

厚生労働科学研究費補助金
(政策科学総合研究事業(臨床研究等 ICT 基盤構築研究事業))
総合研究報告書

Deep Learning 技術を用いた腎生検病理画像の自動分類による
病理診断の効率化と診断補助に関する研究

研究代表者 大江 和彦 東京大学大学院医学研究科 医療情報学分野 教授

研究要旨

腎生検病理画像のデータベースを構築し、深層畳み込みニューラルネットワーク (CNN) を活用した糸球体病理診断手法の開発を目指す。またこのプロセスから得られる知見を糸球体病理画像診断プロセスの標準化に役立てるとともに、腎病理診断の効率化と診断補助に資することを目指す。本年度は、これまでに行った研究結果を総括する。1) 研究参加施設より 112,210 枚の蛍光抗体画像と 7,123 枚の光学顕微鏡画像(Whole Slide Image: WSI)の提供を受けデータベースを構築した。2) CNN による IgG 蛍光抗体により染色された糸球体画像の蛍光強度の推定精度(F1)を評価したところ、メサンギウム領域が 75.6%、糸球体壁領域が 67.9%、その他領域が 60.5%であった。3) 光学顕微鏡糸球体病理画像に関し、Faster R-CNN を用いて高解像度の WSI から微小な糸球体を検出する手法を開発し、先行研究に比べ高い検出精度を達成した。本研究班を構成する 5 名の腎病理の専門家の協議のもと PAS 染色における 12 の所見項目を作成し、その所見一致度を評価したところ、 κ 値は 0.04~0.43 であり、所見項目によるばらつきがみられた。1 枚の糸球体画像を入力として、糸球体所見を分類する CNN の精度を評価したところ、Extracellular matrix sclerosis/Collapsing obsolete と Crescent の ROC-AUC は 0.9 を超えた。また、CNN の着目点を可視化することで誤りの分析を行った。4) 本研究成果を利用し、WSI から糸球体領域を検出する WEB アプリケーションシステムを開発し、インターネットに向けた公開する予定である。

<研究分担者>

河添 悦昌	東京大学大学院医学研究科 医療 AI 開発学講座 特任准教授
堂本裕加子	東京大学医学部附属病院 病理部 特任講師
南学 正臣	東京大学医学部附属病院 腎臓・内分泌内科 教授
松尾 豊	東京大学大学院工学系研究科 特任准教授
中山浩太郎	東京大学大学院工学系研究科 特任講師
清水 章	日本医科大学医学部 病理学 教授

長田 道夫 筑波大学医学医療系 教授
宇於崎 宏 帝京大学医学部 病理学講座 教授
柏原 直樹 川崎医科大学 腎臓・高血圧内科学 教授

<研究協力者>

鈴木 祐介 順天堂大学医学部 腎臓内科 教授
塚本 達雄 公益財団法人 田附興風会医学研究所 北野病院 部長
長洲 一 川崎医科大学 腎臓・高血圧内科学 講師
嶋本 公徳 東京大学医学部附属病院 企画情報運営部 特任研究員
山口 亮平 東京大学大学院医学研究科 医療 AI 開発学講座 特任助教

A. 研究目的

本研究は、本研究チーム構成員の研究機関、日本腎病理研究会、腎臓病学会所属の関連する研究者の所属機関が分散所有する生検腎組織のデジタル病理画像を収集し、これらのデータベースを構築するとともに、人工知能手法のひとつで深層学習の手法である CNN(Convolutional Neural Network: 畳み込みニューラルネットワーク)による画像識別を活用した腎病理診断手法を開発する。またこのプロセスから得られる知見を腎糸球体病理画像診断プロセスの標準化に役立て、腎病理診断の効率化と診断補助に資することを目指す。

B 研究方法

深層学習による画像識別を活用した腎病理診断手法の開発を行った。また、病理診断支援システムのプロトタイプとして、生検腎の光学顕微鏡病理画像(Whole Slide Image: WSI)から糸球体領域を検出する WEB アプリの開発に取り組んだ。以下、本研究班で取り組んできた各研究サブセクションについて述べる。

1. 生検腎病理画像の収集と DB の構築 (H28, H29 年度)

研究分担者が所属する 5 施設において研究倫理申請を完了すると共に、各施設が所有する生検腎組織の蛍光抗体病理画像と光学顕微鏡病理画像を匿名化した上で提供を受けた。

2. 蛍光抗体法により染色された糸球体病理画像の所見分類 (H28, H30 年度)

H28 に引き続き蛍光抗体法により染色された糸球体病理画像の所見分類を行った。

3. 光学顕微鏡病理画像の所見分類 (H29, H30 年度)

本研究は一般的に行われる腎病理の臨床的診断方法とは異なり、1 枚の WSI に含まれる全ての糸球体を検出し、各糸球体の画像的特徴から、複数の所見項目のスコアを分類タスクにより決定し、その結果を元に臨床診断分類するアプローチを取る。そのため、本セクションを 1) WSI から糸球体を検出する手法の開発、2) 糸球体の所見項目の定義と妥当性の評価、

3)糸球体所見を分類する CNN の開発と評価、4)糸球体内部構造領域の抽出手法の開発と評価、の4つのサブセクションとして行った。

4. 光学顕微鏡画像から糸球体領域を検出する WEB アプリケーションシステムの開発(H30年度)

これまでに開発した手法を応用して、1枚のWSIを入力として、それに含まれる全ての糸球体を検出して結果を返却するWEBアプリを開発する。

C. 研究結果

1. 生検腎病理画像の収集と DB の構築(H28, H29年度)

