

原 著

正解病名を必ずしも含まない症例群が提示された場合の 類似画像提示による胸部CT診断支援効果の検討

小谷知也, 伊藤哲哉, 町田和隆, 岡本敏幸, 吉野祐樹, 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：類似画像提示による胸部CT診断支援において，出題症例に対して提示された類似画像症例群に正解病名が含まれるとは限らず，かつ，その事実が明示された場合，診断にどのような影響を与えるのか評価するために読影実験を行い，その影響を検討した．

対象と方法：被検者は医師7人，出題症例を6症例選定し，出題症例ごとに計10症例の「正解病名を含む類似画像症例群」，「正解病名が含まれない類似画像症例群」を作成した．

被検者に対して4症例は「正解病名を含む類似画像症例群」，2症例は「正解病名が含まれない類似画像症例群」が提示されるような組み合わせを作成し，被検者を無作為に割り付けた．被検者は類似画像症例群の提示前後とも鑑別疾患を3つまで記載し，鑑別疾患ごとに確信度を記載した．被検者には類似画像症例群に正解病名が含まれているとは限らないことを明示した．正解数，正解率，確信度に関して集計し，評価を行った．

結果：「正解病名を含む類似画像症例群」が提示された場合，正解率は71%から79%に上昇した．「正解病名が含まれない類似画像症例群」が提示された場合でも，正解病名が削除されるなどの重大な悪影響は観測されなかった．

結論：類似画像症例群に正解病名が含まれない可能性が明示されても，類似画像症例群の提示は診断支援として有効であると考えられた．

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は著しく，人工知能を様々な分野へ応用する動きが高まっている．放射線科領域ではコンピュータ支援診断への人工知能の応用は，画像情報から特定の医療情報を検出し存

在診断を主に支援するCADe(Computer Aided Detection)と，得られた医療情報を解析し質的診断を行うCADx(Computer Aided Diagnosis)に大きく分けられる¹⁾．他にも，画像情報以外の画像特徴量と患者臨床情報を統合し，診断支援の精度を向上させようとする試みも始まっている²⁾．

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，診断時に遭遇した特定の肺病変に対し，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，

2018年3月19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)
Tel: +81-66992-1231, Fax: +81-66992-4845
E-MAIL: kotani.tomoya@jp.panasonic.com
利益相反：なし

びまん性肺疾患を含む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CADxシステムを開発中である³⁻⁵⁾。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病名症例だけを的確に導くことは現実的に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、正解病名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が多い。また、可能性としては正解病名が含まれない検索結果が返ってきてしまうことも考えられる。

我々は前回の報告で、被検者9人に3出題症例を用意し、(画像から挙げるべき鑑別疾患病名、以下適切病名):(画像から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)の割合が約3:7、合計約20症例を類似画像症例群として提示した場合で読影実験を行い、正解率は59%から67%に上昇、適切病名が削除される症例はなかったことを示した⁶⁾。

前回の読影実験では被検者に対し、提示する類似画像症例群の中には必ず「臨床的な確定診断病名」が含まれていることは非公表としていたが、被検者が類似画像症例を盲目的に信頼して鑑別疾患を増やすなど、バイアスが存在していた可能性も否定できない。また、現在の技術では類似画像検索システムが「臨床的な確定診断病名」を挙げられない可能性も想定され、提示する類似画像疾患群の中に正解病名が含まれるとは限らないことを明示した時でも診断支援として有効かどうかを示す必要がある。また類似画像検索システムが正解病名を含まない類似画像症例群を読影者に提示した場合、読影者がその類似画像症例群を参考にし、本来挙げていた鑑別疾患から正解病名を削除してしまい誤診を招く、不必要な鑑別疾患が追加される、確信度が下がるといった悪影響が出る可能性も想定されるが、その影響について実験し、考察した報告は乏しい。

今回の実験では、不確実性のあるCADxを想定し、正解病名が含まれない可能性があることを明示された類似画像症例群を提示した場合でも被検者に

対する診断支援としての有用性を確認できるか、正解病名が含まれない類似画像症例群が提示されたとき、診断への悪影響がどの程度認められるかを確認するための読影実験を行った。

対象と方法

対象

被検者は医師7人(うち3人は放射線科専門医、以下、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、単発嚢胞、多発嚢胞、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、空洞性陰影、蜂巣肺の合計12の病変カテゴリに独自に定義した上で教師付けし、データベースを作成した。(図1)

