

令和5年度卒業論文

胸部X線画像を使用した
深層学習によるマルチラベル付き画像分類の検証

宮崎大学 工学部 情報システム工学科

永井 真美

指導教員 棕木 雅之 教授

概要

本論文では、深層学習によるマルチラベル付き画像分類の検証を行う。深層学習による画像分類問題には、2値分類、マルチクラス分類、マルチラベル分類がある。2値分類はあるクラスに属するか属さないかの2つに分ける分類問題、マルチクラス分類は2つ以上のクラスのいずれかのみに分ける分類問題、マルチラベル分類は1つの対象物が複数のクラスに同時に分けられる分類問題である。2値分類やマルチクラス分類は、マルチラベル分類と区別してシングルラベル分類と呼ばれる。従来の研究では、単一胸部疾患のみを含む胸部 X 線画像を用いて作成した ChestX-ray8 と呼ばれるデータセットを使用して、シングルラベル分類を行っている。また、データセットのラベル付けの方法や使用モデルの違いによる研究は行われているが、複数の胸部疾患を含んだ胸部 X 線画像を使用したマルチラベル分類や、出力層の違いによる比較を行った検証は十分ではない。

本研究では、複数の胸部疾患を含む胸部 X 線画像を使用したマルチラベル付き画像分類を、2値分類のコーディング法、マルチラベル分類のコーディング法、損失関数と出力層の活性化関数の組み合わせの、3つの観点から検証する。1つ目の2値分類のコーディング法では、マルチクラス分類、マルチラベル分類で最も単純なクラス数が2の場合について評価する。2つ目のマルチラベル分類のコーディング法では、マルチラベル分類問題をマルチクラス分類問題に置き換えた手法について評価する。3つ目の損失関数と出力層の活性化関数の組み合わせでは、損失関数である2値交差エントロピー(BCE)、交差エントロピー(CE)、平均2乗誤差(MSE)と、出力層の活性化関数である sigmoid、softmax、linear のそれぞれの組み合わせ毎に比較評価を行う。結果として、総合的な精度のみの観点でいえば2出力のコーディングを用いた学習、Multi-Hot ベクトルを用いて交差エントロピーと softmax の組み合わせを用いた学習が優れていることが分かった。

目次

1. はじめに	1
2. 深層学習による胸部 X 線画像の画像分類	3
2.1. 深層学習による画像分類	3
2.2. 胸部 X 線画像のデータセット	4
2.3. 深層学習による胸部 X 線画像分類の従来研究	5
2.4. マルチラベル分類	7
3. マルチラベル付き画像分類の検証	9
3.1. マルチラベル付き画像分類	9
3.2. 検証項目	9
3.3. 2 値分類のコーディング法	10
3.4. マルチラベル分類のコーディング法	10
3.5. 損失関数と出力層の活性化関数の組み合わせ	13
4. 実験	15
4.1. 実験 1 2 値分類のコーディング法の検証	15
4.1.1 実験データ	15
4.1.2 評価方法	16
4.1.3 結果比較	17
4.1.4 考察	19
4.2. 実験 2 マルチラベルコーディング法の検証	20
4.2.1 実験データ	20
4.2.2 結果比較	21
4.2.3 考察	22
4.3. 実験 3 損失関数と出力層の活性化関数の組み合わせ	23
4.3.1 結果比較	23
4.3.2 考察	27
5. 結論	29
謝辞	30
参考文献	31

1. はじめに

深層学習は機械学習の手法の1つであり、現在では画像分類や音声認識、翻訳など様々な分野で大きな成果を生み出している。近年では、医療分野においても深層学習による画像分類を利用した自動画像診断支援の研究が盛んに行われている。その中で胸部 X 線画像の画像分類は重要な課題の1つである。胸部 X 線画像の画像分類では、分類結果の精度の向上を見込み、データセットやニューラルネットワークの改良が行われてきた。

例えば、Wang ら[1]はテキストマイニングとよばれる手法によって胸部疾患に関連する放射線レポートから疾患を探し出し、胸部 X 線画像に自動で疾患名のラベル付けを行い、ChestX-ray8 と呼ばれるデータセットを作成した。更に、作成したデータセットを使用することで既存のデータセットを使用した場合と比較して分類精度が向上することを示した。

Majdi ら[2]の研究では、DenseNet121 とよばれるニューラルネットワークを用いて胸部 X 線画像から特徴を抽出し、2種類の疾患を分類した。この際に、疾患ごとの画像データの不均衡に対処するために、重み付き 2 値交差エントロピーを使用した。これにより分類精度が向上した。

本研究では複数の胸部疾患を含む胸部 X 線画像を使用した、深層学習によるマルチラベル付き画像分類について検証する。画像分類とは何らかの画像から、その画像自身またはその画像の中にある対象物が何なのかをクラスごとに分類することである。深層学習による画像分類には 2 値分類、マルチクラス分類、マルチラベル分類がある。2 値分類はあるクラスに属するか属さないかの 2 つに分ける分類問題、マルチクラス分類は 2 つ以上のクラスのいずれかのみに分ける分類問題、マルチラベル分類は 1 つの対象物が複数のクラスに同時に分けられる分類問題である。2 値分類やマルチクラス分類は、マルチラベル分類と区別してシングルラベル分類と呼ばれる。マルチラベル分類はシングルラベル分類と比較して、分類結果の可能な組み合わせが多く、学習が難しい。更に、マルチラベル分類では学習時に使用する損失関数と出力層の活性化関数をどのように組み合わせるべきか明確ではなく、マルチラベル付き画像をどのよう

に扱い、深層学習に利用するか十分な検証がなされていない。例えば、Wang ら [1]の研究では単一疾患画像を利用したシングルラベル分類、Majdi ら [2]の研究では2つの疾患画像を比較した2値分類に関して実験されており、複数の疾患を含む画像を使用したマルチラベル分類に対する検証は十分ではない。医療画像は複数の疾患を含むことが多く、十分な量のマルチラベル付き画像データが存在することに着目し、これらのデータセットを利用してマルチラベル分類に関する検証を行う。

本研究では、2値分類のコーディング法、マルチラベル分類のコーディング法、損失関数と出力層の活性化関数の組み合わせという3つの観点から比較評価を行う。1つ目の2値分類のコーディング法では、マルチクラス分類、マルチラベル分類で最も単純な、クラス数が2の場合について比較評価する。2つ目のマルチラベル分類のコーディング法では、マルチラベル分類問題をマルチクラス分類問題に置き換えた手法と比較評価する。3つ目の損失関数と出力層の活性化関数の組み合わせでは、2値分類やマルチクラス分類で使用されている種々の組み合わせをマルチラベル分類に適用して、比較評価する。

本論文の構成は次の通りである。第2章では、深層学習による画像分類の解説と従来研究、胸部 X 線画像のデータセットと使用したニューラルネットワークモデル ResNet50 について述べる。第3章では、マルチラベル付き画像分類の検証について具体的に述べる。第4章では、検証したマルチラベル付き画像分類の評価について言及する。最後に第5章では、結論と今後の課題を述べる。

2. 深層学習による胸部 X 線画像の画像分類

2.1. 深層学習による画像分類

深層学習とは人間の神経細胞の仕組みを再現したニューラルネットワークを用いた機械学習の手法の 1 つであり、特徴として多層構造であることが挙げられる。深層学習は多くの分野で活用されており、画像分類に関する研究も盛んに行われている。

