

## 一般口演8 画像情報

2021年11月19日(金) 16:30 ~ 18:00 H会場 (2号館3階234)

### [2-H-2-03] 二次元画像分類モデルによる間質性肺炎代表断面像抽出手法の開発

\*張 家豪<sup>1</sup>、和田 聖哉<sup>2,1</sup>、杉本 賢人<sup>1</sup>、武田 理宏<sup>1</sup>、岡田 佳築<sup>2,1</sup>、真鍋 史朗<sup>1</sup>、小西 正三<sup>1</sup>、松村 泰志<sup>3,1</sup> (1. 大阪大学大学院医学系研究科 医療情報学, 2. 大阪大学大学院医学系研究科 変革的医療情報システム開発学寄附講座, 3. 国立病院機構大阪医療センター)

\*Jiahao Zhang<sup>1</sup>, Shoya Wada<sup>2,1</sup>, Kento Sugimoto<sup>1</sup>, Toshihiro Takeda<sup>1</sup>, Katsuki Okada<sup>2,1</sup>, Shirou Manabe<sup>1</sup>, Shozo Konishi<sup>1</sup>, Yasushi Matsumura<sup>3,1</sup> (1. 大阪大学大学院医学系研究科 医療情報学, 2. 大阪大学大学院医学系研究科 変革的医療情報システム開発学寄附講座, 3. 国立病院機構大阪医療センター)

キーワード : Deep Learning, Computed Tomography, Interstitial Pneumonia

#### 【背景】

間質性肺炎は肺の間質を中心に炎症を来して線維化が生じる疾患の総称であり、多様な病型を含む。本邦のガイドラインでは、典型的な特発性間質性肺炎の臨床像ではない場合、予後や治療方針決定のために外科的肺生検による病理組織的診断を行うことが推奨されているが、より非侵襲的な検査による診断補助ツールとして、胸部CTによる深層学習モデルの活躍が期待されている。

#### 【目的】

間質性肺炎患者の胸部CT画像から重症度・予後予測を行うこと。そのために本研究ではまず、軽量の機械学習モデルの学習に使用する代表断面像の抽出器を構築する手法を提案する。

#### 【方法】

当院倫理審査委員会の承認を得て、間質性肺炎と診断された患者472例の胸部CTを用いた（400例：学習データ、72例：評価用データ）。代表断面像抽出器には、公開されている事前学習済みモデルの重みが利用可能な、二次元画像の入力をベースとする深層学習アルゴリズムを採用した。各症例について、肺尖部、横隔膜直上部、気管支分岐部の断面像を手作業で選択して抽出対象クラスとした。残りのスライス別のクラスに割り振ることでそれぞれを独立した二値分類タスクと定義し、個別に3種類の学習モデルを構築した。不均衡データによる学習を支援するために、data augmentationと損失関数のサンプル重み付けを適用した。精度評価では、抽出対象スライスに加えて、その上下1枚を学習モデルが抽出した場合でも正しく抽出出来たものと判定して計算した。

#### 【結果】

本研究で評価した深層学習アルゴリズムの中では、EfficientNet-B4が最も性能が良かった。適合率/再現率は、肺尖部モデル、気管支分岐部モデル、横隔膜直上部モデルでそれぞれ95.1%/97.2%、94.2%/95.8%、93.7%/93.0%であった。

#### 【結語】

胸部CTから代表断面像を自動的に抽出するモデルを構築した。

# 二次元画像分類モデルによる間質性肺炎代表断面像抽出手法の開発

張 家豪<sup>\*1</sup>, 和田 聖哉<sup>\*2,\*1</sup>, 杉本 賢人<sup>\*1</sup>, 武田 理宏<sup>\*1</sup>, 岡田 佳築<sup>\*2,\*1</sup>,  
真鍋 史朗<sup>\*1</sup>, 小西 正三<sup>\*1</sup>, 松村 泰志<sup>\*3,\*1</sup>

