

【要約】

Apparent diffusion coefficient analysis for differentiating pleomorphic adenoma and carcinoma ex pleomorphic adenoma: influence of radiologists' performance on inter-observer reliability and diagnostic performance of one-point measurement, histogram analysis, texture analysis and radiomics machine learning approach

(多形腺腫と多形腺腫由来癌を鑑別するための見かけの拡散係数解析: 放射線科医の能力が一点計測法、ヒストグラム解析、テクスチャ解析、Radiomics 機械学習的アプローチにおける検者間一致率と診断能に及ぼす影響)

千葉大学大学院医学薬学府
先端医学薬学専攻
(主任：宇野 隆教授)
和田 武

【目的】

唾液腺腫瘍の治療法は良悪性を含めた組織型、腫瘍の進展範囲などを考慮して決定されるため、術前診断は非常に重要である。MRI による見かけの拡散係数 (ADC: apparent diffusion coefficient) 解析は細胞密度の多寡を定量的に評価でき、唾液腺腫瘍の術前診断に有用な手法である。しかし ADC 解析を用いても、最も頻度の高い唾液腺腫瘍である多形腺腫 (PA: pleomorphic adenoma) と、PA から生じる高悪性度唾液腺腫瘍である多形腺腫由来癌 (CXPA: carcinoma ex pleomorphic adenoma) との鑑別は困難なことがある。これまで ADC 解析は、腫瘍内に検者が特定の関心領域を設定し、関心領域内の ADC 平均値を代表値として解析する方法が主流であった。近年では腫瘍全体を関心領域として設定し、ヒストグラム解析やテクスチャー解析を使用して、多数の画像特徴量を抽出することが可能になり、さらにはそれらの画像特徴量を機械学習的に統合して予測モデルを構築する Radiomics approach が報告されており、従来の方法より高い診断能を示すとされている。一方で、Radiomics approach による予測モデルの構築には信頼性の高いデータが必要であり、全腫瘍計測から抽出された画像特徴量や予測モデルの診断能は特に 1) 腫瘍の計測方法、2) 腫瘍の部位、3) 検者間一致率などに影響されうるとされている。一方で、検者の放射線科医としての経験や技術が、これらの信頼性や診断能にどのように影響するかについては十分に検討されていない。この研究では、PA と CXPA の ADC 解析において、経験の異なる 4 人の検者が一点計測法と全腫瘍計測法をそれぞれ行い、抽出された画像特徴量と Radiomics approach により構築された予測モデルの信頼性と鑑別診断能がどう変化するかを検討した。

【方法】

2010 年 2 月から 2017 年 5 月まで、唾液腺 MRI の検査をうけた 496 人を対象とし、そのうち、切除による組織学的な確定診断が得られた PA 115 例、CXPA 22 例の計 137 例について解析を行った。MRI は 1.5 テスラの同一機種 (Signa HDxt, GE Healthcare, Milwaukee, Wisconsin) で同一プロトコル (T1 強調像軸位断、脂肪

抑制 T2 強調像軸位断、T2 強調像冠状断、造影後脂肪抑制 T1 強調像軸位断・冠状断、拡散強調像軸位断) を施行し、拡散強調像は下記のパラメーターで撮影した (b value = 0, 1000 s/mm²; repetition time, 7000 ms; echo time, 80 ms; FOV, 280 × 280 mm; matrix, 256 × 256; slice thickness, 5 mm; gap, 1 mm)。MRI 機器上で自動的に作成された ADC map を経験年数の異なる 4 人の放射線科医が解析を行った。検者 1 は放射線科医として 15 年間、頭頸部領域画像診断を 9 年間、検者 2 は放射線科医として 7 年間、頭頸部領域画像診断を 3 年間の経験がある。検者 3 は放射線科医として 7 年間、検者 4 は放射線科医として 2 年間の経験があるが、それぞれ頭頸部領域画像診断の経験はない。これらの検者 4 人が、撮影された他の画像を参照して、ADC map 上に一点計測法と全腫瘍計測法の 2 通りの ROI を設定した。一点計測法の ROI は造影効果の良好な非変性部に円形に設定し、10 mm² 以上の大きさとした。全腫瘍計測法では腫瘍の全スライスで輪郭を描画して関心領域とした。一点計測法からは ADC 平均値を、全腫瘍計測法からは 27 個のヒストグラム解析による画像特徴量を Matlab 2015b (Mathworks, Inc., Natick, MA, USA) を用いて、32 個のテクスチャー解析による画像特徴量と 4 個の形態に関する画像特徴量を LIFEx (<https://www.lifexsoft.org/>) を用いて、合わせて 63 個の画像特徴量を抽出した。全ての統計解析と機械学習は R version 3.5.0 (R Foundation for Statistical Computing, Vienna, Austria) を用いて行った。検者間関心領域の類似度を Dice index の中央値、検者間一致率を級内相関係数 (ICC: intraclass correlation coefficient) で、画像特徴量と予測モデルの診断能を AUC (area under the receiver operator characteristic curve) で評価した。ICC は 0.8 以上を信頼性が高いと定義した。Radiomics approach では、過学習を避けるために二重交差検証法を用いて Random forest と Elastic net という二つのアルゴリズムで予測モデルを構築した。二重交差検証法は 5000 回反復し、その平均値を用いて予測モデルの診断能として AUC を計算した。また、予測モデルの構築者を他人のデータセットで検証した場合を想定し、二重交差検証法の変法を考案してこの診断能も検証した。

【結果・考察】

各検者間の Dice index は 0.71–0.83 を示し、全腫瘍計測法から抽出された 63 個の画像特徴量のうち、各検者間の ICC が 0.8 を超える画像特徴量の個数と Dice index に強い相関 (単相関係数: 0.96) が認められた。このことより、検者間の ROI が物理的に類似するほど、より多くの信頼性の高い画像特徴量を抽出できると考えられた。各検者間の ICC と AUC を検討したところ、ヒストグラム解析から抽出された画像特徴量は ICC が高いが AUC は比較的 low、テクスチャ解析から抽出された画像特徴量は ICC、AUC とともに低値から高値まで様々な値をとるものの、ICC と AUC がともに 0.8 を超える、信頼性と診断能が双方とも非常に高い画像特徴量も認められた。全ての検者の組み合わせにおいて、経験年数の高い検者の方が AUC の高い画像特徴量の個数は多かった。解析法ごとの診断能を比較すると、一点計測法による ADC 平均値の AUC は 0.66–0.79、全腫瘍計測法から抽出された画像特徴量である GLZLM_ZLNU (Gray-Level Zone Length Matrix_Zone Length Non-Uniformity) の AUC は 0.79–0.88、Radiomics 予測モデルの AUC は 0.82–0.87 を示し、同一検者においては一点計測法による ADC 平均値の AUC より、全腫瘍計測法から抽出された画像特徴量や Radiomics 予測モデルの AUC が常に高値を示した。Radiomics 予測モデル作成と診断能の検証とを異なる検者のデータを用いて行くと、AUC は 0.81–0.88 を示し、検者の組み合わせによって診断能が変化した。

【結論】

全腫瘍計測法や Radiomics 予測モデルを用いた ADC 解析は PA と CXPA の鑑別診断に有用であり、その診断能は一点計測法による診断能より高い。また、検者の経験が全腫瘍計測法から抽出された画像特徴量や Radiomics 予測モデルの診断能と信頼性に影響するほか、Radiomics 予測モデルの診断能は、誰のデータでモデルを構築し、誰のデータで検証するかで変化する。