



# 深層学習による臓器損傷検出AIシステムの開発

2024 / 3 / 29

Neuro Dive横浜

野田真

# 目次

P2

はじめに

P3~5

制作の動機・背景・目的

P6~11

制作物の仕様と制作の概要

P12~13

精度指標

P14~16

得られた効果とその評価

P17

考えられる応用例

P18~21

精度向上のために考えられる施策

P22~27

Appendix

P28

参考文献

## はじめに

- 制作者プロフィール：

氏名：野田真

経歴：高校卒業後、接客、介護、テレホンオペレーター、  
企業・省庁での事務職・IT事務職を経て休職。

障害：うつ病

訓練歴：2023年4月Neuro Dive横浜利用開始

- 制作期間：約3ヶ月（265時間）

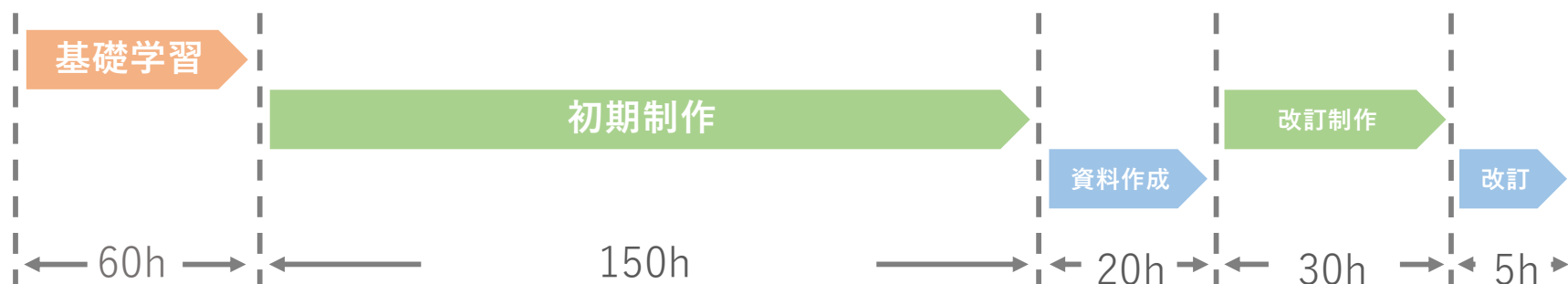
基礎学習：2023年8月22日～9月13日（3週間）

初期バージョン制作：9月13日～11月13日（8週間）

発表資料作成：11月13日～11月30日（2週間）

改訂バージョン制作：2023年12月11日～2024年1月15日（4週間）

発表資料改訂：1月15日～1月18日（1週間）



## 制作の動機：テーマ選びの個人的な動機

個人のウェルビーイングへの貢献：

学んだ技術・知識をなんのために役立てるかを考えたとき、  
人のQOLの向上に直接貢献することが重要だと考えた。  
制作したものが人の暮らしや人生を良くすることが明らかであることが  
制作の内発的な動機として必要だった。  
それに合致するものとして、個人的に親しみのある医療分野から制作テーマを選んだ。



医用診断AIの開発

## 制作の背景：社会的背景

CAD技術の発展と救急医療における迅速診断の必要性：

- CT撮影装置の高性能化とCNNを代表とする深層学習による画像認識技術により、コンピューター診断支援/検出システム（CAD: Computer Aided Detection/Diagnosis）が大きく発展した。
  - 高精度であることに加え迅速な画像診断の可能性が開かれている。
- 一方救急医療においては、初期診断が遅くなるほど、救命率の低下、入院期間の長期化、医療費の増加が起こることが指摘されている。[1]



当成果物では、これらの状況を受け、  
深層学習を用いた迅速かつ高精度なCADe-AIの開発を目指す。

## 制作の目的：提供したい価値

当成果物における目的：

- 医師による傷病状態の診断を補助し、  
診断精度を犠牲にせずに治療方針の決定プロセスを短縮する。  
それによる救急医療における患者の死亡率の低下と予後の改善を目的とする。
- 医療者の業務負担を減らすことで、  
災害など緊急時における医療資源の効率利用の助けとなることを目的とする。



今回の目標はここ

## 制作物の仕様と制作の概要（1）

- AIの仕様：

5つの検査対象（腸、腹腔内出血、肝臓、腎臓、脾臓）について  
損傷の有無を確率値（0~1の間の小数値）で出力する。

- データセット：

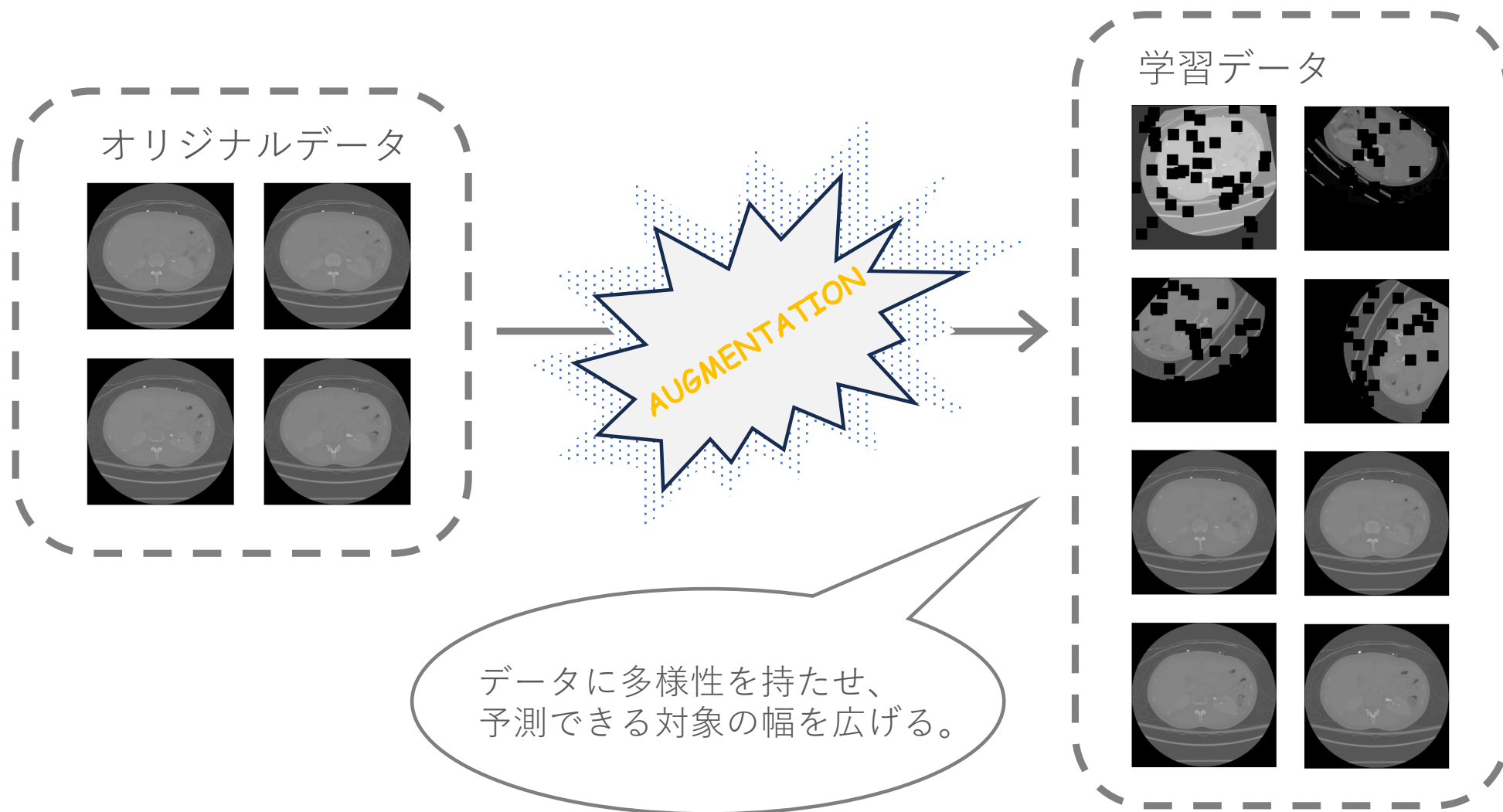
医師による読影で腹部鈍的外傷による臓器損傷が認められたCT画像データから、  
患者221名・9846枚の画像を学習データ、  
患者20名・2183枚の画像をテストデータとした。

