

令和 4 年 6 月 25 日現在

機関番号：32629

研究種目：基盤研究(C) (一般)

研究期間：2018～2021

課題番号：18K11190

研究課題名(和文) 新たなデータ融合型深層学習手法に基づくびまん性肺疾患診断法の確立

研究課題名(英文) Establishment of diagnosis of diffuse lung diseases based on a data-integrated deep learning method

研究代表者

小森 理 (Komori, Osamu)

成蹊大学・理工学部・准教授

研究者番号：60586379

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：びまん性肺疾患データ解析において以下の研究成果を得た。1. 準線形モデリングのクラスター構造を、k-means, fuzzy c-means, 正規混合モデルの観点から考察し、その統計的性質を明らかにした。2. コロナの影響で、臨床データ、病理データの解析は事実上進めることができなかったが、主要データである画像データを用い、深層学習、転移学習、アンサンブル学習を組み合わせることで従来の判別精度より10%ほどの改善を達成した。3. 深層学習法の可視化の1つのGrad-Camを用いることで、画像データのどの部位がどの病変と深く関係があるかの知見を一定程度得ることができた。

研究成果の学術的意義や社会的意義

本研究では深層学習、転移学習、アンサンブル学習を組み合わせることで、先行研究による分類精度を10%ほど改善することに成功した。分類がうまく行かない原因の考察や、分類結果の可視化による医学的な知見の獲得までには至らなかったものの、肺疾患の病変分類の精度改善には貢献しており、社会的な意義も大きいと思われる。

研究成果の概要(英文)：I was able to study a diffuse lung disease analysis study.(1) Applying the structure of quasi-linear options from behind k-means, fuzzy c-means, and normal mixed models to make statistical changes.(2)With Corona access, you can analyze clinical data and Easter data from above. The accuracy of the evaluation of bull learning is 10%.(3)By doing one Grad-Cam of visualization of deep learning method, it is said that the arrangement of the image is obtained.

研究分野：データサイエンス

キーワード：医療統計

## 1. 研究開始当初の背景

急速な情報化社会の中で、医療の分野でも多種多様なデータを利用し、病気のリスク軽減や診断精度向上の取り組みが行われてきた。形態も性質も異なるさまざまなデータを適切に組み合わせることで、1つのデータからでは抽出できない病気固有の特徴量を捉えることができるようになってきた。多様な要因を持ち診断が難しいとされるびまん性肺疾患においても、ATS/ERSの2002年のconsensus reportで指摘されているように臨床、画像、病理の三者を有機的に総合して判断することが必要とされているが、医師の長年の経験や勘に頼ることが多く、これまで客観的で効果的な診断法は確立されてこなかった。そこで今回我々は統計的機械学習の分野で注目されている深層学習の手法を用い、臨床、画像、病理データを融合した新たな解析手法の開発を行う。そして高分解能CTの読影技術で世界レベルを誇る福井大学高エネルギー医学研究センターの研究グループの協力を得て、福井大学病院で蓄積されたびまん性肺疾患のデータ解析を行い、びまん性肺疾患のevidence-basedな早期診断法の確立を目指す。

## 2. 研究の目的

臨床、画像、病理のデータを総合的に解析する手法を構築することで、びまん性肺疾患の診断精度がどれほど改善されるのか(診断精度)?またそれぞれのデータから抽出された特徴量がどのようにびまん性肺疾患の診断に寄与しているのか(結果の解釈可能性)?この2点を明らかにしたい。

## 3. 研究の方法

初年度に臨床、画像、病理のデータの整備を進めた。しかしながらデータの機密性から全て作業は福井大学病院内で行う必要があり、また研究途中でコロナウィルスの蔓延もあり、実質的なデータの整備は画像データのみとなった。そこで画像データに対し、以下の手順で研究を進めた。

びまん性を含む肺疾患の12病変の分類精度を向上させるために、深層学習、転移学習、アンサンブル学習を組み合わせる手法を提案した。深層学習の手法としてはDenseNet, ResNet, VGG16, VGG19の方法を比較検討し、転移学習としては膨大な画像データベースであるImagenetから事前に抽出された画像分類に有用な情報を用いた。通常は深層学習法で事前学習した畳み込み層の情報を用い、全結合層のみを手元のデータに合わせる形でパラメータを再度推定することが行われるが、今回の肺疾患の病変分類の解析ではImagenetで事前に学習したパラメータの初期値のみを用いた。これはImagenetの中には今回の肺疾患に関わる画像は含まれていないものの、パラメータの初期値に含まれる画像分類に重要で本質的な構造は今回のような医療画像データにおいても有用であることを示している。また診断に重要な部位をGrad-Camを用いて可視化し、実際の医学的な知見と比較検討を行った。

## 4. 研究成果

本研究では準線形モデリングのクラスター構造を援用し、臨床、画像、病理のデータを用い、びまん性肺疾患を含む画像診断の精度を向上させ、新たな診断法の開発を目指した。また用いたデータのどの特徴量がどのように疾患に関連しているかを明らかにすることを目的としていた。具体的な成果としては

1. 準線形モデリングのクラスター構造を、k-means, fuzzy c-means, 正規混合モデルの観点から考察し、その統計的性質を明らかにした。
2. コロナの影響で、臨床データ、病理データの解析は事実上進めることができなかったが、主要データである画像データを用い、深層学習、転移学習、アンサンブル学習を組み合わせることで従来の判別精度より10%ほどの改善を達成した。
3. 深層学習法の可視化の1つのGrad-Camを用いることで、画像データのどの部位がどの病変と深く関係があるかの知見を一定程度得ることができた。

