様式６

申請先学部　**医**学部　採択番号　No.1

令和２年度大阪大学未来基金「学部学生による自主研究奨励事業」研究成果報告書

ふりがな 氏名	あべまさとし 安部政俊	学部 学科	医学部医学科	学年	２年
ふりがな 共同 研究者氏名	うめだけんたろう 梅田健太郎	学部 学科	医学部医学科	学年	２年
	とみたいつき 冨田一輝		医学部医学科		２年
					年

アドバイザー教員 氏名	三木 篤也	所属	視覚先端医学寄附講座

研究課題名	網膜眼底画像と深層学習を用いた緑内障を始めとする眼科領域疾患の検出
研究成果の概要	研究目的、研究計画、研究方法、研究経過、研究成果等について記述すること。必要に応じて用紙を追加してもよい。（先行する研究を引用する場合は、「阪大生のためのアカデミックライティング入門」に従い、盗作剽窃にならないように引用部分を明示し文末に参考文献リストをつけること。）

1.研究目的

緑内障は、世界で最も一般的な失明の原因の一つで、早期治療できなければ視野狭窄により視力や生活の質の低下は免れない重大な疾患である。加えて眼科専門医の不足は顕在化しており、弘前市立病院の眼科が今年四月から医師不足により休診に追い込まれるなど、近年の眼科医不足や地方の医療格差により十分な診断が全国的に行われているとは言い難い。画像解析分野での深層学習の発展はめざましく、その影響は医療分野にも及んでいる。特定の分野では、専門医による診断精度を上回る[1][2]までに発展している。また去年二月には広角眼底画像を学習した深層学習を用いた増殖糖尿病網膜症の診断が世界で初めて行われ、精度 97%を達成した。[3]

今回我々は、一般に公開されている眼底画像のデータセットを用いて緑内障の診断精度が一般臨床医と同等以上となることを目指す。さらに本研究の特色として、深層学習の分野で最先端のモデルや手法を眼底画像に適用し、どの手法が使用するデータセットと疾患の検出に最適かを検証することも目標としている。深層学習を用いた網膜眼底画像での緑内障診断の精度が一般臨床医と同等以上となれば、たとえ医師が少なくてもコンピューターさえあれば日本中どこでも質の高い診断を受けられるようになり、他の疾患へと応用していくことで日本の医療格差問題の解決の一助となるであろうと予想する。

2．研究計画

一般公開されているデータセットを用い、眼底写真の直接学習により緑内障と非緑内障を分類するタスクを機械学習を用いて行う。言語は「python」を用いる。使用するモデルとして畳み込みニューラルネットワークベース（ResNeXt,ResNeSt,EfficientNet など）のモデルと attention 機構を主としたモデル（ViT）試す。

まず、先行研究や眼底画像を用いた機械学習コンペティションの手法などを調べ、使用の可能性のある手法を列挙する。次に、疾患の予備知識とデータセットの分布を確認したのち、上記機械学習を用いるベースラインを作る。精度の評価は AUC を用い、予めデータセットは訓練用とテスト用に分

割する。

機械学習を用いる際に問題になる過学習とは、訓練データに過剰に適合してしまい、テストデータに対して正しい予測ができない状態のことである。本研究では、過学習を避けモデルの汎化能力を担保するために交差検証を行い、検証用データに対する精度とテスト用データに対する精度に乖離が生じず、過学習をしないようにベースラインの改善を進めていく。

3.研究方法

3-1:データセット

使用するデータセットとして、機械学習、データ分析コンペティションのためのプラットフォームである kaggle 上にあった 4 つのデータセットを統合して用いた。

1 つ目のデータセットは[4]から入手したもので緑内障 460 枚、非緑内障 511 枚であった。

2 つ目のデータセットは[5]から入手したもので、5000 人の患者の左右の目からのカラー眼底画像と医師がつけた診断キーワード（緑内障疑い、正常　など）が紐づけられたものである。正常、糖尿病、緑内障、白内障、AMD、高血圧、近視、その他の疾患、の 8 つのラベルがつけられていた。

3 つ目のデータセットは[6]から入手したもので緑内障 1711 枚、非緑内障 3143 枚であった。

4 つ目のデータセットは[7]から入手したもので緑内障 188 枚、非緑内障 420 枚であった。

2 つ目のデータセットは、全データに対する緑内障画像の割合が低いためランダムにサンプリングしたものを用いた。これによりデータの極端な不均衡性の是正を図り、緑内障 168 枚、非緑内障 482 枚とした。これによりサンプリングした 2 つ目のデータセットと残り 3 つのデータセットを統合した結果、緑内障 2428 枚、非緑内障 4525 枚となった。

3-2:交差検証

全データのうち、30%ないし 90%を test データ、残りを train/validation データとし、train/validation データを Stratify K-fold(k=4)によって 4 分割した。test データと train/validation データの緑内障：非緑内障の割合は等しくなるようにした。validation データに対する精度の改善の有無をみながら各条件を変えた。test データに対する予測は交差検証によって生じる 4 つモデルの予測値を平均したものを用いた。