蛍光抗体画像 DB の構築

各施設より提供を受けた蛍光抗体画像の画像フォーマットはJPEG形式であり、ファイルサイズは900KB~1,400KB程度、ピクセルサイズは幅が2,048、高さが1,536程度の画像であった。提供を受けた蛍光抗体画像の合計枚数は112,210枚であり、その内訳を末尾資料の表1に示す。

光学顕微鏡画像(WSI)DBの構築

各施設より提供を受けた光学顕微鏡病理スライドをデジタル画像化するための装置は施設によって異なるため、異なる画像フォーマットによりWSIが提供された。ファイルの拡張子はそれぞれ、NDPI(浜松ホトニクス)、VSI(オリンパス)、SCN(ライカ)であり、これらを汎用的なフォーマット(Portable Network Graphics: PNG)に変換した上で、画像データベースを構

築した。提供を受けた光学顕微鏡画像(WSI)枚数の集計と、異なる病理番号および患者番号の集計をそれぞれ、末尾資料の表2と表3に示す。

2. 蛍光抗体法により染色された糸球体病理画像の所見分類(H28, H30年度)

本年度に行った研究の結果を報告する。

目的

蛍光抗体法により染色された糸球体病理画像は、1枚の幅が2,000~2,500ピクセル、高さが1500~2,000ピクセルの4:3で構成されるJPEGフォーマットの画像であり、1枚の画像に1個~数個の糸球体が含まれる。各画像には、(1)メサンギウム (2)係蹄壁(血管内皮細胞) (3)その他、の3種類の領域における3段階もしくは4段階の蛍光強度がラベルとして付与される。本サブセクションでは、画像分類を行うCNNによって、未知の蛍光抗体糸球体画像に対する3種類の領域の蛍光強度の推定精度を評価する。

データセット

東大病院より提供を受けた腎臓糸球体病理画像(IgG 蛍光抗体画像)3,791枚を用いた。データセットの要約を末尾資料の表4に示す。

実験方法

ベースラインの方法として、入力画像のサイズを幅256、高さ256として、畳み込み層5層、全結合層3層のAlexNetを用いて領域ごとにCNNを作成した。ベースラインからの精度向上を目指し、下記

4 つのネットワーク構造を逐次追加して、精度向上に寄与するかを検討した。末尾資料の図 1 にベースラインとなる AlexNet の構造と、以下で追加するネットワーク構造を記す。

1) マルチタスク構造の追加

全結合層を 3 つに分岐し、各領域のラベルを同時に予測する構造とすることで、領域間の相関情報を利用する。

2) クラス重みの追加

クラス間の学習画像枚数の偏りを補正するために損失関数にクラスごとの重みの項を導入する。

3) 入力画像の高次元化

蛍光強度の微細な違いを特徴として抽出するために、入力画像をより高解像度(幅 1024 ピクセル、高さ 1024 ピクセル)にリサイズした画像を用いる。

4) スキップコネクションの追加

蛍光強度という単純な特徴量が識別に重要であると考えに基づき、2 層目の畳み込み層で抽出した特徴量を全結合層に結合する Skip Connection を導入する。

精度の評価は CNN の学習用データセットを 60%、評価用データセットを 40%としてランダムに分割して行い、評価用画像枚数のカテゴリ間の偏りの影響を受けないようにカテゴリごとの F 値の平均値(macro-F1)を評価指標として用いた。

評価結果

末尾の資料図 2 に各領域の蛍光強度の推定結果を示す。最終的に、メサンギウム領域はベースラインの 43.3%から、32.2point 精度向上し、75.6%となり、係蹄壁はベースラインの 67.9%から、8.7point 精度向上し、75.6%となり、その他領域はベースラインの 43.2%から、17.3point 精度向上し、60.5%の精度を示した。

3. 光学顕微鏡病理画像の所見分類

3.1. WSI から糸球体を検出する手法の開発(H29 年度)

目的

一般的な画像からの物体検出タスクと比較して、対物レンズで 40 倍に拡大された光学顕微鏡病理画像からの糸球体検出は、長辺が最大で 20 万ピクセルを超える画像から直径が約 800 ピクセル程度の微小な糸球体を検出する点が大きく異なる。本サブセクションでは、1 枚の WSI をスライディングウィンドウ法により走査し、各ウィンドウに含まれる画像を Faster R-CNN の入力として糸球体を検出し、ウィンドウの境界に位置する糸球体の重複検出を考慮して、1 枚の WSI 上で糸球体の領域を検出する方法を開発する。

Faster R-CNN による糸球体検出手法

対物レンズで 40 倍に拡大された標本の WSI は長辺が最大で 20 万ピクセルを超え、このサイズの画像を直接扱うことは効率的ではないため、目視によって糸球体を十分に検出可能な解像度を検討し、対物レンズ倍率で 5 倍相当にダウンサンプルした画像を用いた。

