

腎臓系球体病理画像の Deep Learning による所見分類手法の検討

The classification of renal biopsy images by deep learning

山口亮平¹ 嶋本公德² 河添悦昌^{1,2} 堂本裕加子³ 宇於崎宏⁴ 大江和彦^{1,2}

Ryohei Yamaguchi¹, Kiminori Shimamoto², Yoshimasa Kawazoe^{1,2},
Yukako Domoto³, Hiroshi Uozaki⁴, Kazuhiko Ohe^{1,2}

¹ 東京大学大学院 医学系研究科 医療情報学分野

¹ Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo

² 東京大学医学部附属病院 企画情報運営部

² Department of Healthcare Information Management, The University of Tokyo Hospital

³ 東京大学大学院医学系研究科人体病理学・病理診断学分野

³ Department of Pathology, Graduate School of Medicine, The University of Tokyo

⁴ 帝京大学医学部病理学講座

⁴ Department of Pathology, Teikyo University School of Medicine

Abstract: It is often difficult to diagnose kidney diseases only by noninvasive examination such as blood collection and ultrasound examination. When we make the accurate diagnoses, understand the pathological state and predict prognosis of the disease, pathological examination is required. Digitized pathological images have been more available from the view of preservation of images and secondary use for research, and it is hoped that making use of these digitized images will aid diagnosis assistance by computer and acquisition of new knowledge on pathological findings. In this paper we report the classification method and classification accuracy of pathological findings using Convolutional Neural Network of renal glomerular images stained by fluorescent antibody method.

1. 背景

腎生検とは、腎臓の組織を採取し、その病理スライドを作成し、顕微鏡でその病態を確認する検査である。腎臓疾患は尿検査や採血検査、超音波検査などの非侵襲的検査のみでは診断に至らない事が多く、腎生検による病理画像を確認して初めて、正確な診断、治療方針決定、予後予測判定ができる。実際、腎生検施行により 92%の患者に診断がついたという報告[1]や、腎生検をすることで 34%の患者で治療方針が変更になったという報告[2]がある。

今回我々は腎臓の中で、体内の毒素を排出するのに中心的な役割を果たす糸球体と呼ばれる部分に注目し、腎臓系球体病理画像のうち、蛍光抗体法によって撮影された画像の分類を Deep Learning を用いて行った。蛍光抗体法とは、1940 年以降 coon らに

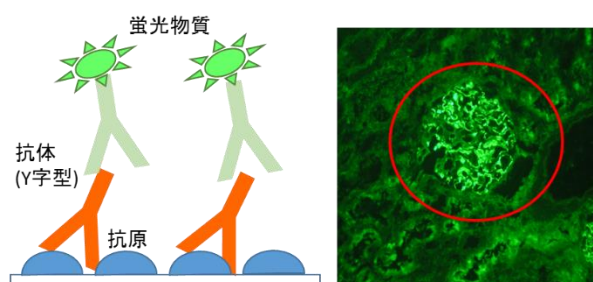


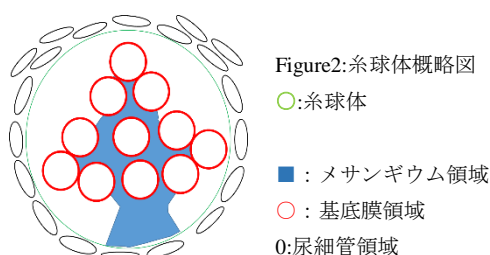
Figure1: 蛍光抗体法概略図及び写真例

左: 蛍光抗体法概略図。抗原に反応して蛍光する物質を加える事で、抗原と呼ばれる物質が存在する場所を光らせる事ができる。

右: 実際に蛍光抗体法で腎臓の糸球体を染色した画像。糸球体の部分が強く光っているのが分かる。(糸球体は赤丸で囲った部分)

よって確立された手法で、抗原(病気の誘因となることが多い物質)に対して抗体(抗原に対する反応する物質)を反応させ、その部位を蛍光させる手法である。Figure1 にその概略及び写真を示す。

