

Masatoshi Abe Self-Supervised Learning for Feature Extraction from Glomerular Images and Disease Classification with Minimal Annotations

Journal of the American Society of Nephrology 2024/2/6

link:https://journals.lww.com/jasn/abstract/2025/03000/self_supervised_learning_for_feature_extraction.17.aspx

安部政俊,新岡宏彦,松本あゆみ,勝間勇介,今井淳裕,奥嶋拓樹,尾崎晋吾,藤井直彦,
岡一雅,坂口悠介,井上和則,猪阪善隆,松井功

深層学習は腎生検病理画像解析に有効であるということが知られています。しかし、深層学習モデルを十分に学習させるにはラベルつきデータを大量に集める必要があり、ラベル付きデータが乏しいことが腎生検画像の解析における深層学習の普及を妨げています。

大阪大学医学部附属病院腎臓内科で腎生検を受けた 384 例の PAS 染色画像中の 10,423 枚の糸球体画像に自己教師あり学習の手法の一つである DINO(self-distillation with no labels)を適応しました。DINO 学習済みモデルによって生成された特徴マップを可視化するために主成分分析(PCA)^{※2}を用いると、糸球体の構成要素ごとに色が分かれ、異なる組織には異なる主成分の要素が強く出ていることが確認できました(図 2)。そして、DINO 学習済みモデルまたは従来の ImageNet^{※3} 学習済みモデルを用いて分類タスクを学習させ、受信者動作特性曲線下面積(ROC_AUC)^{※4}などの指標を用いて性能を評価しました(図 3)。分類タスクとして微小糸球体病変、メサンギウム増殖性糸球体腎炎、膜性腎症、糖尿病性腎症の 4 疾患分類と、高血圧、蛋白尿、血尿などの臨床パラメータ分類の 2 つを用いました。疾患分類では、DINO 学習済みモデル(ROC_AUC=0.934)が ImageNet 学習済みモデル(ROC_AUC=0.892)を上回りました。ラベル付きデータが制限された場合、ImageNet 学習済みモデルの ROC_AUC は 0.763[95%信頼区間:0.724-0.802]に低下しましたが、DINO 学習済みモデルは優れた性能を維持しました(ROC_AUC=0.882 95%信頼区間:0.862-0.903)。DINO 学習済みモデルはいくつかの臨床パラメータにおいてもより高い ROC_AUC を示しました。