出題症例の選定

症例はインターネット上に公開されている6症例(squamous cell carcinoma, multiple aspergillomas, hypersensitivity pneumonitis, pulmonary embolism and infarct, silicosis, pulmonary Langerhans cell histiocytosis)を出題症例として選定した⁷⁾。(図2)

類似画像症例群の選定

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出題症例ごとに1症例の正解病名を含む11症例を胸部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から診断名の重複無く選定し、そこから不正解病名を一つ取り除き1症例の正解病名を含む10症例による「正解病名を含む類似画像症例群」、正解病名を取り除いて正解病名を含まない10症例による「正解病名が含まれない類似画像症例群」を作成した。

表1のように、6つの出題症例のうち4症例は「正解病名を含む類似画像症例群」、2症例は「正解病名が含まれない類似画像症例群」が提示され

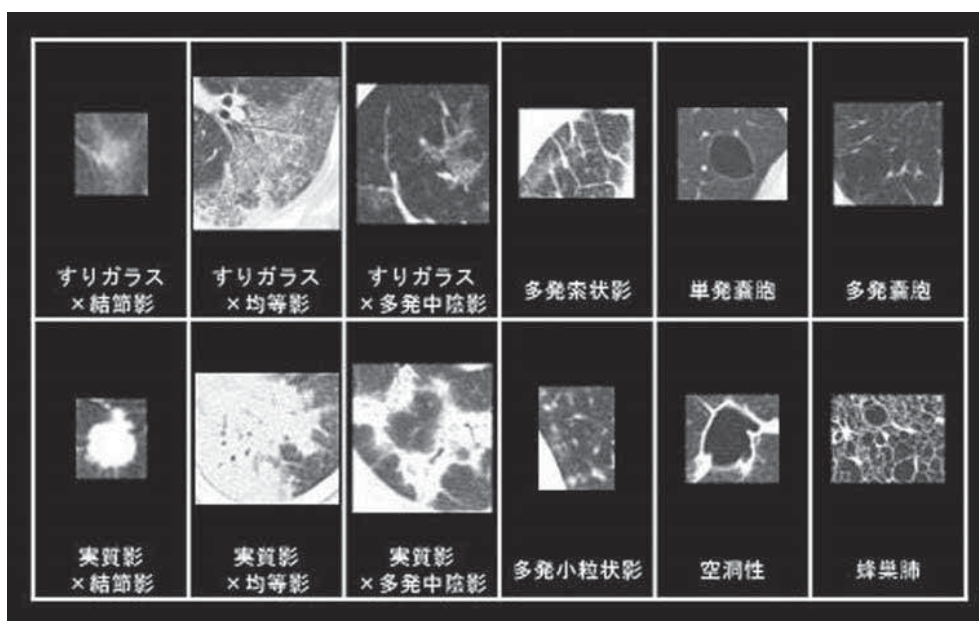


図1 胸部CT類似症例システムデータベースの病変カテゴリ分類

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。

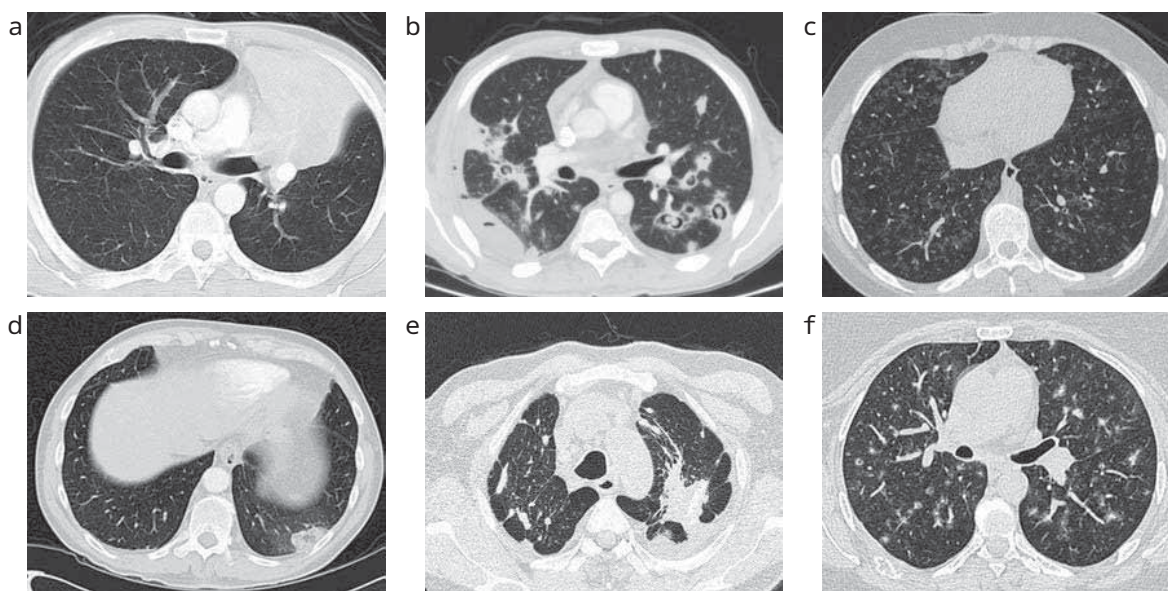


図2

a. Squamous cell carcinoma, b. Multiple aspergillomas, c. Hypersensitivity pneumonitis, d. Pulmonary embolism and infarct, e. Silicosis, f. Pulmonary Langerhans cell histiocytosis

るようにA, B, Cパターンを作成し, 被検者に対していずれかを無作為に割り付けた. 被検者には, 類似画像症例群には正解病名が含まれているとは限らないことを伝えた.

実験方法

被検者はインターネットに接続されたパーソナルコンピュータを使用し, 出題症例の年齢, 性別,