この蛍光抗体法は、腎臓病理の分野で広く使われており、その蛍光パターンを判断する事で、腎臓病の診断に寄与するものである。蛍光パターンには、蛍光している領域「メサンギウム領域、基底膜領域、尿細管領域」の3パターンがあり、それぞれの領域における蛍光の強さが「(-)、(±)、(+)、(++)、(+++)」の5パターン存在する。Figure2 に、領域パターンの概略図を示す。



糸球体の蛍光抗体画像には、基本的に一枚の画像ファイルに1つの糸球体が入っていることが多いが、その中には複数の糸球体が入っている事もしばしばみられる(Figure3)。その為、領域抽出を行った方がよいかどうかについて考察した。

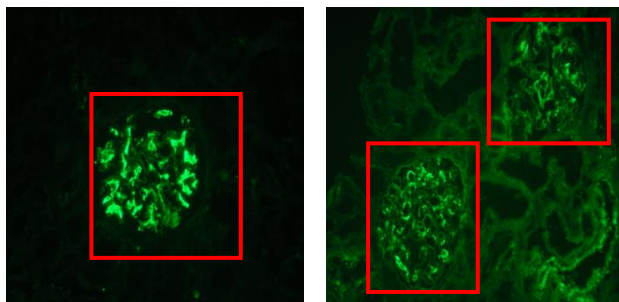


Figure3:糸球体画像例
 左:1画像内に糸球体が1つ含まれる。
 右:1画像内に糸球体が2つ含まれる。

今回は、腎臓糸球体に対してIgGという物質で染色した蛍光抗体法に対して、メサンギウム領域の蛍光強度を5段階分類させるタスクを設定した。Deep Learningのうち、Fukushimaらのネオコグニトロン[5]をルーツに持つ、Convolutional Neural Network(以下、CNN)を用いた。

2. 目的

腎臓糸球体の蛍光抗体病理画像に対して、CNNによる分類タスクを行う際に、領域抽出が画像認識精度に与える影響を評価する。また、ハイパーパラメーターの一つとして、ドロップアウト率を調整することで、ネットワークの性能に与える影響を評価する。

3. 実験手法

3.1 データセット

東大病院にて2001年から2016年までに行われた腎生検画像のうち蛍光抗体法の画像を利用した。3903枚のIgG蛍光抗体画像(画像サイズは様々)のうち1874枚の画像に対して糸球体の領域抽出の後に切り出しを手動で行った。切り出しの際は糸球体が一個収まるような形で、長方形で切り出した。

切り出し前の画像、切り出し後の画像、ともに200*200ピクセルにリサイズしたものをCNNの入力とした。各種データセットのクラス分布はTable1に示すとおりである。

Table1:使用したデータセットの分布

所見	(-)	(±)	(+)	(++)	(+++)	合計
個数	1135	314	313	108	4	1874

3.2 CNNの構造

今回使用するネットワークとしては、AlexNetを参考にした。AlexNetは2012年、ImageNet Large Scale Visual Recognition Challenge(ILSVRC)2012にて物体の多クラス(1000クラス)分類にて優勝したネットワーク[7]である。AlexNetをベースに、今回の実験に使用したネットワークFigure4に示す。

演算上のメモリの関係で、畳み込みフィルタの枚数を全てオリジナルのAlexNetの半数にした。活性化関数には畳み込み層(Conv)/全結合層(Fc)共にReLUを用い、最終層はsoftmax cross-entropyを用いた多クラス分類を行った。Fc層には、ドロップアウトを加えた。また、AlexNetではLocal Response Normalization層を採用しているが、今回は代わりにBatch Normalizationを使用した。Batch Normalizationは、バッチ毎に平均が0、分散が1になるように正規化を行う事で、学習効率を上げ、さらには初期値依存性を軽減するとされている[6]。今回ハイパーパラメーターとしてのドロップアウト率の調整による精度比較を行ったが、その際に初期値による精度変