→ データ量は少なく、過学習を起こしやすい。

- データの拡張：

オリジナル画像に対して画像加工を施し、  
加工を施した画像にオリジナル画像を合わせて学習データとした。  
（アフィン変換、明度・コントラスト変更、ドロップアウト）

## 制作物の仕様と制作の概要（2）





## 制作物の仕様と制作の概要（3）

- アーキテクチャ：

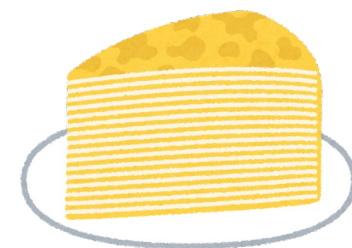
画像認識に用いられるCNNには様々なアーキテクチャ（層構造）が存在する。  
今回は少ないデータ量への対応と速度（計算負荷の小ささ）が必要。



比較検討の結果、MobileNetV2アーキテクチャを採用。  
ImageNetデータセットで学習済みのモデルに転移学習を行った。

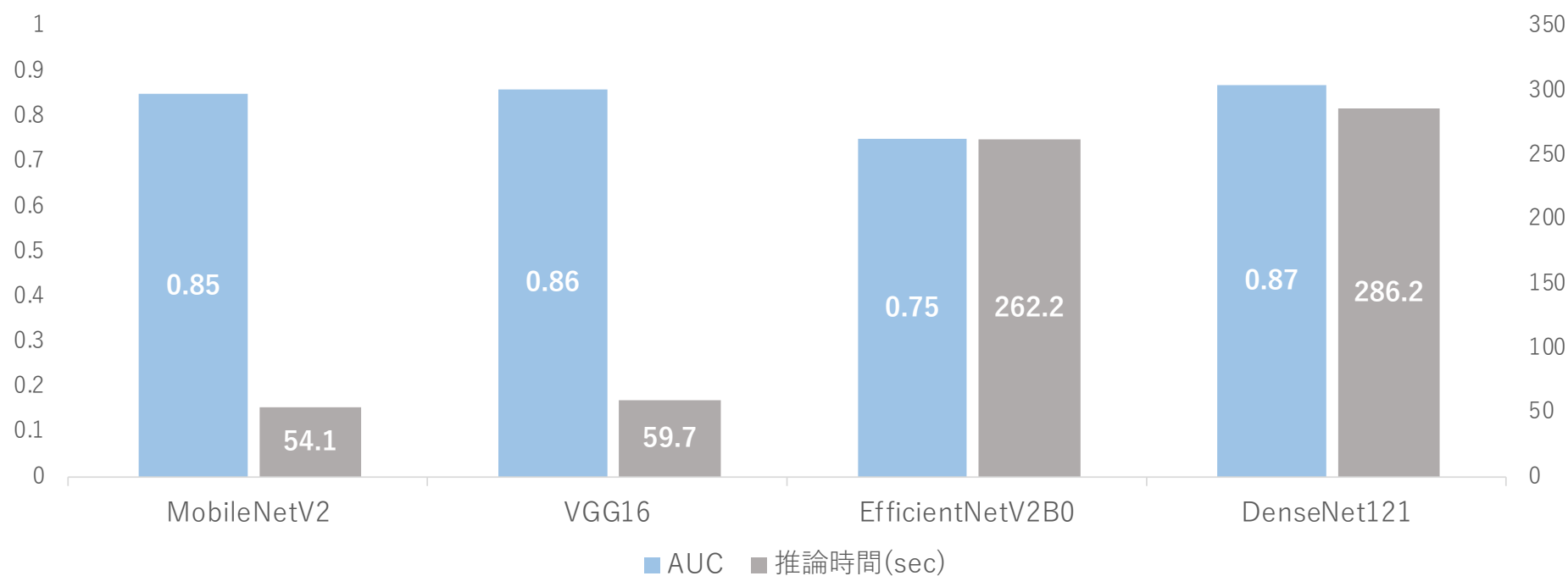
→ MobileNetとは：

