

原 著

類似症例提示による胸部CT診断支援効果の検討

小谷知也, 伊藤哲哉, 辻 悠佑, 小原 雄, 野口直希, 牛嶋 陽, 高田和豊*,
若杉健介*, 小塚和紀*, 近藤堅司*, 清野正樹*, 坂井豊彦**, 木村浩彦**

松下記念病院 放射線科
*パナソニック 先端研究本部
**福井大学医学部 放射線科

要旨

目的: 我々は胸部Computed tomography (CT)の類似性を基準に医学論文や医学書籍等のCTから類似症例を選択して医師に提示する診断支援システムを開発中である。

本論文では「画像の類似性を基準とする類似症例」の提示が診断にどのような影響を与えるのか評価するため読影実験を行い、効果を検討した。

対象と方法: 被検者は医師9人(うち5人は放射線科専門医, 以下, 専門医, 4人は非専門医), 症例は松下記念病院の3症例を選定した。被検者でない診断専門医が出題症例ごとに「画像から挙げるべき鑑別疾患病名」(以下, 適切病名)とそれ以外の「画像から鑑別に挙げるべきではないと考えられる疾患病名」(以下, 不適切病名)を定義し, 適切病名:不適切病名の割合が約3:7となるように約20症例を類似症例として選定した。被検者は出題症例に対して鑑別疾患を挙げた後, 類似症例を参考にして鑑別疾患を再考した。類似症例提示前後で鑑別疾患の変化, 正診率を検討した。

結果: 類似画像提示により鑑別疾患記載数は1症例当たり1.0個追加され, 不適切病名が0.2個削除された結果, 1症例当たり0.8個増加した。適切病名が削除されたケースはなく, 正診率は8%向上した。

結論: 類似症例の提示が適切病名を追加する効果, わずかに不適切病名を削除する効果, 正診率を向上させる効果を確認した。類似症例の提示は診断支援として有用である。

キーワード: lung, computed tomography, image retrieval, support effects, clinical benefits

緒 言

近年人工知能の発達が著しいが, 医用人工知能による完全自動診断は現在のところ実用化には至っておらず, 医師の診断支援を行うことを目的とするものが開発, 普及段階にある¹⁻³⁾。例えば医用人工知能による病変の存在診断の分野ではcomputer assisted detection (CADe)の開発が進

み, マンモグラフィなど一部の領域では普及しているソフトウェアが複数存在する¹⁾。また存在診断から評価に踏み込んだものとしては, 骨シンチ内の高集積部位を検出して異常確率を算出したり骨転移を定量化したりするソフトウェアBONENAVIや, 脳magnetic resonance imaging (MRI)から脳萎縮の度合いを評価する画像処理・統計解析ソフトウェアVSRADなどが存在し, それぞれ「骨シンチから異常集積を検出し, 定量化する」, 「脳MRIから特定の脳領域を検出し, 萎縮の程度を定量化する」といったモダリティを限定し, さらに特定の目的に絞ったものに関しては普及しているものが存在している^{2,3)}。類似症例提示システム

2017年3月1日受付

連絡先: 〒570-8540

大阪府守口市外島町5番55号

松下記念病院

放射線科(小谷知也)

Email: kotani.tomoya@jp.panasonic.com

Tel: +81-66992-1231, Fax: +81-66992-4845

の分野ではこれまでも教育，研究，診断用途を目指して開発が行われている⁴⁾。国内においてはEndoらが放射線科専門医1名，研修医2名を被検者として胸部CT上の孤立性肺結節の良悪性の判断に関して被検者に類似症例検索を提示することにより正診率が向上したことを報告している⁵⁾。

