

令和元年度  
学士学位論文

機械学習による胸部 X 線画像からの疾患  
検出

A Study on Disease Detection from Chest X-ray  
Images by Machine Learning

1200280 東 聖

指導教員 吉田 真一

2019 年 2 月 28 日

高知工科大学 情報学群

# 要 旨

## 機械学習による胸部 X 線画像からの疾患検出

東 聖

近年、畳み込みニューラルネット (CNN) を用いた胸部 X 線画像の認識が研究されている。これらの研究では複数疾患のマルチラベル分類を行なっている。しかし、先行研究では、十分な識別精度を得られていない。この理由として考えられるのがラベル数が多いという点、クラス間のデータ数の差が多いという点である。そこで、本研究では学習時のラベル数に注目する。一度に学習するラベル数を減らすことで疾患の検出精度が向上すると考えた。よって、本研究では検出精度の向上を目的とした、学習時のラベル数を減らす 2 段階検出を用いた胸部疾患の検出を行う。提案手法ではまず、健常画像と疾患画像の 2 クラス分類を行う。その後、疾患画像のみを用いて 14 クラスマルチラベル分類を行う。

検証実験の結果、2 クラス分類は高い精度となり、その後の 14 の疾患の分類も 15 クラスまとめて行う場合より精度、平均再現率ともに上場することを示す。また、クラス数が多くなるほど平均再現率が低下する。データ数が多いほど平均再現率が高くなる傾向がある。

キーワード 胸部 X 線, CAD, マルチラベル分類, 再現率

# Abstract

## A Study on Disease Detection from Chest X-ray Images by Machine Learning

Hijiri AZUMA

Recently, medical image recognition of chest X-ray images using convolutional neural networks (CNN) has been studied. In these studies, multi-class, multi-labeled classification of multiple diseases is performed. However, in the previous study, accuracy has not been sufficient. The reason for this is that there are a lot of labels, and that there is a lot of difference in the number of data between classes. Therefore, in this study, we propose to equalize the number of labels for training. Also we propose to reduce the number of labels to be trained at one time to improve the detection accuracy of diseases. Therefore, in this study, chest disease is detected using two-step detection to reduce the number of labels at the time. In the proposed method, first, two-class classification of healthy and disease images is performed. Then, 14-class multi-labeled classification is performed using only disease images. As a result of the experiment, it is shown that the two-class classification has higher accuracy, and that the subsequent classification of 14 diseases achieves the higher accuracy and recall than that of 15 classes classification. The higher number of classes affect the lower accuracy and the lower the average recall and the higher number of data contributes the higher the average recall.

*key words*    Chest X-ray, CAD, multi-label classification, recall

# 目次

<b>第 1 章</b>	<b>序論</b>	<b>1</b>
1.1	研究背景 . . . . .	1
1.2	研究目的 . . . . .	2
1.3	本論文の構成 . . . . .	3
<b>第 2 章</b>	<b>関連技術</b>	<b>4</b>
2.1	胸部 X 線検査 . . . . .	4
2.2	CAD(Computer-Aided Diagnosis) . . . . .	6
2.3	マルチラベル分類 . . . . .	6
2.3.1	2 値分類・多クラス分類 . . . . .	6
2.3.2	マルチラベル分類 . . . . .	7
2.4	VGG16 . . . . .	9
<b>第 3 章</b>	<b>提案手法</b>	<b>10</b>
<b>第 4 章</b>	<b>実験</b>	<b>11</b>
4.1	データセット . . . . .	11
4.2	実験設定 . . . . .	11
4.2.1	学習率 . . . . .	12
4.3	実験手順 . . . . .	13
<b>第 5 章</b>	<b>結果</b>	<b>18</b>
5.1	実験 1 . . . . .	18
5.1.1	15 クラス分類 . . . . .	18
5.1.2	2 クラス分類 . . . . .	18

## 目次

5.1.3	14 クラス分類 . . . . .	19
5.2	実験 2 . . . . .	20
<b>第 6 章</b>	<b>考察</b>	<b>23</b>
6.1	実験 1 . . . . .	23
6.2	実験 2 . . . . .	23
<b>第 7 章</b>	<b>結論</b>	<b>25</b>
	謝辞	27
	参考文献	28

# 目次

1.1	臓器の重なった疾患画像の例 . . . . .	2
2.1	胸部 X 線画像の例 . . . . .	6
2.2	CAD の概要 . . . . .	7
2.3	2 値分類の例 . . . . .	8
2.4	多クラス分類の例 . . . . .	8
2.5	マルチラベル分類の例 . . . . .	8
2.6	VGG16 モデルの概略図 . . . . .	9
3.1	提案手法の概要 . . . . .	10
5.1	14, 15 クラスマルチラベル分類の再現率 . . . . .	19
5.2	15 クラスマルチラベル分類の正確度と損失 . . . . .	20
5.3	2 クラス分類の正確度と損失 . . . . .	21
5.4	2 クラス分類の混同行列 . . . . .	21
5.5	14 クラスマルチラベル分類の正確度と loss . . . . .	21
5.6	2,3,4 クラスマルチラベル分類の平均再現率 . . . . .	22
6.1	n 段階での分類の概要 . . . . .	24

# 表目次

1.1	胸部疾患検出手法の比較 . . . . .	2
4.1	2 クラス分類 (健常, 疾患) の 20 エポックにおける学習率の比較 . . . . .	13
4.2	2 クラス分類 (健常, 疾患) の 100 エポックにおける学習率の比較 . . . . .	14
4.3	15 クラス分類の 20 エポックにおける学習率の比較 . . . . .	14
4.4	15 クラス分類の 100 エポックにおける学習率の比較 . . . . .	15
4.5	14 クラス分類の 20 エポックにおける学習率の比較 . . . . .	15
4.6	14 クラス分類の 100 エポックにおける学習率の比較 . . . . .	15
4.7	2 クラス分類 (肺浸潤, 胸水) の 20 エポックにおける学習率の比較 . . . . .	16
4.8	2 クラス分類 (肺浸潤, 胸水) の 100 エポックにおける学習率の比較 . . . . .	16
4.9	3 クラス分類 (肺浸潤, 胸水, 無気肺) の 20 エポックにおける学習率の比較 . . . . .	16
4.10	3 クラス分類 (肺浸潤, 胸水, 無気肺) の 100 エポックにおける学習率の比較 . . . . .	17
4.11	4 クラス分類 (肺浸潤, 胸水, 無気肺, 結節) の 20 エポックにおける学習率の比較 . . . . .	17
4.12	4 クラス分類 (肺浸潤, 胸水, 無気肺, 結節) の 100 エポックにおける学習率の比較 . . . . .	17
5.1	分類クラス数の結果比較 . . . . .	19
5.2	分類クラス数の結果比較 . . . . .	22