スマートフォンなど計算速度が遅い環境に向けて開発された、  
CNNアーキテクチャの一つ。



## 制作物の仕様と制作の概要（４）

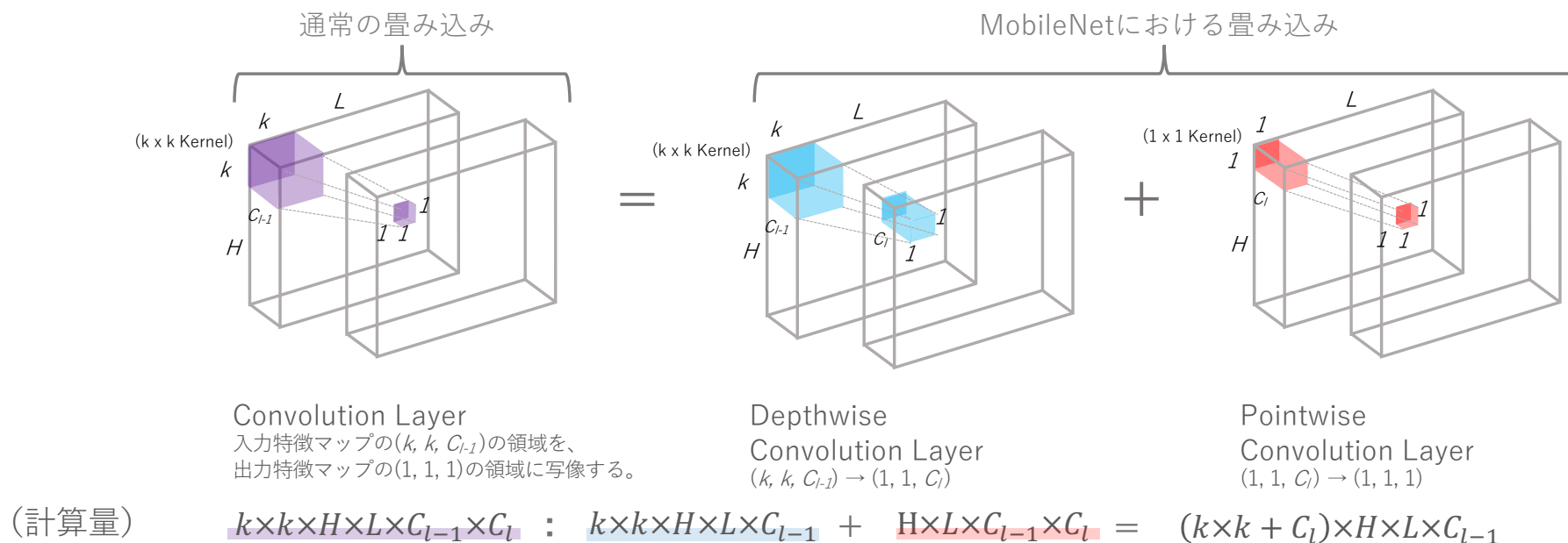
精度と速度のアーキテクチャ間比較：



※腸を対象とした検証用モデルに転移学習を行った  
推論時間はテストデータ100セットの平均

## 制作物の仕様と制作の概要（6）

MobileNetにおける畳み込み：



(計算量の比)

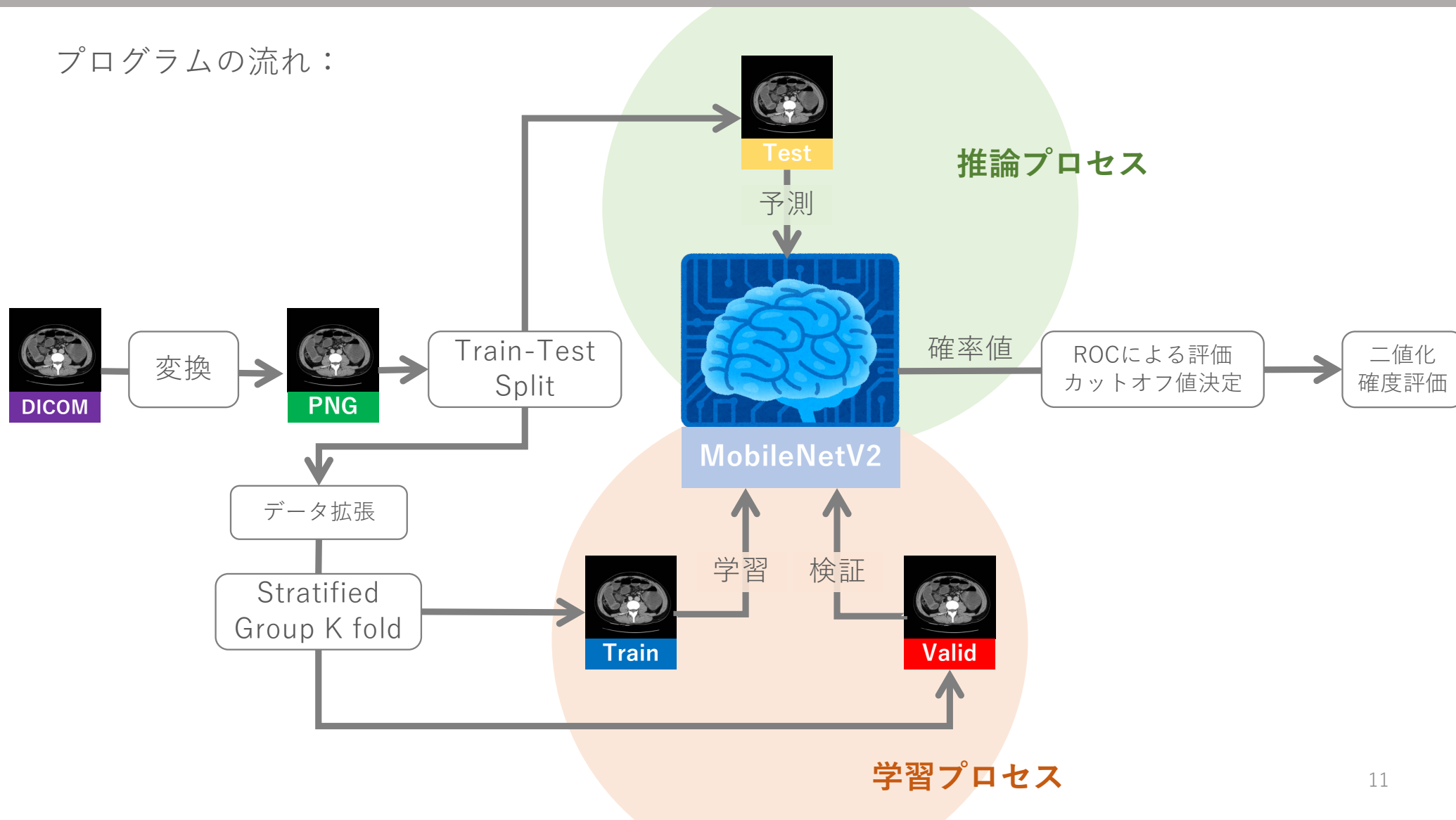
$$\frac{(k \times k + C_l) \times H \times L \times C_{l-1}}{k \times k \times H \times L \times C_{l-1} \times C_l} = \frac{1}{C_l} + \frac{1}{k^2}$$

