

【結果】

AI 単独の正診率は 0.733 で、AI を使用しない 8 人の放射線診断医の精度は 0.696 ± 0.031 だった。複数の放射線診断医による症例分析の結果、COVID vs. 正常+肺炎では AI を使用した放射線診断医と使用しない放射線診断医との間で Area under the curve に統計的に有意な差があった（ $p = 0.0038$ ）。一方で、正常 vs. 肺炎 or COVID と、肺炎 vs. 正常 or COVID では統計的な有意差はなかった。

【議論】

今回の検討では、8 人の放射線診断医が AI を併用しながらの読影と、併用しない読影を行い、その結果を比較・解析した。また、AI 単独の診断能についても調べた。AI は過半数

の放射線診断医よりも正診率・AUC が高かった。さらに、COVID 肺炎の診断において、放射線診断医の診断能は AI を用いることによって有意に改善された。

一方で、“正常 vs. 肺炎+COVID”と“肺炎 vs. 正常+COVID”では有意差がなかった。その理由として考えられるのは、正常や肺炎に比して COVID については読影医の読影経験が乏しいことである。放射線診断医ではなく他科の医者では AI モデルの有用性がより発揮される可能性がある。

今回使用した AI は単独でも過半数の放射線診断医より診断能が高かったので、レントゲン写真における COVID 診断について、スクリーニングなどへの応用が可能となるかもしれない。とくに医者のいない環境においては有用性が期待される。前回の我々の研究でも、我々の AI モデルは COVID の診断において読影医の読影よりも有意に優れていた(p-value = 0.001334)。AI モデルと読影医の診断能を比較した研究には Murphy K や Ramsey M らの報告があるが、一部の放射線診断医を除き、多くの場合で AI モデルの Area under the curve および正診率は放射線診断医を上回る傾向にあった。

RT-PCR に対してレントゲン写真は簡便な画像検査である。ただ一方で診断能が医者の能力に依存するという欠点がある。今回の AI の感度は比較的高いため、レントゲン写真の有用性を高め、RT-PCR の代替あるいは補完する検査へ昇華することができるかもしれない。

外的妥当性を評価した理由の一つに、公開データセットの欠点がある。Beatriz Garcia Santa Cruz らは COVIDx などの公開データセットについて、未知のバイアスが含まれている可能性がある」と指摘している。これらのデータセットを用いて内的妥当性のみを評価する場合、AI の診断能を過大評価してしまう恐れがある。したがって、今回我々は外的妥当性を評価することでこれらのバイアスの軽減を試みた。

他にもこの研究にはいくつかの limitation がある。第一に、今回使用したレントゲン写真はいずれも大きな病院のもので、画質の良い画像のみが用いられている。よって、小さな医療機関の画質の悪い画像については検討できていない。さらに、今回は正常と肺炎と COVID の症例を集めて読影実験を行ったが、肺癌や間質性肺炎などほかの肺疾患については除外しており、検討できていない。

【結論】

外的妥当性を考慮したうえで、正常・肺炎・COVID の 3 分類について、我々の AI は単独で放射線診断医と比較して良好な診断能を示した。また、AI の併用は、COVID vs. 正常 or 肺炎における放射線診断医の診断能を有意に向上させた。

論文審査の結果の要旨			
受 付 番 号	甲 第 3353 号	氏 名	宮崎 亜樹
論 文 題 目 Title of Dissertation	Computer-aided diagnosis of chest X-ray for COVID-19 diagnosis in external validation study by radiologists with and without deep learning system COVID-19 の胸部 X 線診断における、深層学習の人工知能を使用した場合と使用しない場合の放射線科医による診断能についての検討		
審 査 委 員 Examiner	主 査 溝 渕 知 司 Chief Examiner 副 査 森 保 子 Vice-examiner 副 査 小 谷 穂 治 Vice-examiner		

(要旨は1, 000字～2, 000字程度)