# 第 1 章

## 序論

### 1.1 研究背景

胸部疾患の代表的な診断手法として、CT 検査と胸部 X 線検査がある。これらの検査より得られた画像から疾患の情報を得ることを読影と言う。胸部 X 線検査が身体を平面に撮影するのに対し、CT 検査は体の断面を撮影するため胸部 X 線検査よりも疾患検出に優れている。しかし、CT 検査は胸部 X 線検査よりも撮影時間が数倍かかる。また、医師が診察するにあたって読影を行わなければならない画像の枚数が、胸部 X 線検査が 1 枚から 2 枚なのに対して、CT 検査はその数十倍であり手間がかかる (表 1.1)。そのため、胸部診断において最も一般的な手法と言われているのが胸部 X 線検査である。

しかし、胸部 X 線検査は、身体を平面で撮影するため肺だけでなく心臓や肋骨などの臓器が重なって写る。図 1.1 に疾患画像の例を示す。赤枠で囲われた部分が疾患部位である。例となった画像では胸水を患っている。このように、読影が困難な場合があり、疾患を見落とす原因となっている。また、胸部 X 線検査は健康診断の診断項目にも含まれているため、毎年多くの検診者が検査を行なっている。その結果、膨大な数の胸部 X 線画像が医療施設に蓄積され、そのデータの読影を行う医師の数が不足している。このようなことから読影の補助が必要とされており、CAD(Computer-aided detection) が注目されている。

## 1.2 研究目的

表 1.1 胸部疾患検出手法の比較

	撮影時間 (分)	画像枚数
CT 検査	5~15	30~150
胸部 X 線検査	数分	1,2

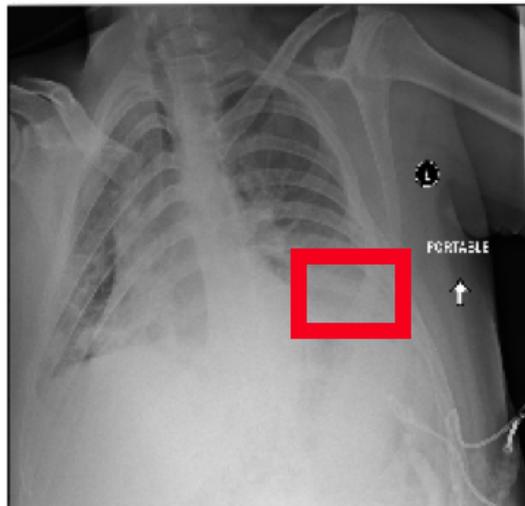


図 1.1 臓器の重なった疾患画像の例

## 1.2 研究目的

昨今では畳み込みニューラルネットワーク (CNN) を用いて、脳画像のみから男女を 97% の精度で識別する研究なども出てきている [1]. 文献 [1] において、人では識別のできない脳画像のみからの男女の識別を、CNN を用いることで実現していることから、胸部疾患の検出においても CNN を用いることで読影の補助に期待できる。

先行研究において、胸部疾患の検出に CNN を用いたものがある [2][3]. これらの研究では複数疾患のマルチラベル分類を行なっている。しかし、先行研究 [2] では、十分な識別精度を得られていない。この理由として、ラベル数が多く、クラス間のデータ数の差が大きいことが考えられる。ラベル数が多くなると分類する数が増えるため、精度が低下したと考える。また、クラス間のデータ数の差が大きいため、データ数の多いラベルに偏った学習をしてしまったと考える。そこで、本研究では、疾患検出精度の向上を目的とした、学習時のラ

### 1.3 本論文の構成

ベル数を減らす 2 段階検出を用いた胸部疾患の検出を行う。

### 1.3 本論文の構成

本論文の構成は，2 章で胸部疾患の診断手法や分類手法，関連研究について記す．3 章では，本実験で用いたデータセットや，実験時のパラメータの設定，実験手順について記す．4 章では，各実験での結果について記し，5 章では，4 章で得られた結果について考察し，その考察から行なった実験について記す．6 章では，本研究のまとめと今後の課題について記す．

## 第 2 章

# 関連技術

本研究は，機械学習を用いた胸部疾患の 2 段階検出である．よって，本章では胸部疾患の検出手法である胸部 X 線検査，CAD(Computer-Aided Diagnosis)，分類手法であるマルチラベル分類の順で記述する．

### 2.1 胸部 X 線検査

胸部 X 線検査は，X 線を照射し体の平面撮影を行う検査である．これによって撮影された画像を胸部 X 線画像という．また，胸部 X 線検査は健診では省くことのできない検査項目の一つである [4]．その理由として挙げられるのが，検査が比較的簡単かつ被験者への負担が少ない検査であること．そして，検査によって得られる情報が多いということである．

胸部 X 線画像の読影には，グラフや数字などの読み取りとは異なる手法や，多くの疾患に対する知識と読影の経験が要求される．そのため，検診での所見やそれに対応する事後指導の標準化が必要とされている [4]，しかし，呼吸器専門医を中心として試みられているが，未だに標準化は発表されていない [4]．

また，胸部 X 線検査において，検査のルーチン化を行うことで，経過観察の際に比較のしやすい画像を取得できる．胸部の評価を最適に行うには，立体での後-前方向 (PA) と側方向のルーチン撮影が最適である．また，胸部の評価を正しく行うためには，適切な撮影を行うことが重要である．胸部 X 線画像を正しく撮影するには以下の 4 つが重要とされている．

- 撮影体位
- 呼吸

## 2.1 胸部 X 線検査

- 照射時間
- 管電圧

適正な撮影を行うためには、X 線ビームがフィルムの中心にただしく入射し、身体がしっかりと正面を向き、肩甲骨肺野外に十分に外れて位置するような体位を取らなければならない [5][6]。そのようにして撮影することで、正しい撮影体位で X 線写真を撮影することができる。また、撮影時の呼吸については深吸気位で停止した状態で撮影を行う。照射時間においては、後-前方向で胸椎及び椎間板がかすかに認識できる程度の照射時間で撮影し、肺血管影がはっきりと可視化できるようにする。つまり、できるだけ短時間での撮影が望ましいとされている。管電圧においては 115~150kVp の高電圧での撮影が推奨されている。

また、胸部 X 線写真は撮影した画像が読影をするにあたって適当であるか、ILO 国際分類の 4 段階で評価される [7]。読影不可に分類された画像を無理に読影すると誤診につながるため、再度撮影を行う必要がある。

胸部 X 線撮影には従来型のフィルム・スクリーン型とデジタル型の 2 種類が存在する。フィルム・スクリーン型は取り扱いが容易で高感度であり、ある程度の良質な画像が得られるという特徴がある [5]。ただし、診断情報をフィルムに写す際にダイナミックレンジが狭いという特徴がある。胸郭中の X 線減衰度は、ほぼ完全に X 線が透過してしまう肺のような臓器から、X 線が強く減衰してしまう縦隔、肋骨、脊椎などの構造物まで種々異なる [5]。以下に、これらをすべて可視化するための手法を 3 つ示す。