$C_l = 32, k = 3$  なら、 $\frac{1}{C_l} + \frac{1}{k^2} = \frac{41}{288}$

計算量が約  $\frac{1}{7}$  に

## 制作物の仕様と制作の概要（7）

プログラムの流れ：

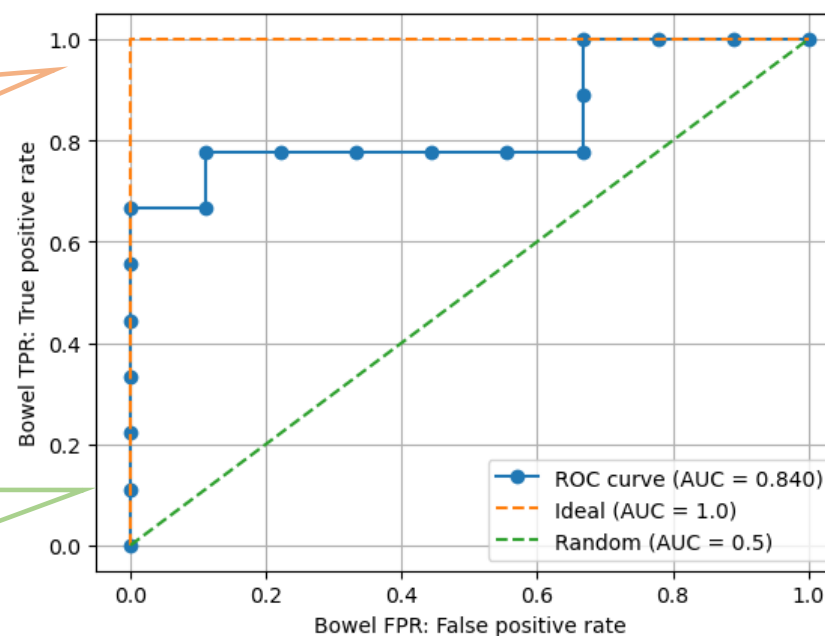


## 精度指標 (1)

- ROC曲線 (Receiver Operating Characteristic curve) :  
偽陽性率をx軸、真陽性率をy軸、閾値を媒介変数としてプロットした点を結ぶ線。
- AUC (Area Under ROC) :  
ROC曲線が囲む面積。予測精度の指標となる。

TPR (真陽性率) が1、  
FPR (偽陽性率) が0となる  
カットオフ値が存在することが理想。  
このときAUC = 1

予測がでたらめなとき、  
AUC = 0.5



## 精度指標（2）

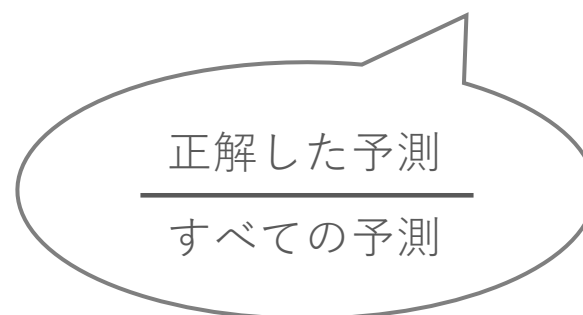
- 閾値（カットオフ値）と確度：

ROCから偽陰性数が0（真陽性率が1）となる閾値で最大の確度を算出した。

	真の値が正	真の値が負
予測が正	真陽性 TP: True Positive	偽陽性 FP: False Positive
予測が負	偽陰性 FN: False Negative (これを0にする)	真陰性 TN: True Negative

$$\text{真陽性率 (感度・Recall)} = \frac{TP}{TP + FN}$$

$$\text{確度 (Accuracy)} = \frac{TP + TN}{TP + FP + TN + FN}$$


$$\frac{\text{正解した予測}}{\text{すべての予測}}$$

## 得られた効果とその評価（１）

精度：

- AUCによる判別対象臓器別の精度評価は、  
腸：0.84、腹腔内出血：0.55、腎臓：0.68、肝臓：0.34、脾臓：0.55であった。
- 偽陰性数が0となる最大のカットオフ値によって陽性/陰性を判定した場合の確度は、  
腸：66%、腹腔内出血：83%、腎臓：50%、肝臓：33%、脾臓：50%であった。



評価：精度は人間の医師による診断に及ばない。

放射線医の読影における見逃し率は通常（15分程度）で10%（確度90%）、  
急いだ読影（15分以下）の場合で27%（確度73%）程度。[1]

→実用性を満たすためには精度向上が必要である

## 得られた効果とその評価（2）

速度：

臓器外傷判別の実行時間は、  
テストデータとしたCTスキャン画像100セットの平均で3分58秒であった。  
(Apple M2 8core GPU Metal performance shaders 環境)



評価：診断のスピードは早い

放射線科医が精度の高い読影をおこなうために必要な読影時間は  
1件あたり15分程度。[2]  
救急医療においてCT検査の発注から検査完了までには  
100分前後の時間がかかる。[1]

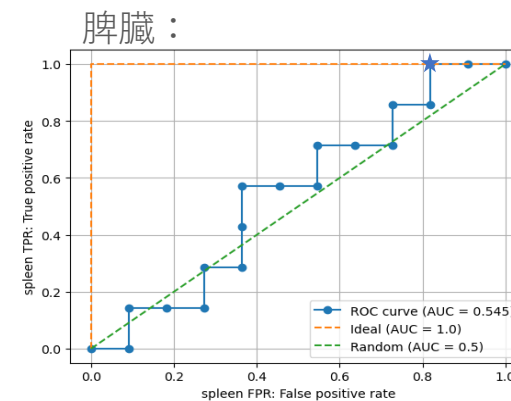
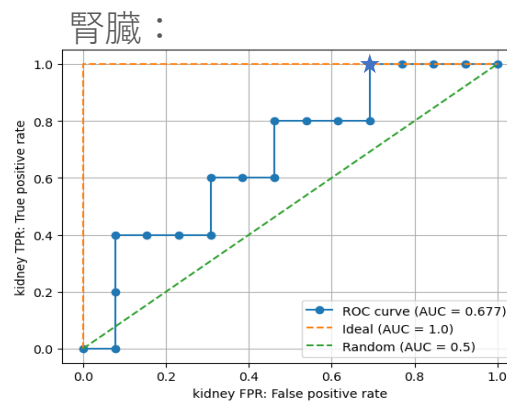
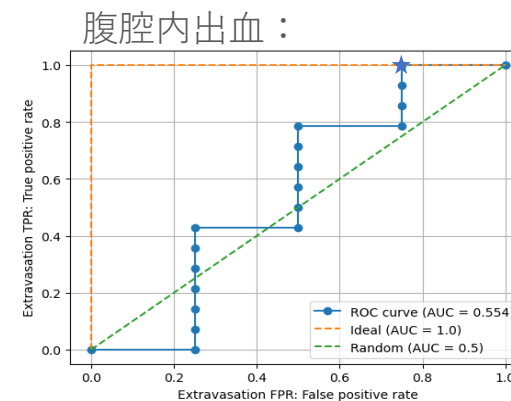
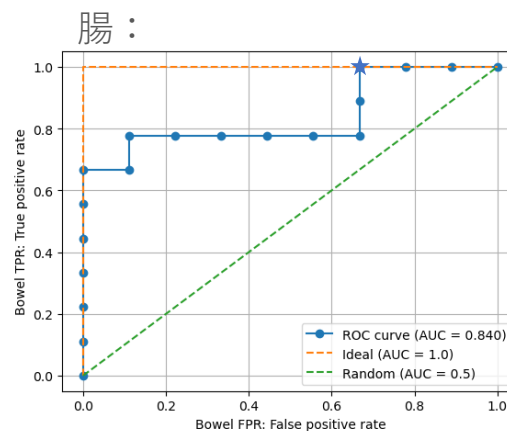
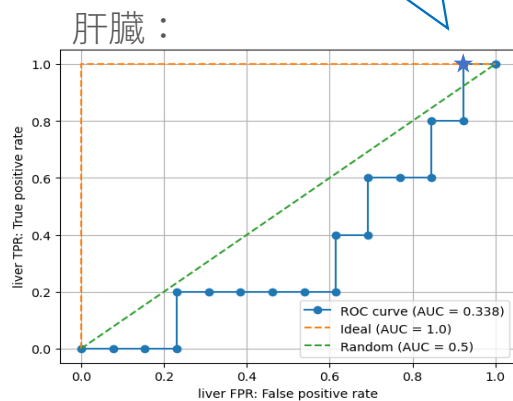
→医療用画像管理システム（PACS）からAIへ、  
データパイプラインを構築し、高速なコンピュータで処理を行うことで、  
さらに迅速な一次診断が可能になるのではないかと。



