

Data-driven prediction of prolonged air leak after video-assisted thoracoscopic surgery for lung cancer: Development and validation of machine-learning-based models using real-world data through ePath system

藤, 沙織

<https://hdl.handle.net/2324/7363638>

出版情報 : Kyushu University, 2024, 博士 (医学) , 課程博士
バージョン :
権利関係 : Creative Commons Attribution 4.0 International



氏 名： 藤 沙織

論文名： Data-driven prediction of prolonged air leak after video-assisted thoracoscopic surgery for lung cancer: Development and validation of machine-learning-based models using real-world data through ePath system

(肺がんのビデオ支援胸腔鏡手術後の遷延性肺瘻のデータ駆動的予測:ePathシステムを介したリアルワールドデータを用いた機械学習ベースのモデルの開発と検証)

区 分： 甲

論 文 内 容 の 要 旨

実臨床におけるデータ駆動型の臨床予測の信頼性には依然として不明確さが残る。本研究では、電子クリニカルパス (ePath) システムから得られたリアルワールドデータを用いて臨床アウトカムを予測するための機械学習モデルを開発、検証することを目的とした。

2つの独立した病院においてビデオ補助胸腔鏡下手術 (video-assisted thoracic surgery: VATS) を受けた肺がんの入院患者を対象とし、ePathシステムを活用してデータを収集した。臨床アウトカムは遷延性肺瘻 (prolonged air leak: PAL) として、術後3日以降のドレーン抜去と定義した。データ駆動型予測モデルは、大学病院における患者314人を開発コホートとして、スパース線形回帰モデル (LASSO、Ridge、Elastic Net) および決定木アンサンブルモデル (Random Forest、XGBoost) を用いて構築した。予測モデルの性能は、三次病院における患者154人を検証コホートとして、受信者動作特性曲線下面積 (area under the receiver operating characteristic curve: AUROC) およびキャリブレーションプロットにより評価した。

バイアスを軽減するために、PALに関連して欠測するデータや高い欠測率を持つ変数はデータセットから除外した。内的妥当性検証 (5分割交差検証) 後の結果では、リストワイズに削除されたデータのみならず欠測値補完後のデータであっても、主要変数に絞った場合にAUROCが向上することが示された。また、連続変数を二値化することで、特に変数を絞った場合に決定木アンサンブルモデルの性能が向上した。最終的に、完全ケース分析で7つの主要変数を組み込んだ回帰モデルが、内的妥当性検証 (AUROC: 0.77-0.84) および外的妥当性検証 (AUROC: 0.75-0.84) とともに優れた識別能を示した。これらのモデルは両コホートで十分な較正能を示した。

ePathシステムを利用したデータ駆動型予測モデルは、実臨床で得られた医療データであっても変数を最適化し集団特性を考慮することで、VATS後のPAL予測において十分な性能を示すことが明らかとなった。