He[3]らの研究では、従来よりも大幅に深いニューラルネットワークの学習を可能にした ResNet (Residual Network) を提案した。画像分類問題においてネットワークの層の深さは重要であり、深いほど精度が向上すると考えられているが、実際には精度劣化が生じる問題があった。しかし、ResNet は残差を用いた学習を行い、スキップ接続と呼ばれる変換元の情報を保持するような接続を用いることで深いネットワークの学習に成功し、2015 年度の ILSVRC と呼ばれる国際コンペティションにおいて大幅な精度改善を達成した。評価には ImageNet[4]のデータを用いており、18 層と 34 層のネットワークを比較した結果、34 層のネットワークの方が 2.8%精度向上した。

また、深層学習においては損失関数と出力層の活性化関数の組み合わせが重要となっている。損失関数は、機械学習により算出した予測値と、実際の正解値のズレを計算するための関数である。実際の正解値と予測値の距離や差、違いの総量や平均を表す関数であり、最適なパラメータを決定する際に使用される。活性化関数は、ニューラルネットワーク内で次の層に遷移する際に値を変換させる関数である。出力層は最終的に出てきた値が求める値かどうかを分類する層となっているため、それまでとは違う活性化関数を利用する。この損失関数と出力層の活性化関数の組み合わせが適切でない場合、学習が進まなかったり学習の速度が遅くなったりする。そのため、学習する問題に合わせて適切な組み合わせを選ぶ必要がある。

2.2. 胸部 X 線画像のデータセット

画像分類は様々な分野において活用されている。その内の 1 つは医療分野における胸部 X 線画像の疾患画像分類診断である。胸部 X 線画像は多くの疾患の診断で広く利用されており、毎年の健康診断で何百万枚もの画像が取得されている。このような膨大な数の X 線画像と、その画像に対する所見データを含んだ放射線検査レポートが蓄積されている。その中で、匿名化された胸部 X 線画像を特定の胸部疾患についてまとめて、多くの人が利用できるデータセットとして公開されているものがある。従来の研究では CheXpert[5]、PadChest[6]、Chest-Xray14[7]と呼ばれる胸部 X 線画像で構成されているデータセットが使用されている。

本研究では、Chest-Xray14 と呼ばれる、アメリカ国立衛生研究所 (NIH) から提供されている約 100,000 枚の胸部 X 線画像データの入ったデータセットを使用し、深層学習による画像分類の検証を進めていく。Chest-Xray14 には疾患のない正常状態を表す画像である「No Finding」を含め、Atelectasis(無気肺)、Cardiomegaly(心肥大)、Consolidation(浸潤影)、Edema(浮腫)、Effusion(胸水)、Emphysema(肺気腫)、Fibrosis(線維症)、Hernia(ヘルニア)、Infiltration(浸潤)、Mass(腫瘍)、Nodule(結節)、Pleural-Thickenig(胸膜肥厚)、Pneumonia(肺炎)、Pneumothorax(気胸)の 14 個の疾患を含んだ胸部 X 線画像が含まれている。また、これらの疾患を複数個持つ胸部 X 線画像も存在し、マルチラベル付き画像として提供されている。本研究ではマルチラベル付き画像を使用した画像分類を主として研究を進めていくので、Chest-Xray14 のデータセット使用が適切であると考えた。表 1 に Chest-Xray14 における No Finding を含めた各疾患のシングルラベル付き画像枚数、表 2 にマルチラベル付き画像枚数 (一部の組み合わせ) を示す。

表 1 : シングルラベル付き画像枚数

ラベル名 (疾患名)	画像枚数	ラベル名 (疾患名)	画像枚数	ラベル名 (疾患名)	画像枚数
No Finding	60,362	Effusion	3,955	Mass	2,139
Atelectasis	4,215	Emphysema	892	Nodule	2,705
Cardiomegaly	1,093	Fibrosis	727	Pleural- Thickenig	1,126
Consolidation	1,310	Hernia	110	Pneumonia	322
Edema	628	Infiltration	9,547	Pneumothorax	2,194

表 2 : マルチラベル付き画像枚数 (一部の組み合わせ)

ラベル名 (疾患名)	画像枚数	ラベル名 (疾患名)	画像枚数	ラベル名 (疾患名)	画像枚数
Atelectasis Infiltration	1,350	Effusion Mass	402	Edema Infiltration	392
Atelectasis Effusion	1,167	Infiltration Mass	420	Atelectasis Consolidation	398
Effusion Infiltration	1,604	Mass Nodule	394	Cardiomegaly Effusion	484
Infiltration Nodule	829	Infiltration Pneumothorax	345	Edema Effusion	189
Consolidation Infiltration	441	Emphysema Pneumothorax	337	Effusion Pneumothorax	403

2.3. 深層学習による胸部 X 線画像分類の従来研究

従来研究として、Wang ら[1]のテキストマイニングによるラベル付けが行われた画像を使用した ChestX-ray8 と呼ばれるデータセットの提案がある。この ChestX-ray8 には、Atelectasis(無気肺)、Cardiomegaly(心肥大)、Effusion(胸水)、Infiltration(浸潤)、Mass(腫瘍)、Nodule(結節)、Pneumonia(肺炎)、Pneumothorax(気胸)の 8 つの胸部疾患を持つ胸部 X 線画像が含まれている。

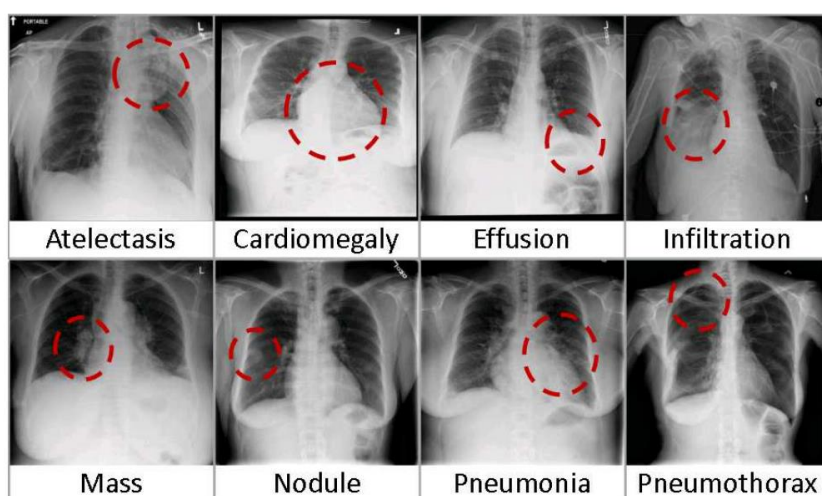


写真 1：8つの胸部疾患の胸部 X 線画像 [1]

テキストマイニングとは、文章データから自然言語処理を使用して出現頻度や相関関係を分析することで有益な情報を取り出す手法である。これを利用し、胸部疾患に関連する放射線レポートから抽出したい疾患を探し出し、胸部疾患画像により適切な疾患名のラベル付けを行った。更に、疾患画像が少ないものに関しては手作業でラベル付けを行った。また、ヒートマップ図を作成して疾患位置の特定を可能にした。これにより既存のデータセットを使用した場合と比較して分類精度が向上した。

Wangらは更に、上記の技術で作成したChestX-ray8データセットに対して、種々のニューラルネットワークによる分類実験を行い、性能を比較している。AlexNet、GoogLeNet、VGGNet-16、ResNet-50の4つのニューラルネットワークを使用して実験を行った。結果として、MassのみはAlexNetを使用した学習結果が最も高い結果を示したが、それを除く7つの疾患においてはResNet-50を使用した学習が最も分類精度が高くなることが分かった。

