Vision Transformer を用いた胸部 X 線画像の肺疾患分類モデルの構築

1250288　植木 涼太　【 知能情報学研究室 】

1　はじめに

じん肺は，微細な粉塵を長期間吸引することで発症する肺疾患の一種であり，畳み込みニューラルネットワーク (CNN) を用いた画像診断の研究が進められている．しかしながら，CNN は局所的な特徴の獲得には有利であるものの，画像の離れた部分に関わる特徴の獲得は困難である．そこで，Transformer 型の言語モデルを画像に応用した Vision Transformer(ViT)[1] が提案され，画像分類や物体検出において高い性能を示しており，肺炎の分類への応用も研究されている [2]．しかし，じん肺の分類に ViT を適用した研究はこれまでに行われていない．そこで本研究では，ViT を用いた胸部 X 線画像のじん肺判定モデルを提案する．

2　提案手法

ViT を用いた胸部 X 線画像のじん肺判定モデルを提案する．ViT には事前学習済みのモデルを用いる．ViT には，DeiT(Data-efficient image transformers)[3] を含む 3 種を使用する．これを画像処理を施したデータセットでファインチューニングを行い，分類モデルを構築する．

3　実験

本研究では，NIHCC，NIOSH，高知大学医学部から収集した胸部 X 線画像データを用いる．各画像には NF(No Finding)，じん肺の 2 種類のラベルが付与されている．肺野領域を抽出したマスク画像 (mask) と，抽出していない非マスク画像 (nonmask) の 2 種類を用意する．すべての画像に対して標準化処理をした後，訓練データに対してのみ，ランダムな左右反転処理を行う．さらに, これらの画像に対し，正規化を適用したもの (norm) としないもの (nonnorm) を入力データとして使用する．ViT では 3 種類の事前学習済みモデルを使用し，CNN では VGG16 を用いる．比較として CNN では活性化関数に softmax，誤差関数に categorical cross-entropy を用いる．それぞれのモデルにおいて，NF とじん肺の 2 値分類を実施する．学習率は 10^{-4}，エポック数は 100 および 300，バッチサイズは 16 および 64 とする．各条件につき 10 回試行し，性能評価には平均正答率を用いる．

4　結果・考察

図 1 は 300 エポックごとの各モデルの正答率の平均を示しており，左の棒グラフが従来手法の CNN である．図 1 では，全ての条件において DeiT ベースモデルが最も高い正答率であることが確認できる．100 エポックの場合，CNN では学習を失敗することがあったのに対し，

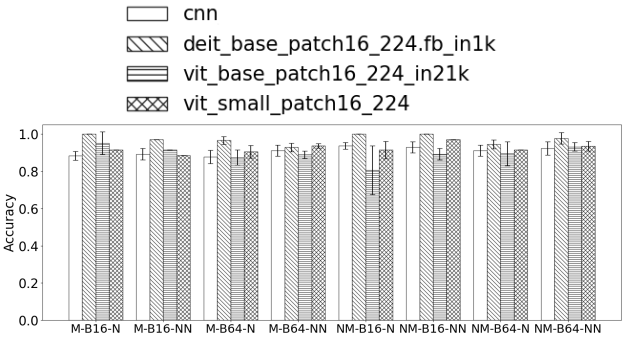

図 1　エポック数が 300 エポックの正答率.
M: マスク画像，NM: 非マスク画像
B16: バッチサイズ，B64: バッチサイズ 64
N: 正規化処理あり，NN: 正規化処理なし

ViT の 3 モデルでは，学習がうまくいった．実行時間について，300 エポックでは CNN で約 700 秒，ViT は約 680 秒となり，学習時間に差がないことが確認できる．CNN と比較した場合，事前学習モデルの種類に関係なく，ほとんどの条件で ViT の正答率が高いことが確認できた．これは ViT の自己注意機構 (Self-Attention) により，CNN よりも浅い層から大局的な表現を捉えることができるためと考えられる．また，ViT ではパッチをエンコーダ部分に入力する際，位置情報を埋め込むことで空間情報を保持することができるため，高精度な分類が可能である．しかし，ViT は大量のデータで学習することが前提としているため，小規模なデータセットでは学習が十分に進まず，正答率が伸び悩む可能性がある．

5　おわりに

本研究では 3 種類の ViT の事前学習済みモデルと CNN を用い，さまざまな条件下でじん肺分類の精度比較を行った．その結果，ViT は CNN よりも高い正答率を達成できることが確認された．また，学習時間に関しては，少ないエポック数では ViT が短時間で効率的な学習が可能であることが確認された．一方，エポック数を増やすと正答率は向上したものの，学習時間に差が見られなくなった．このことから ViT は少ないエポック数でも CNN より効率的な学習を進めることが可能であり，従来手法である CNN よりも高い正答率を獲得することができた．

参考文献

[1] Alexey, D. et al., ICLR, 2021.

[2] Singh, S. et al., Scientific Reports 14, 2024.

[3] Touvron, H. et al., PMLR 139, 2021.