序論

新型コロナウイルスは 2019 年に中国で初めて発見され、今日まで世界で膨大な感染者や死者を出している新しい感染症である。新型コロナウイルス感染症の診断方法として多くの医療機関で用いられているのが RT-PCR(Real-time polymerase chain reaction)検査だが、感度は 70-80%程度であり、CT の感度の方が高かったとする報告もある。さらに、RT-PCR には結果が出るまで一定の時間がかかるのに対して、CT やレントゲン検査は数分で結果が出るため、より迅速に新型コロナ感染症を診断する手段となり得る。

深層学習による人工知能 (AI) の臨床応用が注目されている。AI とレントゲンの組み合わせによる新型コロナの診断が可能となれば、医者のいない施設でも新型コロナ肺炎のスクリーニングが可能となるかもしれない。

この研究の目的は、新型コロナ肺炎の AI 診断が臨床的に有用であることを確認するために、放射線診断医が AI の結果を踏まえて実際に診断を変更するか、またその診断能が変化するかについて調べることである。

方法

今回の研究に利用する AI のレントゲン写真のデータセットには三つの疾患のレントゲン写真が含まれており、正常・COVID-19 でない肺炎、COVID-19 肺炎の三つである。今回は二つの公開されたデータセットと以前に我々が用意したデータセットに一部追加データを合わせたものを使用し、合計 26,213 例のレントゲン写真を開発セット(development set)として使用した。また、今回新規に development set に追加されたレントゲンのうち、一つの医療機関から収集した 168 例(正常 80・肺炎 37・COVID51 例)を AI の内的妥当性(internal validation)チェックおよび放射線診断医の読影練習用として採用した。AI の診断能をより正しく評価するために、前述のデータセットとは異なる医療機関から読影実験用のデータ 180 例を収集し、外的妥当性の検討を行った。このデータセットにおいて COVID は男性 46 例、女性 17 で、年齢 53.4 ± 38.6 歳であった。診断時期 は 2020 年 2 月 29 日から 2020 年 5 月 2 日の間だった。AI の結果は AI 単独の成績の計算および読影実験の際に放射線診断医が参照する情報として主に使用した。

読影実験は、合計 8 人の放射線診断医(経験年数:5~14 年)が一か月以上の期間をあけて合計二回の読影実験を行った。放射線診断医はレントゲン写真を参照し、正常・肺炎・COVID らしさを 100 点満点で採点した。

結果

AI 単独の正診率は 0.733 で、AI を使用しない 8 人の放射線診断医の精度は 0.696 ± 0.031 だった。複数の放射線診断医による症例分析の結果、COVID vs. 正常+肺炎では AI を使用した放射線診断医と使用しない放射線診断医との間で Area under the curve に統計的に有意な差があった ($p = 0.0038$)。一方で、正常 vs. 肺炎 or COVID と、肺炎 vs. 正常 or COVID では統計的な有意差はなかった。

考察

今回の検討では、8人の放射線診断医がAIを併用しながらの読影と、併用しない読影を行い、その結果を比較・解析した。また、AI単独の診断能についても調べた。AIは過半数の放射線診断医よりも正診率・AUCが高かった。さらに、COVID肺炎の診断において、放射線診断医の診断能はAIを用いることによって有意に改善された。

一方で、“正常 vs. 肺炎+COVID”と“肺炎 vs. 正常+COVID”では有意差がなかった。その理由として考えられるのは、正常や肺炎に比してCOVIDについては読影医の読影経験が乏しいことである。放射線診断医ではなく他科の医者ではAIモデルの有用性がより発揮される可能性がある。

本研究のlimitationとして以下が挙げられる。第一に、今回使用したレントゲン写真はいずれも大きな病院のもので、画質の良い画像のみが用いられている。よって、小さな医療機関の画質の悪い画像については検討できていない。さらに、今回は正常と肺炎とCOVIDの症例を集めて読影実験を行ったが、肺癌や間質性肺炎などほかの肺疾患については除外しており、検討できていない。

結論

正常・肺炎・COVIDの3分類について、今回のAIは単独で放射線診断医と比較して良好な診断能を示した。また、AIの併用は、COVID vs. 正常 or 肺炎における放射線診断医の診断能を有意に向上させた。

以上

今回、AIの使用は有用であることを学術的な検討で示唆したことから価値ある業績と認める。よって、本研究者は、博士（医学）の学位を得る資格があると認める。