\*1 大阪大学大学院医学系研究科 医療情報学

\*2 大阪大学大学院医学系研究科 変革的医療情報システム開発学寄附講座

\*3 国立病院機構大阪医療センター

## Development of a Method for Extracting Representative CT Slices of Interstitial Pneumonia using 2D Image Classification Model

Jiahao Zhang<sup>\*1</sup>, Shoya Wada<sup>\*2,\*1</sup>, Kento Sugimoto<sup>\*1</sup>, Toshihiro Takeda<sup>\*1</sup>, Katsuki Okada<sup>\*2,\*1</sup>, Shirou Manabe<sup>\*1</sup>, Shozo Konishi<sup>\*1</sup>, Yasushi Matsumura<sup>\*3,\*1</sup>

\*1 Medical Informatics, Osaka University Graduate School of Medicine,

\*2 Department of Medical Informatics, Osaka University Graduate School of Medicine,

\*3 National Hospital Organization Osaka Medical Center

**Background:** Interstitial pneumonia is a general term for diseases that cause inflammation and fibrosis, mainly in the interstitium of the lung, and includes a variety of disease types. Although histopathological diagnosis by surgical lung biopsy is required to determine the prognosis and treatment strategy, we expect deep learning chest CT models to play an essential role as a diagnostic tool for non-invasive examination.

**Purpose:** To obtain positional markers from chest CT images of patients with interstitial pneumonia to determine representative slices used in training models to predict severity and prognosis.

**Methods:** We used chest CT images of 472 patients diagnosed with interstitial pneumonia. We manually selected apical lung, tracheal bifurcation, and the upper diaphragm as the target class for extraction for each case. By assigning the remaining slices to different classes, we defined each as an independent binary classification task and constructed three separate models. Data augmentation and loss function sample weighting were applied to support training on imbalanced data.

**Results:** In the apical lung model, EfficientNet-b4 achieved a precision of 95.1%. In the tracheal bifurcation model, DenseNet-169 achieved a precision of 95.6%. In the upper diaphragm model, EfficientNet-b4 achieved a precision of 93.7%.

**Keywords:** Deep Learning, Computed Tomography, Interstitial Pneumonia

## 1 緒論

間質性肺炎は、複数の原因によって引き起こされる間質性肺の炎症性・線維性疾患である。病変は主に肺胞上皮細胞、毛細血管内皮細胞、基底膜、血管周囲やリンパ管組織などの間質性肺や肺胞内腔に浸潤し、最終的には間質性肺の線維化を引き起こし、肺胞-毛細血管機能の低下をもたらす。臨床症状や胸部 CT の特徴から確定診断できない間質性肺炎に対しては、通常、病理学的変化を明らかにするために、経気管支鏡的肺生検、経皮的穿刺肺生検、外科的肺生検が必要であるが、より簡便な非侵襲的診断補助ツールの開発も期待されている。

ルーチンで行われる画像検査に最先端の機械学習を適用し、従来以上の疾患解釈を目指す研究が進められている。特に深層学習の発展は目覚ましく、Deep Convolutional Neural Networks (DCNN) <sup>1)</sup>を画像に適用したところ、人間の目の認識精度を超える分野もみられるようになった。また、深層学習モデルのアルゴリズムやハードウェア設備の発達により、深層学習で培われた画像技術を医用画像などの分野に適用することも可能になってきた。胸部エックス線画像を 2D CNN で診断し、結節などの病変を分類してラベル付けする研究はすでに多数存在し、良好な結果が得られている。

CT 画像に関する研究では、通常、画像の空間的特徴を抽出するために 3D CNN が使用される。例えば、肺炎の診断に

関連する研究を考える。まず肺の領域をセグメント化し、次に肺炎の有無を診断的に分類し、最後に肺炎の感染領域の視覚的な 3D モデルを再構築する。3D 画像である CT 画像には、スライス間の空間情報が多く含まれているため、モデルの精度に影響を与える可能性がある。そのため、3 次元量込みニューラルネットワークを用いることで、CT 画像の特性をうまく利用し、画像に含まれる有効な情報を最大限に引き出すことに成功した。しかし、3DCNN の欠点は、モデルの学習プロセスに長時間を要すること、多くの GPU メモリを使用すること、2D モデルで広く公開されている事前学習済みの重みを利用できないことが挙げられる。特に、3D CNN に使用できる一般に利用可能な 3D CT 画像データセットは現時点では少ない。そのため、CT 画像を用いた深層学習モデルに 3D CNN を使用することは、理想的なアプローチとは言い難い。