Majdiら[2]の研究では、DenseNet121とよばれるニューラルネットワークを用いて胸部X線画像から特徴を抽出し、肺結節と心肥大の2種類の疾患を分類した。DenseNetはネットワークの各層の出力を後続するすべての層に接続し、ネットワークの各層が先行するすべての層の出力を入力として受け取る構造になっている。これにより、勾配が消失する問題が緩和され、非常に大きな

データセットを学習するのに必要なメモリ量が大幅に削減できる。この DenseNet の全結合層をグローバル平均プーリングに置き換えたことで2種類の疾患分類に適応させた。また、Majdi らは、Irvin[5]が CheXpert における肺結節のデータ数が他の疾患と比べて少なく、データの不均衡が生じる事を理由に研究対象から除外した点に着目した。Majdi らは、疾患ごとの画像データ数の不均衡への耐性を評価するために、画像データ数に差がある心肥大と肺結節の2値分類を行った。実際に使用した訓練用と評価用の画像データ数は、肺結節が合わせて 8,365 枚なのに対し、心肥大は合わせて 30,493 枚である。画像データ数の不均衡への対処法として、重み付き損失関数の使用を提案し、分類精度が向上することを確認した。

2.4. マルチラベル分類

深層学習による画像分類問題には、2値分類、マルチクラス分類、マルチラベル分類がある。2値分類はあるクラスに属するか属さないかの2つに分ける分類問題、マルチクラス分類は2つ以上のクラスのいずれかのみに分ける分類問題、マルチラベル分類は1つの対象物が複数のクラスに同時に割り当てられる分類問題である。また、2値分類やマルチクラス分類は、マルチラベル分類と区別してシングルラベル分類と呼ばれる。シングルラベル分類は総じて、1つの対象物を必ず1つのクラスだけに分ける分類問題である。

図1と図2はそれぞれシングルラベル分類とマルチラベル分類の例である。縦軸を画像名、横軸をクラス名としている。「0」のときはそのクラスに属さないことを示し、「1」のときはそのクラスに属することを示している。図1のシングルラベル分類では、画像 X1 はクラス y2 のみ、X2 は y4 のみ、X3 は y1 のみ、というようにただ一つのクラスに分類される。図2のマルチラベル分類では、画像 X1 はクラス y2 のみだが、画像 X2 はクラス y2 とクラス y4、X3 は y1 と y4、というように、複数のクラスに分類される。

画像	y1	y2	y3	y4
X1	0	1	0	0
X2	0	0	0	1
X3	1	0	0	0

図1：シングルラベル分類

画像	y1	y2	y3	y4
X1	0	1	0	0
X2	0	1	0	1
X3	1	0	0	1

図2：マルチラベル分類

3. マルチラベル付き画像分類の検証

3.1. マルチラベル付き画像分類

マルチラベル分類では2値分類をクラスの数だけ繰り返すことが基本的な解法となっている。つまりクラス数を N とした時、 2^N 通りの組み合わせが存在する。しかし、シングルラベル分類では複数のクラスの中から1つの正解を選べばよいので、 N 通りの組み合わせで済む。このように、マルチラベル分類はシングルラベル分類と比較して、分類結果の可能な組み合わせが多く、学習が難しい。更に、マルチラベル分類では、学習時に使用する損失関数と出力層の活性化関数をどのように組み合わせれば、より効率良く学習が上手く進むのか明確ではなく、マルチラベル付き画像をどのように扱い、深層学習に利用するか十分な検証がなされていない。

3.2. 検証項目

本研究では、以下の3つの観点から、複数の胸部疾患を含む胸部X線画像を使用したマルチラベル付き画像分類の検証を行う。

- 2値分類のコーディング法
- マルチラベル分類のコーディング法
- 損失関数と出力層の活性化関数の組み合わせ

1つ目の2値分類のコーディング法では、マルチクラス分類、マルチラベル分類で最も単純なクラス数が2の場合について評価する。2つ目のマルチラベル分類のコーディング法では、マルチラベル分類問題をマルチクラス分類問題に置き換えた手法について評価する。3つ目の損失関数と出力層の活性化関数の組み合わせでは、2値分類やマルチクラス分類で使用されている種々の組み合わせをマルチラベル分類に適用して、比較評価する。

以上の手法毎の詳細については、次項より順次解説していく。

3.3. 2値分類のコーディング法

2値分類のコーディング法として、以下の2つを比較する。

- 1出力でのコーディング
- 2出力でのコーディング

1出力でのコーディングでは、出力が「1」の場合は疾患あり、「0」の場合は疾患なし、というように、1出力で疾患の有無を判定する方法を利用して2値分類を行う。マルチラベル分類では一般的に、1出力を複数クラス分行って分類するという解法になっているが、出力の数を1つにすることで、最も単純なマルチラベル分類を行い、結果として2値分類に適応させている。

2出力でのコーディングでは、出力を2次元ベクトルで表す。第1要素を疾患あり、第2要素を疾患なしとすると、出力が(1,0)の場合は疾患ありのクラス、(0,1)の場合は疾患なしのクラスとなる。マルチクラス分類では一般的に、この出力の数をクラス数分だけ用意して分類するという解法になっている。ここではクラス数を2として2値分類を行う。

まとめると、マルチラベル分類の手法とマルチクラス分類の手法を、それぞれ2クラスの場合に適用して二値分類を行い、比較評価を行う。

3.4. マルチラベル分類のコーディング法

マルチラベルのコーディング法として以下の2つを比較する

- Multi-Hot ベクトル
- One-Hot ベクトル

マルチラベル分類では、Multi-Hot ベクトルを用いたコーディング法と、One-Hot ベクトルを用いてマルチラベル分類問題をマルチクラス分類問題に置き換えたコーディング法の、それぞれを用いて学習を行った際の比較評価を行う。

出力をベクトルとし、ベクトルの各要素が各クラスに対応していると考え。この時、要素が「1」の時はそのクラスに属し、「0」の時は属さないとする。

マルチラベル画像は複数の要素が「1」となるため、これを Multi-Hot ベクトルとよぶ。図3のように、「疾患名」という要素を Cardiomegaly、Edema、Pneumothorax の3クラスに分けて、0か1かでベクトル表現する。つまり、3.3節で述べた1出力のコーディング法を、3クラス数分について同時に出力し、3つの疾患のマルチラベル分類を行う。図3では、鈴木はEdema、田中はCardiomegaly、佐藤はPneumothorax、樋口はCardiomegalyとEdema、諸星はEdemaとPneumothoraxに罹患しているという事になる。なお、本研究では、使用するマルチラベル付き胸部X線画像が例えばCardiomegaly|EdemaやPneumothorax、というように2つ以下のラベルがついた画像のみであり、「1」の出力は最大でも2つまでとなる。

