

腎臓病理診断 AI の開発

大阪大学大学院医学系研究科 特任助教 松本 あゆみ *

1. はじめに

腎生検病理診断は疾患の診断や治療方針決定、予後の評価に不可欠である。腎疾患の診療では、血液検査、尿検査、画像検査、臨床経過を総合的に判断する必要があるが、腎生検組織から得られる病理所見は、疾患活動性や慢性化の程度を直接反映する重要な情報である。特に、糸球体病変は、治療介入の必要性や腎予後の推定に関わるため、正確かつ再現性の高い評価が求められる。

一方で、腎病理診断には高度な専門性が必要である。また、同一標本を評価した場合でも、観察者間で所見の判定に差が生じることがある^[1]。さらに、腎生検病理画像は施設ごとに標本作製法、染色条件、スライドスキャナ、画像形式が異なり、画像データ自体も極めて大容量である。そのため、人工知能（artificial intelligence：AI）を用いた病理診断支援には大きな期待が寄せられているものの、実臨床で多施設に展開するためには、施設間差異に堅牢なモデルを構築する必要がある^[2]。

加えて、腎病理 AI では、希少病変の検出が特に重要な課題となる。糸球体病変のなかでも、半月体形、分節性硬化や全節性硬化は、疾患活動性や重症度、慢性化の程度を反映し、治療方針に影響し得る重要な病変である。しかし、これらの病変は全糸球体所見の中で出現頻度が低く、AI 学習に必要な十分量の教師データを確保しにくい。一般に、深層学習モデルは多数の正例を用いて特徴を学習するため、出現頻度の低い病変では、見逃しが生じやすい（クラス不均衡）。このため、単純に教師あり学習を行うだけでは、臨床的に重要な希少病変を安定して検出することは難しい。

AI 開発においても一つ重要な問題が、ドメインシフトである。これは、学習に用いたデータ分布と、実装後に入力されるデータ分布が異なることにより、AI の性能が低下する現象である^[3]。病理画像では、同じ染色が施された標本であっても、染色濃度、色調、明るさ、スキャナ固有の画像特性が施設ごとに異なるため、単一施設のデータで学習した AI を他施設に適用すると性能低下が生じる可能性が高い。腎病理 AI を社会実装するためには、単に一施設内で高精度なモデルを作成するだけでは不十分であり、多施設でも安定して機能する方法論の確立が不可欠である。そこで本研究では、腎生検 Periodic acid Schiff (PAS) 染色画像を対象に、施設間差異を補正しつつ、希少な腎病理所見を自動検出する AI の開発を行った。

2. 研究方法と結果

2.1. 多施設腎生検 PAS 染色画像データセットの構築

国内 22 施設で施行された腎生検症例の PAS 染色組織をデジタルデータ化した whole slide image (WSI) を用いた。対象は 2014 年 1 月から 2018 年 12 月までに腎生検を受けた 16 歳以上の患者であり、解析には開発コホート 2498 WSI、テストコホート 2178 WSI、合計 4676 WSI を用いた。各施設で通常診療として作製された PAS 染色標本を、Aperio ScanScope、Hamamatsu NanoZoomer、VS120 virtual slide microscope のいずれかのスキャナでデジタル化し、SVS、NDPI、VSI の 3 種類の画像形式を含むデータセットを構築した。

*（助成採択時） 大阪大学医学部附属病院 腎臓内科 医員

2.2. 糸球体病変アノテーションと希少病変の設定

複数の腎臓専門医により、PAS 染色画像上の糸球体領域に対してアノテーションを付与した。分類対象は、全節性硬化および半月体、分節性硬化を除く糸球体、全節性硬化、半月体、分節性硬化の4カテゴリとした (Fig.1)。

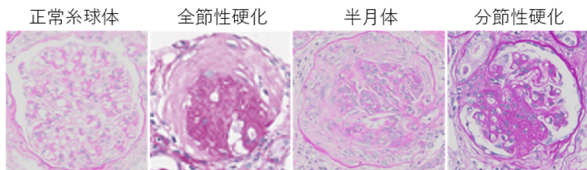


Fig. 1 分類対象とした糸球体病変

全節性硬化は全アノテーションの約 16%を占めた一方、半月体および分節性硬化はそれぞれ約 2~3%にとどまった。半月体や分節性硬化はデータ数が少なく、AI にとって検出が難しい希少病変であることが確認された。