- 適正露光域が広く寛容度が大きいフィルムを使用する
- 患者ごとにオーダーメイドしたフィルタを使用する
- フィルムの黒化度を均等化することで各構造物を適切な濃度でフィルムに病がする撮影法

デジタル型は多くの面で従来型のフィルム・スクリーン型よりも優れている。デジタル型は、フィルム・スクリーン型の 10 倍~100 倍広域のダイナミックレンジを有する [5]。また、電子画像であるため従来型と比べて保管が容易である。図 2.1 に胸部 X 線画像の例を示す。

## 2.2 CAD(Computer-Aided Diagnosis)



図 2.1 胸部 X 線画像の例

## 2.2 CAD(Computer-Aided Diagnosis)

CAD はコンピュータが画像情報の定量化，分析を行い，医師の画像診断を補助する意思決定支援システムである．現在胸部 X 線検査において，読影時の所見や事後指導の標準化が行われていない [4]．医師毎に基準がバラバラであるため，医師がより良い診断を行うためには標準化は必須である．そこで，CAD を用いることで読影時に定量的な尺度を提供することができ，読影の補助や疾患の見落としの減少が期待できる．図 2.2 に CAD の概要を示す．

## 2.3 マルチラベル分類

本研究では，胸部 X 線画像がどのクラス (疾患) に属するか分類する．使用する胸部 X 線画像には複数のラベル付がされているため，マルチラベル分類を使用した．分類には大きく分けて，2 値分類，多クラス分類，多ラベル分類 (マルチラベル分類) が存在する．

### 2.3.1 2 値分類・多クラス分類

2 値分類とは，入力された画像が A と B のどちらに属するか分類するものである．図 2.3 では入力された胸部 X 線画像を肺浸潤，胸水の 2 つに分類している．2 値分類では分類時に

## 2.3 マルチラベル分類

A, B それぞれに属する確率を A, B の合計が 1 になるように算出し, 多クラス分類は, 分類する個数が 3 つ以上のものを指す. 2 クラス分類と多クラス分類ともに, 入力された画像は 1 つのクラスにだけ分類される. 図 2.4 はラベル付した例である.

### 2.3.2 マルチラベル分類

多クラス分類が一つの入力画像に一つだけラベル付されていたのに対して, マルチラベル分類は, 一つの入力画像に対して複数のラベル付がされたものを分類するものである. 図 2.5 はラベル付した例である.

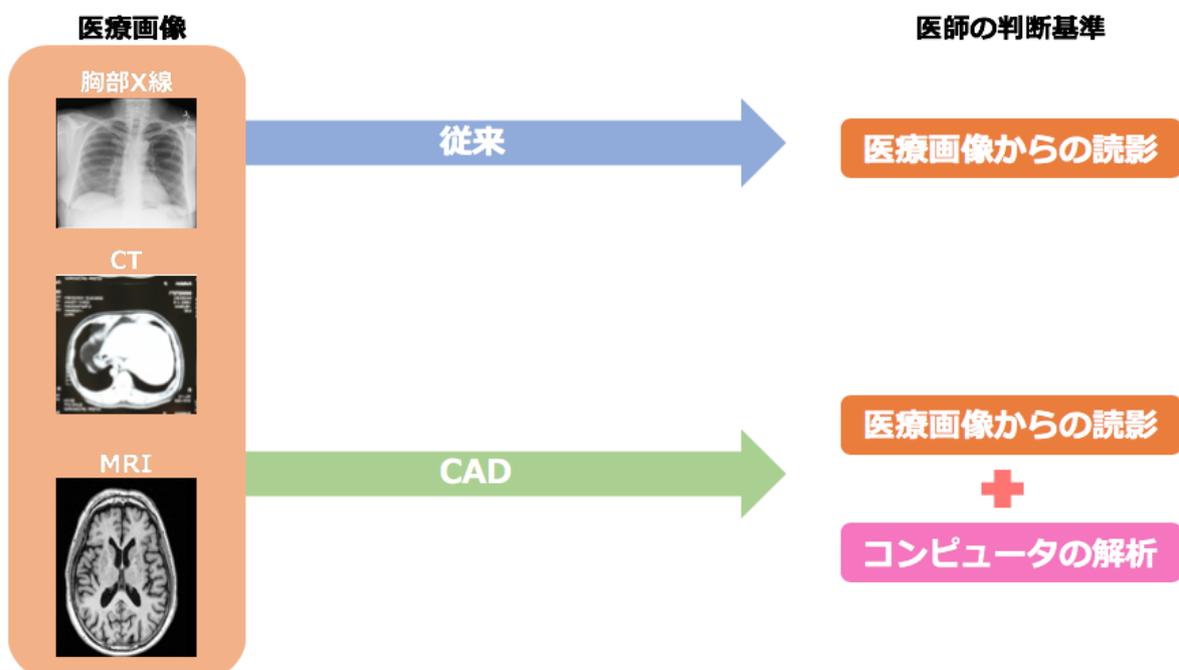


図 2.2 CAD の概要

## 2.3 マルチラベル分類

画像	ラベル1	ラベル2
	1	0
	0	1

図 2.3 2 値分類の例

画像	ラベル1	ラベル2	ラベル3	ラベル4
	1	0	0	0
	0	1	0	0

図 2.4 多クラス分類の例

画像	ラベル1	ラベル2	ラベル3	ラベル4
	1	0	1	0
	0	1	0	1

図 2.5 マルチラベル分類の例

## 2.4 VGG16

### 2.4 VGG16

本研究では、CNN を用いて胸部 X 線画像の疾患分類を行う。そこで、有名な CNN モデルの 1 つである VGG16 を用いる。VGG16 は ImageNet の 100 万以上のデータで学習された、学習済みモデルの 1 つである。学習済みモデルには、他にも ResNet, AlexNet などがある。また、VGG16 は、畳み込み層が 13 層、全結合層が 3 層の合計 16 層の CNN である。主な特徴として 3 かける 3 の小さな畳み込みフィルタが特徴となっている。概略図を図 2.6 に示す。