## 得られた効果とその評価（3）

臓器別ROC曲線：

真陽性率が1で、  
偽陽性率が最小となる点の  
閾値で確度を算出



## 考えられる応用例

- 医師不足・地域間偏在への対応として
  - 通常行われる読影のダブルチェック・トリプルチェックのプロセスの一部をAIによる読影に代替することで医師の作業工数を削減し、へき地等医師が不足する地域での医療の質を確保する。
  - 人間の医師が1日12時間勤務した場合、可能な読影数は48件/日。AIであれば24時間稼働し、最大288件/日の読影を行える。
  - すべての医療機関で同水準の診断が可能。診断精度における属人性の排除。
- 検査コストを下げ、医療費の削減と予防的検査の受診者のQOL向上に貢献する。
  - 日本の放射線科医の平均的な時給はおおよそ1万1千円。[4]  
AIの使用によって医師による読影時間を半分にできれば  
CT検査1件あたり1350円程度の費用削減が可能になる。  
→ 国内のCT検査件数は年間約200万件。[2]  
これらにかかる人件費をすべて半分にできれば年間27億円の医療費を削減できる。
  - 人間ドッグなどヘルスケア産業の拡大への貢献
- 製造業における、X線CTを用いた部品内部欠陥の非破壊検査への応用。

## 精度向上のために考えられる施策（1－1）

- 対象臓器別に異なるモデルを作成する：

今回は実行速度を重視し5つの判別対象を1つのモデルで同時に判別した。  
学習経過を検証した結果、腸に関しては過学習である一方、  
肝臓・腎臓に関しては検証データに対する予測精度が学習データに対する予測精度を上回る  
異常な経過であった。  
後者の現象は学習データの外傷あり/なしのクラス比率が不均衡であったことが  
原因であると考えられる。

→ 過学習対策の学習時パラメータ調整とともに、  
各臓器に対して個別に均衡データでの学習を行うことが必要。

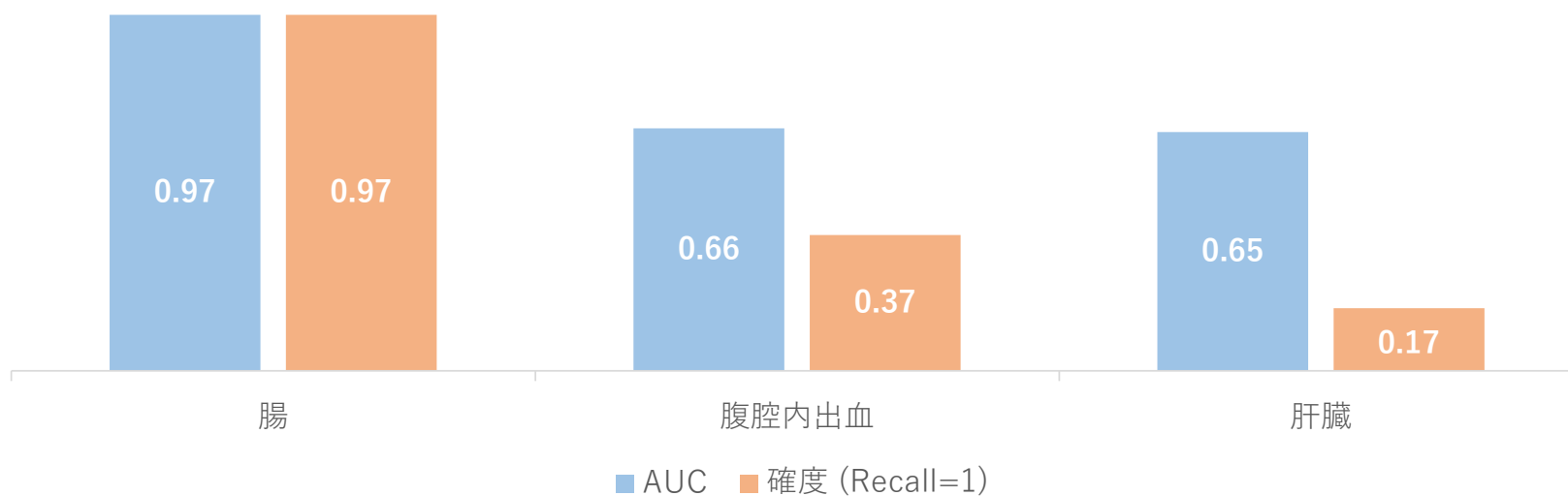


## 精度向上のために考えられる施策（１－３）

- 均衡データを用いて作成した臓器別モデルの精度：
  - 判別対象の各臓器について個別にモデルを作成。  
損傷あり/損傷なしのラベル比率が 1:1 になるよう学習データを取り、  
ファインチューニングを行った。