2.3. 施設間・スキャナ間ドメインシフトの可視化と定量評価

まず、PAS 染色糸球体画像の特徴量解析を行った。ImageNet で事前学習された ResNet50 を用いて糸球体画像から特徴量を抽出し、t-SNE により分布を可視化した。

その結果、画像特徴は施設およびスキャナ形式により明確にクラスター化し、スキャナ由来の画像特性が AI 性能に影響し得る主要因であることが示された (Fig.2)。

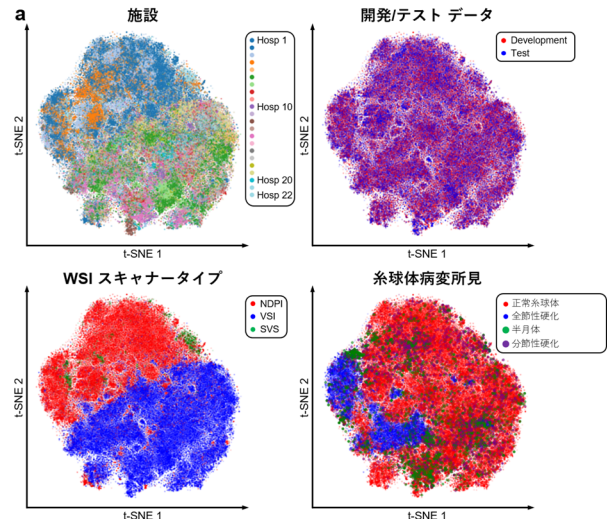


Fig. 2 多施設腎生検 PAS 画像における施設間・スキャナ間差異

2.4. YOLOv8 によるベースライン病変検出モデルの構築

次に、糸球体病変検出モデルとして YOLOv8 を採用した。NDPI、VSI、SVS を代表する 3 施設を選び、各施設のデータを用いて教師あり学習による Baseline YOLO を作成した。性能評価においては、同一施設の未使用画像、同一スキャナ形式だが別施設の画像、異なるスキャナ形式の画像という 3 種類のテスト条件を設定し、施設差およびスキャナ差に対する頑健性を検証した。

その結果、一般的な糸球体検出では比較的高い性能が得られたが、希少病変である半月体および分節性硬化では検出性能が低く、さらに他施設画像、特に異なるスキャナ形式の画像に適用した場合に性能低下を認めた。例えば、VSI 形式施設で学習したモデルを NDPI 形式画像に適用した場合、半月体検出の AP50 は 0.64 から 0.19 へ低下し、希少病変検出がドメインシフトの影響を強く受けることが示された。これは、腎病理 AI の実用化において、施設間差異の補正が必須であることを示す重要な結果であった。

2.5. 半教師あり学習による未ラベル画像の活用

第一の対策として、半教師あり学習を導入した^[4]。単一施設の専門医アノテーション付き画像で作成した Baseline YOLO を用いて、他施設の未ラベル画像に疑似ラベルを付与し、一定以上の信頼度を示す画像パッチを元の教師データに加えて再学習した。疑似ラベル採用の信頼度閾値として 50%、70%、90% を比較した。90% 閾値では採用される疑似ラベルが少なく、改善効果は限定的であった一方、50% 閾値ではより多様な疑似ラベルを取り込むことができ、特に希少病変の検出性能改善に有用であった。このため 50% 閾値で疑似ラベルを用いる Semi-Supervised YOLO を最適モデルとして選択した。

Semi-Supervised YOLO は、同一施設内評価において希少病変の検出性能を改善した。これは、専門医による追加アノテーションを大幅に増やさなくても、未ラベル画像を活用することで AI 性能を高められることを示している。ただし、半教師あり学習単独では、スキャナ形式が異なる施設への一般化には限界が残った。すなわち、未ラベルデータの活用には有効であるが、画像スタイル自体が大きく異なる場合には、ドメインシフト補正を併用する必要があると考えられた。

