

論文内容の要旨

論文提出者氏名 金子 正大

論文題目

Urine cell image recognition using a deep learning model for an automated slide evaluation system

論文内容の要旨

尿細胞診はほぼ無侵襲的に繰り返し施行可能で安価な、初発膀胱癌の検出スクリーニング検査であり、膀胱癌再発のフォローアップにも利用可能である。しかし、その精度は細胞検査士の経験・技量に依存し、観測者間再現性にも課題を抱えている。病理医 1 人が新たに診断する癌症例の数は世界的に増加を認めており、尿細胞診に要する病理医療従事者の労力を軽減し、かつ、精度と再現性の高い手法が必要とされている。近年、ディープラーニング(deep learning, DL)を用いた人工知能(Artificial Intelligence, AI)による画像認識技術は飛躍的な進歩を遂げており、医療分野においても有望な適応結果が報告されている。

我々は、DL を用いた画像認識技術を尿細胞診に適応し、高い精度と高い再現性を持ち、細胞検査士と病理医の尿細胞診診断プロセスを支援する、自動画像識別システムの開発を行った。

まず、京都府立医大病院で 2016 年 1 月から 2017 年 12 月の間に根治手術によって組織学的に尿路上皮癌と診断され、かつ、術前 3 カ月以内に尿細胞診を提出していた、連続した 195 例の患者の尿細胞診ガラスライドを収集した。尿細胞診ライドは、世界標準の尿細胞診診断様式であるパリ分類に基づき熟練病理医のレビューを受けた。尿細胞診診断結果の内訳は不適正 1 枚、高異型度尿路上皮癌陰性 78 枚、異型尿路上皮細胞 33 枚、高異型度尿路上皮癌疑い 20 枚、高異型度尿路上皮癌 63 枚であった。尿細胞診ガラスライドを浜松ホトニクス製の NanoZoomer S360 を用いデジタルデータ化した。このデジタルデータから各ライド毎に 20 個以上、合計 4987 個の細胞画像を抽出した。次に 2 人の日本臨床細胞学会認定細胞検査士が独立して細胞種類のラベル付けを行い、それぞれのラベルが一致した 4637 細胞画像を正解データとした。これら細胞種類の内訳は、良性細胞 3128 個、異型細胞 398 個、尿路上皮癌細胞 1111 個であった。病理医の確認が必要な細胞のスクリーニングにおける使用を想定して、良性細胞を陰性、異型細胞あるいは尿路上皮癌細胞を陽性と定義した。

機械学習工程はデータ収集、モデル訓練、モデル評価の 3 工程で行った。

データ収集工程では、40 倍の倍率で抽出された正解データ付きの細胞画像を枠の長径に 20 ピクセルの余白を加えた長さを 1 辺とした正方形で切り取った。このうち 80%(155 スライド、3684 細胞)をさらに 4:1 に分割し AI の training と validation に使用した。残る

20%(39 スライド、953 細胞)はモデル評価に用いた。

モデル訓練では、モデルのロバスト化を図るためデータ拡張を行った。本研究では、事前に全ての画像を 256*256 ピクセルにリサイズした上で、2 つの画像を合わせることでデータ拡張を行った。一方の画像の円形部(半径 256*2/5)を切り抜き、別の画像の中心に切り抜いた円形の画像を貼り付け、Blending により円の外周を滑らかにし、中心の画像のラベルを割り当てた。これにより背景に映り込んだ画像ではなく、中心部の細胞画像のみを効率的に学習させることを意図した。また、実行順序はランダムで、左右反転(実行確率 50%)、上下反転(同 50%)、GaussianBlur(同 50%)または AverageBlur(同 50%)、CLAHE(50%) 0,90,180,270 度回転(いずれか一つ)といった基本的な画像処理でもデータ拡張を行った。モデルアーキテクチャは、特徴抽出には高精度かつ高効率のモデルである EfficientNetB6 を使用し、判定層に ArcFace を使用した。このモデルを 5 分割交差検証で訓練させた。

モデル評価では学習に使用されていない 39 スライド・953 細胞を用いて精度を確認した。正解ラベルを作成した者とは別の 3 人の日本臨床細胞学会認定細胞検査士も、臨床情報を参照することなく独立して AI と同じ細胞画像を評価し、良性・異型・尿路上皮癌細胞の別を判定し、AI の診断結果と比較した。この際、AI と全く同じ画像を用いての評価を T1、AI と同じ画像を評価するが当該細胞の含まれるライド全体の参照が許される設定での評価を T2 とした。3 人の中で最高の精度を達成した細胞検査士と AI の比較においては、ベイズ推定法を用い比較した。最高精度の細胞検査士の診断能を X、AI の性能を Y としたとき、事後確率 $\Pr(Y>X)>0.8$ のとき AI が細胞検査士の成績を上回ったと判断した。

陰性陽性の 2 値分類において、AI の Area under the receiver operating characteristic (AUROC) curve は 0.99、精度は 95%、感度は 97%、特異度は 95%であった。一方で最高精度の細胞検査士の精度は 94% (T1)、90% (T2)、感度は 97%(T1)、98%(T2)、特異度は 92%(T1)、85%(T2)であった。精度の事後確率は $\Pr(Y>X) = 0.95$ (T1)、1.0 (T2)であり、AI は最高精度の細胞検査士を上回る精度を示した。細胞のサブタイプごとの解析においても AI は一貫して 90%以上の精度を示し、組織学的グレード、初発再発、pathological ステージ、膀胱内注入療法後群などのサブグループ解析においても 89-97%の高い精度を示した。診断に要する時間についても細胞検査士が T1 の評価に 57-138 分、T2 の評価に 156-233 分を要したのに対し、AI は 28 秒で全テスト細胞評価を完了した。

我々の今回開発した AI を発展させることで、迅速かつ高精度・高再現性の尿細胞診診断支援システムの構築に寄与しうる。