- 腸で大幅な改善が見られた  
腹腔内出血と肝臓ではAUCは改善したが確度は低下した

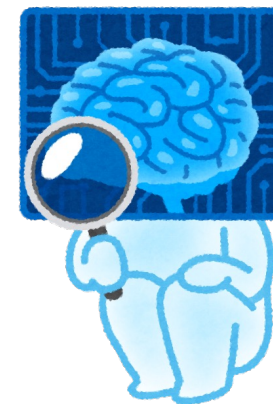


## 精度向上のために考えられる施策（2－1）

- 学習データの画像解像度を高くする：

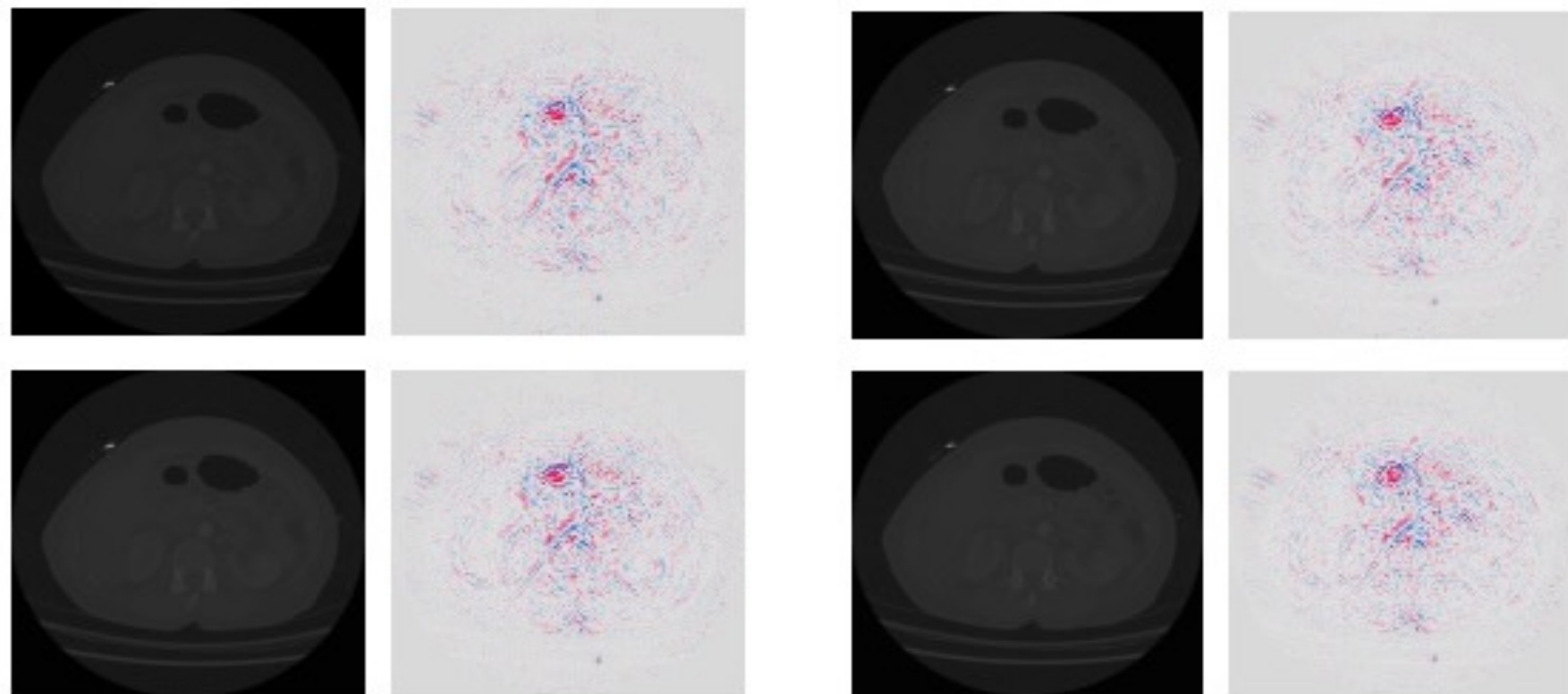
今回は用意できた計算機環境の制限から、  
学習データ・テストデータの画像サイズを 224x224 pixelとした。

→画像中の微細な特徴を識別する必要があるタスクであることから、  
画像サイズをより大きく（512x512, 1024x1024 など）  
することで、精度が向上する可能性がある。



## 精度向上のために考えられる施策（2－2）

- 特徴量部分の可視化：AIは画像のどこを見ているのか



● ： 損傷判定に正の影響を与えている部分    ● ： 損傷判定に負の影響を与えている部分

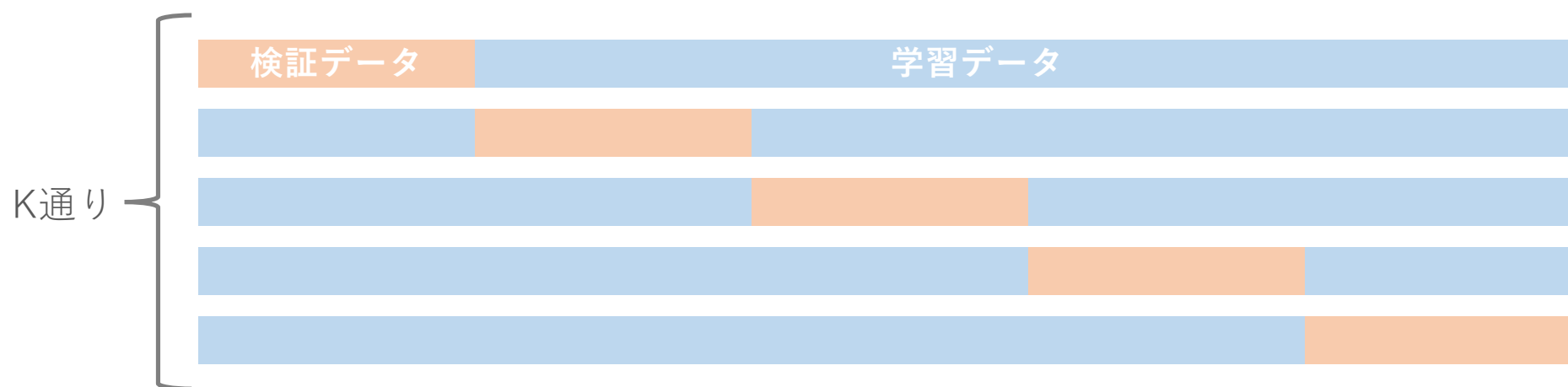
EOF



## Appendix（１）：精度向上の工夫（１）

- クロスバリデーションによるアンサンブル学習：

グループ付き層化K分割によって12通りのパターンで学習データと検証データの分割を行い、それぞれの組み合わせについてモデルを作成して交差検証を行った。  
作成した12個のモデルにテストデータを入力し、出力された確率値の平均を最終出力とした。



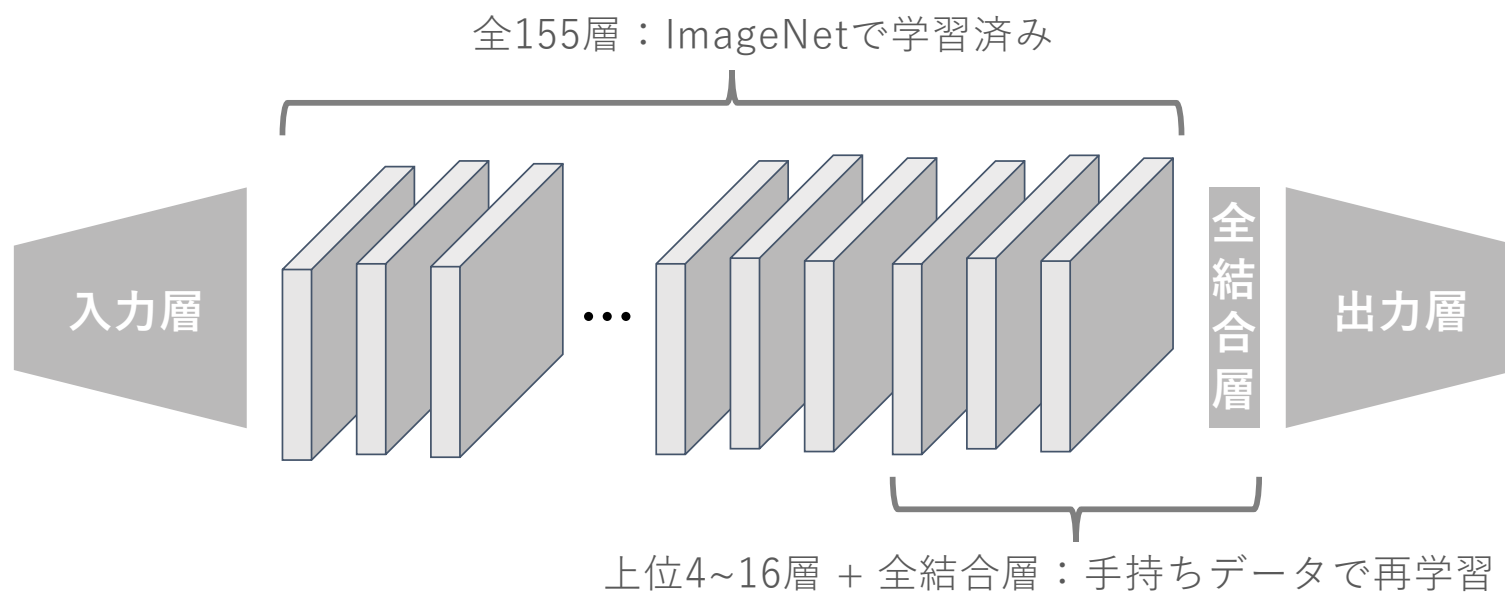
- 同じ患者IDを持つ画像は同じデータに含まれる。
- 検証データと学習データに含まれるラベルの割合はほぼ同じになる。