3-2:画像の前処理

一部の眼底画像は余白の黒い部分が多かったため、これを除くために閾値を用いて余白の除去を試みた。その後、取りきれない余白の除去のため画像の中央の何%かだけを切り出した。また、実験によっては先行研究[8]により提示されている手法を用いて画像中の血管を消去する前処理をすることもあった。上記の処理を終えたのち、深層学習のモデルに入力するため縦横それぞれ 512 ピクセルに resize した。モデルによっては縦横のピクセルが 512 では使用できなかったものがあったのでその場合はモデル固有の画像の大きさに揃えた。

3-3:モデル

深層学習用のフレームワークとして広く使用されており、かつ前述の kaggle での使用者も多い pytorch を用いた。モデルの実装は自分では行わず、timm というライブラリにて既に実装されているものを用いた。全て imagenet 用いたモデルは EfficientNetb0/b2/b6, ResNeXt50_32x4d, ResNeSt50d, deit_base_patch16_224 であり、ハイパーパラメーターの探索にはモデルのパラメーター数が少なく高速に作動する EfficientNetb0 を使用した。

EfficientNet[9]は 2019 年に発表されたモデルで、深層学習モデルの深さ(層数)と

広さ（1 層あたりのパラメーターの数）と入力画像の大きさをバランス良く調整されているモデルで、高速に作動するモデルの中では imagenet を含む複数のデータセットでの精度が良いとされるモデルである。特徴として図のように imagenet の精度が同等のモデルに比べてパラメーター数が少ない。ResNeXt50_32x4d、ResNeSt50d は EfficientNet が発表される前後のモデルであるが、比較のために試行した。

図 2 : [9]より引用。従来のモデル（黒点線）よりもパラメーター数が少ない（赤点線）ことがわかる

ViT(Vision Transfomer)は 2020 年に発表されたモデルで、図のように画像を分割したのち分割したそれぞれを固定長のベクトルにして attention 機構を備えた Transofomer に各ベクトルを入力するというものである。従来モデルとの違いとして畳み込みを主な演算に使用していない点が挙げられる。deit_base_patch16_224 は図のように既存の ViT より高速に作動する。

図 3：[10]より引用。自然言語分野にてよく利用される Transfomer の構造が使用されている。

図 4：[11]より引用。速度と精度はトレードオフであると考えられているが、既存のモデルよりも両方優れているとされている。

3-4:画像拡張

画像拡張は画像認識の問題に頻繁に用いられるもので、画像に摂動（上下反転、左右反転）を加えることで使用すればデータ数を疑似的に増やすことを目指したものである。実験ごとに使用する手法は変えたが主には上下反転、左右反転、Augmix を用いた。Augmix[12]は 2020 年に発表された手法で図のように複数の画像拡張を合成したものである。どれだけ画像拡張を合成するかを 0~10 の 11 段階で変更できるようなっているので摂動の度合いをこれによって適宜変更した。

図 5：[12]より引用。複数の画像拡張を組み合わせている

3-6:正解ラベルについて

緑内障を 1、非緑内障を 0 として学習させた。ここでモデルが高い確信度を持っているのにもかかわらず誤った予測を出力してしまうのを防ぐため緑内障を 1*(1-α)、非緑内障をα/2 をとする label_smoothing[13]を用いた。

3-7:予測の後処理等

半教師あり学習の手法の一つとして、test データに対する予測を疑似ラベルとして訓練用データ(train/validation)に加える psuedo_labeling を試行した。このとき test データに対する予測が 0.98 以上ないし 0.02 以下のもののみを使用することで訓練用データにノイズが入らないようにした。今回は疑似ラベルは 0 か 1 の hard_label を用いた。また、test 画像に上下反転、左右反転などの摂動を加えたものを学習済みモデルに入力し、出力として得た各予測値を平均してオリジナルの test 画像に対する予測とする TestTimeAugmentation を試行した。

4.結果

以下の図が各条件と実験番号の対応表である。

図 6：各実験の条件とその結果

図 5 にて、validation データに対する評価指標の値を AUC(CV)、test データに対する評価指標の値を AUC(inner_test)と表記している。以下ではこの表記に倣う。全体的に、AUC(CV)と