ダウンサンプルした WSI をスライディングウィンドウ法により走査し、各ウィンドウに含まれる画像を Faster R-CNN の入力として糸球体を検出した。糸球体の直径を $200\ \mu\text{m}$ (110 ピクセル) 程度と見積もり、複数の糸球体と十分な背景画像を含めることができるように、 $2,000 \times 2,000\ \mu\text{m}$ のウィンドウサイズを設定した。Faster R-CNN モデルの学習時には、アノテーションされた各糸球体を中心として、設定したサイズのウィンドウで切り出した画像と、その画像に含まれる領域情報を Faster-RCNN の入力とした。モデルの評価時には、WSI の全領域をスライディングウィンドウ法で走査し、ウィンドウに含まれる画像を Faster-RCNN の入力とした。ここで、ウィンドウの境界に位置する糸球体であってもより完全な形でウィンドウに含めることができるよう、隣接するウィンドウを上下左右にそれぞれ $200\ \mu\text{m}$ 重複して走査した。このことにより、糸球体形状のバリエーションを減らすことができる一方、ウィンドウの重複領域に含まれる糸球体を複数回検出する可能性が生じる。そのため、検出された糸球体領域が他の糸球体領域と重なる場合には、重複する面積が自身の面積の一定割合以上のものを、重複する他方とマージして一つの糸球体領域とする処理を行った。

データセット

4 種類 (PAS, PAM, MT, Azan) で染色された WSI を対象として、各染色 200 枚、計 800 枚の WSI に対して、合計約 33,000 個の糸球体領域をアノテーションした教師

付きデータセットを用いて評価を行った。各染色 200 枚の WSI には平均して約 8,300 個の糸球体が含まれ、1 枚の WSI に含まれる糸球体個数は平均して約 42 個であった。検出された糸球体の例を末尾資料の図 3 に示す。

評価結果

糸球体領域の検出を、教師データとして用意した糸球体領域を、Faster-RCNN が正しく検出できたかどうかを評価する 2 値分類のタスクとして設定し、感度と陽性的中率、ならびにその調和平均である F 値のマイクロ平均を評価指標とした。教師領域と提案領域が十分な重なりを持つかどうかの指標は Intersection over Union (IoU) を用い、その閾値を 0.5 とした。モデル学習用、モデル選択用、モデル評価用に用いる WSI をそれぞれ 120 枚、40 枚、40 枚として 5 分割交差検定により評価した、PAS、PAM、MT、Azan の各染色画像における検出精度 (F 値) はそれぞれ、0.93、0.93、0.90、0.88 であった。また、感度はそれぞれ 0.93、0.94、0.92、0.90、感度はそれぞれ 0.92、0.92、0.88、0.85 であった。

3.2. 糸球体の所見項目の定義と妥当性の評価 (H29, H30 年度)

目的

本研究のアプローチは、WSI から切り出した各糸球体画像に対して、分類タスクによりその画像的特徴から所見項目のスコアを決定するものであるが、分類タスクを教師付き学習により行うためには、教師ラベルを付けるための基準が必要と

なる。ここで、腎臓病理の分野において、このようなラベル付けの基準に相当するものはなく、近いものとして、IgA 腎症に対する Oxford 分類や移植腎に対する Banff 分類、巣状糸球体硬化症に対する Columbia 分類といった、特定の疾患を対象とした所見項目の分類基準は存在するものの、これらは教師付き学習のラベル付けを想定した分類基準ではない。また、本研究では疾患を限定しない病理画像が入力されることを想定するため、特定の疾患を前提とした分類基準を用いることでは不十分であると考えた。そこで、本サブセクションでは糸球体画像を分類するための所見項目を専門家の協議のもとに決定し、各項目のスコアリング基準を作成しその妥当性を評価する。腎病理の分野では PAS 染色、PAM 染色など複数の染色画像が用いられるが、ここでは最も一般的に用いられる PAS 染色を対象とし所見項目の定義を作成した。

所見項目定義の作成と一致率の評価

所見項目の定義は、糸球体画像を分類する上で有用な所見項目と、各項目のスコアリング基準からなる。まず、Oxford 分類や病理学の成書を参考とし、たたき台となる定義案を腎臓内科医師で博士課程の山口亮平氏が作成した。次に、東大病院の腎病理デジタルスライド (WSI) から切り出した糸球体画像を 100 枚選択し、作成した所見項目のスコアリング基準に基づく所見付け (スコアリング) を、山口氏と本研究班の腎病理の専門家 (長田 道夫 教授、清水 章 教授、宇於崎 宏 教授、堂本 裕加子 特任講師) ら 5 人の医師

が行った。5 人の医師のスコアリングの一致度を Cohen の κ 係数により算出した上で目合わせの会議を開催し、所見付け定義の改定を行った。100 枚の画像に対するスコアリングと一致度の算出、会議の開催、所見項目の改定を計 3 回繰り返した結果得られた所見付け定義と、3 回目のスコアリングの一致度の要約を以下に示す。

所見付け定義

末尾資料の表 5 に所見付け定義の最終案を示す。所見項目は計 12 項目となり、各項目は取りうる値のスコアと、そのスコアリング基準から構成される。例えば Mesangial hypercellularity は Normal、Mild、Moderate、Severe、判別不能、の 5 つのスコアからなり、そのスコアリング基準は、「糸球体内の血管極付近を除く、もっとも細胞増殖の多いメサンギウム領域におけるメサンギウム細胞をカウントし、以下の基準で所見をつける。4 未満: Normal、4-5: Mild、6-7: Moderate、8 以上: Severe。メサンギウム領域が存在しな、もしくはメサンギウム領域にメサンギウム基質が完全に残っていない場合は、判断不能とする。また、Sclerosis を生じているメサンギウム領域は対象外とする。」とした。ここで、スコアに含まれる「判別不能」はその所見のスコアを判断するために必要な糸球体構造物が画像中に含まれない場合に選択する値として留意した。