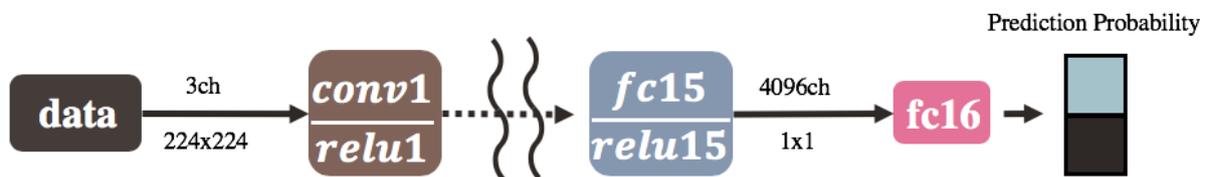


図 2.6 VGG16 モデルの概略図

## 第 3 章

# 提案手法

胸部 X 線画像から疾患を分類する場合、より多くの疾患を分類できることが望ましい。そのため、それを実現するためには一度の学習で多くのラベルを用いる必要がある。しかし、これまでの研究 [2] では十分な識別精度を得ることはできていない。原因として、既存の手法ではラベル数やクラス間のデータ数の差が大きく、それが原因で精度が低下していると考えられる。そこで本研究では、疾患検出精度の向上を目的とした、学習時のラベル数を減らす 2 段階検出を用いた胸部疾患の検出を提案する。1 度に学習するラベル数を減らすことで疾患の検出精度が向上すると考えた。図 3.1 は提案手法の概要である。推定モデルには VGG16 を使用し、この時の最適化には Adam を使用している。手順 1 で健常画像と疾患画像の 2 クラス分類を行う。手順 2 で疾患画像のみを入力として 14 クラスマルチラベル分類を行う。このように 2 段階で学習することで精度の向上を目指す。

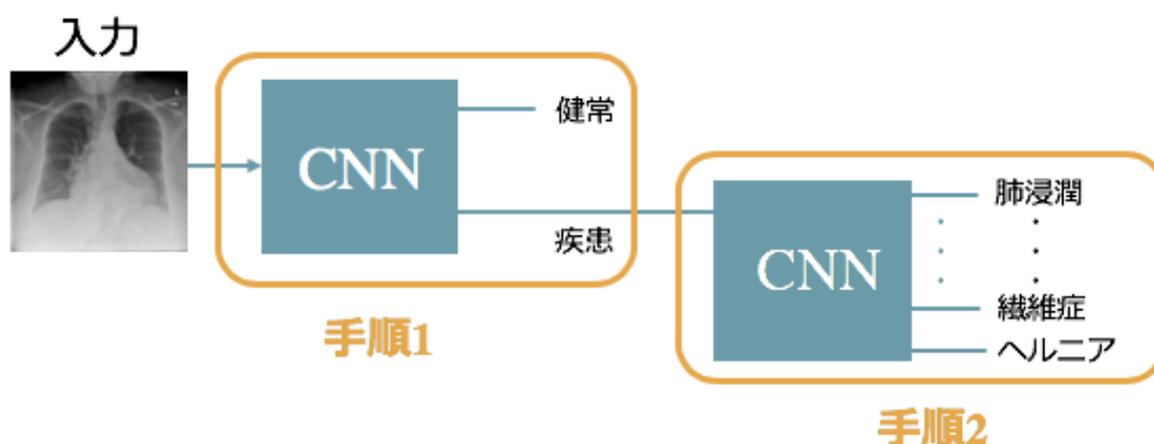


図 3.1 提案手法の概要

## 第 4 章

# 実験

本章では、本研究で行なった実験のデータセット、実験設定、使用したモデル、実験手順について記述する。今回の実験では実験 1 として全ラベルを識別する 15 クラスマルチラベル分類、健常と疾患の 2 クラス分類、疾患画像のみを入力とした 14 クラスマルチラベル分類を行う。この実験でラベルが多い場合と少ない場合の比較を行う。そして、実験 2 としてデータ数の多い疾患順にラベル数を 2 つ、3 つ、4 つと増やした、2, 3, 4 クラスマルチラベル分類を行う。これらの比較により、識別精度に対するラベル数とデータ数の差の影響を検証する。

### 4.1 データセット

本研究では NIHCC(National Institutes of Health Clinical Center) の胸部 X 線画像のデータセットを使用した。これはアメリカ国立衛生研究所のもので、112120 枚の胸部 X 線画像で構成されており、14 種の疾患のラベル付けがされている。画像サイズは 1024 × 1024、画像の DPI は 72dpi である。

また、このデータセットには疾患である無気肺、心肥大、胸水、肺浸潤、腫瘍、結節、肺炎、気胸の 8 種類の疾患部位を示したバウディングボックスも存在する。

### 4.2 実験設定

今回は 2 クラス分類と多ラベル分類を行う。そのため、各疾患の確率を求めるために活性化関数はシグモイド関数、損失関数ではバイナリークロスエントロピーを使用する。

## 4.2 実験設定

評価指標には、平均再現率、平均適合率、F-measure の 3 つを用いる。本研究で行う実験では、実験毎にデータ数が異なるため、再現率を求め平均をとった。また、平均適合率、F-measure は実験毎にデータ数が異なることから実験毎の比較はできないが、実験単体で評価する場合に参考になるため用いた。

### 4.2.1 学習率

学習率は、 $10^{-3}$ ,  $10^{-4}$ ,  $10^{-5}$ ,  $10^{-6}$ ,  $10^{-7}$ ,  $10^{-8}$  の 6 種類をエポック数 20 で学習し、混同行列を確認して学習を行っていたものみを、100 エポックで学習する。そして、その中で最も平均再現率が高いものを選択する。

表 4.1 は疾患と健常の 2 クラス分類を 6 種類の学習率で 20 エポック学習した時の再現率の比較である。この結果より、学習率が  $10^{-5}$ ,  $10^{-6}$ ,  $10^{-7}$  の 3 つを 100 エポックで学習した。表 4.2 は上記の 3 つを 100 エポックで学習した時の再現率の比較である。この結果より、学習率  $10^{-6}$  が最も再現率が高くなった。よって、2 クラス分類においては学習率を  $10^{-6}$  とする。

表 4.3 は 15 クラス分類を 6 種類の学習率で 20 エポック学習した時の平均再現率の比較である。この結果より、学習率が  $10^{-4}$ ,  $10^{-5}$  の 2 つを 100 エポックで学習した。表 4.4 は上記の 2 つを 100 エポックで学習した時の平均再現率の比較である。この結果より、学習率  $10^{-5}$  が最も再現率が高くなった。よって、15 クラス分類においては学習率を  $10^{-5}$  とする。

表 4.5 は 14 クラス分類を 6 種類の学習率で 20 エポック学習した時の平均再現率の比較である。この結果より、学習率が  $10^{-5}$ ,  $10^{-6}$  の 2 つを 100 エポックで学習した。表 4.6 は上記の 2 つを 100 エポックで学習した時の平均再現率の比較である。この結果より、学習率  $10^{-5}$  が最も再現率が高くなった。よって、14 クラス分類においては学習率を  $10^{-5}$  とする。

表 4.7 は肺浸潤、胸水の 2 クラス分類を 6 種類の学習率で 20 エポック学習した時の再現率の比較である。この結果より、学習率が  $10^{-5}$ ,  $10^{-6}$  の 2 つを 100 エポックで学習した。表 4.8 は上記の 2 つを 100 エポックで学習した時の再現率の比較である。この結果より、学習率  $10^{-5}$  が最も再現率が高くなった。よって、肺浸潤、胸水の 2 クラス分類においては学

### 4.3 実験手順

習率を  $10^{-5}$  とする.

