

# Deep learning による眼底画像診断に関する研究

S18105 島野 郁人

## 1. はじめに

本研究では診断が難しい場合が存在する緑内障診断について画像認識を扱い、どれほどの精度で認識できるかを検証する。本研究では画像解析をするツールとして weka<sup>1)</sup>を扱い、weka に含まれているディープラーニングのアルゴリズムを幅広くサポートしたフレームワークである deep learning4j<sup>2)</sup>を利用する。weka 内の deep learning4j は、全体で10個のネットワークがあり、全部実行した結果、8種類のもので実行可能であった。本研究では、実行ができた8種類の中でも有名なアーキテクチャでありかつ数値として比較をしやすい3種類のネットワークで比較をしていく。

## 2. 使用するディープラーニングネットワーク

本研究で比較を行なった3種類のネットワークの特徴を説明する。

### ● Model 1: AlexNet

AlexNet は初めて深層学習の概念および畳み込みニューラルネットワークの概念を取り入れたアーキテクチャである。また人が特徴量を設定しなくても十分なデータさえ存在すれば、機械自身が特徴量を見つけ出すことを証明した。Weka 内の AlexNet の構造は全体で畳み込み層が5層、サブサンプリング層が3層、全結合層が2層、そして出力層という構造であるため、11層から成り立っている。

### ● Model 2: inceptionResNetV1

ResNet は shortcut connection という機能を導入し、手前の層の入力を後ろの層に直接足し合わせることで勾配消失問題を解決したのが特徴で

ある。Weka 内の inceptionResNetV1 の構造は全体で畳み込み層が127層、activation層が41層、サブサンプリング層が4層、全結合層が1層、そして出力層という構造であるため、174層から成り立っている。

### ● Model 3: LeNet

LeNet は畳み込みニューラルネットワークの最初の提案および実装であり、初めての deep learning 化を行った研究であった。Weka 内の LeNet の構造は全体で畳み込み層が2層、サブサンプリング層が2層、全結合層が1層、そして出力層という構造であるため、6層から成り立っている。

## 3. 各ネットワークの実行と評価

実験内容は健康な目と緑内障の眼底画像をネットワークに学習させ、学習後評価をし、ネットワークとの関係性を考察する。学習させる画像データは、本研究室で収集されたもので、健康な目の眼底画像が55枚、緑内障の眼底画像が104枚の合計159枚の画像を利用した。

評価方法は10-fold cross validation (10-分割交差検証)を扱い、訓練データセットを10個のサブセットに分割し、そのうち9個で学習し残り1個で検証する方法である。そのサイクルを10回繰り返すものである。

それぞれのネットワークの評価は、正しい分類が出来ているかを調べる分類精度、二乗平均平方根誤差、および複数回の判定の一致度が高いほど値が大きくなり、値が大きいほど信頼性が高いとされる k 統計量、予測の結果と実際の結果の関係を示す混同行列によって評価を行う。

- AlexNet の実行結果

分類精度は、62.8931%であり、100 枚の画像を正当に分類することが可能であった。また、二乗平均平方根誤差の値は 0.6091 で、k 統計量は 0.0155、また、表 1 には、混合行列を示した。

この表 1 から緑内障を緑内障と判断できた割合が 94/104 であり、健康な目を健康な目と判断できた割合が 6/55 であるため、後者の割合が低すぎる微妙な結果となった。また、二乗平均平方根誤差の値が大きい点からも誤差が大きいモデル(精度が悪いモデル)ということが分かった。

表 1: AlexNet の混合行列

	緑内障	健康な目
緑内障	94(TP)	10(FN)
健康な目	49(FP)	6(TN)

- inceptionResNetV1 の実行結果

分類精度は、65.4088%であり、104 枚が画像を正当に分類しているように見える。また、二乗平均平方根誤差の値は 0.4806 で、k 統計量は 0、また、表 2 には、混合行列を示した。

この表 2 より眼底画像をすべて陽性と判断しているため、正しく評価ができなかったことが分かる。観察者間の結果の一致度を評価する k 統計量の値が 0 という数値は、観察者同士の一致がない、もしくは観察者同士の一致率を偶然の一致率が超えてしまう場合しかないため、この評価は信頼性がないものだと考えられる。また、ネットワークの大きさに対して、入力したデータが少ないことから学習結果側に問題があったと考えている。

表 2: inceptionResNetV1 の混合行列

	緑内障	健康な目
緑内障	104(TP)	0(FN)
健康な目	55(FP)	0(TN)

- LeNet の実行結果

分類精度は、83.6478%であり、133 枚の画像を

正当に分類することが可能であった。また、二乗平均平方根誤差の値が 0.3682 で、k 統計量は 0.6386、また、表 3 には、混合行列を示した。

この表 3 から緑内障を緑内障と判断できた割合が 91/104 であり、健康な目を健康な目と判断できた割合が 42/55 であるため良い結果となった。また二乗平均平方根誤差の値が一番低い点や k 統計量の値が 6 割を超えている点から精度が良く、観察者間の一致度が高い、つまり評価への信頼性が高い値にあることが分かる。

表 3: LeNet の混合行列

	緑内障	健康な目
緑内障	91(TP)	13(FN)
健康な目	13(FP)	42(TN)

#### 4. おわりに

本研究では緑内障診断を様々なネットワークを利用してどのくらいの精度を出すことが出来るのか検証を行った。その結果、一番古いネットワークであり、層の数が少ない LeNet が一番良い結果となった。これは階層数とデータ数が比例関係にあることからデータ数が少なかったため、低階層のネットワークで良い結果が出たのではないかと考えられる。データ数に依存する結果となってしまったが、精度を上げるにはデータ数を増やし、層ごとの与える影響を理解し適切な畳み込み層数などを考える必要があるのではないかとと思われる。

#### 参考文献

- 1) Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition.
- 2) Home - WekaDeepLearning4j, <https://deeplearning.cms.waikato.ac.nz/> (Access to 2022.1.12)