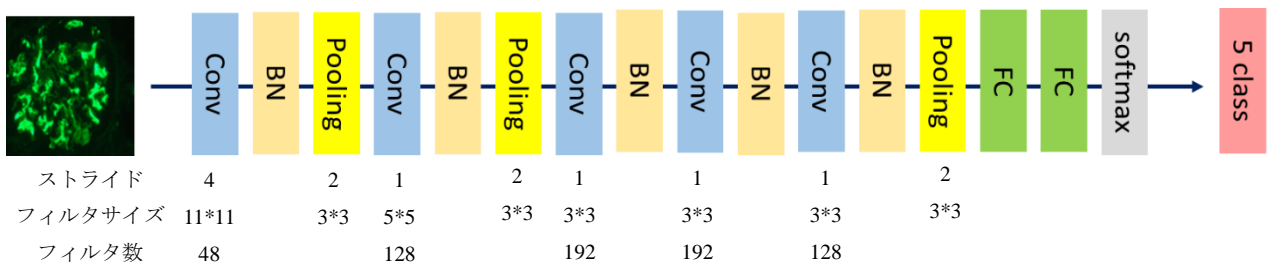


Figure 4:今回使用した CNN の構造

Conv: convolutional layer(ReLU) BN: Batch Normalization layer

Pooling: Pooling layer(max pooling) FC: Fully Connected layer(ReLU)

化の影響を極力避けるため、Batch Normalization 層を導入している。

学習の際は、batch size=100 のミニバッチ学習にて行い、最適化手法としては、Adam optimization を使用した。

3.3 実験設定

3.3.1 領域抽出が精度に与える影響の評価

元々のデータセットから 1874 枚をランダムに選び、領域抽出ならびに切り出しを行った。その 1874 枚の画像に対して、切り出す前の元画像(non-cropped images)、および、切り出した後の画像(cropped images)それぞれに対して、train data と test data を 4:1 とし分割し、5 分割交差検定を行い、accuracy 及び画像一枚あたりの loss の平均値を評価指標とした。

3.3.2 Dropout 率の変化による精度への影響の評価

Fc 層に Dropout 率をそれぞれ 20%,40%,60%,80% に設定し、そのネットワークへの性能の影響を評価した。train data と test data を 4:1 とし分割し、5 分割交差検定を行い、accuracy 及び画像一枚あたりの loss の平均値を評価指標とした。

4. 結果

4.1 領域抽出が精度に与える影響の評価結果

train loss および test loss のグラフを Figure5 に示す。

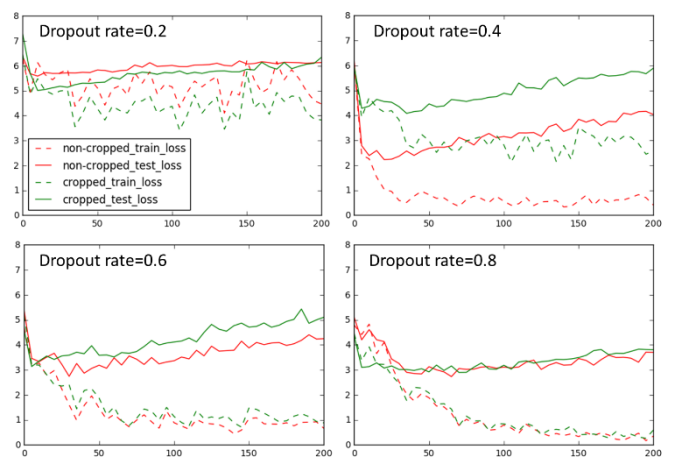


Figure5:領域抽出が精度に与える影響の実験結果

縦軸は 1 入力あたりの loss の値を、横軸は epoch 数を表す。

Loss の値は、dropout rate が 20% のもので、cropped のほうが non-cropped に比べて test loss は良い値を示した。一方で、Dropout rate が 40% もしくは 60% のものでは、むしろ逆に non-cropped 群の方が良い test loss の値を取った。Dropout rate が 80% のものでは、cropped と non-cropped 群では殆ど差は見られなかった。