所見付けフローチャートの作成

所見定義に加えて、実際に所見付けを行

う上で所見項目間の依存関係を明確にするためにフローチャートを作成した。このフローチャートは、「まず A という所見をつける。その後、A という所見が a の時は、B という所見は判別不能とする」というように、所見項目間に依存関係がある場合のスコアリング基準を明確にしたものである(論文未発表のため詳細は未掲載)

医師間のスコアリング一致度の要約

次に、各所見項目における、3 回目のスコアリングの一致度の要約を末尾資料の表 6 に示す。「判別不能」を一つのスコアとして扱った場合と、誰か一人でも「判別不能」とした画像を除外した場合とに分けて κ 値を算出した。先行研究を参考に κ 値が 0.4 以上のものをそこそこの一致が見られるものとする、判断不能を含む場合に κ 値が 0.4 を超えた所見項目は、Increased Mesangial Matrix, Crescent (Fibrous), Collapsing Obsolete, Mesangiolytic, Polar Vasculosis の 5 項目であった。これら項目は、異なる医師がスコアリングしたとしても、そこそこの一致度を示す所見項目であると解釈できる。

3.3. 糸球体所見を分類する CNN の開発 (H29, H30 年度)

本サブセクションでは、畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を用いた糸球体所見項目の分類結果について示す。

全体データセットの構築

研究班に参加する、東大病院、北野病院、筑波大学病院より提供された WSI を対象として、研究補助者が手作業で切り出した糸球体画像 8369 枚を目視し、ピンぼけている等の 213 枚の不適切画像を除外し、8156 枚の画像を対象とした。博士課程の山口亮平氏が、所見項目定義に含まれる全 12 項目について、本研究で作成した所見付け定義とフローチャートとに従って、計 8156 枚の画像に対する所見付けを一人で行った。全体データセットの要約を末尾資料の表 7 に示す。

CNN 開発用データセットの選択

表 7 に含まれる全 12 項目のうち、判断不能とされた画像を除外した場合の平均 κ 値が 0.2 以上であり、学習に 1000 枚以上の画像が利用できる項目を対象とした。この基準により、Mesangiolytic、Polar Vasculosis、Afferent/Efferent Arteriolar Hyalinosis の 3 項目が除外された。残りの 9 項目のうち、3 種類の Crescent はスコアの論理和をとり 1 つの項目として集約し、Extracellular Matrix Sclerosis と Collapsing Obsolete も同様にスコアの論理和をとり 1 つの項目として集約した。この操作により、9 項目が 6 項目に集約された。更に、この 6 項目についてスコアを 2 値化したものを CNN 開発用のデータセットとして用いた。末尾資料の表 8 に CNN の開発に用いたデータセットの要約を示す。また、各所見項目において典型的な陽性所見を有する糸球体画像の例を末尾資料の図 4 に示す。

実験設定

各所見項目について2値分類を行うCNNを作成した。ImageNetで学習済みのVGG16モデルをベースとして、CNN開発用のデータセットでファインチューニングする方法を採用した。データセットの分割に関して、糸球体画像の単位で分割すると、同じ病理番号に由来する異なる糸球体が、学習用データと評価用データに含まれることになり、汎化性能が損なわれると思われたため、病理番号の単位で分割した5分割交差検定によって精度を評価した。各検定におけるデータの分割は、学習用データを80%、モデル選択用データを10%、評価用データを10%として、評価用データの5回平均値を最終評価値とした。プライマリの評価指標として陽性例のAUC-ROCを用いた。病院施設によって受診患者の疾患分布が異なることが想定されるが、感度と特異度から算出されるAUCはデータセットの陽性例の比率に影響を受けないため、本研究の結果を他施設での適用事例に外挿しやすいためである。また、セカンダリの評価指標としてF1スコアを用いた。これは、感度と陽性適中率の調和平均によって算出される値であり、トレードオフの関係にある感度と陽性適中率がまんべんなく高い場合に高値を示す。

CNNの性能結果とCNN着目点の可視化

評価用データを対象とした評価値の5回平均値を末尾資料の表9に示す。また、CNNの着目点をGrad-Cam(Gradient-weighted Class Activation Mapping)とLIME(Local Interpretable Model-

agnostic Explanations)により可視化した糸球体画像の例を末尾資料の図5(a)-(d)に示す。Grad-Camは誤差逆伝播時に計算される勾配情報を用いて、確率スコアへの影響が大きい画像の位置をヒートマップによって表現する。LIMEは非線形のモデル(CNNなど)を、各サンプルの近傍においてロジスティック回帰モデルで近似し、線形モデルの回帰係数を可視化するものである。

3.4. 糸球体内部構造領域の抽出手法の開発と評価(H30年度)

前述の、CNNを用いたクラス分類によって、比較的粗大な病変であるCrescentとSclerosis / Collapsing Obsoleteを比較的高い精度(ROC-AUC: 0.91, 0.96)で分類できることが判明した。一方、メサンギウム細胞の増殖を始めとする、糸球体内部のごく狭い領域に注目する必要のある病変の検出精度は十分ではなかった。そこで、糸球体内部の構造を複数の領域として抽出した上で病理所見をスコア化する方法を検討する。本サブセクションでは糸球体内部の領域分割を、画像処理におけるSemantic Segmentationのタスクとしてとらえ、畳み込みエンコーダ・デコーダモデルを適用した結果を報告する。

領域抽出のCNN開発用データセット

北野病院より提供されたPAS染色のWSI 15枚から、手作業によって切り出した糸球体画像404枚を対象とし、サイズを高さ513、幅513ピクセルにリサイズした画像を用いた。抽出対象とする領域クラ

