

Vision Transformer と Focal Loss を用いた MRI 画像における卵巣がんの診断

王 夢雨[†] 濱上 知樹[†] 宮城 悦子[‡] 長内 奈々[‡] 宇都宮 大輔[‡]

[†] 横浜国立大学 [‡] 横浜市立大学

1 はじめに

卵巣がんは、初期段階に症状が出にくいいため、発見された時は進行していることが多いがんである。さらに、患者数は1970年代に比べ、約5倍に増えており、大きな社会問題になっている。診断にあたっては、医師がMRI検査の画像を用いてがんの種類と性質を推測する。そして、患者が適切な治療を受けられるよう、卵巣がんの種類に応じて治療方針が選択される。しかし、画像分析医の人手不足に加え、医師の経験差により診断結果に違いが出てしまうという問題が存在する。

以上で述べた問題を解決するため、人工知能を用いたMRI画像の分類による卵巣がんの支援が注目されている。

卵巣がんを正しく診断するには、(1) 画像内卵巣両側の病変と周辺組織の特徴、(2) スライス間腫瘍の形状変化と関連性、(3) 異なるモダリティ (T1, T2 など) の画像における病変の特徴を組み合わせることが必要である。

一方、MRIなど医療画像は患者のプライバシーに関わるため、大量のデータを収集することが難しく、データ量不足とデータ不均衡という問題が生じる。そのため、少数データからデータ不均衡の影響を受けにくい学習方法が求められる。

そこで本研究では、卵巣がんの診断に Vision Transformer[1] と Focal loss[2] を用い、MRI 画像内病変の特徴とスライス間の空間情報の活用する

ことで、データ不均衡の緩和をはかる手法を提案する。

2 関連技術

2.1 Vision Transformer (ViT)

ViT は画像認識に自然言語処理分野の Transformer を利用したモデルである。画像をパッチに分割し、次元に平坦化する。位置情報を付与した入力を Transformer encoder に与える。Attention 機構により、異なる特徴間の依存関係を捉えることができる。

2.2 Focal Loss

データの不均衡がある場合、単純にデータが多いクラス (easy examples) の損失を合計すると、少数派クラス (hard examples) の損失が無視されて、少数派クラスの特徴を捉えられない可能性がある。式1の損失関数 Focal Loss を使用し、Easy examples の重みを減らすことができる。ここで、 p_t は正解ラベルに対する推定確率、 γ はフォーカスパラメータである。

$$FL(p_t) = -(1 - p_t)^\gamma \log p_t \quad (1)$$

3 提案手法

提案手法の手順を図1に示す。

まず、隣接する3枚のスライス画像を1枚の3チャンネル画像に合成し、空間情報を含ませる。次に ViT を特徴抽出器として用い、病変の局所的な特徴と異なる領域間の関連情報を抽出する。これらの特徴を使って、良性内膜症性嚢胞、明細胞癌、類内膜癌の3分類を行う。損失関数として Focal Loss を用い、データが少ないクラスに注目するようにモデルを学習させることにより、データ不均衡の解決が期待される。

Diagnosis of Ovarian Cancer on MRI Images using Vision Transformer and Focal Loss

[†] Wang Mengyu, Hamagami Tomoki, Yokohama National University

[‡] Miyagi Etsuko, Osanai Nana, Utsunomiya Daisuke, Yokohama City University

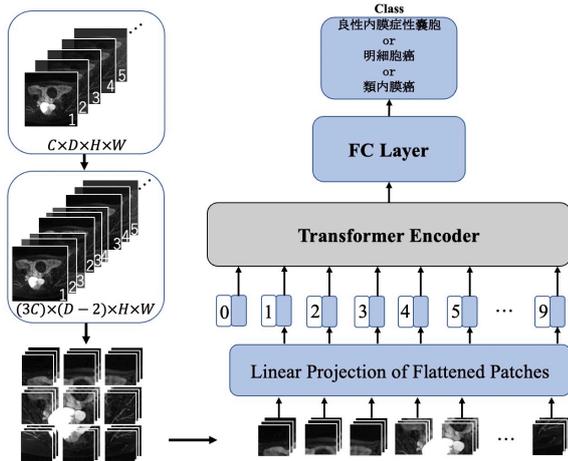


図1 提案手法の手順

4 実験

4.1 データセット

実験に使用するデータセットは、17名の卵巣がん患者の T2 画像 (合計 358 枚) から構成される。画像サイズは 512×512 である。その内、良性内膜症性嚢胞の患者 10 名、明細胞癌 4 名、類内膜癌 3 名である。

4.2 実験概要

提案手法の有効性を検証するため、以下三つの実験を行う。1. 卵巣癌の分類に対して ViT の効果を評価する。2. Focal loss を損失関数として ViT で卵巣がんを分類する。3. 隣接する 3 枚のスライスを fusion し、focal loss を用い、診断の精度を向上できるかを確認する。

画像を学習データ、検証データ、テストデータ (7:1:2) に分割し、バッチサイズ 8、エポック数 100 とする。実験 3 では、データセットを分割する前に 3 チャンネル画像を作成する。事前学習済みの ViT を用いて分類し、テストデータの正解率と F1-macro で評価し、各実験同士の精度比較を行う。

4.3 実験結果

各実験結果を表 1 に示す。図 2 の Attention Map の可視化から、モデルがスライス画像内の病変部を適切に捉えていることが確認された。そのため、ViT で根拠のある診断ができることが明らかになった。また、実験 2 の精度は実験 1 と比べて高くなった。それぞれの癌の種類と精度を比較

表 1 各実験結果

	Accuracy	F1-Macro
ViT	0.7571	0.5505
ViT+Focal loss($\gamma = 2$)	0.8000	0.6673
Fusion+ViT+Focal loss($\gamma = 2$)	0.8515	0.7829

した結果、最もデータが少ない癌 (類内膜癌) の精度が Focal loss によって向上していることが確認できた。すなわち、データ不均衡が緩和されたことがわかった。実験 3 では最も高い精度が得られた。ViT でパッチ間の連関情報を捉え、fusion によりスライス間の空間情報を活用する。これらの特徴を組み合わせることで、診断精度が高まることが明らかになった。

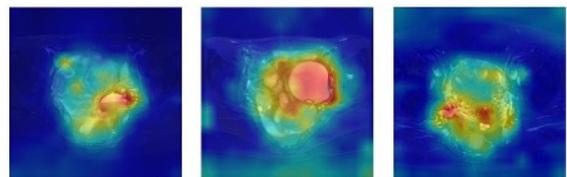


図2 Attention Map 可視化の結果

5 おわりに

本研究では、MRI 画像の隣接スライスの fusion し、卵巣がんの診断に ViT と Focal Loss を用いることを提案した。今後は、データをより効率的に活用するため、MRI 画像の複数モダリティ間の特徴を抽出する方法を検討する。

参考文献

- [1] Alexey Dosovitskiy, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [2] Tsung-Yi Lin, et al. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.