簡単な病歴, 主要な検査所見がまとめられた情報と撮像範囲全てのCTが提示され, 必要に応じて自由に教科書, 論文などの診断補助の検索を行うことを許可され, 鑑別順位の高い順に上限3つまで鑑別疾患を挙げ, 鑑別疾患それぞれに対して確信度を5段階の自己評価として記載した.

その後, あらかじめ選定した類似画像症例群を読影端末とは別のパーソナルコンピュータにて提

表 1

	A	B	C
Q1	×		
Q2	×		
Q3		×	
Q4		×	
Q5			×
Q6			×

：1 症例の正解病名を含む 10 症例による「正解病名を含む類似画像症例群」を提示
 ×：正解病名を含まない 10 症例による「正解病名が含まれない類似画像症例群」を提示とし、被検者に対して A, B, C パターンいずれかを無作為に割り付けた。

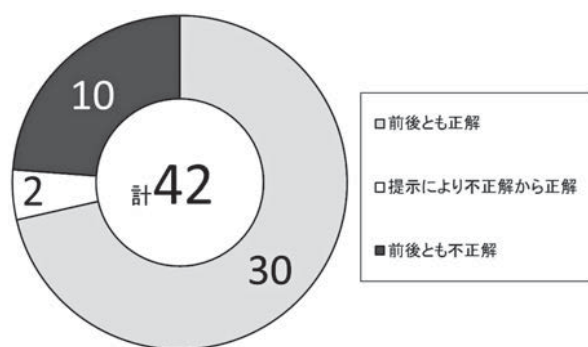


図 3 類似画像症例群提示前後結果

合計 42 回答のうち類似画像症例群提示前後でいずれも正解が 30 回答、いずれも不正解が 10 回答、提示前には不正解であったが提示後に正解となったものが 2 回答となった。

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。

回答時間は類似画像症例群の提示前、後それぞれ 15 分の制限時間を設けた。

回答の評価

上限 3 つの鑑別疾患の中に正解病名が存在すれば正解、無ければ不正解として正解率を算出した。

$$\text{正解率 (\%)} = (\text{正解数} / \text{出題症例数}) \times 100$$

全体、専門医群、非専門医群で正解率、正解病名の確信度の平均を算出した。

「正解病名を含む類似画像症例群」提示群、「正解病名を含まない類似画像症例群」提示群に分けた集計、類似画像症例群提示の前後で正解/不正解、鑑別順位、正解病名を挙げていた場合はその確信度の集計を行った。

表 2 正解率

	専門医	非専門医	全体
類似症例提示前	89%	58%	71%
類似症例提示後	89%	67%	76%

表 3 正解病名の確信度の平均

	専門医	非専門医	全体
類似症例提示前	3.8	3.4	3.3
類似症例提示後	4.0	3.1	3.3

上記計算結果は有効数字 2 桁以下を四捨五入した。

結 果

総計

合計 42 回答の結果、類似画像症例群提示前後でいずれも正解が 30 回答、いずれも不正解が 10 回答、提示前には不正解であったが提示後に正解となったものが 2 回答となった。(図 3)