我々はパナソニック先端研究本部，福井大学医学部との共同開発により人工知能の技術を応用し，診断時CT病変の画像類似性を基準にして院内患者や医学論文，医学書籍症例の胸部CTから類似症例を選択，提示することで肺癌の良悪性のみならず，肺の異常陰影を包括する診断支援を行うことを目的とするシステムを開発中である^{5,6)}。現時点での画像認識技術はCT二次元情報の1スライスのみを評価していること，仮に画像認識技術が完成したとしても画像診断は画像類似性だけでなく他の医療情報による重み付けを要することなどの限界があり，画像所見のみから正解疾患症例だけを的確に導くことは難しいと考えられる。診断対象のCTを元に類似症例検索システムに類似症例を選択させると，正解疾患だけでなく放射線科医から見ると類似性に乏しく鑑別に挙げるべきではないと考えられる疾患も複数含んだ検索結果が返ってくることが多い^{5,6)}。従って，類似症

例検索システムにより検索，提示された類似症例は不確実性が存在することを前提として，そのような適切でない検索結果の混在を伴うシステムが放射線診断医の鑑別疾患の挙げ忘れを防ぐことや文献検索の効率化といった診断支援として有用に機能しうるかが問題となる。そこで我々は，胸部CT類似症例検索システムを想定した適切でない検索結果の混在した類似症例を提示した時に読影結果にどのような影響を与えるのかを検討した。

対象と方法

対象

被検者は9人の医師(うち5人は放射線科専門医，4人は非専門医)とした。

胸部CT類似症例システムデータベースと病変カテゴリ分類

医学論文に掲載されている確定診断がついた胸部CTを約2000枚登録した。病変部は病変の濃度・大きさ・テクスチャ・数などによってすりガラス×結節影，すりガラス×均等影，すりガラス×多発中陰影，多発索状影，単発囊胞，多発囊胞，実質影×結節影，実質影×均等影，実質影×多発中陰影，多発小粒状影，空洞性，蜂巣肺