スを、Background(背景)、Glomerulus(糸球体外縁)、Crescent、Collapsing / Sclerosis、Mesangiumとして、研究班の堂本裕加子 特任講師と山口亮平氏、ならびに研究補助者1名で領域のアノテーションを行った。データセットの要約を末尾資料の表10に、領域アノテーションの例を末尾資料の図6に示す。

領域抽出の方法

領域抽出のCNNは、Atrous 畳み込みによりCNNの深い層における位置情報の減損を軽減することと、Atrous 空間ピラミッドプーリングによって多様な視野範囲の特徴を利用することを特徴とするエンコーダモジュールと、シンプルで効果的なデコーダモジュールで構成される、一般画像(PASCAL VOC)で学習済みのDeepLab v3+をベースとして用い、領域アノテーションを付与したデータセットを用いてファインチューニングする方法で行った。画像データ数が十分ではなく、多クラス領域を一度に抽出するモデルを学習することは困難であったため、各領域クラスとbackgroundを二値分類するモデルを作成し、各抽出結果を合成して擬似的に多クラス領域を抽出する方法をとった。抽出方法の概要を末尾資料の図7に示す。

実験設定

15検体から得られた404枚の糸球体画像を病理番号の単位で5分割し、そのうちの3セットを学習に用い、別の1セットを開発セットとして用いて学習の停止を判断し、残りの1セットで評価を行う検証を5回行った。上記5回の試行の精度

を、ピクセル単位のIoU(Intersection over Union)指標で評価し、5回の試行のIoUの平均値を最終評価値とした。

評価結果

末尾資料の表11に糸球体内部構造領域の抽出精度を示す。二クラスの領域抽出精度(IoU)は、Backgroundが0.96、Glomerulusが0.90、Crescentが0.47、Sclerosisが0.46、Mesangiumが0.35であり、微小な構造物ほど抽出精度が低下する傾向にあった。擬似多クラスの領域抽出精度も同様の傾向にあり、5領域のMean IoUは0.59、Frequency Weighted IoUは0.86であった。Segmentationは画素レベルのクラス分類であるが、例えば、Crescentに属する画素数が一定の閾値を超えた場合に、Crescent陽性と判断する、といったような要領で所見項目のクラス分類のタスクとして扱った場合の、分類精度を末尾資料の表12に示す。

4. 光学顕微鏡画像から糸球体領域を検出するWEBアプリケーションシステムの開発(H30年度)

これまでに開発した手法の応用システムとして、1枚のWSIからすべての糸球体領域を検出するWEBアプリケーションシステムを開発した(末尾資料の図8(a)-(d))。このシステムは、1) ユーザーがWEBブラウザ上でWSI(ndpiフォーマット)を選択し、2) 提示された染色の種類に間違いがないかを確認し、3) 染色の種類とWSIをサーバにアップロードする。4) サーバ側システムは、アップロードされたWSIを染色の種類に応じた学習済み

Faster R-CNN の入力として糸球体の検出処理を行い、5) 結果ファイルを返却する、という 5 つのステップによって動作する。結果ファイルには、切り出されたすべての糸球体画像と、WSI をダウンサンプルした画像(糸球体領域が四角形で囲まれる)、ならびに、この 2 種類の画像を表示し、WEB ブラウザ上で簡便な操作で切り出した糸球体を次々に表示する機能を持った HTML ファイルが含まれる。

D. 考察

1. 生検腎病理画像の収集と DB の構築 (H28, H29 年度)

研究参加施設より 112,000 枚の蛍光抗体画像と 7123 枚の生検腎光学顕微鏡画像 (WSI) の提供を受けた。蛍光抗体画像に関しては、IgG による染色画像が最も多く 19,955 枚あり、次いで IgA が 18,065 枚、IgM が 15,517 枚であった。東大病院と北野病院より提供を受けた蛍光抗体画像には、アノテーション情報(メサンギウム領域、糸球体領域、その他領域の 3 種類の領域に対する蛍光強度)が付与されていたが、他の 3 施設よりの画像には付与されていなかった。光学顕微鏡画像 (WSI) は、PAS 染色によるものが最も多く 1,794 枚あり、次いで PAM 染色が 1,527 枚、HE 染色が 994 枚、EM 染色が 709 枚であった。WSI は単一の画像ファイルではなく、ファイルのメタ情報として個人情報記録される場合があり、匿名化する上でこの情報を削除する点に注意が必要であった。また、WSI の画像サイズや光学レンズの観察倍率はファイルのメタ情報として記録されるため、プログラム処理によって

抽出可能であったが、染色の種類は記録されておらず、ファイル名の一部に記載される場合と、画像を目視しなければ判断できない場合、とがあった。多数の画像を目視して染色の種類を判断することは現実的ではないため、メタ情報もしくはファイル名称に染色の種類を判別できる情報を残すことが望まれる。

2. 蛍光抗体法により染色された糸球体病理画像の所見分類 (H28, H30 年度)

ベースラインとなる AlexNet のネットワークに、マルチタスク構造の追加、クラス重みの追加、入力画像の高次元化、スキップ接続の追加を行うことで、それぞれ精度の上乗せ効果が見られた。ここで、糸球体領域の損失関数にクラス重みを追加した場合に、若干の精度低下が見られたが、これはタスク間で重みの正規化を行っていないため、重みが大きくなったメサンギウムとその他領域については学習が進み、相対的に重みが小さくなった糸球体領域については学習が進まなかったためと考えられた。最終的に各領域の蛍光強度の推定精度 (macro-F1) はメサンギウム領域が 75.6%、糸球体領域が 67.9%、その他領域が 60.5%を示した。