専門医、非専門医で分けて集計を行うと、正解率は専門医では 18 回答のうち類似画像症例群提示前後どちらも 16 回答(89%)、非専門医では 24 回答のうち類似画像症例群提示前は 14 回答(58%)、提示後は 16 回答(67%)であった(表 3)

正解病名の確信度の平均を類似画像症例群提示前後で集計すると全体で提示前 3.3、提示後 3.3、専門医では提示前 3.8、提示後 4.0、非専門医では提示前 3.4、提示後 3.1であった。

類似画像症例群提示前後で正解していたときの正解疾患の確信度について集計すると、専門医では正解していた計 16 回答のうち、確信度に影響のないものが 14 回答(88%)、確信度が上昇したものが 2 回答(12%)であった。非専門医では正解していた計 14 回答のうち確信度に影響のないものが 6 回答(43%)、確信度が上昇したものが 3 回答(21%)、確信度が低下したものが 5 回答(36%)であった。(図 4)

「正解病名を含む類似画像症例群」提示の効果

「正解病名を含む類似画像症例群」が提示された合計 28 回答を集計すると、提示前後いずれも正

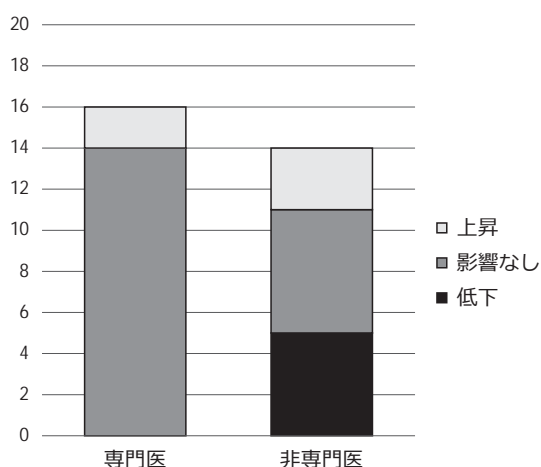


図4 類似画像症例群提示前後で正解病名確信度の影響

類似画像症例群提示前後で正解していたときの正解疾患の確信度について集計した。専門医では正解していた計16回答のうち、確信度に影響のないものが14回答(88%)、確信度が上昇したものが2回答(12%)であった。非専門医では正解していた計14回答のうち確信度に影響のないものが6回答(43%)、確信度が上昇したものが3回答(21%)、確信度が低下したものが5回答(36%)であった。

解であったものが20回答、いずれも不正解であったものが6回答、類似画像症例群の提示前は不正解であったが、提示により正解病名が追加されたものが2回答(いずれも非専門医)あった。

類似画像症例群の提示前は正解であったが提示後に不正解となった回答はなく、「正解病名を含む類似画像症例群」の提示があった場合、正解率は7.1%向上した。

類似画像症例群提示前後いずれも正解であった合計20回答のうち、確信度、鑑別順位とも影響がなかったものが11回答、正解病名の確信度が上昇したものが5回答、鑑別疾患が追加されたものが1回答、鑑別疾患を増減したものが1回答、正解病名の確信度が低下したものが2回答あった。また、類似画像症例群の提示により正解病名の確信度が上昇した5回答のうち1回答は鑑別順位の上昇も認められた。(表4)

類似画像症例群提示前後いずれも不正解であった6回答のうち、提示により確信度、鑑別順位とも影響がなかったものが3回答、鑑別疾患を増減したものが3回答あった。(表5)

「正解病名を含まない類似画像症例群」提示の影響
「正解病名が含まれない類似画像症例群」が提示

表4 正解を含む類似画像症例群提示前後いずれも正解であった回答への影響

	鑑別順位 不変	鑑別順位 上昇	鑑別順位 低下	鑑別疾患 追加	鑑別疾患 増減	計
確信度 不変	11	0	0	1	1	13
確信度 上昇	4	1	0	0	0	5
確信度 低下	2	0	0	0	0	2
計	17	1	0	1	1	20