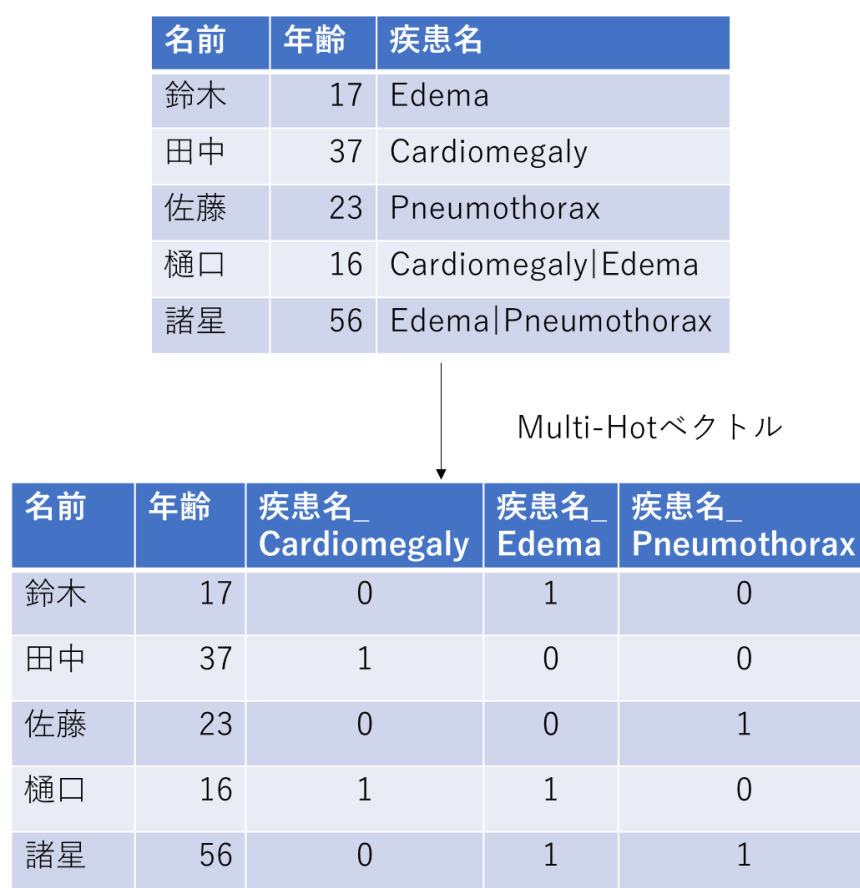


図3：Multi-Hot ベクトル

One-Hot ベクトルは、ベクトルの要素のうち1つのみが「1」となるため、

マルチクラス分類で一般的によく用いられている、この One-Hot ベクトルを利用してマルチラベル分類を行う。One-Hot ベクトルに対応するために、複数の疾患を持つマルチラベル付き画像を1つのクラスとして、マルチクラス分類を行う手法を使用し画像分類を行う。図4のように、「疾患名」というクラスを Cardiomegaly、Edema、Pneumothorax、Cardiomegaly|Edema、Edema|Pneumothorax の5クラスに分けて、0か1かでベクトル表現する。つまり、3.3節で述べた2出力のコーディング法を、クラス数分の5出力に置き換えて、5クラスのマルチクラス分類を行う。図4では鈴木はEdema、田中はCardiomegaly、佐藤はPneumothorax、樋口はEdemaとPneumothorax、諸星はCardiomegalyとEdemaに罹患しているという事になる。

名前	年齢	疾患名
鈴木	17	Edema
田中	37	Cardiomegaly
佐藤	23	Pneumothorax
樋口	16	Cardiomegaly Edema
諸星	56	Edema Pneumothorax

↓ One-Hotベクトル

名前	年齢	疾患名_ Cardiomegaly	疾患名_ Edema	疾患名_ Pneumothorax	疾患名_ Cardiomegaly Edema	疾患名_ Edema Pneumothorax
鈴木	17	0	1	0	0	0
田中	37	1	0	0	0	0
佐藤	23	0	0	1	0	0
樋口	16	0	0	0	0	1
諸星	56	0	0	0	1	0

図4：マルチラベル付き画像に対応した One-Hot ベクトル

本研究では、前述したマルチラベル付き胸部 X 線画像が含まれるデータセットを使用して、Multi-Hot ベクトルを利用した手法と One-Hot ベクトルを利用した手法のそれぞれでマルチラベル付き画像分類を行い、どちらの手法を利用した場合が高い性能が出るのかを確認していく。

3.5. 損失関数と出力層の活性化関数の組み合わせ

損失関数とは、機械学習が算出した予測値と実際の正解値の差を計算するための関数である。活性化関数は、ニューラルネットワーク内で次の層に遷移する際に値を変換する関数である。深層学習においては損失関数と出力層の活性化関数の組み合わせが重要となっている。この損失関数と出力層の活性化関数の組み合わせが適切でない場合、学習が進まなかったり学習の速度が遅くなったりするため、学習する問題に合わせて適切な組み合わせをする必要がある。学習時の基本的な考え方として、低コストで計算できるような「損失関数」と「出力層の活性化関数」の組み合わせが頻繁に用いられる、というものがある。これを踏まえた上で、損失関数と出力層の活性化関数の基本的な組み合わせが存在する[8]。

- 交差エントロピー (CE) + softmax
- 2値交差エントロピー (BCE) + sigmoid
- 平均2乗誤差 (MSE) + linear

以上の3つが多く活用されている組み合わせとなっている。

交差エントロピー (CE) と softmax の組み合わせは、マルチクラス分類を行う際に適した組み合わせであると言われている。交差エントロピーは予測値と正解値の2つの確率分布の差を表した損失関数である。softmax は確率的なベクトル出力をする際によく用いられる活性化関数である。合計が1になるように、ベクトル中の各次元の変数を正規化する関数であり、多変数向き、つまりマルチクラス分類向きである。

2値交差エントロピー (BCE) と sigmoid の組み合わせは、2値分類を行う際に適した組み合わせであると言われている。2値交差エントロピーは交差エントロピーと似ているが、予測値の計算が p か $1-p$ に必ずなるという点において違いがある。sigmoid は S 字型のシグモイド曲線を描く関数であり、入力を0か1かのどちらかに近い値に変換して出力できる。softmax の2値出力版に相当する。また、マルチラベル分類は2値分類をクラス数分だけ繰り返すことから、2値交差エントロピー (BCE) + sigmoid が適した組み合わせであると

予想される。

平均 2 乗誤差 (MSE) と linear の組み合わせは、重回帰 CNN で学習する際に適した組み合わせであると言われている。平均 2 乗誤差は、予測値と正解値の差を 2 乗した値の平均値をとったものである。linear は出力層での値をそのまま出力する関数で、値の変換を行わない。重回帰 CNN など出力層が任意の値を近似する場合に使われる。

上記で述べた損失関数と出力層の活性化関数の組み合わせが一般的であるとされている。そのためか損失関数と出力層の組み合わせを変えて画像分類を行うといった検証や、胸部 X 線画像を使用して、組み合わせ方による分類精度の比較を行うといった検証は十分でない。また、マルチラベル分類ではどの組み合わせ方が最も分類精度が高くなるのか明確ではない。したがって、本研究では、損失関数と出力層の組み合わせ方を変えてマルチラベル付き画像分類を行い、比較評価を行う。

4. 実験

学習の手順として、最初にシングルラベルのみの胸部 X 線画像を使用し 2 値分類を行った後、マルチラベル付き画像を含んだマルチラベル分類を行っていく。

4.1. 実験 1 2 値分類のコーディング法の検証

3.3 節で述べた 2 つの手法による 2 値分類の比較評価を行う。2 値分類では、シングルラベル付き画像のみを用いて「No Finding」か「疾患」であるかの画像分類を行う。

- 1 出力でのコーディング
- 2 出力でのコーディング

以上 2 つの手法を用いて、それぞれ学習を進める。どちらの手法も、使用したニューラルネットワークモデルは ResNet-50、バッチサイズは 32、学習率は 0.005、エポック数は 20 とした。また、損失関数と出力層の活性化関数は、1 出力でのコーディングでは 2 値交差エントロピーと sigmoid、2 出力のコーディングでは交差エントロピーと linear を使用している。

4.1.1 実験データ

2.2 節で説明した Chest-Xray14 の胸部 X 線画像のデータセットから、極端に画像枚数に差のある疾患を除外し、4 つの胸部疾患を選択して 2 値分類を行う。学習用に使用した胸部疾患名と対応する画像枚数は表 3 の通りである。