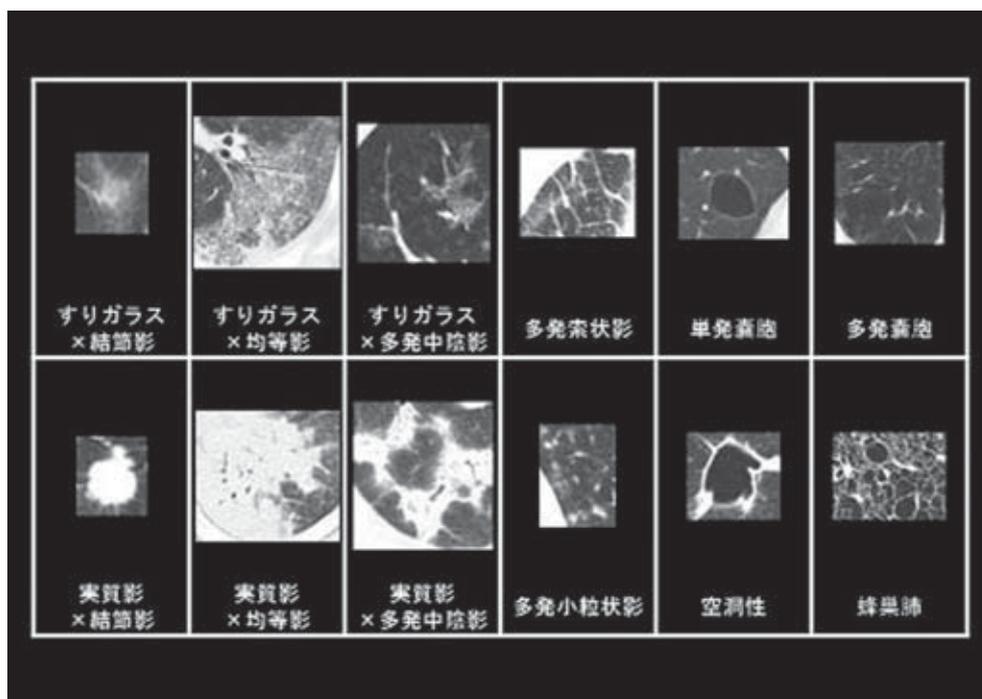


図1 胸部CT類似症例システムデータベースの病変カテゴリ分類

肺異常陰影は全て病変の濃度・大きさ・テクスチャ・数などによって12のカテゴリに分類され登録されている。

陰影，多発小粒状影，空洞性陰影，蜂巣肺の合計12の病変カテゴリに手動で教師付け(分類)し，データベースを作成している．病変カテゴリ名は独自に定義した(図1)．

出題症例の選定

松下記念病院の症例から臨床的な確定診断病名のついている3症例すなわち，すりガラス×均等影の病変カテゴリからびまん性すりガラス陰影を呈したニューモシスチス肺炎症例，多発索状影の病変カテゴリから小葉間隔壁肥厚を呈した肺水腫症例，多発小粒状影の病変カテゴリから多発小葉中心性粒状影を呈した結核症例を選定した．

臨床的な確定診断病名と適切病名の定義

出題症例それぞれに対して，被検者でない放射線科診断専門医2名の合議によって「画像から挙げるべき鑑別疾患病名」(以下，適切病名)とそれ以外の「画像から鑑別に挙げるべきではないと考えられる疾患病名」(以下，不適切病名)を定義した．

出題症例の臨床的な確定診断病名は適切病名の中に含めた．

類似症例の選定

胸部CT類似症例検索システムデータベースに登録されている画像の中で，被検者でない放射線科診断専門医が出題症例と同じ病変カテゴリの中から適切病名：不適切病名の割合が約3：7となるように症例データベースの中から診断名重複のない約20症例の画像を出題症例ごとに設定した．

実験方法

被検者は通常の読影環境で出題症例の年齢，性別と撮像範囲全てのCTのみ提示され，鑑別疾患を列挙した．その後，読影端末とは別のコンピュータにてあらかじめ選定した類似症例を提示し，被検者は類似症例を参考にして鑑別疾患を再考した．

実験全体を通して被検者には臨床情報を非提示とし，教科書，論文などの診断補助の検索は許可しなかった．

被検者に対して読影時間は無制限，鑑別疾患の評価方法は非公表とし，鑑別疾患の記載数には制限を設けなかった．

類似症例提示後に追加または削除された病名の妥当性評価，正診率の評価

評価は専門医，非専門医，全体に分けて1症例当たりの鑑別疾患の平均記載数を集計した．

被検者が類似画像を提示された後に追加された，または削除された疾患病名に対して，上記の適切病名，不適切病名を用いて妥当性を判定し，1症例当たりの増減数を集計した．

追加された疾患の中の適切病名の割合を以下のとおり算出した．

$$\text{追加された疾患の中の適切病名の割合(\%)} = (\text{類似画像提示後に追加された適切病名数} / \text{類似画像提示後に追加された疾患数}) \times 100$$

また，適切病名を不適切病名に対してどの程度選択的に選べたかの指標として適切病名の選択性を以下のように定義した．適切病名：不適切病名の比率は3：7として

$$\text{適切病名の選択性} = 7/3 \times (\text{類似画像提示後に追加された適切病名数} / \text{類似画像提示後に追加された不適切病名数})$$

被検者が挙げた鑑別疾患の中に「臨床的な確定診断病名」が入っていれば正診とし，正診率を算出した．

$$\text{正診率(\%)} = (\text{正診となった症例数} / \text{出題症例数}) \times 100$$

上記計算結果は有効数字3桁以下を四捨五入した．

結 果

1症例当たりで平均した類似画像提示前後の鑑別疾患記載数は専門医で2.3個から2.9個，非専門医で2.5個から3.5個，全体で2.4個から3.2個となり，専門医0.6個，非専門医1.0個，全体では0.8個増加した(表1)．

表1 鑑別疾患病名記載数／症例(標準偏差)