## 2 目的

本研究の目的は、3 次元 CT 画像データセットの中から代表的なスライスを用いて、間質性肺炎の診断と重症度予測を実現する 2D CNN モデルの構築手法を提案することである。これを達成するために、今回はまず代表スライスの抽出アルゴリズムに焦点を絞り報告する。

## 3 方法

### 3.1 重要なスライスの選択基準

まず、同一症例から、肺尖先端部、気管分岐部、横隔膜直上部という、明確な特徴を持つ3つのスライスを、抽出対象スライスのポジションマーカースとして選択する。これらの3つのポジションマーカースを胸部の正面像で示したものを図1に示す。肺の最上部の領域が先端部に現れたときに選択されたスライスを肺尖部スライスの選択基準とした。気管分岐部は、CTスライスで気管支がちょうど分岐する位置を基準とした。横隔膜直上部の選択は、CT画像の肺野の形状と幅に基づいて、横隔膜の直上部ではほぼ同じ形態のスライスを選択した。一般的に、横隔膜の直上部のスライスは、今回使用したデータセットでは、肺野が消失したスライスから6枚目あたりに位置していた。症例のばらつきが大きいので、このベンチマークをもとに慎重な判断と選択が必要であった。3つのポジションマーカースの位置で選択した横断像の一例を図2に示す。このようにしてポジションマーカースを画像認識の機械学習モデルで取得後に、間質性肺炎の診断や予後予測に用いる肺実質の代表スライスを同定するアルゴリズムを考案した。

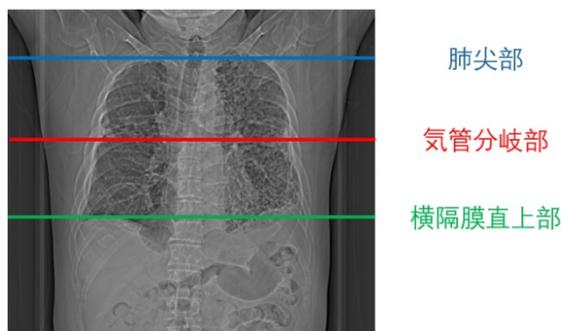


図1 3つのポジションマーカース



図2 ポジションマーカースの横断像

### 3.1 データセット

本研究では、病院の倫理委員会で研究利用が認められた、間質性肺炎の胸部CT画像を対象とした。このデータセットから、472例を本研究のデータセットとして選択した。データセットの70%をトレーニングセット、15%を検証セット、15%をテストセットとして使用した。データセットの元のファイル形式はDICOM形式であるが、前処理や学習プロセスで共通の画像形式として使用できるように、まずDICOMデータセット形式をPNG形式に変換する処理を行った。DICOMファイルフォーマットには、患者情報、スライス間隔、CTウィンドウなど、数多くの情報が含まれている。データセットには0.6mmと4mmの2つのデフォルトのスライス間隔があり、同じ症例の場合、それぞれ約140~160枚と70枚のスライスの合計数に対応する。本研究では代表的なスライスを選別する必要があるため、データセット内の各症例に存在するスライスの総数が多すぎると、データの不均衡度が悪化し、モデルの精度に影響を与えることになる。そこで、本研究では、4mm間隔のデータセットを選択した。