## Appendix（2）：精度向上の工夫（2）

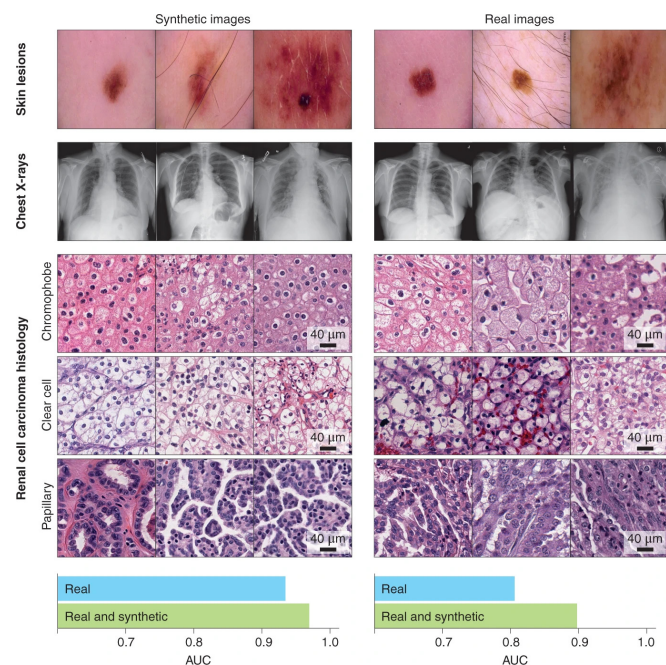
- ファインチューニング：

汎用データセット（ImageNet）で事前学習済みのMobileNetV2アーキテクチャに対し、全155層のうち出力層に近い4～16層のレイヤーを手持ちのデータで再学習させた。



## Appendix (3) : 将来的な精度向上のための施策案 (1)

- 生成AIによるシンセティックデータの利用：
  - データ不足への対策として、生成モデルによって作られた合成データを学習データとして利用することが考えられる。
  - 既存の研究によると、生成AIによる合成データを学習に用いた画像診断モデルではAUCで0.03~0.1程度の精度向上が認められている。[4]



- 医用画像分野においてデータ不足は典型的な問題
- 生成AIによる学習データ生成は近年の技術トレンド

## Appendix（４）：将来的な精度向上のための施策案（２）

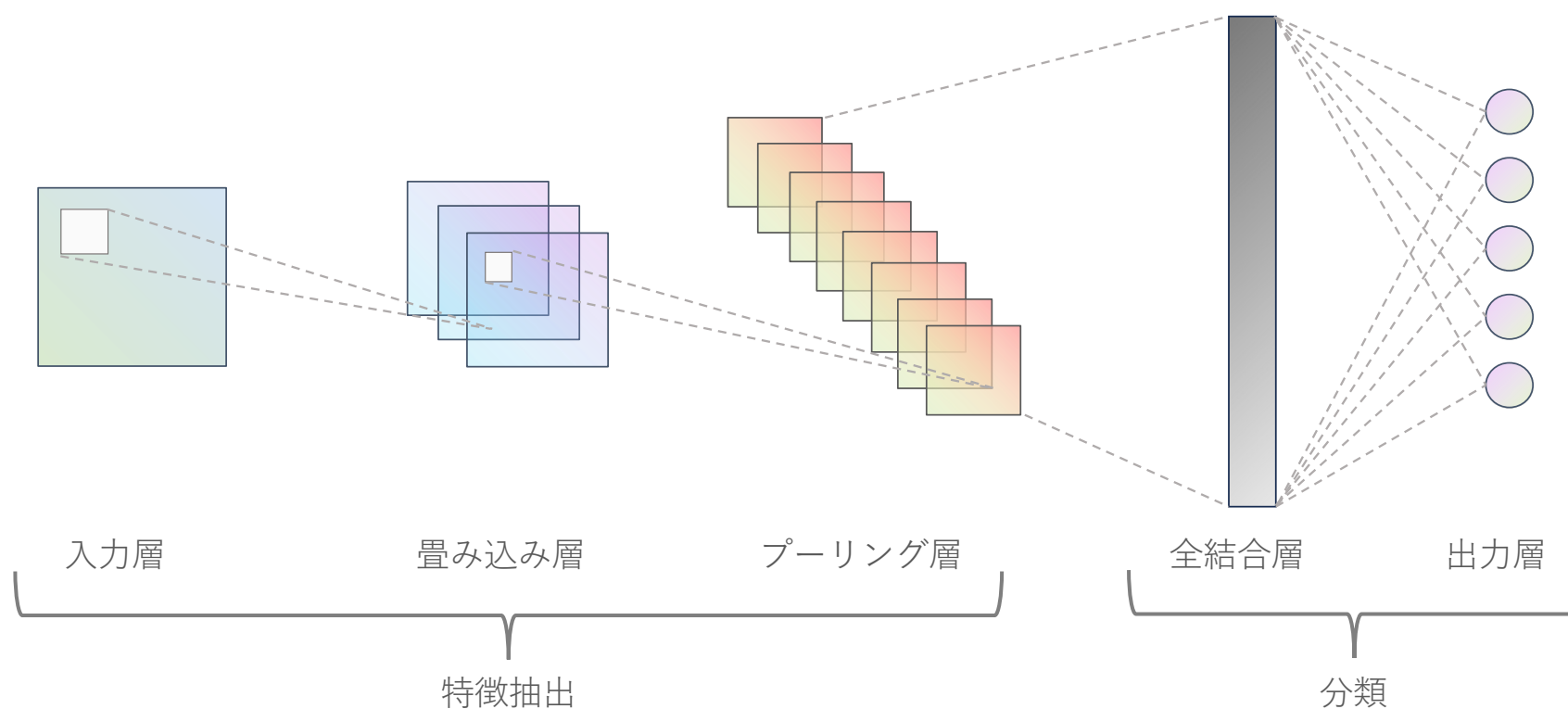
- 連続するCT画像を一つの3次元画像として扱い、  
3次元畳み込みニューラルネットワークによる識別を行う：
  - 肺がん検査において2D-CNNと3D-CNNを比較した研究によると、  
3D-CNNの精度は2D-CNNをAUCで0.007~0.024、確度で2%~6%上回る。[5]
  - 臓器損傷も腫瘍と同じく3次元的な特徴を持つ[6]ため、  
3D-CNNの使用は精度向上に資する可能性がある。