表 4.9 は 3 クラス分類を 6 種類の学習率で 20 エポック学習した時の平均再現率の比較である. この結果より, 学習率が  $10^{-4}, 10^{-5}, 10^{-6}$  の 3 つを 100 エポックで学習した. 表 4.10 は上記の 2 つを 100 エポックで学習した時の平均再現率の比較である. この結果より, 学習率  $10^{-5}$  が最も再現率が高くなった. よって, 3 クラス分類においては学習率を  $10^{-5}$  とする.

表 4.11 は 4 クラス分類を 6 種類の学習率で 20 エポック学習した時の平均再現率の比較である. この結果より, 学習率が  $10^{-5}, 10^{-6}$  の 2 つを 100 エポックで学習した. 表 4.12 は上記の 2 つを 100 エポックで学習した時の平均再現率の比較である. この結果より, 学習率  $10^{-5}$  が最も再現率が高くなった. よって, 4 クラス分類においては学習率を  $10^{-5}$  とする.

## 4.3 実験手順

本研究では疾患と健常の 2 クラス分類, 疾患のみを分類する 14 クラス分類, 健常, 疾患を含めた 15 クラス分類を行う. 以下の手順で各実験を行う.

1. 画像サイズを  $1024 \times 1024$  から  $224 \times 224$  に変更
2. 画像を正規化するために画像を 255 で除算

表 4.1 2 クラス分類 (健常, 疾患) の 20 エポックにおける学習率の比較

モデル	学習率	再現率
2 クラス (健常, 疾患)	$10^{-3}$	0
	$10^{-4}$	0
	$10^{-5}$	0.5693
	$10^{-6}$	0.6882
	$10^{-7}$	0.5165
	$10^{-8}$	0

### 4.3 実験手順

表 4.2 2 クラス分類 (健常, 疾患) の 100 エポックにおける学習率の比較

モデル	学習率	再現率
2 クラス (健常, 疾患)	$10^{-5}$	0.5993
	$10^{-6}$	0.6481
	$10^{-7}$	0.5466

表 4.3 15 クラス分類の 20 エポックにおける学習率の比較

モデル	学習率	再現率
15 クラス	$10^{-3}$	0
	$10^{-4}$	0.1004
	$10^{-5}$	0.1781
	$10^{-6}$	0
	$10^{-7}$	0
	$10^{-8}$	0

3. ラベルを one-hot 化

4. 各モデルを学習

学習時には 5 エポック間にバリデーションの正確度が向上しない場合に学習率を減らすコールバック関数の `ReduceLROnPlateau` を用いた.

### 4.3 実験手順

表 4.4 15 クラス分類の 100 エポックにおける学習率の比較

モデル	学習率	再現率
15 クラス	$10^{-4}$	0.0659
	$10^{-5}$	0.1029

表 4.5 14 クラス分類の 20 エポックにおける学習率の比較

モデル	学習率	再現率
14 クラス	$10^{-3}$	0
	$10^{-4}$	0
	$10^{-5}$	0.1451
	$10^{-6}$	0.0149
	$10^{-7}$	0
	$10^{-8}$	0

表 4.6 14 クラス分類の 100 エポックにおける学習率の比較

モデル	学習率	再現率
14 クラス	$10^{-5}$	0.2036
	$10^{-6}$	0.0165

### 4.3 実験手順

表 4.7 2 クラス分類 (肺浸潤, 胸水) の 20 エポックにおける学習率の比較

モデル	学習率	再現率
2 クラス (肺浸潤, 胸水)	$10^{-3}$	0
	$10^{-4}$	0
	$10^{-5}$	0.7431
	$10^{-6}$	0.7549
	$10^{-7}$	0
	$10^{-8}$	0

表 4.8 2 クラス分類 (肺浸潤, 胸水) の 100 エポックにおける学習率の比較

モデル	学習率	再現率
2 クラス (肺浸潤, 胸水)	$10^{-5}$	0.7582
	$10^{-6}$	0.7561

表 4.9 3 クラス分類 (肺浸潤, 胸水, 無気肺) の 20 エポックにおける学習率の比較

モデル	学習率	再現率
3 クラス (肺浸潤, 胸水, 無気肺)	$10^{-3}$	0
	$10^{-4}$	0.5802
	$10^{-5}$	0.5519
	$10^{-6}$	0.4519
	$10^{-7}$	0
	$10^{-8}$	0

### 4.3 実験手順

表 4.10 3 クラス分類 (肺浸潤, 胸水, 無気肺) の 100 エポックにおける学習率の比較

モデル	学習率	再現率
3 クラス (肺浸潤, 胸水, 無気肺)	$10^{-4}$	0
	$10^{-5}$	0.5757
	$10^{-6}$	0.5328

表 4.11 4 クラス分類 (肺浸潤, 胸水, 無気肺, 結節) の 20 エポックにおける学習率の比較

モデル	学習率	再現率
4 クラス (肺浸潤, 胸水, 無気肺, 結節)	$10^{-3}$	0
	$10^{-4}$	0
	$10^{-5}$	0.46
	$10^{-6}$	0.3648
	$10^{-7}$	0
	$10^{-8}$	0

表 4.12 4 クラス分類 (肺浸潤, 胸水, 無気肺, 結節) の 100 エポックにおける学習率の比較

モデル	学習率	再現率
4 クラス (肺浸潤, 胸水, 無気肺, 結節)	$10^{-5}$	0.4737
	$10^{-6}$	0.3448

# 第 5 章

## 結果

本章では本研究で行なった実験の結果を示す。

### 5.1 実験 1

表 5.1, 図 5.1 に実験 1 の結果を示す。

#### 5.1.1 15 クラス分類

既存手法である 15 クラスマルチラベル分類を行なった結果, 5.2 のようなグラフが得られた。また, 得られた混同行列より, 平均再現率, 平均適合率, F-measure を求め, その結果を表 5.1 に示す。

5.1 より, 15 クラスマルチラベル分類が平均再現率, 平均適合率, F-measure 共に最も低くなった。また, 図 5.2 より, 正確度が低下しており, 損失も増加傾向となっていた。

#### 5.1.2 2 クラス分類

健常と疾患の 2 クラス分類を行なった結果, 5.3 のようなグラフが得られた。また, 図 5.4 のような混同行列が得られた。混同行列より, 再現率, 適合率, F-measure が得られた。その結果を表 5.1 に示す。

5.1 より, 健常と疾患の 2 クラス分類が平均再現率, 平均適合率, F-measure 共に最も高くなった。また, 図 5.3 より, 50epoch 辺りから損失が増加していることが確認できた。

## 5.1 実験 1

表 5.1 分類クラス数の結果比較

モデル	学習率	平均再現率	平均適合率	F-measure
15 クラス	$10^{-5}$	0.1029	0.347	0.1588
2 クラス (健常, 疾患)	$10^{-6}$	0.6481	0.6593	0.6537
14 クラス (健常なし)	$10^{-5}$	0.2036	0.3554	0.2589