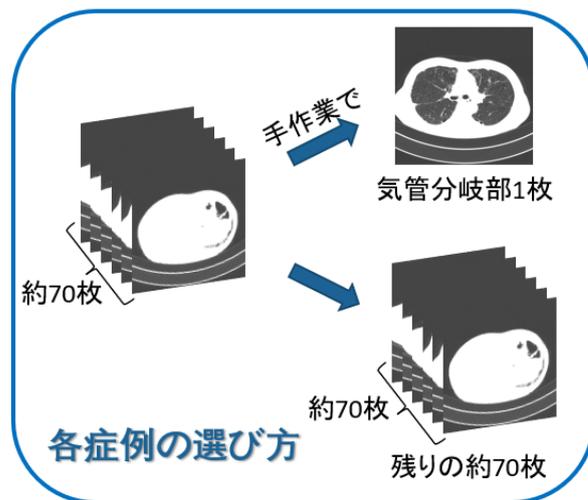


図3 ポジションマーカース抽出器のためのデータセット構築

### 3.2 データセットの前処理

データセットは、4mm間隔のCTデータを使用し、1患者あたりの総スライス数はおよそ70枚であった。

まず、気管分岐部抽出器(図3参照)の構築について説明する。各症例から、気管分岐部の代表的なCT画像1枚を選択し、抽出対象のラベルを付与した。同様に、それぞれのケースで基準を満たさない(気管分岐部ではない)画像は、抽出対象外と分類した(残りの画像約70枚)。計472症例に対してこの判定を手作業で実施し、データセットを構築した。この2値分類はバランスの悪い不均衡データである。抽出器の学習フェーズにおいて上手く分類性能を獲得できるようにしてモデル精度を向上させるためには、2つのカテゴリーを可能な限りバランスよく学習する方法を採用する必要がある<sup>2)</sup>。そこで、学習セットにオーバーサンプリング法の前処理を適用し、サンプル数のバランスをとることとした。オーバーサンプリング法とは、少数派のカテゴリーのサンプル数が多数派のカテゴリーのサンプル数と同じ数になるように、少数派のカテゴリーのサンプルをランダムに選択する方法である。サンプルの選択はランダムであり、少数派のカテゴリーのサンプル数を大幅に増やすことができるため、トレーニングセットの異なるカテゴリーのサンプル数をよりバランスのとれたものにすることができる<sup>2)</sup>。図4に示すように、気管分岐部抽出モデルの構築において、気管分岐部カテゴリーには329枚のCT画像が、非気管分岐部カテゴリーには32916枚のCT画像が含まれている。少数カテゴリー(気管分岐部スライス)にオーバーサンプリングを適用して、合計65832枚のCT画像を含むトレーニングセットを準備した。

同様の方法で、肺尖部と横隔膜の直上部データセットを作成し、肺尖部、気管分岐部、横隔膜直上部をそれぞれ抽出する3つのモデルを構築する。

### 3.3 データ増強とカテゴリーウェイトを追加した損失関数

データ増強を用いることでモデルの精度が大幅に向上することが報告されている<sup>3)</sup>。一般的には、元の画像を歪めたり、回転させたり、切り取ったりする処理をモデル学習時にデータセットに適用することで、学習モデル構築に使用されるサン

ル画像の枚数を増やすことが出来る。

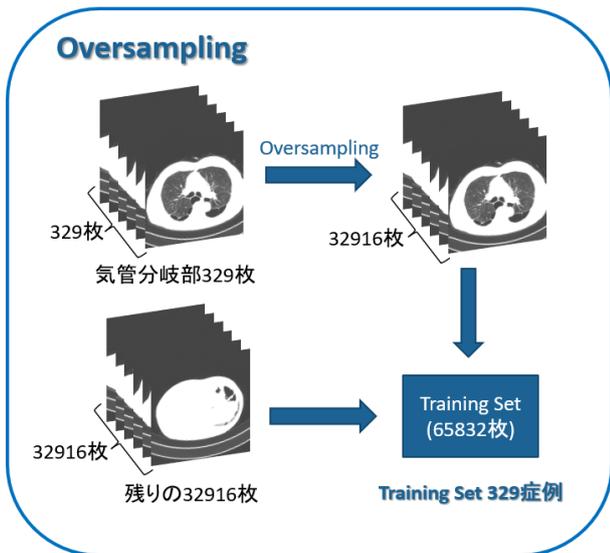


図4 オーバーサンプリング法による気管分岐部データセットの前処理例

損失関数にカテゴリの重みを付ける理由は、モデルの予測結果を評価するために混同行列をベースにした評価指標 (precision) を使用するためである(表.1)。

表1 一般的な混同行列

Confused Matrix		Precision	
		Positive	Negative
True label	Positive	TP	FN
	Negative	FP	TN