	専門医	非専門医	全 体
類似症例提示前	2.3(1.0)	2.5(1.4)	2.4(1.2)
類似症例提示後	2.9(1.7)	3.5(2.3)	3.2(2.0)

類似症例提示後に追加された鑑別疾患は専門医 0.7 個、非専門医で 1.3 個、全体で 1.0 個であり、追加された鑑別疾患のうち適切病名と判定されたのは 1 症例当たり専門医 0.5 個、非専門医で 0.6 個、全体で 0.5 個、不適切病名と判定されたのは 1 症例当たり専門医 0.3 個、非専門医で 0.8 個、全体で 0.5 個であった(表 2)。

類似症例提示後に削除された鑑別疾患は専門医 0.1 個、非専門医で 0.3 個、全体で 0.2 個であり、専門医、非専門医とも適切病名は削除されておらず、不適切病名のみ削除された(表 3)。

追加された疾患の中の適切病名の割合は専門医 63 %、非専門医 43 %であった。

不適切病名に対する適切病名の選択性は専門医で 4.0 倍、非専門医で 1.8 倍となった。

正診率は類似症例提示前後で専門医は 67 % から 73 %、非専門医は 50 % から 58 %、全体で 59 % から 67 % となり、全体で 8 ポイント向上した(表 4)。

考 察

今回の読影実験では、被検者でない 2 名の診断専門医によって CT の情報から鑑別に挙げるべき適切病名を合議によって定義し、類似症例提示前後で読影レポートに追加または削除された鑑別疾

表 2 追加された鑑別疾患病名数/症例(標準偏差)

	専門医	非専門医	全 体
適切病名	0.5(0.9)	0.6(0.7)	0.5(0.8)
不適切病名	0.3(0.6)	0.8(1.1)	0.5(0.8)
合 計	0.7(1.2)	1.3(1.2)	1.0(1.2)

表 3 削除された鑑別疾患病名数/症例(標準偏差)

	専門医	非専門医	全 体
適切病名	0.0(0.0)	0.0(0.0)	0.0(0.0)
不適切病名	0.1(0.5)	0.3(0.5)	0.2(0.5)
合 計	0.1(0.5)	0.3(0.5)	0.2(0.5)

表 4 臨床的な確定診断病名記載率(標準偏差)

	専門医	非専門医	全 体
類似症例提示前	67% (33.3)	50% (19.2)	59% (27.8)
類似症例提示後	73% (28.0)	58% (16.7)	67% (23.6)

患病名が適切病名か不適切病名か判定することによりその妥当性を評価するという点と、正診率の変化の 2 点で、類似症例提示による鑑別疾患に与える影響の評価を行った。

類似画像提示後に適切病名が新たに追加された、または類似症例提示前に挙げられた不適切病名が類似症例提示後に削除された場合、類似画像提示によって鑑別疾患の妥当性が向上したと考えられる。今回の実験では類似症例提示により、平均して不適切病名が 0.5 個追加されたものの、適切病名が 0.5 個追加され、不適切病名が 0.2 個削除される効果が認められ、全体として類似症例提示によって鑑別診断の妥当性が向上したと考えられ、結果として正診率も 8 ポイント向上したという結果になった。

我々が開発中の胸部 CT 類似症例検索システムは日常診療での利用を想定しており、診断に迷う必要のない common disease や、逆に日常診療では遭遇しにくい非常に稀な症例に関しては利用対象となることは少ないと考え、出題症例は胸部 CT 類似症例システムデータベースで採用されている 12 カテゴリー分類の中から異なる 3 カテゴリーを選び、かつ疾患特異性の比較的高い所見が認められ、日常診療レベルよりも難易度が高く、当院で 1 年に 1 ~ 数回遭遇するようなレベルのものを採用し、当院の症例を用意した。