表 3 : 2 値分類時の学習用画像枚数

ラベル名 (疾患名)	画像枚数
Cardiomegaly (心肥大)	928
Edema (浮腫)	512
Effusion (胸水)	2,080
Pneumothorax (気胸)	1,888

評価用データについては学習に使用しなかった画像を用いる。評価用に使用した胸部疾患名と対応する画像枚数は表 4 の通りである。

表 4 : 2 値分類時の評価用画像枚数

ラベル名 (疾患名)	画像枚数
Cardiomegaly (心肥大)	112
Edema (浮腫)	112
Effusion (胸水)	320
Pneumothorax (気胸)	304

また、「No Finding」の画像は、学習用データも評価用データも 2 値分類時に対応する疾患と同数を用いる。

4.1.2 評価方法

本研究では手法毎による画像分類の性能を評価するために、Accuracy (正解率) という評価指標を用いる。指標の定義は以下の通りである。

- TruePositive(TP) = 陽性である疾患に対して陽性であると予測したケース
- TrueNegative(TN) = 陽性である疾患に対して陰性であると予測したケース
- FalsePositive(FP) = 陰性である疾患に対して陽性であると予測したケース
- FalseNegative(FN) = 陰性である疾患に対して陰性であると予測したケース

$$Accuracy(\text{正解率}) = \frac{\text{正解数}}{\text{データ数}} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy は正しく疾患を分類できた割合を示す。Accuracy が高いほど、正確に疾患を分類することができる。この評価指標を使って比較評価を行う。

4.1.3 結果比較

1 出力でのコーディングを使用した 2 値分類の際、閾値設定が必要となる。閾値は、評価用データに対する特定の疾患と No-Finding の Accuracy の差が最小となる値を実験により定める。評価用データを使用し、閾値を変えながら 1 出力でのコーディングによる 2 値分類を行った際の各疾患と No-Finding の Accuracy の変化を図 5 に示す。閾値を 0.0 から 1.0 まで 0.01 ごとに変えた結果をプロットしている。各疾患において、図中に示した最適な閾値を採用した。

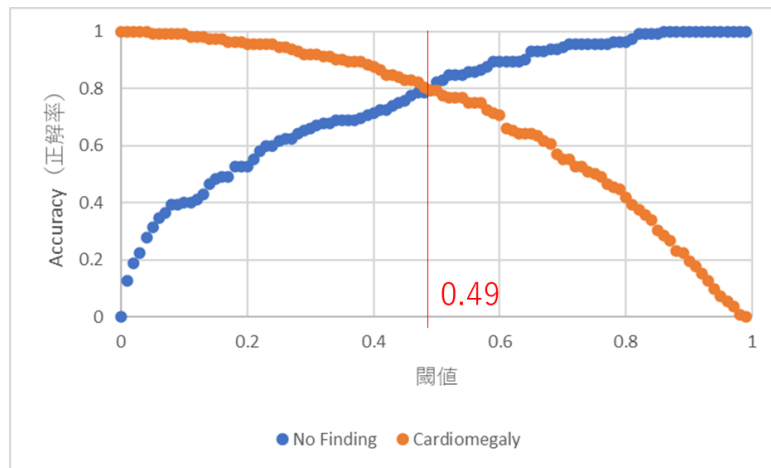


図 5 (a) : Cardiomegaly と No-Finding の Accuracy による閾値設定

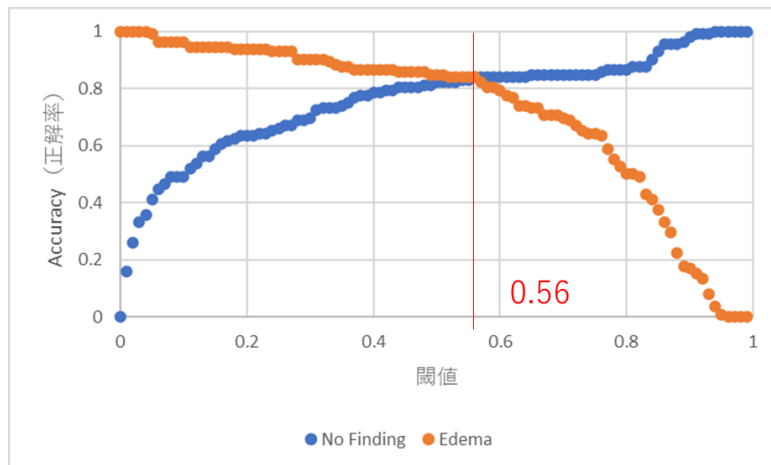


図 5 (b) : Edema と No-Finding の Accuracy による閾値設定

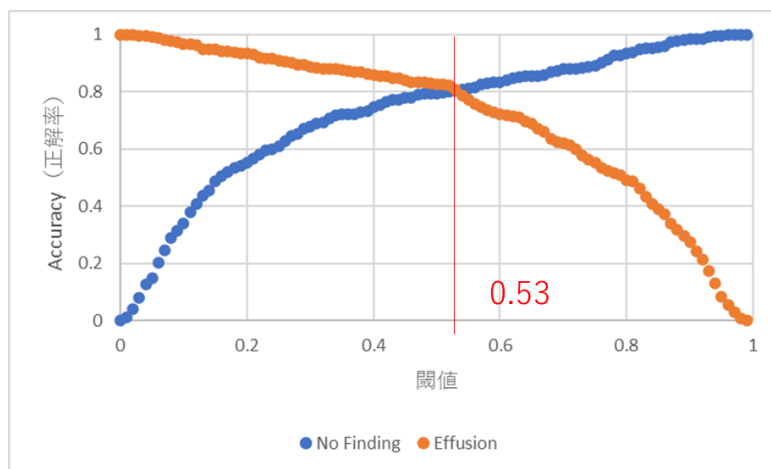


図 5 (c) : Effusion と No-Finding の Accuracy による閾値設定

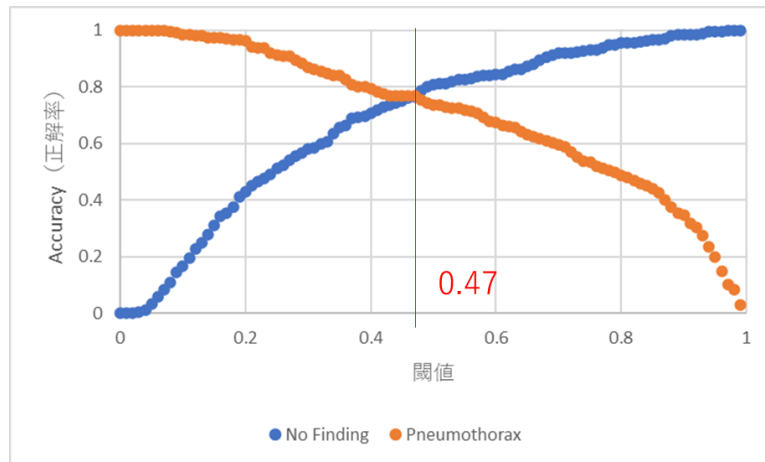


図 5 (d) : Pneumothorax と No-Finding の Accuracy による閾値設定

1 出力でのコーディングによる学習の Accuracy と、2 出力でのコーディングによる学習の Accuracy を表 5 に示す。最終的な Accuracy は疾患と No-Finding それぞれの Accuracy の平均とする。

表 5 : 1 出力での Accuracy と 2 出力での Accuracy 比較

	1出力での Accuracy	2出力での Accuracy
Cardiomegaly	0.795	0.850
Edema	0.839	0.806
Effusion	0.839	0.835
Pneumothorax	0.765	0.800
平均	0.810	0.823