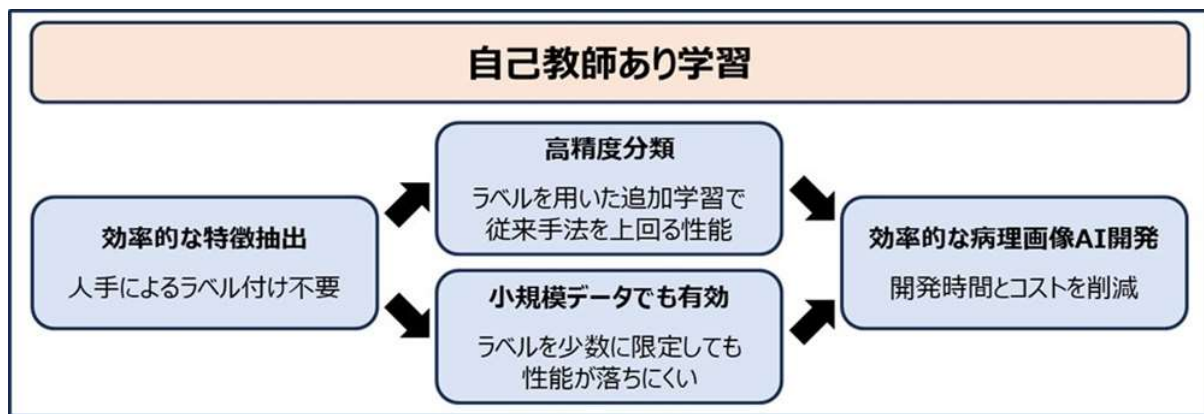


図 1: 自己教師あり学習により効率的な病理画像 AI 開発が可能になる

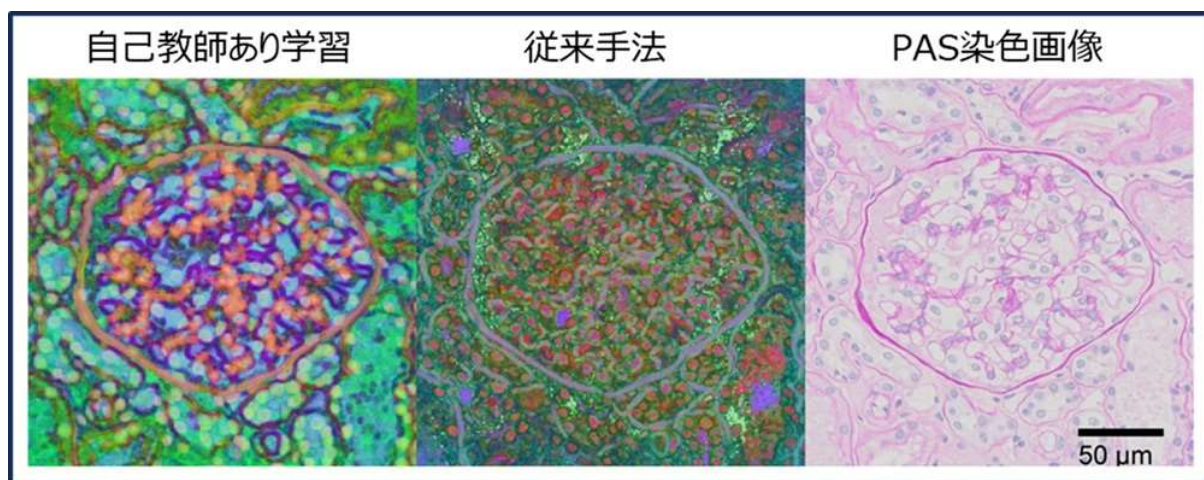


図 2: 自己教師あり学習を用いることで、人間の指示なしで組織学的特徴を抽出 自己教師あり学習および従来手法で抽出した画像特徴の主成分をカラー表示 自己教師あり学習では腎糸球体の構成要素ごとに色が分かれている

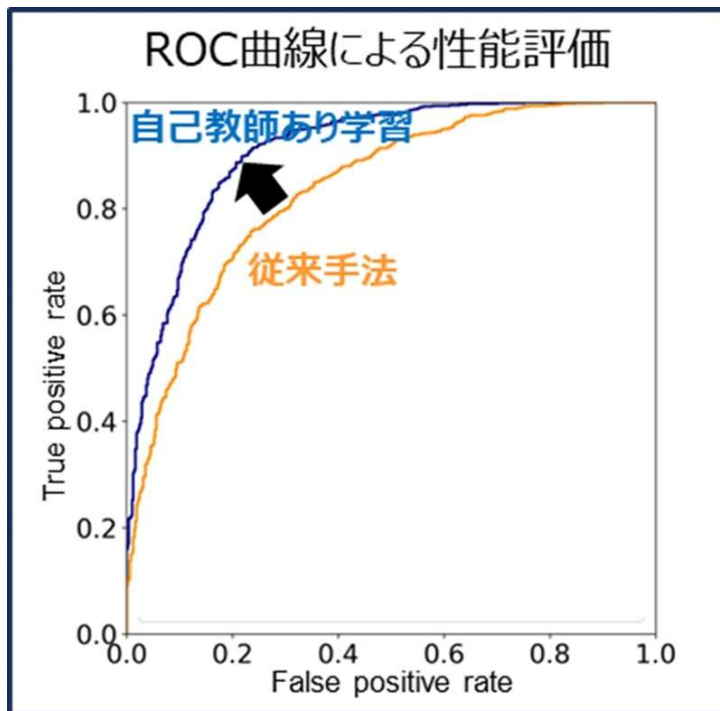


図 3: 自己教師あり学習により疾患分類性能が向上

※1 自己教師あり学習

ラベルのないデータを使用して、下流のタスク（本研究では疾患分類や臨床パラメータ分類）に有用な表現を得るための機械学習の手法の一つ。

※2 主成分分析(PCA)

多数の変数を持つデータから最も重要な情報を抽出し、より少ない新しい変数（主成分）に要約する統計手法で、データの解釈を容易にする。

※3 ImageNet

画像認識の研究で用いるために設計された大規模な一般画像のデータセット。

※4 受信者動作特性曲線下面積(ROC_AUC)

横軸に偽陽性率、縦軸に真陽性率をプロットし、曲線が上に凸なほどモデルの性能が高いことを示し、AUC（曲線下の面積）が1に近いほど、モデルの予測力が優れていることを示す。