表5 正解を含む類似画像症例群提示前後いずれも不正解であった回答への影響

鑑別順位 不変	鑑別順位 上昇	鑑別順位 低下	鑑別疾患 追加	鑑別疾患 増減	計
3	0	0	0	3	6

表6 正解を含まない類似画像症例群提示前後いずれも正解であった回答への影響

	鑑別順位 不変	鑑別順位 上昇	鑑別順位 低下	鑑別疾患 追加	鑑別疾患 増減	計
確信度 不変	5	0	1	1	0	7
確信度 上昇	0	0	0	0	0	0
確信度 低下	0	0	0	3	0	3
計	5	0	1	4	0	10

表7 正解を含まない類似画像症例群提示前後いずれも不正解であった回答への影響

鑑別順位 不変	鑑別順位 上昇	鑑別順位 低下	鑑別疾患 追加	鑑別疾患 増減	計
3	0	0	0	1	4

された14回答を集計すると、提示前後いずれも正解であったものが10回答、いずれも不正解であったものが4回答であり、類似画像症例群の提示前は不正解で、提示により正解病名が追加されたものや、提示前は正解であったが提示後に不正解となった回答はなかった。

類似画像症例群提示前後いずれも正解であった合計10回答のうち、確信度、鑑別順位とも影響がなかったものが5回答であった。鑑別疾患が追加されたものが4回答あり、そのうちの1回答は

正解病名の確信度の変化は無かったが、3 回答は正解病名の確信度の低下があわせて認められた。また 1 回答は類似画像症例群の提示により正解病名の確信度は変化しなかったが正解病名に関して鑑別順位の低下が認められた。(表 6)

類似画像症例群提示前後いずれも不正解であった 4 回答のうち、確信度、鑑別順位とも影響がなかったものが 3 回答、鑑別疾患を増減したものが 1 回答あった。(表 7)

考 察

今回の出題症例は、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いることから、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 カテゴリ分類の中から互いに異なる 6 カテゴリを選び、その中から common disease 及びやや稀な症例が含まれるように選択した。

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る、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ように設計されているが、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。

前回我々が報告した読影実験では、提示する類似画像症例群の選定は類似画像症例群を症例データベースの中から適切病名：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、被験者には明示していないが全て正解病名を含むものを提示し、読影実験中は文献などの検索は不許可とした。

一方今回の読影実験では、提示する類似画像症例群は一律に計 10 症例に限定し、被験者には類似画像症例群の中に正解が含まれているとは限らないことを明示し、被験者には明示していないが正解病名を含む類似画像症例群が提示されるのは 4/6 症例としており、類似画像症例群としては不完全性がより高いと考えられる条件を想定し、また被験者には 15 分の間文献などの検索を許可して実験設計を行った。

専門医は非専門医と比較して正解率、確信度いずれも高く、高い読影スキルを反映しているものと考えられた。正解病名の確信度の平均は、全体では類似画像症例提示後に上昇していたが、非専門医ではわずかに低下していた。非専門医で類似画像症例提示後に正解病名の確信度低下した要因は、非専門医では類似画像症例提示後に 24 回答中に確信度の上昇が認められた回答が 3 回答あったのに対して、低下が認められた回答が 5 回答あり、また 2 回答は正解病名が追加された回答で、低い確信度をつけて正解病名が追加されたためであった。

今回の読影実験では正解病名を含む類似画像症例群が提示された場合に 7.1 % の正解率向上が確認されたことは、前回の読影実験で約 8 % の正解率向上が認められていたことと同様の傾向であり、正解病名が含まれるとは限らないことが明示されている状況でも類似画像症例群の提示は被検者に適切病名の追加を促し、結果的に正解率を上昇させる効果が再確認されたものと考えられた。

専門医では文献などの検索は、自身の記憶を確認する作業が主体であり 15 分という時間制限でも効率よく正解にたどり着くことができる一方、非専門医は初学者になるほど知識が少ないため、文献などの検索が許可されても膨大な情報の中からどこを検索するかという当たりを付けることから難しく、初見で分からなかった症例に対しては制限時間 15 分程度では自力で正解にたどり着くことが難しいと考えられる。類似画像症例群の提示は画像のみという制約はあるものの、膨大な情報の中から画像が類似しているものが提示されることを目的としており、その想起支援効果から効率のよい疾患検索が可能になっている可能性がある。