3. 光学顕微鏡病理画像の所見分類

3.1. WSI から糸球体を検出する手法の開発 (H29 年度)

本手法による PAS, PAM, MT, Azan の 4 種類の染色画像に対する糸球体の検出精度は F 値でそれぞれ、0.93、0.93、0.90、0.88 であった。PAS、PAM に比べ MT、Azan

の精度が若干低い理由として、後者の染色は硬化糸球体を特徴付けることで糸球体の外見を大きく変えるため、検出するべき糸球体のバリエーションが増えたことが原因と考えられた。本研究手法は、ハンドクラフトな特徴量を用いた先行研究に比べて高い検出精度であることを、より大規模なデータセットを使い示した。また、この結果は同じ CNN ベースの手法を用いた先行研究と同等の精度を示すものであったが、我々の実験では、単一の染色のみならず、複数の染色に対しても同等に高い精度が得られることを示した。

3.2. 糸球体の所見項目の定義と妥当性の評価 (H29, H30 年度)

各所見項目における医師間の一致度 (κ 値) の高低について、判別不能画像を除く、除かないに関わらず平均 κ 値が 0.4 を超えたものとして、Increased Mesangial Matrix 及び Crescent (Fibrous)、Collapsing Obsolete の 3 つの所見があった。ここで、Collapsing Obsolete と Crescent (Fibrous) の二つの所見は、腎機能の予後に直結しやすい所見であり、腎病理に関わる多くの医師が日常臨床において特に重要視していることから、参加医師間で高い一致度を示したものと考えられた。

κ 値が低下する要因の一つとして、医師のキャリアの違いが考えられた。IgA Nephropathy Classification Working Group の 2016 年の報告によると、261 人の IgA 腎症のコホート study において、ワーキンググループの中心となる医師に比べて、そうでない医師は、Mesangial

Hypercellularity、Endocapillary Proliferation の所見に対して、2~3 倍の頻度でそれらの項目を陽性にする傾向が報告されている。本研究に参加した 5 人の医師においても、経験年数などのキャリアが大きく異なっていることから、それが kappa 係数の低下に繋がった可能性が考えられた。

3.3. 糸球体所見を分類する CNN の開発 (H29, H30 年度)

CNN による所見項目のスコア分類精度に関して、Extracellular matrix sclerosis / Collapsing obsolete と Crescent の ROC-AUC は 0.9 を超えた。特に Extracellular matrix sclerosis / Collapsing obsolete は、ROC-AUC: 0.96、感度: 0.72、特異度: 0.94 であり、インフルエンザ抗原検出キットの感度: 0.62、特異度: 0.98 (Ann Intern Med 2012, 156: 500-511) と比較しても遜色なく、実用的なレベルと考えられた。この 2 つの所見項目は、陽性例の比率がそれぞれ 19% と 25% であり陽性例と陰性例の比率のバランスが悪いため、学習を行う上での条件が若干悪いと思われたが、着目すべき病変が比較的粗大であることの利点がそれを上回り、高い分類精度につながったと考えられた。次に分類精度が高かった群は、Mesangial hypercellularity と Increased mesangial matrix であり、ROC-AUC はそれぞれ 0.80 であった。この 2 つの所見項目は、陽性例の比率がそれぞれ 32% と 79% であり、前述の 2 項目と比較しても学習の条件が良いにもかかわらず、着目すべき病変の部位が微細であ

ることのデメリットがそれを上回り、分類精度がそれほど上がらなかったと考えられた。最後に、Adhesion と Endocapillary Proliferation の2つの所見項目の ROC-AUC がそれぞれ 0.75 と 0.77 であり、前述の4項目と比べても分類精度が悪かった。特に、Endocapillary Proliferation は陽性例の比率が7%と陰性例との比率のバランスが悪いことに加え、着目すべき病変の部位が極めて微細であることが低い分類精度の原因であると考えられた。

CNN 着目点の可視化に関して、Grad-Cam と LIME による例を示した。全体的な傾向として、CNN が陽性と判断したケース (TP と FP) は着目点が比較的狭い範囲に収まっているが、陰性と判断したケース (FN と TN) は着目点幅広い範囲に散らばる傾向にあった。また、Grad-Cam と LIME の示す着目点は、概ね同様の傾向にあると思われたが、Grad-Cam のほうが直感的に理解しやすい可視化手法と思われた。

次に、所見項目を個別に検討する。Mesangial hypercellularity (図5 a) の FP 例 (左) に関して、正解となるスコアリング基準には、血管極周囲のメサンギウム細胞増殖は除外するという条件をつけているが、CNN はそのような知識を持たないため、血管極周囲のメサンギウム細胞増殖に着目して陽性と判断したことが FP の原因と考えられた。次に、Increased mesangial matrix (図5 b) の FP 例 (右) に関して、CNN の注視点は実際には Sclerosis を生じている領域であるが、これを誤って Increased mesangial matrix と判断していると考えられた。

