

タイトル	Machine learning based gene alteration prediction model for primary lung cancer using cytologic images
別タイトル	細胞診画像を対象とした機械学習による原発性肺癌の遺伝子変異予測モデルの構築
作成者（著者）	石井, 脩平
公開者	東邦大学
発行日	2023.03.14
掲載情報	東邦大学大学院医学研究科 博士論文 内容の要旨及び審査結果の要旨. 6.
資料種別	学位論文
内容記述	主査：澁谷和俊 / タイトル：Machine learning based gene alteration prediction model for primary lung cancer using cytologic images / 著者：Shuhei Ishii, Manabu Takamatsu, Hironori Ninomiya, Kentaro Inamura, Takeshi Horai, Akira Iyoda, Naoko Honma, Rira Hoshi, Yuko Sugiyama, Noriko Yanagitani, Mingyon Mun, Hitoshi Abe, Tetuo Mikami, Kengo Takeuchi / 掲載誌：Cancer Cytopathology / 巻号・発行年等：130(10): 812-823, 2022 /
著者版フラグ	none
報告番号	32661甲第1062号
学位記番号	甲第734号
学位授与年月日	2023.03.14
学位授与機関	東邦大学
メタデータのURL	https://mylibrary.toho-u.ac.jp/webopac/TD90283263

博士學位論文

論文内容の要旨

および

論文審査の結果の要旨

東邦大学

石井脩平より学位申請のため提出した論文の要旨

学位番号甲第734号

学位申請者 : いし 井 しゅう へい
石 井 脩 平

学位論文 : Machine learning-based gene alteration prediction model
for primary lung cancer using cytologic images

(細胞診画像を対象とした機械学習による原発性肺癌の遺
伝子変異予測モデルの構築)

著 者 : Shuhei Ishii, Manabu Takamatsu, Hironori Ninomiya, Kentaro Inamura,
Takeshi Horai, Akira Iyoda, Naoko Honma, Rira Hoshi, Yuko Sugiyama, Noriko
Yanagitani, Mingyon Mun, Hitoshi Abe, Tetuo Mikami, Kengo Takeuchi

公表誌 : Cancer Cytopathology 130(10): 812-823, 2022

論文内容の要旨 :

背景・目的:

肺癌は世界のがん死亡原因の第一位を占めている。肺癌治療は基本的に組織学的分類に依存するが、近年では上皮細胞成長因子受容体(Epithelial Growth Factor Receptor, EGFR)などの代表的遺伝子変異を評価し、分子標的治療が可能かを検索することが推奨されている。また、新たな分子標的治療の候補が年々報告されており、遺伝子検索の頻度が増加している。一方で、遺伝子検索コストが高く、患者によっては治療の選択肢が確保されず、その恩恵が受けられない状況も考えられる。また、微小検体に対して遺伝子パネル検査を行うことは困難であり、信頼性の低い結果につながる可能性がある。このような問題を解決するためには、ターゲットとなりうる遺伝子を絞り込む別の方法が必要である。最近では、深層学習を用いて組織画像から遺伝子変異を予測する人工知能(AI)モデルが開発されている。組織画像はデジタルスライドスキャナで容易に画像を取得できるが、細胞診標本は主に3次元構造が対象となるため、焦点のあった画像を効率的に取得することは容易ではない。したがって、細胞診検体を対象としたAI画像解析による遺伝子変異予測を行うには、まず細胞診標本から効率的に画像を取得するシステム開発が必要となる。今回、汎用デジタルカメラを付設した光学顕微鏡で取得した細胞診画像から、原発性肺癌の代表的な遺伝子変異を予測するAIモデルを構築し、そのモデルによる遺伝子変異推定に強く関連する細胞所見の抽出を試みた。

対象・方法：

公益財団法人がん研究会有明病院で2010年から2021年に組織学的に原発性肺腺癌と診断された99人を対象とし、直接塗抹法 (n=65)及び捺印法 (n=48)により作製された細胞診標本を用いた(以下、癌陽性例)。対照として、悪性細胞を含まない32人の細胞診標本を用いた(以下、癌陰性例)。最初に、顕微鏡付属のカメラから検鏡動作に合わせて連続的に画像を取得し、1片299ピクセルの小画像(以下、パッチ)に分割した上で、ImageNetで事前学習済みの画像分類モデルであるMobileNet-V2をベースモデルとする転移学習によって、癌陽性例と癌陰性例の画像を教師あり学習させ、癌画像の特徴を抽出するAIモデルを生成した(Convolutional Neural Network, CNN1)。CNN1により癌陽性例画像を推論し、最終出力にソフトマックス関数を適用して、推論結果とともに各パッチの確率(Probability)を算出した。次に、癌陽性例を学習用と検証用症例にランダム分割し、いずれの群からもCNN1により癌画像を抽出後、学習用症例の癌画像にEGFR、KRAS、ALK、None(前3者のいずれも検出せず)の教師ラベルを付与し、転移学習(MobileNet-V2)により遺伝子変化を予測するAIモデルを生成した(CNN2)。検証用症例の癌画像に対し、CNN2により遺伝子変化を推論し、各遺伝子予測のパッチ数を基に推定確度(Rank)を付与するアルゴリズムを介して症例単位での遺伝子変化が予測可能か検討した。最後に、CNN2が遺伝子変化を推論する際に関連していた細胞学的特徴を抽出した。