サンプル数の異なるカテゴリ間でバランスを取るためにオーバーサンプリングの方法を学習時に適用したが、実際には少数カテゴリの結果は、大きな改善がなく、しばしばモデルがより多くの False Positive (FP) サンプルを生成することになり、理想的な精度を達成できないことがある。そこで、FP のペナルティを増やして、FP を減らす方向に学習が進むように学習時の損失関数を調整し、モデル精度の向上を目指した。

### 3.4 深層学習モデル

本研究では、様々な画像認識深層学習モデルを活用して、その抽出精度を評価した。DenseNet-121、DenseNet-161、DenseNet-169、DenseNet-201<sup>4)</sup>、ResNet-50、ResNet-101<sup>5)</sup>、Inception-V3<sup>6)</sup>、EfficientNet<sup>7)</sup>シリーズのモデルについて、各事前学習済みウェイトを初期値として我々の準備したデータセットで再学習を行い、モデルを構築した。各モデルのハイパーパラメータは共通のものを使用した(20 エポック、学習率 1e-3;最後の 10 エポックでは学習率を減少させて実施)。オプティマイザには AdamW<sup>8)</sup>を、損失関数には Cross Entropy 損失<sup>9)</sup>を選択した。学習済みモデルはテストセットで評価を行い、DenseNet-169、ResNet-50、Inception-V3、EfficientNet-b4 など、各系列で最も良いモデルの性能を提示した。

評価指標は、テストセットにおける precision とした。ここで、precision に対応する値を本研究の目的に沿うよう調整を行っ

ていることに留意する。すなわち、本研究はカテゴリに属するサンプル数が極端に偏っている不均衡データのために難易度の高い分類タスクであることを考慮し、最終的な precision の計算において、モデルが対象スライスの上下のスライスの位置でスライスを抽出した場合でも、正しく抽出出来たとみなしてモデル評価を行う。この評価基準を本研究では 1-Up-Down と定義した。

具体的には、まず、すべてのモデルを同じテストセットで評価した。一般的に、各モデルは同じケースにおいて、ポジションマーカーの近傍にある複数の疑似正解のスライス (FPs) を得ることができる。モデルが出力した疑似正解を 1 つずつ比較して、そのスライスが 1-Up-Down 基準を満たしているかどうかを確認し、テストセットに対するモデルの精度を算出した。

## 4 結果

テストセットで得られた評価結果を表 2 に示す。EfficientNet-b4 は、肺尖部モデルと横隔膜の直上部モデルの両方で最高の結果を得た。Precision はそれぞれ 95.1%、93.7%であった。気管分岐部モデルでは、DenseNet-169 が 95.6%と最も高い精度を達成した。

各モデルの出力した予測スライス位置と、正解スライス位置との相対距離を図 5 に示す。モデルの出力は、正解スライス(中心スライス)の位置をピークに狭い範囲で分布していた。肺尖部抽出器と気管分岐部抽出器では、±2 の範囲に全ての出力が分布していた。横隔膜直上部抽出器のみ、±2 の範囲外のスライスを予測したものが 1 枚だけ存在した。

表2 本研究における各種モデルの性能評価

Model	肺尖部モデル	気管分岐部モデル	横隔膜直上部モデル
	1-Up-Down Precision	1-Up-Down Precision	1-Up-Down Precision
DenseNet-169	93.6%	95.6%	87.4%
ResNet-50	92.2%	92.1%	86.3%
Inception-V3	94.3%	90.6%	89.5%
EfficientNet-b4	95.1%	94.2%	93.7%