Mesangial matrix の増加の後に Sclerosis が生じる場合があることを考えると、見た目的な特徴は同質であり、人間にとっても判断が難しいケースであると考えられた。このケースは該当しなかったが、CNN の出力する確率値が 0.5 に近いものは分類が難しい例であると考えられるため、確率値もユーザーに提示することで、意思決定の一助になると思われた。最後に、Extracellular Matrix Sclerosis / Collapsing Obsolete (図5 d) の FP 例 (左・右) に関して、これはいずれも Crescent の領域を Extracellular Matrix Sclerosis / Collapsing Obsolete と誤って判断したと考えられた。この原因として、Extracellular Matrix Sclerosis / Collapsing Obsolete と Crescent が共起する頻度が高かったことが原因と考えられた。このようなことを防ぐためには、ネットワークを Extracellular matrix sclerosis / Collapsing obsolete と Crescent を同時に学習するマルチタスクと構造として共起頻度を考慮することで、このようなケースを減少させることが可能であると思われた。以上の検討のように、CNN の着目点を可視化し、どのような理由で誤りが生じたかを検討することで、さらなる精度の向上に貢献すると考えられた。

3.4. 糸球体内部構造領域の抽出手法の開発と評価 (H30 年度)

まず、セグメンテーションの精度に関して、今回採用した方法は、2クラスのセグメンテーション結果を合成して擬似的に多クラスのセグメンテーションを行な

っているため、各クラスを塗り重ねる順番によって多クラス分類の精度が変わる可能性が考えられた。次に、セグメンテーションの結果をクラス分類として扱った場合の精度について、既述の糸球体所見のクラス分類タスクと比較して、Crescent では同等もしくは若干の精度向上、Extracellular matrix sclerosis / Collapsing obsolete では精度が劣っていると思われた。ただし、データセットに含まれる枚数(約8000枚と約400枚)の違いがあり、またアノテーションに要する時間も異なるため、現時点でどちらの手法が良いかの比較は難しい。最後に、今後の展望であるが、今回の実験で用意した領域アノテーション付き画像の数は400枚と少なく、これを更に増やすことで、精度の向上が期待されるためこれを目指す。また、例えば、正常糸球体領域と Crescent や Extracellular matrix sclerosis / Collapsing obsolete との面積比が、腎機能の予後を予測する上で良い特徴量となるなど、画像から得られた情報を予後予測モデルの説明変数に用いることで予測の精度が向上する可能性が考えられ、これを検証するとともに、より精度の高い糸球体内部構造の領域抽出を目指す。

4. 光学顕微鏡画像から糸球体領域を検出する WEB アプリケーションシステムの開発 (H30 年度)

本研究班で開発した光学顕微鏡画像 (WSI) から糸球体を切り出す手法を用いて、切り出した結果を返却する WEB アプリケーションソフトウェアを開発した。

現在のところは、東大病院のネットワーク内のみで公開であるが、今後はインターネットに向けた公開を行う予定である。WSI から全ての糸球体を切り出す計算機負荷は高く、GPU を用いても数分程度時間を要するため、複数のリクエストを同時に処理する上での問題が生じるため、ユーザーからのリクエストに対してリクエストキーを発行し、この情報をあとから入力することで、非同期で結果を返却する実装とした。現在のところは糸球体を切り出す処理のみの実装であるが、今後は切り出した糸球体を、各所見項目のスコア进行分类するネットワークで処理した上で、そのスコア情報等を含めてレポートとして返却することを検討する。

E. 結論

以下、本研究班の研究結果を総括する。

- 1) 生検腎病理画像の収集に関して、研究参加施設より 112,210 枚の蛍光抗体画像と 7,123 枚の光学顕微鏡画像 (WSI) の提供を受けデータベースを構築した。WSI ファイルのメタ情報として個人情報情報が記録される場合があり、匿名化する上でこの情報を削除する点に注意が必要であった。また、WSI ファイルのメタ情報に染色の種類は記録されておらず、ファイル名の一部に記載される場合と、画像を目視しなければ判断できない場合とがあった。多数の画像を目視して染色の種類を判断することは現実的ではないため、メタ情報もしくはファイル名称に染色の種類を判別できる情報を残すことが望まれる。
- 2) 蛍光抗体糸球体病理画像の所見分類

に関して、IgG 蛍光抗体により染色された糸球体画像の蛍光強度の推定精度(F1)はメサングウム領域が 75.6%、係蹄壁領域が 67.9%、その他領域が 60.5%であった。

3) 光学顕微鏡糸球体病理画像の所見分類に関し、次の4つのサブセクションからなる研究を行った。

3-1) 1枚のWSIを入力として糸球体を検出する手法に関し、物体検出を行う Faster R-CNN を用いて高解像度のWSIから微小な糸球体を検出する手法を開発した。PAS、PAM、MT、Azan の各染色画像、計 800 枚のWSIに含まれる約 33,000 個の糸球体領域を人手により注釈付けし手法の評価を行ったところ、検出精度(F1)はそれぞれ、0.93、0.93、0.90、0.88 であり、先行研究より高い精度を達成した。

3-2) 糸球体の所見項目の定義と妥当性の評価に関して、研究班を構成する5名の腎病理の専門家の協議のもと、PAS 染色における12の所見項目を作成した。5人の専門家による所見一致度(κ 値)は0.04~0.43 であり、所見項目によるばらつきがみられた。より高い一致度を得るためには、スコアリング基準並びにスコアリングを行う上でのフローの更なる改定が必要と考えられた。

3-3) 糸球体所見を分類する CNN の開発に関し、約 8000 枚の糸球体画像に対して前項で開発した基準に基づくアノテーションを行い、学習済みの VGG16 をファインチューニングする方法で精度の評価を行ったところ、Extracellular matrix sclerosis / Collapsing obsolete と Crescent の ROC-AUC は 0.90 を超え、特