画像サイズを大きくする  
↓  
シンセティックデータの使用  
↓  
3D-CNN  
の順で効果があると考えられる



## Appendix（5）：畳み込みニューラルネットワークの概要

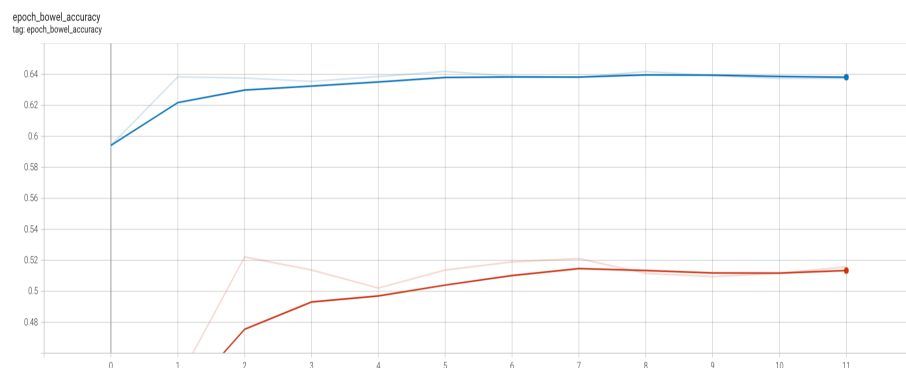
CNN模式図：



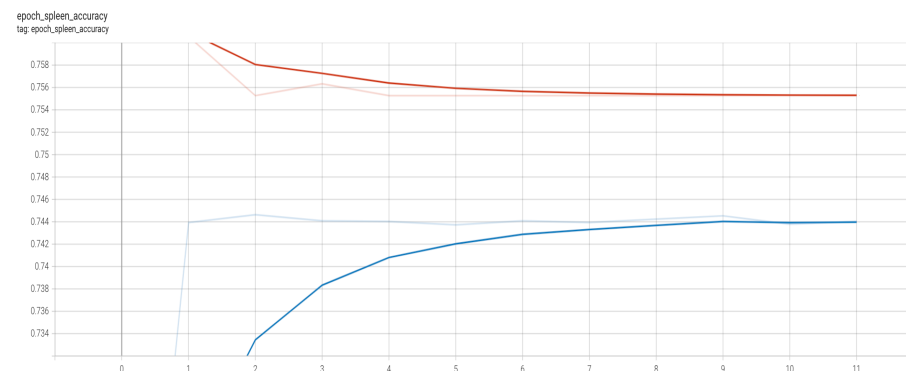
## Appendix（6）：学習曲線・損失曲線

学習曲線・損失曲線：

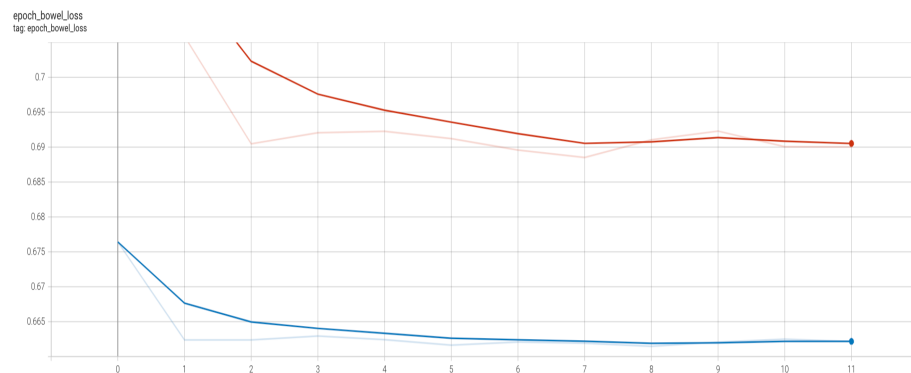
腸：Accuracy



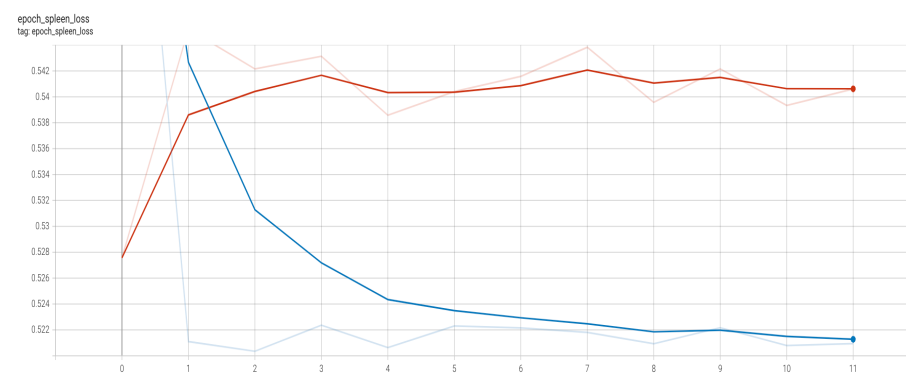
脾臓：Accuracy



腸：Loss



脾臓：Loss



## 参考文献：

- [1] Pratik Rachh, Andrew M Pendley, Phuong-Anh T Duong, Tarek N Hanna, Marta E Heilbrun,  
Decreasing CT Acquisition Time in the Emergency Department through Lean Management Principles,  
RadioGraphics May-June 2021
- [2] 日本放射線科専門医会・医会,  
【提言】医療の質と安全を担保するための読影量について,  
<https://jcr.or.jp/news/recommendations20220216/>
- [3] マイナビDOCTOR,  
放射線科医の年収事情を解説!やりがいや専門医になる方法、年収を上げるコツとは?,  
<https://doctor.mynavi.jp/column/workstyle031/>
- [4] Richard J. Chen, Ming Y. Lu, Tiffany Y. Chen, Drew F. K. Williamson, Faisal Mahmood,  
Synthetic data in machine learning for medicine and healthcare,  
Nature Biomedical Engineering June 2021
- [5] Juezhao Yu, Bohan Yang, Jing Wang, Joseph Leader, David Wilson, Jiantao Pu,  
2D CNN versus 3D CNN for false-positive reduction in lung cancer screening,  
Journal of Medical Imaging September 2020
- [6] 日本外傷学会臓器損傷分類2008