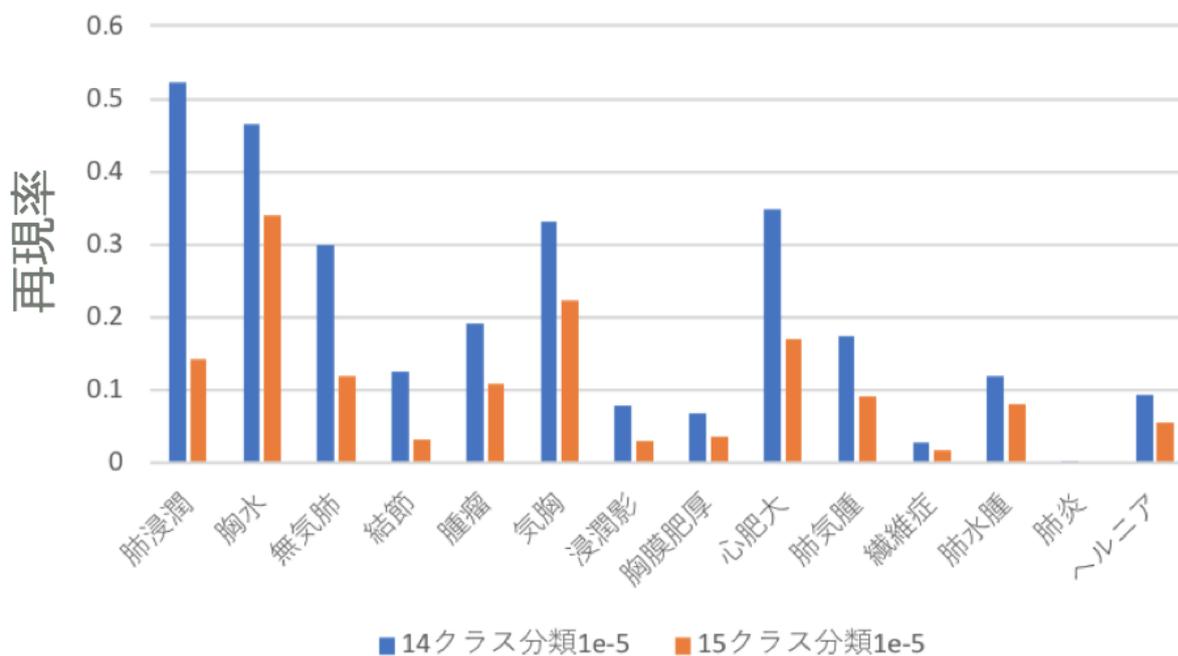


図 5.1 14, 15 クラスマルチラベル分類の再現率

### 5.1.3 14 クラス分類

疾患画像のみを用いた 14 クラス分類を行なった結果, 5.5 のようなグラフを得られた. また, 得られた混同行列より, 平均再現率, 平均適合率, F-measure を求め, その結果を表 5.1 に示す.

5.1 より, 既存手法の 15 クラスマルチラベル分類と, 提案手法である 14 クラスマルチラベル分類を比較すると, 提案手法の方が平均再現率が約 0.1 増加した.

図 5.1 より, 全ての疾患において, 14 クラスマルチラベル分類の再現率が高くなった. ま

## 5.2 実験 2

た、疾患の種類毎に検出のし易さが異なっていることが確認できた。そして、データ数が少なくなるほど再現率が低下する傾向が見られた。

## 5.2 実験 2

表 5.2, 図 5.6 に実験 2 の結果を示す。結果として、2 クラス (肺浸潤, 胸水) 分類が全ての実験の中で最も高い精度となった。また、クラス数が増加するほど平均再現率が低下する傾向が確認できた。また、データ数が多いほど平均再現率が高くなる傾向があることが確認できた。