上記1の成果に関しては、各クラスターをKolmogorov-Nagumo平均で結ぶことで、クラスター分析で広く用いられてきたk-means, fuzzy c-means, 正規混合モデルを統一

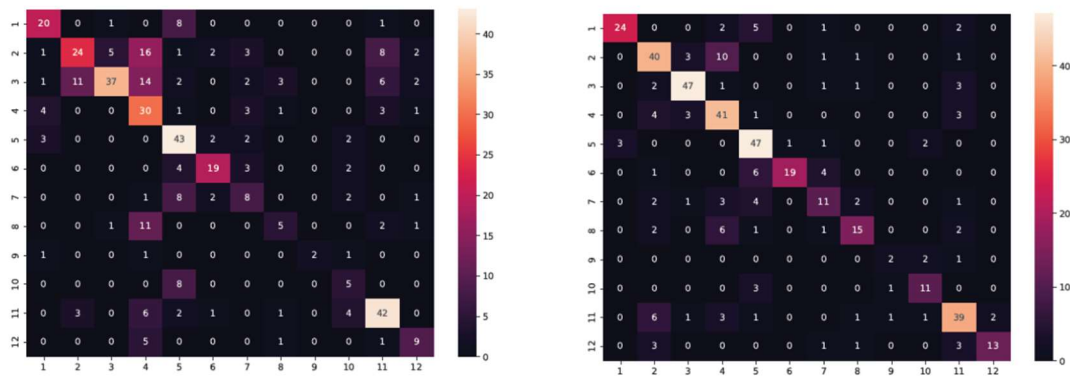
的に扱えることを示した. また各クラスターの平均と分散構造を規定するパラメータをEM アルゴリズムを用いて効率よく推定するアルゴリズムを提案し, クラスターデータベースでのその性能を評価した. 学会発表と論文執筆で研究成果を公表した.

上記2の成果に関しては, 深層学習法ではDenseNet, ResNet, VGG16, VGG19らを, 転移学習ではImageNetを, アンサンブル学習ではAdaboostなどによる重み付き多数決手法を用いた. 深層学習法に含まれる膨大なパラメータをファインチューニングすることで10%以上の精度改善を達成した(表1).

表 1: 各モデルアーキテクチャ毎のテストデータに対する Top1 正答率

アーキテクチャ	転移学習なし	転移学習あり※1	転移学習あり※2	転移学習あり※2+アンサンブル学習
小さな CNN	42.9 %	NA	NA	NA
DenseNet	42.5 %	62.1 %	68.9 %	NA
ResNet	44.6 %	42.9 %	45.7 %	NA
VGG16	59.1 %	65.4 %	68.6 %	72.4 %
VGG19	57.9 %	71.7 %	72.1 %	73.3 %

また12病変に関する詳細な結果を図1に示す. どちらも「単発嚢胞(9)」の正答率が低いことが分かる. また、「すりガラス×多発小粒状影(4)」を「すりガラス×均等」と取り違えているケースも相対的に多いことが分かる.

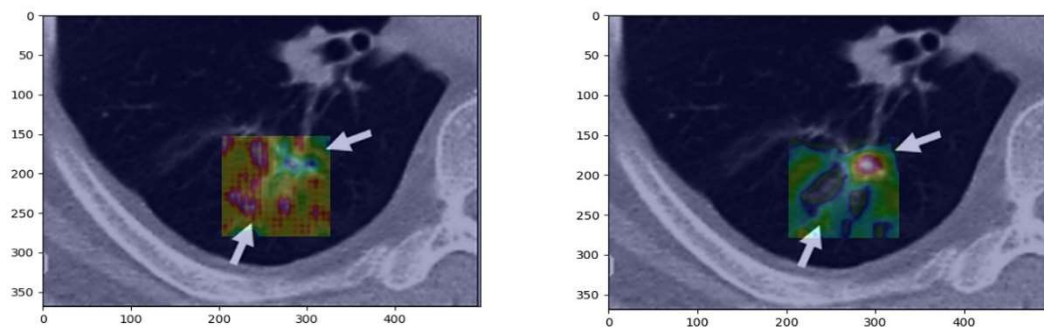


(a)従来法

(b)提案法

図 1 12病変における従来法と提案法の判別結果(混同行列)

上記3の成果に関し. 図2はその1つの病型を持つ画像データに深層学習法であるCNNを適用した結果である. 肺疾患のどの部位が病型に密接に関連しているかを可視化しており, 従来法と比較して関連部位の同定がより明確になった. また判別精度においても改善することができた.



(a)先行研究の手法

(b)今回の手法

図 2 Grad-Camによる病変部位の可視化

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 1件）

1. 著者名 Osamu Komori and Shinto Eguchi	4. 巻 23
2. 論文標題 A Unified Formulation of k-Means, Fuzzy c-Means and Gaussian Mixture Model by the Kolmogorov-Nagumo Average	5. 発行年 2021年
3. 雑誌名 Entropy	6. 最初と最後の頁 1-21
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/e23050518	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計3件（うち招待講演 1件/うち国際学会 0件）

1. 発表者名 小森理
2. 発表標題 G-computation に関するoverview
3. 学会等名 統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 小森理
2. 発表標題 一般化エネルギー関数に基づくクラスター分析
3. 学会等名 統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 小森理
2. 発表標題 統計的機械学習法とデータの異質性に注目した解析法
3. 学会等名 SICE制御部門 データ科学とリンクした次世代の適応学習制御調査研究会（招待講演）
4. 発表年 2018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	江口 真透  (Eguchi Shinto)  (10168776)	統計数理研究所・数理・推論研究系・特任教授    (62603)	

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
---------	---------