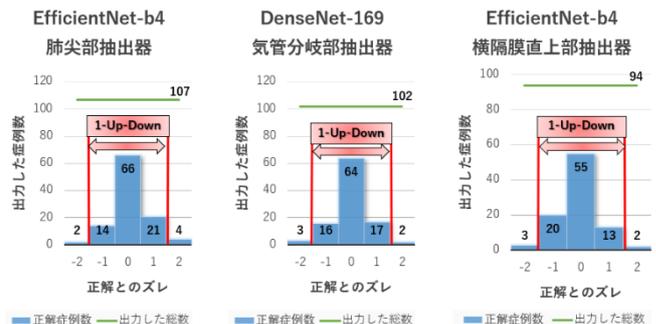


図5 肺尖部モデル、気管分岐部モデル、横隔膜直上部モデルの予測分布

(横軸) 0: 中心として選出したスライス、±1:中心スライスからの差の上下1枚、±2:中心スライスからの差の上下2枚

## 5 考察

深層学習の2D CNN モデルを用いて、胸部 CT シリーズ画像からポジションマーカーとして利用することの出来るスライスを取得する抽出器を構築した。データセットのカテゴリ分布が極端に不均衡であるという問題を解決し、高い精度でポジションマーカー候補を同定することが出来ることを示した。

図 5 の分析結果によると、モデルが選択したスライスの多くがテストセットのターゲットスライス 0 と、その上下 1 枚以内に集中している。そのため、「1-Up-Down」基準による precision は、肺尖部、気管分岐部、横隔膜直上部の 3 つの位置すべてで 95%前後の精度を達成していた。本研究ではまず、間質性肺炎の予後予測に使用するスライスを決定するためのポジションマーカ―選択を目指した。図 5 及び「1-Up-Down」基準による precision から、本研究で構築した各抽出器はポジションマーカ―選択の点において有効に機能することが期待できる。

深層学習モデルの構造毎に認められる分類精度の傾向について検討する。肺野の画像は患者ごとに異なるため、トレーニングセットのラベルを作る際に、肺尖部と横隔膜の直上部の特徴は明らかでないことが多い。しかし、気管分岐部のスライスは特徴的な外観を有しており、またその特徴は異なる患者間でも比較的類似しているため、非常に容易に区別することができると思われる。各モデルの性能を見ると、判別が容易ではないタスクでは EfficientNet-b4 が優れた性能を發揮したが、比較的単純なタスクでは従来の DenseNet ファミリーのモデルが優れていた。

現在のデータセットのオブジェクトスライスの選択基準は厳格すぎるという制約がある。その結果、各症例に対して 1 枚のスライスしか目的スライスとして選択されないため、抽出対象データの多様性を損ねてしまっている可能性が懸念される。今後は、目的スライスに対してより多くのスライスをオブジェクトスライスの基準を満たすスライスとして判定されるよう調整を行い、データセット内のカテゴリー間の不均衡をある程度解消して精度向上が目指せるかどうか、検証する予定である。

## 6 結論

胸部 CT 代表スライス位置決定に利用可能なポジションマーカ―スライスを選択する抽出モデルを構築した。極端にクラスバランスの悪いデータセットでも、オーバーサンプリングや損失関数への重み付けを行うことで、モデルの精度を向上させることが出来る。

## 参考文献

- 1) Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, 25: 1097-1105.
- 2) Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem//2018 international interdisciplinary PhD workshop (IIPhDW). IEEE, 2018: 117-122.
- 3) Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning. Journal of Big Data, 2019, 6(1): 1-48.
- 4) Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
- 5) He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks//European conference on computer vision. Springer, Cham, 2016: 630-645.
- 6) Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks//International Conference on Machine Learning. PMLR, 2019: 6105-6114.
- 7) Chollet F. Xception: Deep learning with depthwise separable convolutions//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.
- 8) Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101, 2017.

- 9) Zhang Z, Sabuncu M R. Generalized cross entropy loss for training deep neural networks with noisy labels//32nd Conference on Neural Information Processing Systems (NeurIPS). 2018.