4.1.4 考察

表 5 より、Edema と Effusion については 1 出力でのコーディングを使用して 2 値分類を行った際の Accuracy の方が高い。一方、Cardiomegaly と

Pneumothorax については 2 出力でのコーディングを使用して 2 値分類を行った方の Accuracy が高い。2 値分類を行った疾患の違いによって、1 出力でのコーディングを用いた学習か 2 出力でのコーディングを用いた学習か、どちらが良いのかが変わっている。また、1 出力でのコーディングでは疾患ごとの Accuracy の差が目立つが、2 出力でのコーディングでは全体的に Accuracy が 0.800 を超えている。それぞれの手法毎の Accuracy の平均を見ると、1 出力でのコーディングの場合は 0.810、2 出力でのコーディングの場合は 0.823 となっており、2 出力でのコーディングの方が 0.013 程 Accuracy は高くなっていることが分かる。しかし結果としては Accuracy に大きな差は見られなかった。

4.2. 実験 2 マルチラベルコーディング法の検証

3.4 節で述べた 2 つの手法によるマルチラベル分類の比較評価を行う。マルチラベル分類では、シングルラベル付き画像と共にマルチラベル付き画像を用いて画像分類を行う。

- Multi-Hot ベクトル
- One-Hot ベクトル

以上 2 つの手法を用いて、それぞれ学習を進める。どちらの手法も、使用したニューラルネットワークモデルは ResNet-50、バッチサイズは 32、学習率は 0.005、エポック数は 100 とした。また、損失関数と出力層の活性化関数は、Multi-Hot ベクトルでは 2 値交差エントロピーと sigmoid、One-Hot ベクトルでは交差エントロピーと linear を使用している。

4.2.1 実験データ

2 値分類では 1 クラスごとに画像枚数を統一して学習を行ったが、マルチラベル分類を行う際には画像枚数に差があることで学習に支障が出るのを防ぐため、画像枚数の調整を行う。最終的に学習用に使用した胸部疾患名と対応する画像枚数は表 6 の通りである。

表 6 : マルチラベル分類時の学習用画像枚数

ラベル名 (疾患名)	画像枚数	ラベル名 (疾患名)	画像枚数
No Finding (正常)	769	Cardiomegaly Effusion	411
Cardiomegaly (心肥大)	556	Edema Effusion	160
Edema (浮腫)	559	Effusion Pneumothorax	342
Effusion (胸水)	533		
Pneumothorax (気胸)	672	総計	4,002

評価用データについては学習に使用しなかった画像データを用いる。評価用に使用した胸部疾患名と対応する画像枚数は表 7 の通りである。また、4.3 節で使用するデータも同様のものとする。

表 7 : マルチラベル分類時の評価用画像枚数

ラベル名 (疾患名)	画像枚数	ラベル名 (疾患名)	画像枚数
No Finding (正常)	136	Cardiomegaly Effusion	73
Cardiomegaly (心肥大)	99	Edema Effusion	29
Edema (浮腫)	99	Effusion Pneumothorax	61
Effusion (胸水)	95		
Pneumothorax (気胸)	119	総計	711

4.2.2 結果比較

Multi-Hot ベクトルでのコーディングを使用したマルチラベル分類の際、閾値設定が必要となる。閾値は、評価用データに対する No-Finding を含むすべての疾患の総合的な Accuracy が最大となる値を実験により定める。評価用データを使用し、閾値を変えながら Multi-Hot ベクトルによるマルチラベル分類を行った際の Accuracy の変化を図 6 に示す。閾値を 0.0 から 1.0 まで 0.01 ごと

とに変えた結果をプロットしている。図中に示した最適な閾値を採用した。

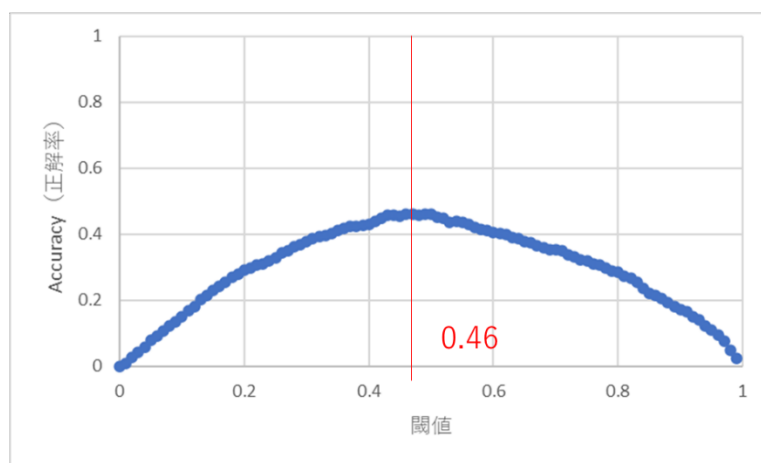


図6：Multi-Hot ベクトルを使用した Accuracy による閾値設定

Multi-Hot ベクトルによる学習の Accuracy と、One-Hot ベクトルによる学習の Accuracy を表8に示す。

表8：Multi-Hot ベクトルと One-Hot ベクトルの Accuracy

	Multi-Hotベクトル	One-Hotベクトル
Accuracy	0.461	0.494

4.2.3 考察

表8より、One-Hot ベクトルを使用してマルチラベル分類を行った際の Accuracy の方が高い。Multi-Hot ベクトルでのコーディングの場合は 0.461、2出力でのコーディングの場合は 0.494 となっており、One-Hot ベクトルでのコーディングの方が 0.033 程 Accuracy は高くなっていることが分かる。3.1 節で述べたように、マルチラベル分類では、クラス数が N の時、出力は 2^N 通りの組み合わせが可能である。しかし、実際のデータですべての組み合わせが生じるとは限らない。本実験でも、疾患のクラス数は 4 であるが、利用したデータセット中でこれらの中の複数のラベルが同時に付与されているデータは 3

通りの組み合わせしかなかった。Multi-Hot ベクトルでは、 $2^4 + 1 = 17$ 通りの組み合わせの出力が可能であるのに対して、One-Hot ベクトルでは使用したデータに合わせて 8 通りの出力のみとなっている。Accuracy の差は、One-Hot ベクトルでの組み合わせの少なさを反映したものと考えられる。

4.3. 実験 3 損失関数と出力層の活性化関数の組み合わせ

3.5 節で述べた損失関数と出力層の活性化関数の組み合わせを変えてマルチラベル分類し、比較評価を行う。実際に実験を行う損失関数と出力層の組み合わせ方は以下の通りである。

- 交差エントロピー (CE) + linear
- 交差エントロピー (CE) + softmax
- 交差エントロピー (CE) + sigmoid
- 2 値交差エントロピー (BCE) + softmax
- 2 値交差エントロピー (BCE) + sigmoid
- 平均 2 乗誤差 (MSE) + linear
- 平均 2 乗誤差 (MSE) + softmax
- 平均 2 乗誤差 (MSE) + sigmoid

以上の組み合わせを全て検証し、どの組み合わせで学習を行った場合が高い分類精度が出るのかを確認していく。すべての組み合わせにおいて、使用したニューラルネットワークモデルは ResNet-50、バッチサイズは 32、学習率は 0.005、エポック数は 100 とした。

4.3.1 結果比較

4.2 節で実験を行った、Multi-Hot ベクトルと One-Hot ベクトルのそれぞれの手法を用いたマルチラベル分類を、活性化関数と出力層の損失関数の組み合わせを変えて比較評価をする。