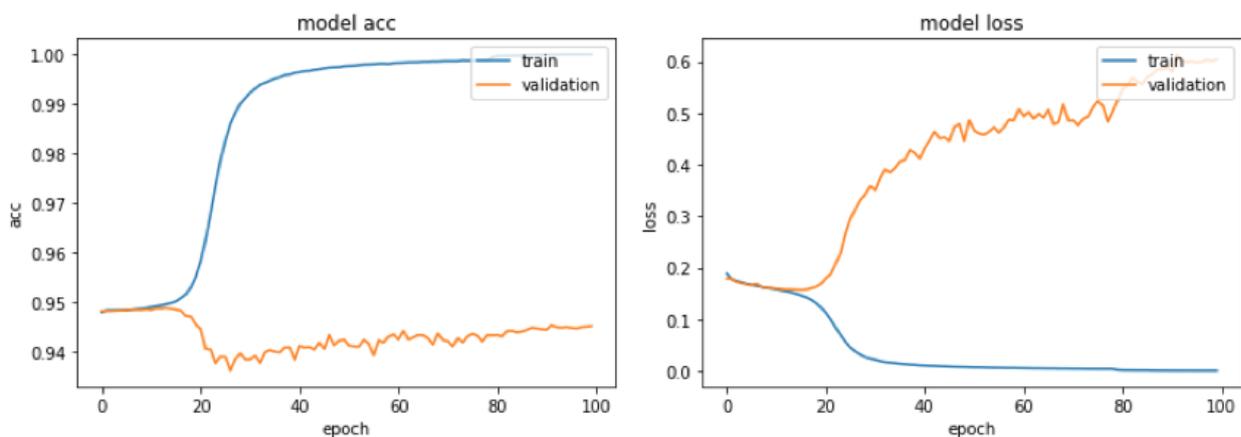


図 5.2 15 クラスマルチラベル分類の正確度と損失

## 5.2 実験 2

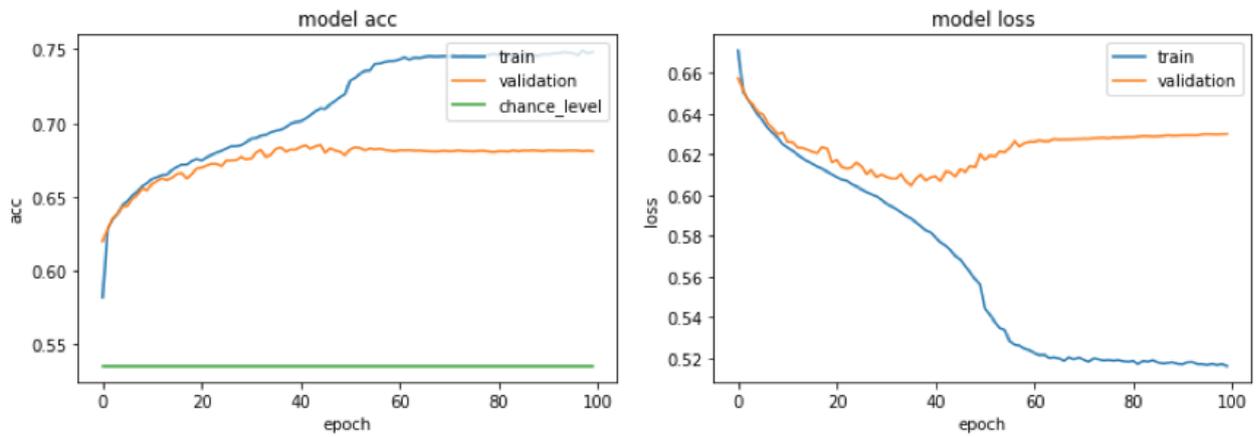


図 5.3 2 クラス分類の正確度と損失

		モデル予測	
		Positive	Negative
実 ク ラ ス	Positive	8439	4583
	Negative	4360	10648

図 5.4 2 クラス分類の混同行列

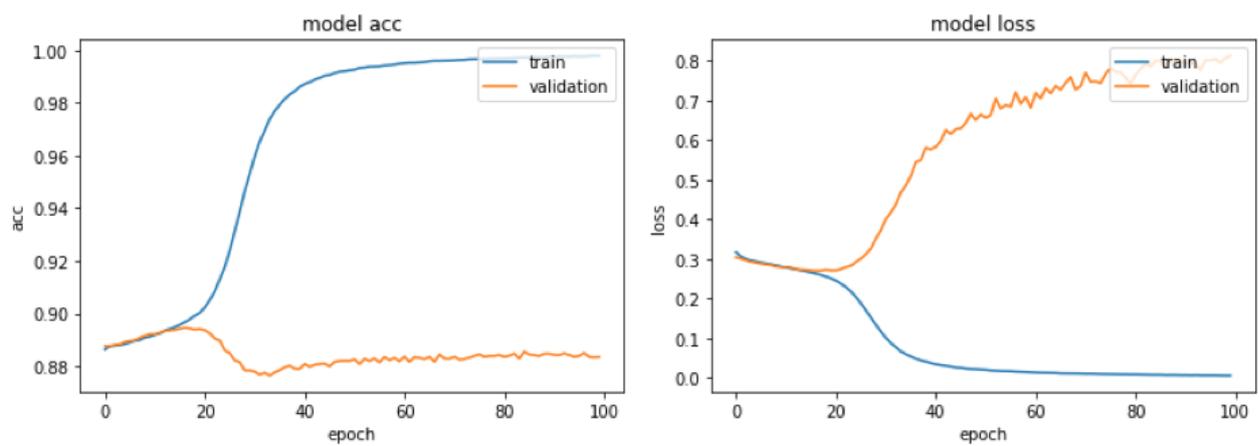


図 5.5 14 クラスマルチラベル分類の正確度と loss

## 5.2 実験 2

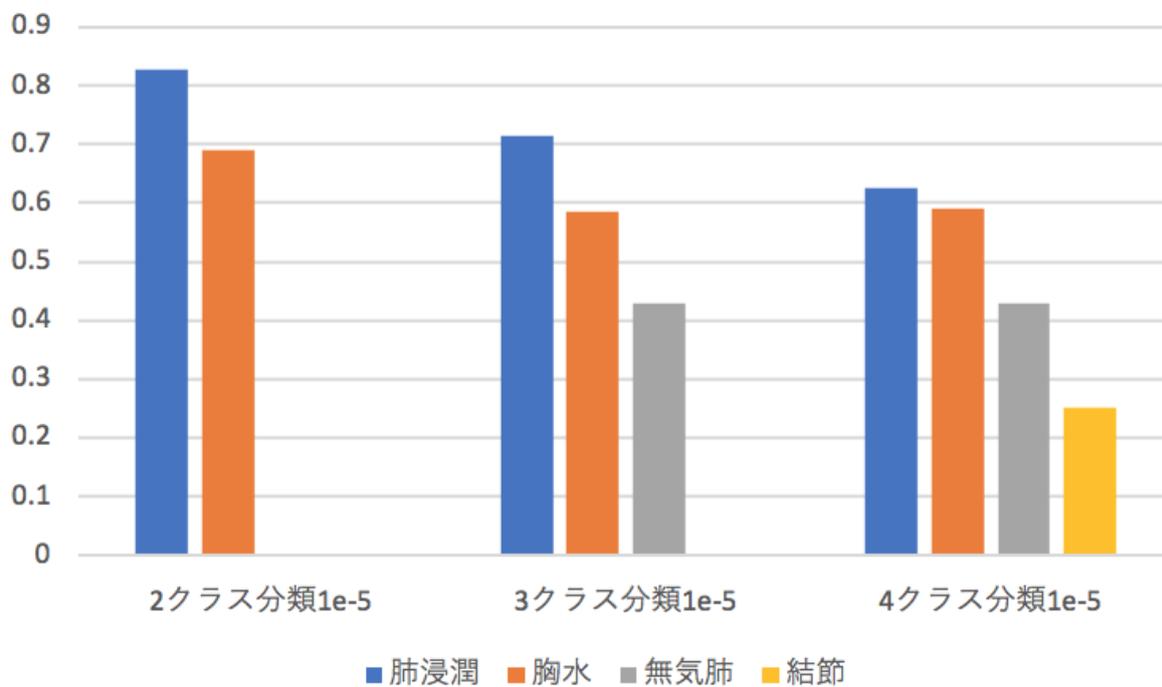


図 5.6 2,3,4 クラスマルチラベル分類の平均再現率

表 5.2 分類クラス数の結果比較

モデル	学習率	平均再現率	平均適合率	F-measure
2 クラス (肺浸潤, 胸水)	$10^{-5}$	0.7582	0.7115	0.7341
3 クラス (肺浸潤, 胸水, 無気肺)	$10^{-5}$	0.5757	0.5819	0.5787
4 クラス (肺浸潤, 胸水, 無気肺, 結節)	$10^{-5}$	0.4737	0.5055	0.4891

## 第 6 章

# 考察

以下に、実験結果からの考察を述べる。

### 6.1 実験 1

本実験では、健常と疾患の 2 クラス分類、既存手法である 15 クラス分類、疾患画像のみの 14 クラス分類を行なった。実験の結果、図 5.1 より、データ数が減少する毎に再現率が低下する傾向が見られた。この結果より、データ量が疾患精度に影響すると考えた。そのため、さらに精度を向上させるにはより多くのデータが必要であると考えた。上記の 3 つの実験の中で、健常と疾患の 2 クラス分類が平均再現率、平均適合率、F-measure 共に最も高くなった。この理由として、分類するクラス数が少ないこと、疾患のデータ数が多いこと。以上 2 つの要因が考えられる。14 クラス分類は、15 クラス分類と比較して、平均再現率、平均適合率、F-measure 共に高く、図 5.1 より、各疾患毎に再現率を比較しても、全て 14 クラス分類が上回っている。これらの結果より、クラス数の増加により、再現率が低下していくことが考えられる。