4.2 Dropout 率の変化による精度への影響の評価結果

Test loss のグラフを Figure6 に示す。

Non-cropped images では、dropout rate が 40%(Figure6 緑線)のものが loss の値としてはもっとも良い結果を示した。Cropped images では、dropout rate が 80%(Figure6 紫線)のものが最も良い結果を出した。200 epoch 内での accuracy はそれぞれ、non-cropped で最大値 69.8%(dropout80%),cropped での最

大値 65.1%(dropout80%)であった。

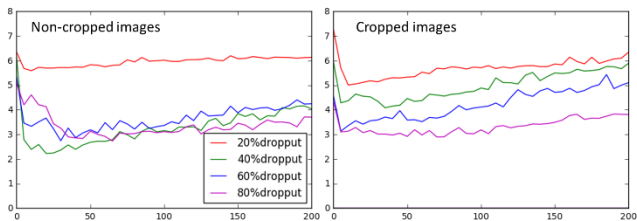


Figure6:Dropout rate の影響の実験結果
縦軸は 1 入力あたりの loss の値を、横軸は epoch 数を表す。

5. 考察

領域抽出をした方がよいかという問題に対しては、今回はむしろ領域抽出をした方が、精度が落ちる傾向にあった。本来、画像認識において、顔認識などでそうであるが、領域を抽出してから画像解析をかけたほうが良いとされる。近年提唱されている R-CNN などの手法は、その考えに基づいている。しかしながら、今回の実験において領域抽出をした画像での精度が低い傾向にあった。その理由として、以下二つのような事が考えられる。

一つの理由としては、領域抽出の範囲の問題が挙げられる。通常、医師は蛍光抗体法の所見をつける際に、その光り方を評価し(-)から(+++)の 5 段階評価を付ける。その際、その 5 段階評価は、周囲とのコントラストを加味して医師が判断している。糸球体とその周囲とのコントラストの違いによる、所見の例を Figure7 に示す。

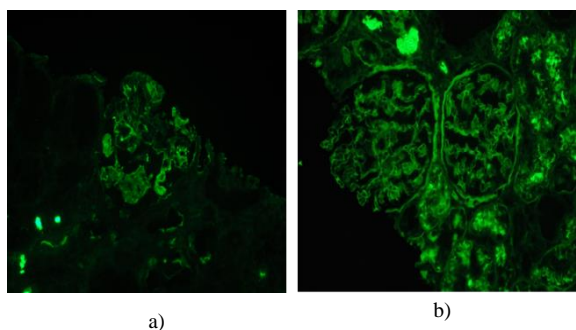


Figure 7:所見付けにコントラストが重要な 2 例
a):メサンギウム領域蛍光 : (+)
b):メサンギウム領域蛍光 : (-)

Figure7 a)の画像は中央の糸球体部はかなり暗いものの周囲とのコントラストを考えると、所見としては(+),つまり 5 段階評価の 3 番目となる。一方で、Figure7 b)の画像は中央の糸球体部はそれなりに明る

いものの、周囲が明るい事を考慮し、「糸球体は相対的に光っていない」として(-)となる。今回、領域抽出を行った画像に関しては、コントラストを判断する為に十分な糸球体周囲の情報が欠損していた為に識別精度が落ちた可能性が考えられる。

二つ目の理由としては、一枚の画像内に Figure8 に示すような発光程度の異なる糸球体が入っていた可能性が考えられる。

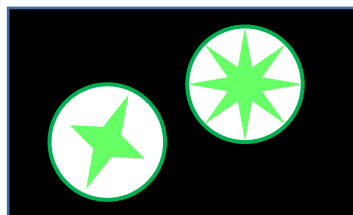


Figure8:発光度の違う糸球体が入っている画像。

上記 Figure8 では、蛍光程度が異なる二つの糸球体が入っており、画像全体に対して蛍光の強さ(5 クラス)が付けられている。この場合、そのどちらの糸球体を取っても、その発光程度とつけられた所見が合致しない可能性がある。基本的に糸球体の蛍光抗体法の発光程度が一人の患者病理検体内であればそれほど異なる事はあまりないが、この可能性が領域抽出による精度低下に寄与した可能性は否定できない。