Multi-Hot ベクトルでのコーディングを使用したマルチラベル分類の際、閾

値設定が必要となる。閾値は、評価用データに対する No-Finding を含むすべての疾患の総合的な Accuracy が最大となる値を実験により定める。評価用データを使用し、閾値を変えながら Multi-Hot ベクトルによるマルチラベル分類を行った際の Accuracy の変化を図7に示す。閾値を 0.0 から 1.0 まで 0.01 ごとに変えた結果をプロットしている。活性化関数と出力層の損失関数の組み合わせ毎に、図中に示した最適な閾値を採用した。

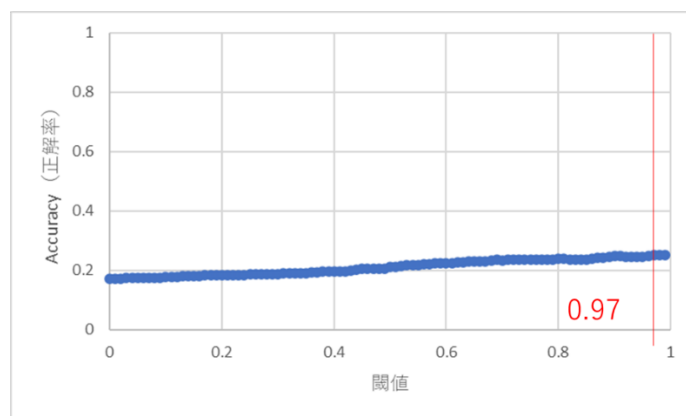


図7(a)：交差エントロピー + linear の場合の Accuracy による閾値

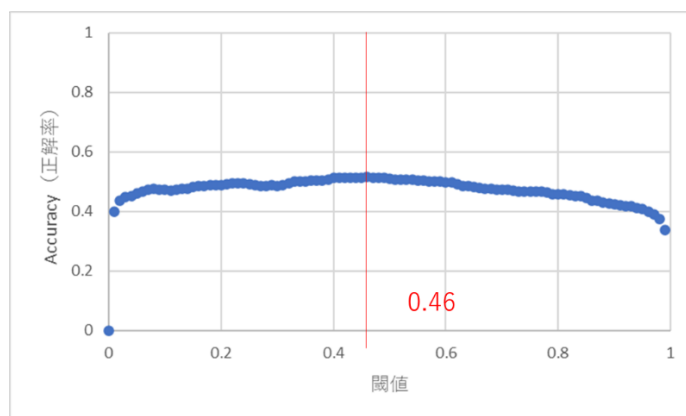


図7(b)：交差エントロピー + softmax の場合の Accuracy による閾値

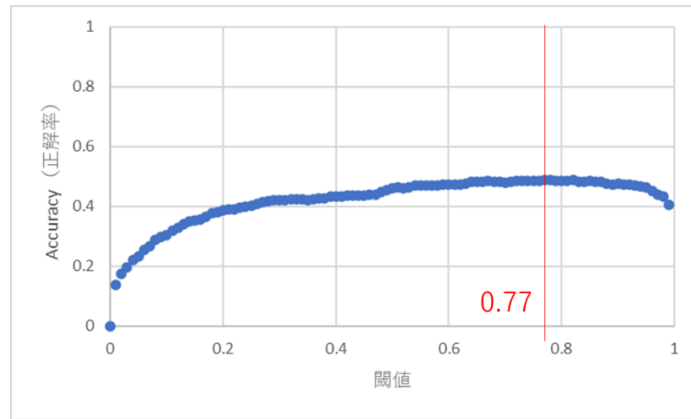


図 7 (c) : 交差エントロピー + sigmoid の場合の Accuracy による閾値

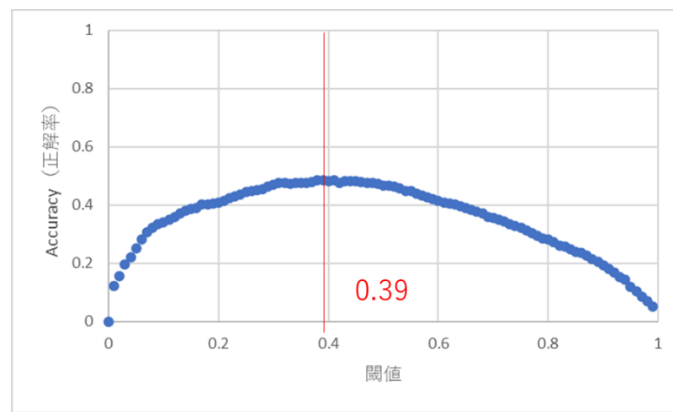


図 7 (d) : 2 値交差エントロピー + softmax の場合の Accuracy による閾値

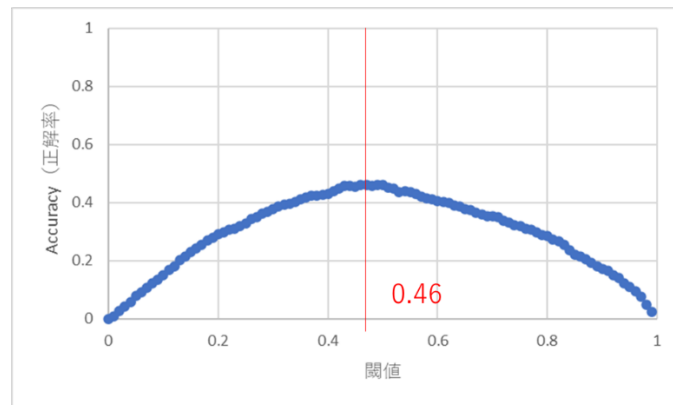


図 7 (e) : 2 値交差エントロピー + sigmoid の場合の Accuracy による閾値

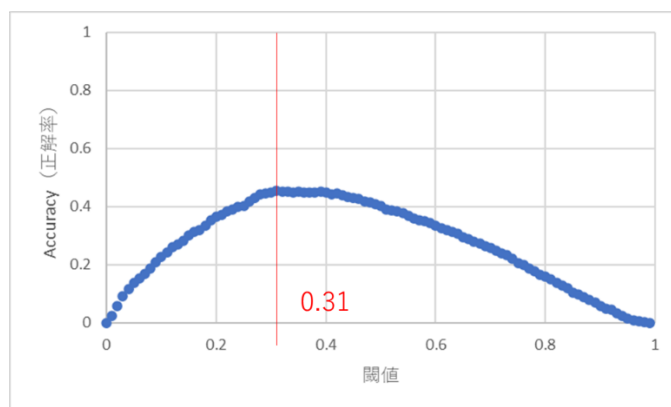


図 7 (f) : 平均 2 乗誤差 + linear の場合の Accuracy による 閾値

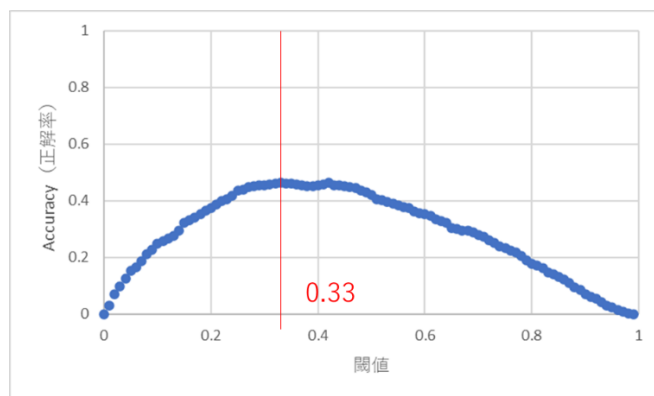


図 7 (g) : 平均 2 乗誤差 + softmax の場合の Accuracy による 閾値

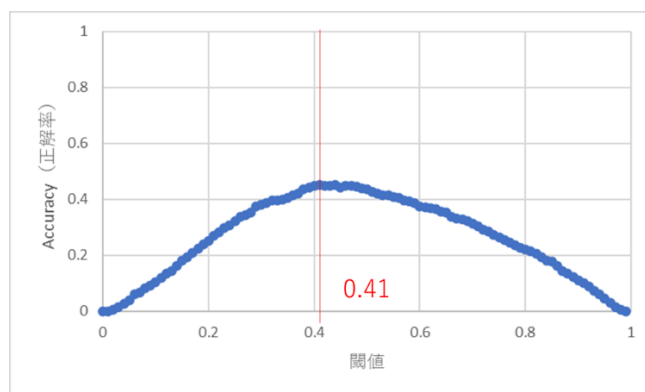


図 7 (h) : 平均 2 乗誤差 + sigmoid の場合の Accuracy による 閾値

また、求めた閾値時点での、Multi-Hot ベクトルを用いたマルチラベル分類の Accuracy を損失関数と活性化関数の組み合わせごとに表 9 に示す。さらに、

One-Hot ベクトルを用いたマルチラベル分類の Accuracy を損失関数と活性化関数ごとに表 10 に示す。

表 9 : Multi-Hot ベクトルを用いた
損失関数と活性化関数の組み合わせごとの Accuracy

損失関数と活性化関数	Accuracy	損失関数と活性化関数	Accuracy
交差エントロピー + linear	0.250	2値交差エントロピー + sigmoid	0.461
交差エントロピー + softmax	0.515	平均2乗誤差 + linear	0.454
交差エントロピー + sigmoid	0.488	平均2乗誤差 + softmax	0.463
2値交差エントロピー + softmax	0.487	平均2乗誤差 + sigmoid	0.453

表 10 : One-Hot ベクトルを用いた
損失関数と活性化関数の組み合わせごとの Accuracy

損失関数と活性化関数hy	Accuracy	損失関数と活性化関数	Accuracy
交差エントロピー + linear	0.487	2値交差エントロピー + sigmoid	0.450
交差エントロピー + softmax	0.494	平均2乗誤差 + linear	0.390
交差エントロピー + sigmoid	0.492	平均2乗誤差 + softmax	0.426
2値交差エントロピー + softmax	0.457	平均2乗誤差 + sigmoid	0.429

4.3.2 考察

表 9 より、Multi-Hot ベクトルを用いた損失関数と活性化関数の組み合わせごとの Accuracy では、交差エントロピーと softmax の組み合わせの Accuracy が 0.515 と一番高い。表 10 より、One-Hot ベクトルを用いた損失関数と活性化関数の組み合わせごとの Accuracy でも、交差エントロピーと softmax の組み合わせの Accuracy が 0.494 と一番高い。比較すると、Multi-Hot ベクトルを用いて、損失関数と活性化関数は交差エントロピーと softmax を使用した場合に最も Accuracy が高く、分類精度が高くなることが分かった。

3.5 節では、画像分類において、マルチクラス分類を行う際には交差エントロピーと softmax の組み合わせが適していると述べた。マルチラベル分類を行