結果：

CNN1の正確度(Accuracy)と損失関数(Loss)は、それぞれ0.859と0.307であり、感度と特異度はそれぞれ0.942と0.962であった。CNN1が癌細胞陽性と判定したパッチをランダム抽出し検証した結果、98%のパッチに癌細胞が含まれていた。CNN2のAccuracyとLossは、それぞれ0.729と0.706であった。検証用症例におけるEGFR群とKRAS群のAccuracyは、それぞれ0.969と0.947であった。症例単位でのAccuracyは0.750であり、Rank Aの症例は高い判定精度であった。CNN2が遺伝子変化を推論する際に関連していた細胞学的特徴は、大半が既報告の細胞所見と合致していたが、EGFR変異を特徴づける特異的な所見も新たに抽出し得た。

考察：

パッチ単位ではなく、推定Rankを付与するアルゴリズムを介した症例単位で評価することは、信頼度が明確になり臨床応用に有用と考えられた。また、細胞診画像を用いた遺伝子変化の初期スクリーニングが可能となれば、可能性の高い対象遺伝子を効率的に検査可能となり、不要な組織検体の採取を避けることができると考える。一方で、判定精度に限界がみられたが、その原因はあるグループの細胞学的特徴が他のグループにも重複して出現するためと考えられ、特にALK群とNone群の出現様式は共に孤立性に出現する傾向が高かったため、両群間での誤判定がより顕著に現れたと考えられた。

結論：

原発性肺癌の細胞診標本画像から遺伝子変異を予測するモデルを構築することができた。このモデルは、実臨床に有用な情報を提供する可能性がある。

1. 学位審査の要旨および担当者

学位番号甲第 734 号	氏 名	石 井 脩 平
学位審査担当者	主 査	澁 谷 和 俊
	副 査	岸 一 馬
	副 査	内 藤 篤 彦
	副 査	高 橋 啓
	副 査	蛭 田 啓 之

学位論文の審査結果の要旨 :

近年、悪性腫瘍に対する分子標的治療に必要な遺伝子パネル検査は、十分かつ良質な組織の確保が求められるばかりでなく、医療経済面でも課題が残されている。本研究では、人工知能 (AI) モデルを用いた癌細胞の遺伝子変異の予測システムを構築し、細胞診標本に関する有用性について検討した。

材料は、原発性肺腺癌 99 症例から作成した細胞診標本 113 検体及び悪性細胞を含まない 32 検体。各検体の画像から 1 片 299 ピクセルのパッチを抽出し、教師あり学習として事前学習済み画像分類モデルに転移学習させ、癌画像の特徴を抽出する AI モデルを生成した (CNN1)。このモデルで癌陽性例画像を推論し、各パッチの確率 (Probability) を算出した。次に、癌陽性例を学習用と検証用の 2 群に分割し、学習用症例の癌画像に EGFR、KRAS、ALK、None (前 3 者のいずれも検出せず) の教師ラベルを付与し、転移学習により遺伝子変化を予測する AI モデルを生成した (CNN2)。検証用症例の癌画像に対し、CNN2 により遺伝子変化を推論し、各遺伝子予測のパッチ数を基に推定確度 (Rank) を付与するアルゴリズムを介して症例単位での遺伝子変化の予想に関する妥当性を検討した。この結果、CNN1 の正確度 (Accuracy)、損失関数 (Loss)、感度及び特異度は、それぞれ 0.859、0.307、0.942 及び 0.962 であった。また、CNN2 の正確度と損失関数は、それぞれ 0.729 と 0.706。検証用症例における EGFR 群、KRAS 群及び ALK 群の Accuracy は、それぞれ 0.969、0.947 及び 0.760 であり、更に ALK 群と None 群間での誤判定が顕著であった。この理由として両群の低発現頻度の影響を推測する。また、CNN2 が EGFR 変異を推論する根拠となる新たな所見として「ライトグリーンに比較的均一に厚く染まる細胞質を有する多辺形～類円形悪性細胞の平面的な出現」が抽出された。本研究により、原発性肺癌の細胞診標本画像から、特定の遺伝子変異を予測する AI モデルを構築することが可能であった。また、EGFR 変異を推論する根拠となる細胞診所見を新たに抽出した。

学位審査会は 2022 年 11 月 22 日、審査委員全員の出席 (書面により事前審査 1 名を含む) の下で開催された。申請者による研究内容についてのプレゼンテーションに続いて、活発な質疑応答が行われた。組織 (形態) による遺伝子変異推定の可能性、症例単位での画像集計でパッチの割合ではなく枚数で分けた理由、遺伝子変異の検索方法、研究の強みと弱みなど多岐にわたる全ての質問に対して申請者は適切に回答した。

当該研究は、細胞診標本を用いた AI モデルによる肺腺癌の遺伝子変異推定の妥当性・有用性を明確に結論づけた研究であり、学位授与に相当するものであることを審査委員全員一致で認めた。