AUC(inner_test)は相関関係にあると言える。よって、AUC(CV)が改善していくように各種パラメーターを探索していくものとした。
実験 000 をベースラインとし、まずは 1 条件ずつ変化させていった。実験 001,002,003 より、label_smoothing を使用するとαの値によっては AUC(CV)が改善した。実験 004,005 より Augmix の摂動の度合いを大きくしていくと AUC(CV)は改善しなかった。実験 006 より psuedo_labeling をすれば AUC(CV)が改善した。実験 010 より、batch_size を 16 から 64 にすると AUC(CV)がわずかに下がった。しかし、1 回の試行にかかる時間は 5/6 になったので可能であればこれ以降の実験では batch_size=64 としている。モデル間の比較として、実験 015~017,027,028 より EfficientNetb6 が最も AUC(CV)が高かった。しかし、AUC(inner_test)は ResNeSt50d の方が deit_base_patch16_224 は精度を犠牲にせず高速に作動した。最終的なモデルとしては AUC(CV)と速度の両方を考慮して resnest50d とした(実験 029)。それ以外の focal_loss,学習率の scheduler、中央での crop は AUC（CV）の改善に寄与しなかった。
実験 0p0~0p6 では test データの全体に対する割合を 0.3 ではなく 0.9 としているが、これらの結果を見ると psuedo_labeling を複数回繰り返すことによって AUC(CV)は段階的に改善していった。しかし、実験 0p6 をみると、test データに対する精度が低くなったとき（実験 0p5）の予測値を疑似ラベルとして加えているので validation データ、test データに対する精度は実験 0p4 に比べて悪化した。

眼底画像に対して疑似ラベルの使用が有効であることは糖尿病性網膜症の重症度を予測するコンペティションでも挙げられていることであり、先行の知見と一致した。また、段階的に改善することは半教師あり学習を用いた先行研究[14]とも一致している。画像に摂動を加える augmentation について摂動の度合いを上げると精度が下がったのは、図 7 のように摂動を加えた眼底画像が現実世界の画像とかけ離れたものになっているからだと考えられる。

図 7：Augmix の摂動（左から度合い 3,5,10）画像によって組み合わさった摂動は様々だが、左から３枚目はオリジナル（図 1）と大きくかけ離れているのがわかる。

5.展望
疑似ラベルの使用を繰り替えして精度が上昇するのであれば、医師による正確なラベリングが必要な枚数が減少し、多量のラベルなしデータと少量のラベルありデータさえあればモデルを学習させることが可能となる。このことは、医療データはラベリングのコストが他のデータよりかさむという事実を鑑みると、有用であると言える。

緑内障の診断に用いられるのは眼底画像だけでなく、OCT などの系列性を兼ね備えた画像データ、検査によって得られる数値データ、医師による問診等によって得られる自然言語データなどがあり、実際に医師が診断をする場合はこれらを統合的に判断している。今後はこれらのデータを統合的に入力として扱うマルチモーダルなモデルを組めるように今後の学習を進めていきたい。

6.参考文献等

[1]: Mark J.J, Bram G, Carel BH, Thomas T, Clara IS. “Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images”. *IEEE Transactions on Medical Imaging*. 2016 Feb; 35(5): 1273-1284.

[2]: Gopi K, Selvakumar J, Amir H.G, Manikandan R, Simon J.F, Rizwan P. “Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier”. *Expert Systems with Applications*. 2019 Nov; 134(15): 112-119.

[3]: Frank D.Verbraak, Michael D.Abramoff, Gonny C.F. Bausch, Caroline K, Giel Nijpels, Reinier O. Schlingemann, Amber A. van der Heijden. “Diagnostic Accuracy of a Device for the Automated Detection of Diabetic Retinopathy in a Primary Care Setting”. *Diabetes Care*. 2019 Apr; 42(4): 651-656.

[4]: https://www.kaggle.com/himanshuagarwal1998/glaucomadataset
[5]: https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k
[6]: https://www.kaggle.com/sreeharims/glaucoma-dataset
[7]: https://www.kaggle.com/sshikamaru/glaucoma-detection

[8]: 中川俊明　林佳典　畑中裕司　青山陽　水草豊　藤田明宏　加古川正勝　原武史　藤田広志　山本哲也“眼底画像支援診断システムのための血管消去画像を用いた視神経乳頭の自動認識及び疑似立体視画像の生成への応用
“電子情報通信学会論文誌 D　Vol.J89-D,No.11,pp.2491-2501,2006/11/01

[9]:Mingxing Tan, Quoc V. Le”EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”ICML 2019 arXiv:1905.11946v5 [cs.LG] 11 Sep 2020

[10]:Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby ” An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”
arXiv:2010.11929v1 [cs.CV] 22 Oct 2020

[11]:Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, Hervé Jégou “Training data-efficient image transformers & distillation through attention”
arXiv:2012.12877v1 [cs.CV] 23 Dec 2020

[12]:Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, Hervé Jégou “Training data-efficient image transformers & distillation through attention”
arXiv:2012.12877v1 [cs.CV] 23 Dec 2020

[13]: Rafael Müller, Simon Kornblith, Geoffrey Hinton “When Does Label Smoothing Help?”
arXiv:1906.02629v3 [cs.LG] 10 Jun 2020

[14]:Qizhe Xie, Minh-Thang Luong, Eduard Hovy, Quoc V. Le “Self-training with Noisy Student improves ImageNet classification” CVPR 2020 arXiv:1911.04252v4 [cs.LG] 19 Jun 2020