う際には2値交差エントロピーと sigmoid の組み合わせが適していると予想した。One-Hot ベクトルはマルチクラス分類と同義なので、3.5節で述べた通り交差エントロピーと softmax の組み合わせが良いことが確認できた。しかし、Multi-Hot ベクトル、つまり2値分類をクラス数分だけ繰り返すマルチラベル分類では、3.5節での予想とは異なり、交差エントロピーと softmax の組み合わせが良いということが実験結果によって示された。交差エントロピーは、出力が「1」となる場合のみを評価し、「0」となる場合は考慮しない。softmax は、出力値の和が1になるような正規化を行っており、かつ、出力が最大となるクラスを強調する効果を持っている。これらは、Multi-Hot ベクトルとの相性が悪いと考えられるが、実験結果からはこの組み合わせが最良となった。本実験では複数のラベルが付与されたデータより、単一ラベルが付与されたデータの方が多かった。Multi-Hot ベクトルでも、単一ラベルが付与されたデータの Accuracy が向上したことで、全体の Accuracy が高くなった可能性がある。データごとの正解率を集計するなど、より詳細な検討が今後必要である。

また、optimizer や学習率等を様々に変えて、最適な値を求めた上での分類精度の比較評価を行う必要がある。

5. 結論

本研究では、複数の胸部疾患を含む胸部 X 線画像を使用した深層学習によるマルチラベル付き画像分類について検証した。2 値分類のコーディング法、マルチラベル分類のコーディング法、損失関数と出力層の活性化関数の組み合わせという 3 つの観点から比較評価を行った。1 つ目の 2 値分類のコーディング法では、1 出力でのコーディングと 2 出力でのコーディングを比較して、2 出力でのコーディングの方の分類精度が高くなっているが、大きな差は見られないことが確認できた。2 つ目のマルチラベル分類のコーディング法では、One-Hot ベクトルでのコーディングと Multi-Hot ベクトルのコーディングを比較して、One-Hot ベクトルでのコーディングの方の分類精度が高くなっていることが確認できた。3 つ目の損失関数と出力層の活性化関数の組み合わせでは、Multi-Hot ベクトルを用いた上で、交差エントロピーと softmax の組み合わせを適用すると分類精度が最も高くなることが確認できた。

今後の課題としては、optimizer や学習率等を様々に変えて最適な値を求め、比較評価を行うことが挙げられる。また、本研究では、Cardiomegaly、Edema、Effusion、Pneumothorax の 4 つの疾患のみを用いて実験を行ったので、その他の疾患画像を学習に使用して比較評価を行うことが今後の課題として挙げられる。

謝辞

本論文の作成にあたり、丁寧な対応とご指導をしてくださった椋木雅之教授に深く感謝いたします。実験や調査で行き詰った際に適切なアドバイスをくださり、更に論文の執筆においても繰り返し添削をしてくださり、誠にありがとうございました。

また、研究室の皆様には研究について日ごろから様々な助言を頂きました。皆様のおかげで充実した時間を研究室で送ることが出来ました。お礼申し上げます。

参考文献

- [1] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald M. Summers. “ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases”, IEEE Conference on Computer Vision and Pattern Recognition, pp.2097–2106, 2017.
- [2] Mohammad S. Majdi, Khalil N. Salman, Nirav C. Merchant, Jeffrey J. Rodriguez, Michael F. Morris. “Deep learning classification of chest x-ray images”, IEEE SSIAP, pp.116-119, 2020.
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. “Deep Residual Learning for Image Recognition”, IEEE Conference on Computer Vision and Pattern Recognition, pp.770–778, 2016.
- [4] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. “Imagenet large scale visual recognition challenge”, arXiv:1409.0575, 2014.
- [5] Jeremy Irvin et al. “CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison”, arXiv: 1901.07031, 2019.
- [6] Aurelia Bustos, et al. “PadChest: A large chest X-ray image dataset with multi-label annotated reports”, arXiv:1901.07441, 2019.
- [7] Imane Allaouzi, Mohamed Ben Ahmed. “A Novel Approach for Multi-Label Chest X-Ray Classification of Common Thorax Diseases”, IEEE Access 2019, 7, pp.64279–64288.
- [8] CNN の損失関数(Loss Function) その(1): 交差エントロピーと MSE | CVML エキスパートガイド (cvml-expertguide.net)