### 6.2 実験 2

この実験では、データ数の多い疾患順にクラス数を 2 つ、3 つ、4 つと増やした、2 クラス (肺浸潤、胸水) 分類、3 クラス (肺浸潤、胸水、無気肺) 分類、4 クラス (肺浸潤、胸水、無気肺、結節) 分類を行なった。実験の結果、2 クラス (肺浸潤、胸水) 分類が全ての実験の中で最も高い識別精度となった。この理由として、分類するクラス数が少ないこと、疾患のデー

## 6.2 実験2

タ数が多いこと。以上2つの要因が考えられる。2クラス(肺浸潤, 胸水)分類が2クラス(健常, 疾患)分類よりも精度が高いのは, 疾患ラベルに14種類もの疾患が含まれていることが原因ではないかと考える。次に, クラス数が多くなるほど平均再現率が低下する傾向があることを確認できた。また, データ数が多いほど平均再現率が高くなる傾向があることを確認できた。これらの結果から, クラス数とデータ数が分類時に強い影響を与えており, 現段階ではデータ数が足りていないということが考えられる。また, データ数だけでなくデータの質にも問題があると考えられる。本研究で使用した画像には大部分が白くぼやけた画像も含まれており, その画像が悪影響を及ぼしている可能性が考えられる。

これら結果より, 分類時にクラス数を増やさずに, 2クラス分類を $n$ 回続ける $n$ 段階での分類を行うことで, 精度の向上が見られるのではないかと考える。しかし, この手法ではクラス数を減らすことができてもデータ数が少ない疾患がいくつかあるため, それらは学習ができず精度が向上しないことが考えられる。そのため, クラス数の少ない疾患を分類することは現状では難しい。図6.1に,  $n$ 段階での分類の概要を示す。

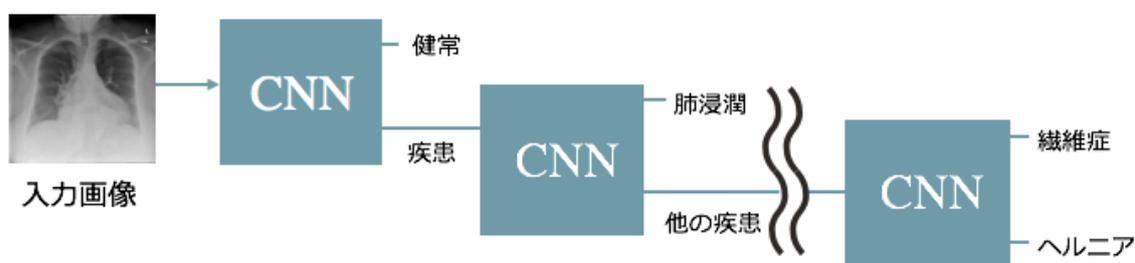


図 6.1  $n$  段階での分類の概要

## 第7章

# 結論

本研究では、疾患検出精度の向上を目的とした、学習時のラベル数を減らす2段階検出を用いた胸部疾患の検出を行なった。提案手法と比較するため、まず、データセットの全データを用いた15クラスマルチラベル分類を行なった。次に、提案手法のステップ1である健康画像と疾患画像の2クラス分類を行なった。そして、提案手法のステップ2である疾患画像のみを用いた14クラスマルチラベル分類を行なった。評価指標には、疾患毎にデータ数が異なるため、平均再現率を用いた。また、混同行列を求め、平均適合率、F-measureを算出し比較した。

結果として、15クラスマルチラベル分類は表5.1より、全ての分類の中で最も低い精度となった。また、5.2の学習曲線より、lossが上昇しており学習ができていないことが分かった。これは、ラベル毎の学習データの数に差があることが原因であると考えられる。最もデータの少ないヘルニアの学習データは100枚程度であるため、データ数の多い肺浸潤や胸水に偏った学習をしてしまったと考えられる。2クラス分類では、5.1と比較すると全ての分類の中で最も高い精度となった。また、5.3の学習曲線より、チャンスレベルよりも精度が高くなった。14クラスマルチラベル分類は、5.1より、15クラスマルチラベル分類の平均再現率0.1029と比較して10%高い0.2036という結果が得られた。これは、分類するクラス数が減ったことで精度が向上したと考えられる。

これらの結果から、クラス数の増加により再現率が低下していくことが考えられた。よって、データ数の多い疾患順にクラス数を2つ、3つ、4つと増やし、分類を行なった。その結果、クラス数が多くなるほど平均再現率が低下することを確認できた。また、データ数が多いほど平均再現率が高くなる傾向があることを確認できた。

今後は考察において提案した  $n$  段階での分類を実装して精度が向上するか確認したい。また、本研究の問題点であるデータ不足の改善や、本研究では確認のできなかった VGG16 以外のモデルを使用した比較を行いたい。

# 謝辞

本研究を進めるにあたり、ご指導して頂いた吉田真一准教授に心から感謝いたします。研究に行き詰まった時には的確なアドバイスを頂き、先生のお力添えにより、無事論文の執筆を終えることができました。研究室生活ではそのユニークなキャラクターも相まって、とても充実した研究室生活を過ごすことができました。また、ITnews, ポトラックなどのイベントで、先生にご指導いただく度に成長を実感することができました。

四宮友貴助教には研究に関わるだけでなく個人的な相談にも乗って頂き、公私ともに多くのことを教えて頂きましたこと大変感謝しております。

同期の4年生とは2年間共に研鑽には励んだ仲であり、楽しく大学生活を過ごせた大きな要因でした。それぞれキャラクターの尖ったメンバーばかりで、退屈のしない賑やかな毎日でした。

同研究室である生意気ながらも可愛げのある後輩達や尊敬できる先輩方、副査を担当して頂いた妻鳥貴彦准教授、敷田幹文教授には大変お世話になりました。

大学生活を送るにあたって、両親、友人、先生方など多くの人に支えられてきました。これまで支えてきてくださった皆様に改めて感謝を申し上げます。

# 参考文献

- [1] Takenobu Sasatani Shinichi Yoshida, Yuki Shinomiya and Kaechang Park. Prediction of human characteristics from brain structural MRI using CNN. In *ISFT-2019*, 2019.
- [2] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [3] Ahmed Tahseen Minhaz<sup>1</sup> Khalid Ashraf Mohammad Tariqul Islam<sup>1</sup>, Md Abdul Aowal<sup>1</sup>. Abnormality detection and localization in chest x-rays using deep convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [4] 後藤由夫, 奈良昌治. 健康判別基準ガイドライン . 株式会社 分光堂 , 2016.
- [5] N.L. ミュラー R.S. フレイザー, P.D. パレ. フレイザー 呼吸器病学エッセンス. 西村書店, 2009.
- [6] 中島啓. 胸部 X 線・CT の読み方やさしくやさしく教えます! 株式会社 羊土社, 2018.
- [7] 労災病院じん肺研究グループ編集委員会. よくわかる じん肺健康診断. 公益財団法人 産業医学振興財団, 2018.