同システムの利用の際には診断医が胸部 CT の病変部位を検索対象として矩形で指定する必要があり、システムはその部分の画像特徴量をもとにしてデータベース内の胸部 CT から類似性の高い症例を提示する。このとき矩形の指定が変わると画像特徴量が変化して検索結果も異なるため、今回は提示する類似症例を被検者でない診断専門医があらかじめ選定し統一した。類似画像検索システムの検索結果には適切病名、不適切病名が混在することが多いため、今回の読影実験では診断支援効果が見込める最低限の目標を想定して適切病名：不適切病名の比率を 3 : 7 とした。また、類似症例提示数を少なくすると適切病名や正解疾患名が提示できなくなるリスクが増大することになり、逆に類似症例提示数が多すぎる場合は診断医が提示症例を吟味する手間が増大してしまうことになるため約 20 症例を提示数の目安とした。

今回の読影実験では適切病名と不適切病名の比

率が約3:7であるにも関わらず1症例あたりの追加疾患は専門医で適切病名0.47個に対して不適切病名0.27個となったことから専門医の適切病名の選択性は4.0倍であった。また、非専門医では適切病名の選択性が1.8倍であり、読影スキル向上により適切病名の選択性が向上することを反映していると思われた。また、専門医、非専門医とも初めに鑑別疾患として挙げた適切病名が類似症例を提示された後に削除されたという事例は今回認められなかった。専門医は類似症例の中に不適切と思われるものが混在していたとしても、自身の判断を加えてより適切な取捨選択を行うことで適切病名を多く選んでいたと考えられる。今回の読影実験では専門医と比較して非専門医は類似症例提示後に追加した疾患の数が多く、また適切病名の選択性は約1.8倍に留まった。今回は被検者数が少ないため断定はできないが、非専門医にとっては画像と疾患病名の情報のみからは適切病名と不適切病名との区別が難しく、安全性をとって鑑別疾患を広めに追加することが多かったのではないかと考えられる。いずれにせよ、結果的に専門医、非専門医とも「臨床的な確定診断病名」を追加する効果があるという点は評価されることが考えられる。今回の実験結果では鑑別疾患を削除する効果はわずかであった。理由は、疾患と画像の対比が出来る専門医の場合は読影対象の画像から不適切病名が想起されることがそもそも少ないためと考えられる。非専門医の場合は病変の所見というよりも臨床的な頻度から想起した病名が挙がりやすい印象を受け、類似症例を提示することにより病変の所見の違いに気づき鑑別疾患から不適切病名を削除していた可能性は考慮される。

今回の読影実験では、被検者には非公表ではあるが類似症例の中に必ず「臨床的な確定診断病名」が含まれていた。被験者の中には自分が挙げた鑑別疾患病名が、提示された類似症例の中に入っていないことを理由として鑑別疾患病名から削除した可能性も否定できない。現実の胸部CT類似症例検索システム検索結果には「臨床的な確定診断病名」が必ずしも入っているとは限らないため、今後「臨床的な確定診断病名」が類似症例に含まなかった場合を想定した読影実験による診断への影響を検討する必要がある。

鑑別疾患を広げること自体は必ずしも医療上の

アウトカムを向上させるとは限らないことが予想されその是非の判断は難しいと考えられるが、類似症例の提示は鑑別疾患想起を支援し適切病名を追加する効果があり、結果として正診率が向上しうると考えられた。

結 語

今回は類似症例検索システムを実際に使用した読影実験ではなく、より一般化された「画像の類似性を基準とする類似症例」を提示することの診断への影響を確認した。

画像の類似性を基準とする類似症例の提示は、被験者が初めに想起した鑑別疾患に対して適切病名を追加する効果、わずかながら不適切病名を削除する効果と正診率を向上させる効果を確認し、類似症例の提示は診断支援として有用であると考えた。

今回の実験結果において、非専門医は専門医と比較して類似症例提示後に不適切病名を追加する割合が多かった。非専門医でも適切な適切病名を選択できる支援が必要であると考えられる。