2.6. Residual CycleGAN による施設間画像差の補正

第二の対策として、Residual CycleGAN に基づくドメイン適応を導入した^[5]。これにより、染色調やスキャナ由来の画像スタイル差を補正しながら、糸球体係蹄、硬化、半月体など診断上重要な形態情報を保持することを目指した。

代表施設ごとに、他 21 施設の画像スタイルへ変換する Residual CycleGAN を構築し、変換画像を用いて GAN-Augmented YOLO を作成した。さらに、このモデルを用いて他施設画像に疑似ラ

ベルを付与し、50%以上の信頼度を示すデータを追加して再学習する GAN-Semi-Supervised YOLO を構築した (Fig.3)。Residual CycleGAN による変換後画像では、施設間の特徴量分布差が縮小し、平均 FID は 55.9 から 20.2 へ低下し、施設およびスキャナ由来の画像分布差が定量的に緩和されたことが示された。

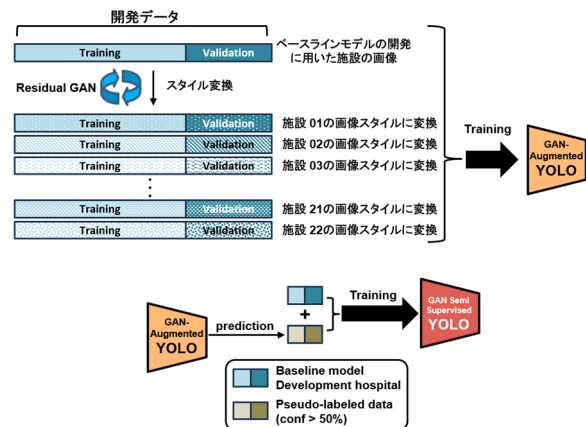


Fig. 3 Residual CycleGAN を用いたドメイン適応と AI 学習戦略

2.7. AI モデルの性能比較と希少病変検出の改善

最終的に、Baseline YOLO、Semi-Supervised YOLO、GAN-Augmented YOLO、GAN-Semi-Supervised YOLO の 4 種類を比較した。一般的な糸球体検出ではいずれのモデルも高性能であったが、半月体および分節性硬化といった希少病変では学習戦略による差が明確であった。同一施設内評価では Semi-Supervised YOLO が有効であり、希少病変検出に 15.2~17.7% の改善を示した。一方、NDPI と VSI のようにスキャナ形式が異なる条件では、GAN-Semi-Supervised YOLO が最も安定して性能を改善し、半月体検出では最大 63.4% の改善を示した (Fig.4)。これにより、半教師あり学習と Residual CycleGAN によるドメイン適応を組み合わせることで、専門医アノテーション不足と施設間差異という二つの課題を同時に緩和できることが示された。

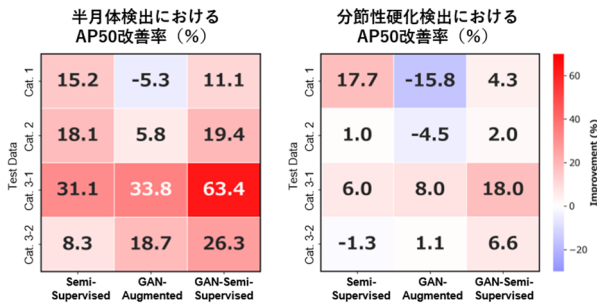


Fig. 4 学習戦略による病変検出性能の改善

2.8. Grad-CAM による AI 判断根拠の可視化

さらに、モデルの判断根拠を検討するため、Gradient-weighted Class Activation Mapping (Grad-CAM) を用いた可視化を行った。Baseline YOLO では萎縮尿細管を全節性硬化と誤認する例や、分節性硬化部位への注意が不十分な例がみられたが、GAN-Semi-Supervised YOLO では実際の糸球体領域、硬化部位、非典型的半月体により適切に注意が向く傾向が確認された。この結果は、性能改善が単なる数値上の改善にとどまらず、病理構造の識別に関わる特徴学習の改善を伴っている可能性が示された。

