

## II 医療AIを加速させる研究開発の動向

# 3. 医療AI開発における Vision Transformerの可能性

小田 昌宏 名古屋大学情報連携推進本部情報戦略室

近年の人工知能 (AI) は深層学習を用いたものがほとんどであり、AIの性能向上には深層学習関連の技術発展が深くかかわっている。画像を対象とした深層学習では、これまで畳み込みニューラルネットワーク (convolutional neural network : CNN) や全層畳み込みニューラルネットワーク (fully convolutional network : FCN) を用いるものが主流であったが、近年 Vision Transformer (以下、ViT) が新たに提案され注目を集めている。一般画像を対象とした研究では ViT が CNN や FCN より高い性能を持つことが報告されており、ViTの医療応用が世界中で進められている。本稿では、CNNやViTを含む深層学習モデル発展の経緯、ViTの医療AIへの適用について述べる。

### AIのコアとしての 深層学習モデル

深層学習では、データからの判断を行う枠組みである“深層学習モデル”を使用し、学習データを用いて深層学習モデルの学習を行うことで、自動的な判断が可能となる。学習データの数も重要であるが、学習データからより適切な判断を構築可能な深層学習モデルを使用することが重要である。画像処理で用いられる深層学習モデルとして、疾患分類や年齢推定などを行う CNN、臓器セグメンテーションなどを行う FCN がある。ここ10年の間に登場したAIはCNNやFCNを用いたものが多いが、2021年に新たな深層学習モデルとして ViT が提案され、注目されている。ViTを含む深層学習モデル発展の経緯を以降で説明する。

### 深層学習モデルの発展

ViTはこれまでの深層学習モデルとどのように異なるのだろうか。過去から現在までの深層学習モデル登場の経緯や特徴を概観しながら ViT の特徴を述べる。時系列に沿った深層学習モデルの発展を図1に、深層学習モデルの特徴やモデル構造を図2にまとめたので併せて参照されたい。

#### 1. 多層パーセプトロン (MLP)

深層学習モデルの最も基本的かつ重

要なものが多層パーセプトロン (multi-layer perceptron : MLP) である。MLPは生物の神経回路網を参考に作られており、簡単な計算や判断を行う要素 (単純パーセプトロンと呼ばれる神経細胞に対応するもの) を多数接続した構造を持つ。MLPは第1次AIブームの頃から研究が行われていたものの、画像処理に直接的に用いるのが困難であり、学習データに過剰適合しやすい (過学習を起こしやすい) など応用上の課題があった。

#### 2. 畳み込みニューラルネットワーク (CNN)、全層畳み込みニューラルネットワーク (FCN)

現在のAIに多く用いられる深層学習モデルがCNNとFCNである。これらは画像処理に効果的であり、医療応用だけでなく、自動運転、ロボット自動制御、さらに、農業、林業、漁業の自動化にも用いられる。CNNとFCNの構造は、基本的にMLPと類似するが、単純パーセプトロン間の接続を画像上の狭い範囲 (3×3や5×5画素程度) に限定している。つまり、画像上の狭い範囲を見て判断する仕組み (局所的な判断) を持ち、これを多数連結させてモデルを構成する。局所的な判断が採用される理由は、①深層学習モデルがシンプルになり、現実的に入手可能な数 (数千～数万枚) のデータで学習が可能となること、②腫瘍など体内の異常構造は画像上の狭い範囲に存在する 경우가多く、腫瘍の検出を行う診断支援AIなど、AIの使用