当研究は松下記念病院、パナソニック先端研究本部、福井大学医学部との共同研究である。

文 献

- 1) 医療機器評価指標ガイドライン画像診断分野(コンピュータ診断支援装置)開発WG報告書 http://www.aist.go.jp/pdf/aist_j/iryokiki/2009/techrep_cad_fy2009.pdf
- 2) 若林大志. BSIからみた骨転移. 核医学画像診断(Imaging) 2013; 28: 36-38.
- 3) VSRAD: <http://www.vsrad.info/index2.html> (Access 2017年2月)
- 4) H Müller, N Michoux, D Bandon, et al. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. Int J Med Inform 2004; 73: 1-23.
- 5) Endo M, Aramaki T, Asakura K, et al. Content-based image-retrieval system in chest computed tomography for a solitary pulmonary nodule: method and preliminary exper-

iments. Int J CARS 2012; 7: 331-338.
doi:10.1007/s11548-011-0668-z

- 6) 小塚和紀, 高田和豊, 近藤堅司, 他. 多様な肺疾患の診断・教育を支える類似症例検索技術の開発. 信学技報 2014; MI2013-81.
- 7) 若杉健介, 高田和豊, 近藤堅司, 他. マルチ解像度マルチサイズ画像を入力としたDeep Learningによる類似症例検索. 第36回医療情報学連合大会(横浜市), 2016年11月24日.

Support effects of similar cases for the diagnosis of lung lesions by using computed tomography

Tomoya Kotani, Ito Tetsuya, Tsuji Yusuke, Ohara Yu, Noguchi Naoki, You Ushijima,
Kazutoyo Takata*, Kensuke Wakasugi*, Kazuki Kozuka*, Kenji Kondo*, Masaki Kiyono*,
Toyohiko Sakai**, Hirohiko Kimura**

Matsushita Memorial Hospital, Radiology, Osaka, Japan

**Panasonic Corporation, Advanced Research Division, Kyoto, Japan*

***University of Fukui, Faculty of Medical Sciences, Fukui, Japan*

Purpose

We previously proposed a similar case retrieval system for computed tomography (CT) images of diverse lung lesion patterns. The purpose of this study was to evaluate the effects of similar cases for the diagnosis of diverse lung lesion patterns on CT.

Methods

Nine subjects participated in the experiments; all of them were medical doctors, of which, five were radiology specialists. Three test cases were prepared as examples of diverse lesion patterns, pneumocystis pneumonia, pulmonary edema, and pulmonary tuberculosis. Age, sex, and all CT slices were presented. In all, approximately twenty similar cases were selected for each test, and 30% had a “*correct*” diagnosis. The *correct* diagnosis included appropriate disease identification while reading the images, in addition to a definitive diagnosis. These *correct* diagnoses were determined via a discussion with multiple radiology specialists. The experiment was conducted in two steps. First, the subjects wrote a radiology report without references. Second, the subjects wrote a report using the similar cases. We then analyzed additional and reduced diagnoses using similar cases.

Results

The number of diagnoses increased by 0.8/case after presentation of similar cases because the *correct* diagnosis added 0.5/case, incorrect diagnosis added 0.5/case, and incorrect diagnosis reduced the number by 0.2/case. The number of correct diagnoses was not reduced. The accuracy rate of definite diagnosis improved by 8 points with similar case retrieval.

Conclusions

We evaluated the supportive effects of similar cases for the diagnosis of diverse lung lesion patterns in the lung by CT. According to the results obtained, we found that the presentation of similar cases increased the rate of correct diagnosis and slightly reduced incorrect diagnoses. As a result, the accuracy of definitive diagnosis improved. Presentation of similar cases is useful for the diagnosis of lung lesions using CT.

Key Words: lung, computed tomography, image retrieval, support effects, clinical benefits