に Extracellular matrix sclerosis / Collapsing obsolete はインフルエンザ抗原検出キットの精度と比べても遜色がなく、実用的なレベルと考えられた。しかしながら、他の所見項目は十分な精度が得られていないため、更なる検討が必要である。

3-4) 前項の課題の一つである、糸球体内部のごく狭い領域に注目する必要がある所見の分類精度の向上を目的として、糸球体内部の構造を複数の領域として抽出した上で病理所見をスコア化する方法を検討した。糸球体所見のクラス分類タスクと比較して、Crescent では同等もしくは若干の精度向上、Extracellular matrix sclerosis / Collapsing obsolete では精度が劣っていると思われる。ただし、アノテーション付きデータセット枚数の違いがあり、またアノテーションに要する時間も異なるため、現時点でどちらの手法が良いかの比較は難しい。今後は領域アノテーション付き画像の枚数を増やし再度実験を行うこと予定している。

4) 光学顕微鏡画像から糸球体領域を検出する WEB アプリケーションシステムに関し、1枚のWSIを入力として切り出した糸球体画像を結果として返却する WEB アプリケーションシステムを開発し東大病院内のネットワークで公開した。今後は切り出した糸球体を、各所見項目のスコアを分類するネットワークで処理した上で、そのスコア情報等を含めてレポートとして返却すること、ならびにインターネットに向けた公開を予定している。

F. 健康危険情報

特になし。

G. 研究発表

1. 論文発表

1. 河添悦昌, 医療における人工知能技術の応用, 医歯薬出版株式会社 医学のあゆみ 264 巻 3 号 p260, 2017
2. 大江和彦: 医療における人工知能の活用と将来展望, 日本腎臓学会誌, 59 巻 7 号, 1060-1063, 2017.
3. 河添悦昌, 大江和彦, AI と ICT が変える医療, 科学評論社 腎臓内科・泌尿器科 7 巻 2 号, 183-187, 2018.
4. Kawazoe Y, Shimamoto K, Yamaguchi R, Shintani-Domoto Y, Uozaki H, Fukayama M, Ohe K. Faster R-CNN-Based Glomerular Detection in Multistained Human Whole Slide Images. Journal of Imaging. 2018; 4(7):91, p1-19.

2. 学会発表

1. 山口亮平, 嶋本公徳, 河添悦昌, 堂本裕加子, 宇於崎宏, 大江和彦, 腎臓糸球体病理画像の Deep Learning による所見分類手法の検討, 第 3 回 JAMI & JSAI AIM 合同研究会, 神奈川県, マホロバマインズ三浦, 2017 年 03 月.
2. ICT と AI がもたらす医療, 口頭, 大江和彦, ICT と腎臓病学: past, present and future (日本医療情報学会合同企画) シンポジウム 6 第 60 回日本腎臓学会学術総会, 2017/5/26
3. ビッグデータと AI がもたらすこれからの

医療, 口頭, 大江和彦, エジソンの会 第 18 回会合, 2018/2/20

4. ビッグデータと AI で医療は変わるだろうか, 口頭, 大江和彦, 次世代産業ナビゲーターズフォーラム, 2018/3/13
5. IoT・ビッグデータ・AI がもたらすこれからの医療, 口頭, 大江和彦, Medical Information Conference in Kawasaki, 2018/3/14
6. 山口 亮平, 河添 悦昌, 嶋本 公徳, 堂本 裕加子, 宇於崎 宏, 清水 章, 長田 道夫, 大江 和彦. AI による画像診断に向けた腎病理糸球体画像の所見一致度の評価. 第 38 回医療情報学連合大会, 2018 年 11 月 25 日.
7. 河添 悦昌, 人工知能技術を用いた腎病理の診断支援に向けた取り組み. 第 154 回日本医学会シンポジウム, 2019 年 1 月 16 日.
8. 中村 一成, 河添 悦昌, 山口 亮平, 堂本 裕加子, 大江 和彦, 深層学習を用いた蛍光抗体法による糸球体病理画像の所見分類. 第 1 回日本メディカル AI 学会学術集会, 2019 年 1 月 26 日.
9. 河添 悦昌, 嶋本 公徳, 山口 亮平, 篠原 恵美子, 堂本 裕加子, 大江 和彦. 腎病理画像の自動診断に向けた畳み込みエンコーダ・デコーダモデルによる糸球体内部構造の領域抽出. 第 1 回日本メディカル AI 学会学術集会, 2019 年 1 月 26 日.
10. 山口亮平, 河添 悦昌, 嶋本公徳, 篠原恵美子, 堂本裕加子, 宇於崎宏, 清水章, 長田道夫, 塚本達雄, 大江和彦. AI 画像診断に向けた、糸球体

PAS 染色画像の医師間所見一致度の
評価. 第 62 回日本腎臓学会総会、
2019 年 6 月 22 日. (採択済)

11. Why current Deep Learning can't
diagnose the glomerular images
correctly? -From perspective of
characteristics of glomerular
images-. Ryohei Yamaguchi,
Yoshimasa Kawazoe, Kiminori
Shimamoto, Emiko Shinohara, Yukako

Shintani-Domoto, Hiroshi Uozaki,
Akira Shimizu, Michio Nagata,
Tatsuo Tsukamoto, Kazuhiko Ohe.
The Secretariat of The 62nd
Annual Meeting of the Japanese
Society of Nephrology. Jun 22,
2019. (Accepted)

H. 知的財産権の出願・登録状況
なし