正解病名を含まない類似画像症例群の提示が原因で鑑別順位が下がることや、確信度が低下することは確認されたが、正解病名が削除されることは確認されなかった。正解病名を含まない類似画像症例群の提示による正解率の向上は認められなかったが、低下も認められないという結果だった。今後、実験数を増やすと類似画像症例群提示により正解病名が削除されてしまう例が確認される可能性がある。その対策として例えば、被験者に対して類似画像症例群の役割を鑑別疾患想起支援の

みと捉えて、鑑別疾患を追加することのみ許可し、削除することを禁止することも有効であると考えられ、正解病名を含まない類似画像症例群の提示が仮になされたとしてもその悪影響は少ないものと思われた。

結 語

類似画像症例群の中に正解病名が含まれるとは限らず、かつ、それを明示される状況下であっても、正解病名が含まれている類似画像症例群の提示があった場合には診断支援効果による正診率向上が期待でき、また正解病名が含まれていなかった場合でも、鑑別に挙げていた正解病名が削除されるような重大な悪影響は確認されなかった。

類似画像症例の提示は診断支援として有用で、仮に正解病名が含まれていなかった場合でも悪影響は限定的であると考えられた。

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。

参考文献

- 1) Ray Cody Mayo , Jessica Leung. Artificial intelligence and deep learning-Radiology s next frontier? Clinical Imaging 2018; 49: 87 - 88 .
- 2) 上野翔子 , 杉山 治 , 西尾瑞穂 , 他 . 患者情報を考慮した胸部CT画像の診断支援の試み . 医療情報学会・人工知能学会AIM合同研究会資料 , 2017 ; SIG-AIMED-004-04 : 04-1-04-4 .
- 3) 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 ; 113 : 139 - 142 .
- 4) 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 , 2016 .
- 5) 若杉健介 , 近藤堅司 , 清野正樹 . Convolutional Neural Networksを用いた類似症例検索 . 医用画像情報学会雑誌 2017 ; 34 : 109 - 111 .
- 6) 小谷知也 , 伊藤哲哉 , 辻 悠佑 , 他 . 類似症例提示による胸部CT診断支援効果の検討 . 松仁会医誌 2017 ; 56 : 1 - 7 .
- 7) 出題症例
Squamous cell carcinoma:
<https://radiopaedia.org/cases/squamous-cell-carcinoma-of-the-lung>
Multiple aspergillomas:
<https://radiopaedia.org/cases/pulmonary-aspergillosis-multiple-aspergillomas>
Hypersensitivity pneumonitis:
<https://radiopaedia.org/cases/hypersensitivity-pneumonitis-1>
Pulmonary embolism and infarct:
<https://radiopaedia.org/cases/pulmonary-embolism-and-infarct>
Silicosis:
<https://radiopaedia.org/cases/silicosis-with-progressive-massive-fibrosis>
Pulmonary langerhans cell histiocytosis:
<https://radiopaedia.org/cases/pulmonary-langerhans-cell-histiocytosis-14>
(Access 2018 年 5 月)

Support effects of similar cases which include diagnostic uncertainty in lung computed tomography

Tomoya Kotani, Tetsuya Ito, Kazutaka Machida, Toshiyuki Okamoto, Yuki Yoshino,
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

To evaluate the support effects of similar cases in lung computed tomography, even if it is informed that the similar cases do not necessarily contain correct diagnosis.

Methods

Subjects were seven doctors, of whom three were radiologists. We prepared six test cases as examples of having diverse lesion patterns, ten duplicate similar cases set with/without one correct diagnosis for each test case.

We made three pattern of similar cases set for four test cases contained correct diagnosis and for two test cases contained no correct diagnosis, and randomly assigned to subjects.

The experiment was conducted in two steps. First, each subject diagnose up to three disease for six test cases with reference to usual medical texts. Second, each subject modified his diagnosis with reference to similar cases set. We informed the subjects that the similar cases presented do not necessarily contain correct diagnosis. We evaluated number, accuracy rate, and confidence of correct diagnosis.

Results

The accuracy rate of definite diagnosis improved from 71% to 79%. Further, similar cases containing no correct diagnosis had few negative influence on the reports.

Conclusions

We evaluated the support effect of similar cases for the diagnosis of lung CT. According to the results obtained, we found that the presentation of similar cases increased the rate of correct diagnosis. Presentation of similar cases is useful for the diagnosis of lung lesions by using CT, even if it is informed that the similar cases do not necessarily contain correct diagnosis.

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