

Deep Learning for Echo Analysis, Tracking, and Evaluation of Mitral Regurgitation (DELINEATE-MR)

—ディープラーニングによる僧帽弁逆流の評価—

Long A, Haggerty CM, et al. *Circulation*. 2024;17

DOI: 10.1161/CIRCULATIONAHA.124.068996.

背景

人工知能、特にディープラーニング (Deep learning; DL) は、経胸壁心エコー図検査 (transthoracic echocardiography; TTE) の解釈を改善する大きな可能性を秘めている。ただし僧帽弁逆流 (Mitral Regurgitation; MR) などカラードプラーで評価される病変については、DL での解析はほとんど報告されていない。

目的

本研究の目的は、①TTE 画像から MR の重症度を正確に評価できる DL システムを開発すること、②多断面の TTE 画像（例えば心尖部四、三、二腔像）が、単一断面のモデル（例えば心尖部四腔像のみ）に比べ診断精度を改善するかを検証すること、③独立した外部データセットでこのモデルを検証すること、④心臓専門医と DL モデルの MR 評価の違いを詳細に分析すること、である。

方法

2015 年から 2023 年にかけて複数施設における一連の TTE 画像を内部データとし、別施設で 2019 年から 2021 年に撮像した TTE 画像を外部データセットとした。先天性心疾患患者、左室補助装置、僧帽弁に対する手術後や経カテーテル的治療後の患者は除外とし、また TTE データセットが十分でないものも除外した。

MR 評価結果は、米国心エコー図学会のガイドラインに基づき、視覚的評価、定量的ドプラー法、PISA 法の組み合わせによって評価・記載された TTE レポートより抽出した。さらに正確な MR 分類のため 200 件以上の TTE レポートが研究者によりレビューされた。TTE 画像は患者ごとに train (60%), validation (20%), test (20%) に分けられ、DL システム (DELINEATE-MR) により僧帽弁の存在を含むカラードプラー画像がまず識別され、その中から MR の程度が最も高いものを組み合わせて最終的に MR 重症度を判定した。また多

断面画像を使用する MR モデル（本研究の主なモデル）と心尖部四腔像のみを使用する MR モデルの診断精度を比較し検証した。MR モデルの結果と心臓専門医による臨床的な MR 診断を重み付け κ 係数で評価し、モデル予測と心臓専門医の解釈が一致する割合を検証した。

AI モデルと心臓専門医による MR 重症度評価の不一致が多かった症例については、複数の高度トレーニングを受けた心臓超音波医師が手動で検証し、コンセンサスを得て MR 評価を行った。

結果

44,628 人の患者から計 61,689 件の TTE 画像が内部データセットに含まれた。内部テストセットの患者の年齢は 65.1 ± 16.8 歳で、MR 評価結果は none-trace (trivial と同義) ($n=5895$; 65.6%), mild ($n=1878$; 20.9%), moderate ($n=1034$; 11.5%), severe ($n=180$; 2.0%)であった。内部テストセットでは DL モデルと心臓専門医読影の MR 重症度の一致は 7411 件 (82%)、重み付け κ 係数は 0.84 であった。また不一致のほとんど ($n=990$; 63%) が none-trace MR と mild MR 間のものであった。AUC, PR-AUC, F1 スコア (いずれも機械学習モデルの性能を評価するための指標で、1 に近い程高精度を示唆する) はそれぞれ 0.98, 0.90, 0.80 であった。外部データセットにおける DL モデルの MR 分類の診断精度も高かった (精度 79%; 重み付け $\kappa=0.80$)。心尖部四腔像のみを使用した DL モデルでは、多断面画像を使用した DL モデルと比べ診断精度が低下した (80%対 82%)。内部テストセットでは、心臓専門医と DL モデルによる MR の重症度判定で有意な不一致 (6 段階スケールで 2 段階以上の相違) が合計 106 件 (テストセットの 1.2%) 同定された。これらのうち再評価可能だった 104 例において、前述の通り複数の高度トレーニングを受けた心臓超音波医師による合議制で MR 再評価を行い、DL モデル結果との比較を追加した。

その結果、30 件 (29%) は正確に MR の重症度判定まで可能であった。また、46 件 (44%) では両者の MR の重症度評価のずれが 6 段階スケールで ± 1 段階のみだった。モデルが失敗した 28 件 (27%) のうち、6 件 (6%) は適切な MR 画像を前処置で識別できなかったことに起因し、残りの 22 件 (21%) は適切な画像で誤った MR 分類を行っていた。

結論

DL システム (DELINEATE-MR) は、TTE 画像を使用しての MR 分類において汎用性のあるパフォーマンスを示した。このシステムは人間のみでの解釈と比較し MR 診断能を改善させる可能性がある。

COMMENTS

TTE は心機能や弁膜症評価において最も広く用いられる確立したツールである一方で、画像の不均一性や、視覚的な定性評価に依存する部分も多いため評価や解釈にばらつきが出やすい検査であることも事実である。TTE における DL の研究としては、これまで左室駆出率測定や大動脈弁狭窄など特定の疾患の検出などが報告されているが、MR を含む弁逆流の評価は困難とされてきた。逆流のメカニズムは症例毎に異なっており、適切に逆流の重症度を判断できる断面がわずかしかないことなどがその理由だが、本研究ではデータの入力から最終的な出力までの全プロセスを 1 つのモデルで学習するエンドツーエンドアプローチを取り入れ、それを可能とした。また、DL モデルと初期診断が不一致だった症例を再評価し、約 3 割は心臓専門医の初期診断が誤りだった点も興味深い。

今後他の弁膜症への応用も期待されるところであり、より正確な弁膜症診断のために必須のツールとなる可能性を秘めている。一方で臨床医の過度の依存や、限られた施設のみのツールとなることへの懸念も示されており、克服すべき課題も多い。

文責： 江口 紀子