Dropout rate に関しては、今回は Dropout rate が高いものほど良いスコアを出した傾向にあった。Dropout はニューロンをランダムに落とすことで、その汎化性能が上がると思われる[3]。最近使われているネットワークにおいて一般的に Dropout rate は 50% とするものが多いが、今回は dropout 率がそれを大きく上回る 80%とした時に、最も test loss が良い結果を示した。実際、認識精度を保ちつつ CNN のパラメーター削減に取り組んだ論文[4]では、CNN の中で層によっては約 80%のニューロンを削減しても、精度に大きな影響は出なかったとされる。Dropout rate が 80%で最も良い結果を示した事は、ネットワークが全体的に過学習であった事が影響していると思われる。

なお、全体的に今回の CNN は過学習になっている理由は、以下の二つの理由が考えられる。①用いた画像の枚数が少なかった事、②AlexNet が今回のタスクには表現力が強すぎた事、である。AlexNet は、ImageNet の画像の分類にて 2012 年の ILSVRC にて優勝した際、数十万枚から百万枚単位の画像を学習

させていることから考えても、今回対象とした画像の枚数はかなり少なかったと言える。その為、3903枚の non-cropped image を用いた追加実験を行った。データの分布を Table2 に、結果を Figure9 に示す。

Table2 : データセットの分布

所見	(-)	(±)	(+)	(++)	(+++)	合計
個数	1135	314	313	108	4	1874
個数	2419	678	586	213	7	3903

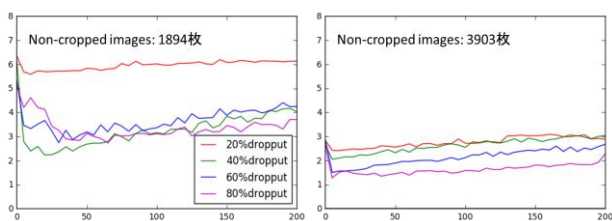


Figure9: データ数影響の実験結果

縦軸は 1 入力あたりの loss の値を、横軸は epoch 数を表す。

Figure9 に示されるように、明らかに 1 入力あたりの test loss の値はデータ数を増やすことで大幅に低下しており、ネットワークの性能は向上していると考えられる。また、過学習の程度に関しても改善していると思われる。このことから、今後はさらにデータ数を増やして行くことが必要と考えられる。

6. 結論

今回 AlexNet をベースとした CNN を構築し、腎生検蛍光抗体法による画像所見の識別精度を評価した。その中では、Dropout rate が 80% のものが最も良い結果を出した。領域抽出を行った画像群の方が、行わない画像群に比べて、却って精度が低下する傾向にあった。これには、コントラスト情報が領域抽出画像に十分に含まれていなかった可能性、また、一枚の画像内に異なる蛍光程度の糸球体が含まれていた可能性、などが理由として考えられた。今回の AlexNet を利用した学習は全体的に過学習傾向であり、今後さらに大量のデータセットを用意し、その精度を上げていく必要があると思われた。

参考文献

[1] Pfister, M. et al., 1999. Judgment analysis in clinical nephrology. American journal of kidney diseases: the official journal of the National Kidney Foundation, 34(3), pp.569-75.

[2] Kitterer, D. et al., 2015. Diagnostic impact of percutaneous renal biopsy. Clinical Nephrology, 84(6), pp.311-322

[3] Hinton, G., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. , Journal of Machine Learning Research 15 (2014) 1929-1958, pp.1929-1958

[4] Han, S., Mao, H. & Dally, W.J., 2016. Deep Compression - Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. ICLR,2016

[5] Fukushima, K. & Miyake, S., 1982. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. Pattern Recognition, 15(6), pp.455-469.

[6] Szegedy, C. & Com, S.G., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML,2015.

[7] Krizhevsky, A. & Hinton, G.E., ImageNet Classification with Deep Convolutional Neural Networks. NIPS,2012.