

論文内容要旨

Content-based CT image retrieval system using deep learning: Preliminary assessment of its accuracy for classifying lesion patterns and retrieving similar cases among patients with diffuse lung diseases

(深層学習を用いた CT 画像検索システムによる、びまん性肺疾患患者の病変性状分類および類似症例検索精度の予備評価)

Hiroshima Journal of Medical Sciences, 71(1-2):9-22, 2022.

主指導教員：栗井 和夫 教授

(医系科学研究科 放射線診断学)

副指導教員：服部 登 教授

(医系科学研究科 分子内科学)

副指導教員：田代 聡 教授

(原爆放射線医科学研究所 細胞修復制御)

寺田 大晃

(医歯薬保健学研究科 医歯薬学専攻)

一般に、患者の画像データは本人の経過観察や治療計画に利用されているものの、他の患者の診断支援等には活用されていない。その理由は画像データベースを活用するための実用的な画像検索システムが現時点では存在しないためである。画像診断医は診断困難な症例に遭遇した場合、しばしば、自身が過去に経験した症例と比較をしたり、確定診断のついた類似画像症例を文献で検索したりするため、画像データベース上で類似した形態の画像を検索できるシステムが実現できれば迅速な診断および業務の効率化につながる可能性がある。びまん性肺疾患は胸部 CT 等の画像検査でびまん性の多彩な肺異常陰影を示す疾患群であり、多数の異なる疾患を含んでいる。今回、我々は深層学習技術（ディープラーニング）を用いて、びまん性肺疾患を対象として形態類似性に着目した CT 画像検索システムを開発し、病変パターンの分類精度と類似症例の検索精度を検討した。

対象は胸部 CT 症例 503 例で、うち 328 例は CT 画像上でびまん性異常陰影を含んでいた。まず、2 名の放射線診断専門医が、各症例における優位な肺野所見を 5 種類の肺異常陰影（浸潤影、気腫性陰影、蜂巣肺、すりガラス影、粒状影）あるいは正常に合議により分類し、CT 画像検索システムの参照基準（正解）とした。病変性状の分類については、precision（適合率）、recall（再現率）および F-score（precision と recall の調和平均）を評価した。次に 503 例の中から 5 種類の異常陰影のそれぞれにつき 3 症例ずつ（計 15 症例）を検索元画像として選択した。残りの 488 例（=503 例-15 例）から病変の類似性に基づいて検索を行い、検索された画像について、別の 3 名の放射線診断専門医が検索元画像との類似性を視覚的に 5 段階で評価した。病変の性状と分布の両方がほぼ一致する場合は Grade5、両方とも比較的類似している場合は Grade4、病変の性状は非常に類似しているが分布が異なる場合は Grade3、病変の性状が比較的類似しているが分布が異なる場合は Grade2、病変の性状と分布がいずれも異なる場合は Grade1 とした。3 名の観察者間の一致は Cohen kappa 係数を用いて評価した。Kappa 値が 0-0.20 の場合は「悪い」、0.21-0.40 の場合は「良い」、0.41-0.60 の場合は「中程度」、0.61-0.80 の場合は「良い」、0.81 以上の場合は「非常に良い」とした。

結果は F-score は浸潤影で 0.71、気腫性陰影で 0.74、蜂巣肺で 0.61、すりガラス影で 0.63、粒状影で 0.15、正常肺で 0.67 であった。3 名の放射線診断専門医全員が浸潤影、気腫性陰影、蜂巣肺を認める症例の 68%に病変の性状および分布が類似している Grade4 もしくは Grade5 と評価した一方で、すりガラス影、粒状影を認める症例の 68%では病変の分布が異なる Grade2 もしくは Grade3 と評価した。観察者間の一致は非常に良好であった。

CT 画像検索システムの検索精度は浸潤影、気腫性陰影、蜂巣肺を認める症例については概ね満足できるものであったが、すりガラス影、粒状影を認める症例については満足のいく結果ではなかった。臨床的に有用であるためには症例を収集し学習症例をさらに増やして検索性能を向上させる必要がある。