3. 考察

本研究により、腎病理画像 AI の社会実装における主要課題である、専門医アノテーション不足と施設間ドメインシフトに対して、半教師あり学習と Residual CycleGAN を組み合わせた方法が有効であることが示された。特に、半月体および分節性硬化のような希少病変は、臨床的には重要である一方、アノテーション数が少なく、AI 学習において不利である。本研究では、疑似ラベルを用いて未ラベル画像を活用することで、このデータ不足を部分的に補える可能性を示した。

一方、半教師あり学習単独では、異なるスキャナ形式への一般化には限界があった。この結果は、病理 AI においては学習データ数を増やすだけでなく、画像取得条件の違いそのものを補正する必

要があることを示している。Residual CycleGAN によるドメイン適応は、施設間の画像特徴量差を縮小し、NDPI と VSI のような異なるスキャナ形式間で希少病変検出性能を改善した。これは、当初計画で掲げた「施設間差異を標準化した腎生検病理画像を用いることで AI 性能を向上させる」という目的に合致する成果である。

本研究成果は、腎生検診断の均てん化に資する可能性がある。腎病理専門医に限られる医療環境において、AI が糸球体や主要病変を自動検出し、定量的な情報を示唆できれば、診断補助、教育支援、病理カンファレンスの効率化に寄与し得る。今後は、より多様な施設・染色法・病変を含む外部検証を進め、臨床情報との統合や予後予測への応用を検討する必要がある。

4. 結論

本研究では、腎生検画像を対象に、腎病理診断 AI の実用化に向けた基盤技術を開発した。多施設・多スキャナ環境では明確なドメインシフトが存在し、特に半月体や分節性硬化などの希少病変検出で AI 性能が低下することを明らかにした。

これに対し、半教師あり学習により未ラベル画像を活用し、さらに Residual CycleGAN によるドメイン適応を組み合わせることで、専門医アノテーション不足と施設間画像差の双方を補正し、希少病変検出性能を改善できることを示した。

本成果は、腎病理 AI の社会実装に向けた有用な方法論であり、今後は多様な染色法、病変、施設への展開を通じて、病理画像診断 AI 全般への応用が期待される。

5. 参考文献

- [1] Tizhoosh, H. R. et al. Searching images for consensus: can AI remove observer variability in pathology? Am. J. Pathol. 191(2021), 1702–

1708.

- [2] Satturwar, S. & Parwani, A. V. Artificial intelligence-enabled prostate cancer diagnosis and prognosis: current state and future implications. *Adv. Anat. Pathol.* 31 (2024), 136–144.
- [3] Stacke, K., Eilertsen, G., Unger, J. & Lundstrom, C. Measuring domain shift for deep learning in histopathology. *IEEE J. Biomed. Health Inf.* 25 (2021), 325–336.
- [4] Cheplygina, V., de Bruijne, M. & Pluim, J. P. W. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med. Image Anal.* 54 (2019), 280–296.
- [5] Ade Bel, T., Bokhorst, J. M., van der Laak, J. & Litjens, G. Residual cyclegan for robust domain transformation of histopathological tissue slides. *Med. Image Anal.* 70 (2021), 102004.

6. その他成果発表等

6.1. 論文発表

- ・ Matsui I, Matsumoto A, Imai A, et al, Domain-adaptive semi-supervised learning for efficient rare pathological lesion detection with minimal annotation. *NPJ Digit Med.* 23-8(1), (2025), 778.

6.2. 学会発表

- ・ 松本あゆみ, 松井功, 今井淳裕ら, 異なる医療機関における病理画像差異を校正する手法の確立, 第 67 回 日本腎臓学会学術総会, 横浜, 2024 年 6 月 28 日～30 日.
- ・ 松井 功, 腎生検画像における稀少糸球体病変検出のための半教師あり学習と Residual GAN を用いたデータ拡張の統合的アプローチ, 第 7 回 日本メディカル AI 学会学術集会, 東京, 2025 年 6 月 27 日～28 日. 優秀一般演題賞

7. 謝辞

本研究は、公益財団法人 島津科学技術振興財団の助成金を受けて実施され、成果の報告に至りました。心より感謝申し上げます。