

医用画像からの Radiomics 特徴量抽出と、 時系列での状態遷移の可視化

Radiomics feature extraction from medical images and visualization of time-series state transition

大矢めぐみ
理化学研究所

1. 研究目的

医用画像を網羅的に解析して定量的な特徴を多く抽出する Radiomics 解析が近年注目を集めている。医用画像の特徴には、人間の目では視覚的には認識できない疾患特有の情報が含まれていると考えられており、そのような情報を画像から抽出し、臨床データや遺伝子データとの関連をみつけるための様々な研究が行われている。

Radiomics 解析では病変の大きさや濃度、形状、テクスチャなど多様な画像特徴量が用いられる。抽出された特徴量に AI 解析等を応用することで、脳、肺、心臓、肝臓、腎臓、副腎、前立腺など様々な臓器における診断や予後予測のモデル作成が試みられてきた。

一方、多変量の時系列データを解析し経過を可視化するための手法としてエネルギーランドスケープ解析という手法がある。この手法では、特徴量ごとの発現の組み合わせによって状態を表現し、エネルギーランドスケープという地形図上の動きとして状態遷移をマッピングすることができる。状態は、地形図上の下り坂を下りてエネルギーの低いところに留まりやすいが、ランダムなゆらぎによって小さな峠を超え、別のエネルギーの低い地帯に移動することもある。こうした移動の軌跡をみることで、時系列データの可視化と意味付けを行うことができる。

Radiomics 解析では数多くの特徴量を抽出できるため、各特徴量の発現組み合わせパターンにエネルギーランドスケープ解析を適用することで、患者ごとの状態を可視化し、予後についての示唆を得ることができると考えられる。そこで本プロジェクトでは乳がん患者のデータセットに対して、①Radiomics 解析を通じて特徴量抽出を行い、②得られた特徴量に対しエネルギーランドスケープ解析を適用し、状態の可視化に取り組むことを目的としている。

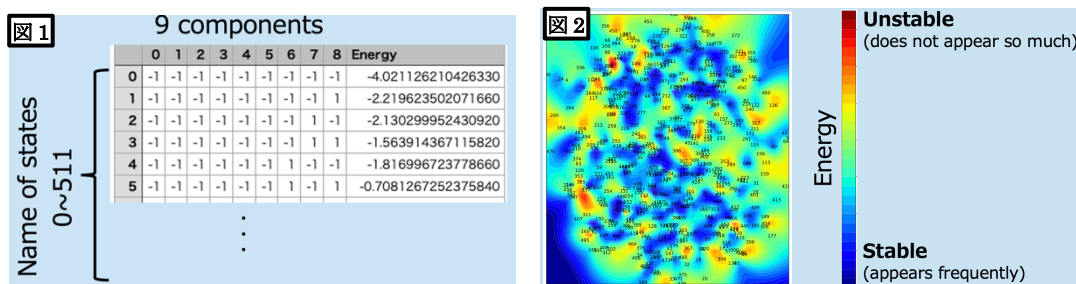
2. 研究成果の内容

医用画像から病変部位における Radiomics 特徴量を抽出するためには、画像の中で病変部位がどこにあたるのかを明確にする必要がある。そのため、Cygnus を用いたこれまでのプロジェクトにおいて我々が開発した、乳がんに対する放射線治療における治療ターゲットである乳腺組織を高精度に検出（セグメンテーション）するモデルを応用し、乳がん患者のデータセットから病変部位の Radiomics 特徴量の抽出を行う

た。今回は 192 人の患者の乳がん病変部位の画像をセグメンテーションし Radiomics 解析を適用した結果、88 の Radiomics 特徴量が抽出された。今回用いた 192 人の乳がんデータセットは、各患者に紐づく臨床情報がないデータセットであった。そのため、抽出された 88 の Radiomics 特徴量を用いて教師なしクラスタリングを行い、患者のグループ分けを行った。その結果、K-means 法を用いたクラスタリングで、患者は 8 つのクラスターに分類された。

次に、エネルギーランドスケープ解析手法を、本データセットに適用した。エネルギーランドスケープ解析では、各特徴量の値を $-1/1$ に 2 値化し、各特徴量における $-1/1$ の組み合わせの出現パターンから地形図を描出する【図 1】。本来であれば 88 のすべての特徴量から直接地形図を描くことを目指していたが、そのためには 2 の 88 乗といった膨大な組み合わせパターンの発現頻度を求める必要がある一方、今回のデータセットは 192 人の患者数に限られていたため、88 の特徴量を用いて解析するのは理論上困難であった。そのため、まず Radiomics 特徴量を主成分分析によって圧縮し、データの 9 割以上を説明できる 9 つの主成分の値から地形図を描出した。その際、各患者における 9 つの主成分の値を、値の中央値との大小の比較で $-1/1$ に置き換え、各患者が 2 の 9 乗の組み合わせパターンのうちどのパターンにあてはまるか、分析した【図 2】。

描かれた地形図上で、各クラスターの患者が地形図のどのあたりに配置されるかを検証した。クラスターによって所属する患者数が異なることもあり、クラスターと地形図の関連を説明するのは困難であるが、クラスターによっては一定の傾向が見られた。この傾向が表す意味については臨床データと画像の特徴との関連を探ることで明らかにできる可能性がある。



3. 学際共同利用プログラムが果たした役割と意義

エネルギーランドスケープ解析を行う際、使用する特徴量の数を増やすためには 2 の [特徴量の数] 乗通りの特徴量パターンの組み合わせを計算する必要があるため、豊富な計算資源を利用可能な本学際共同利用プログラムは非常に貴重な研究の場であった。2023 年度の研究では、使用することができた患者数が 192 人と小さいデータセットであったため、理論上の制約で少ない患者数に対応可能な範囲まで特徴量の数を極端に減らして計算を実施した。そのため、Cygnus のリソースを使用すると短時間

でジョブが終了しており、リソース消費は少ない状況であったが、今後患者数とともに使用する特徴量の数を増やしても対応可能な環境があることでプロジェクトの推進の後押しとなっている。

本学際共同研究によって、計算の高速化技術について理解を深める機会もいただいている。年に1度開催されるシンポジウムでは日常の業務ではなかなか聞くことができない様々な分野の研究発表に触れることができ、見識を広げ、新たな研究の可能性を見いだせる重要な場となっている。

4. 今後の展望

今回用いた 192 人の乳がんデータセットについては、各患者に紐づく臨床情報が取得されていないものだったため、描かれた地形図と患者の状態の臨床的な意味付けが難しかった。更に、今回は特徴量を主成分分析の成分に変換してしまったため説明可能性が低下してしまい、地形図を解釈する上での困難さが増してしまった。今後、画像データと臨床データが紐づいたデータセットを取得することで、予後をはじめとしたアウトカムと地形図との臨床的な意味付けや、Radiomics 特徴量の関連性の分析がエネルギーランドスケープ解析を通じてできる可能性を示すことができた。疾患の発症や進行パターンを細かく分類することで、将来的には各患者の体質や疾患の個人差に合わせた個別化医療の実現にも貢献できることが期待される。

今後、公開されているデータセットなどを複数組み合わせた活用などを通じて、地形図の描写に限らず、直感的には理解が難しい Radiomics 特徴量の可視化手法の幅広い開発に取り組んでいく。

5. 成果発表

- (1) 学術論文
- (2) 学会発表
- (3) その他

使用計算機	使用計算機に ○	配分リソース※	
		当初配分	追加配分
Cygnus	○	5000	
Wisteria/BDEC-01			
※配分リソースについてはノード時間積をご記入ください。			