

# 「正解があるとは限らない」 診断課題のための支援AI

量子科学技術研究開発機構(QST)

量子医科学研究所 分子イメージング診断治療研究部

立花 泰彦

# この技術でどんな医療用AIを作りたいか

現在の医療AIデザインは

- ・ 「雑用の補助」 「ミスの防止」 にしか役立たない
- ・ スペシャリストの評価はあまり高くない（作業効率×）
- ・ 開発用データ収集に莫大な時間とコスト



**スペシャリストが欲しがる機能に特化したAIを、  
現実的なデータから開発したい**

- ・ 診断医の診断効率を向上
- ・ 専門医が少ない病院での診断力向上
- ・ 医療AI開発コストの大幅な削減

# 医療診断へのAI応用の進展と限界

## 近年の医療診断(画像診断)へのAI導入

(例)

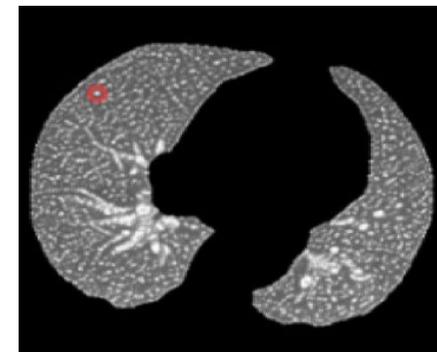
- ・ 肺CTからの結節検出
- ・ マンモグラフィの異常検出と分類

Rodriguez-Ruiz, A. et al. 2019 J Natl Cancer Inst.

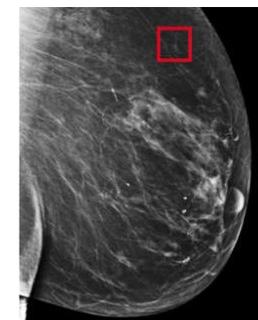


比較的 「単純な課題」 に限定

答えがあるもの、医師が迷わないもの



Shaukat, F., et al. 2019 J. Ambient Intell. Humaniz. Comput.



<https://tokuteikenshin-hokensidou.jp/news/2019/008395.php>

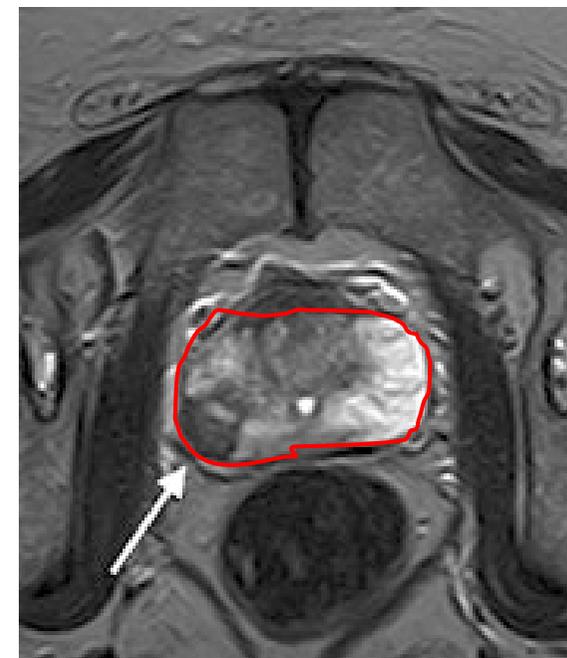
# 医療診断へのAI応用の進展と限界

「複雑な課題」 = 「明確な答えがない課題」  
「医師が迷う課題」

(例) 腫瘍が浸潤しているかどうか

- 画像のみで断定は不可能な場合も
- 医師ごとに意見は異なりうる

(様々な正しい考え方がありうる)



[https://www.teramoto.or.jp/teramoto\\_hp/kousin/sinryou/gazoushindan/case/case206/index.html](https://www.teramoto.or.jp/teramoto_hp/kousin/sinryou/gazoushindan/case/case206/index.html)

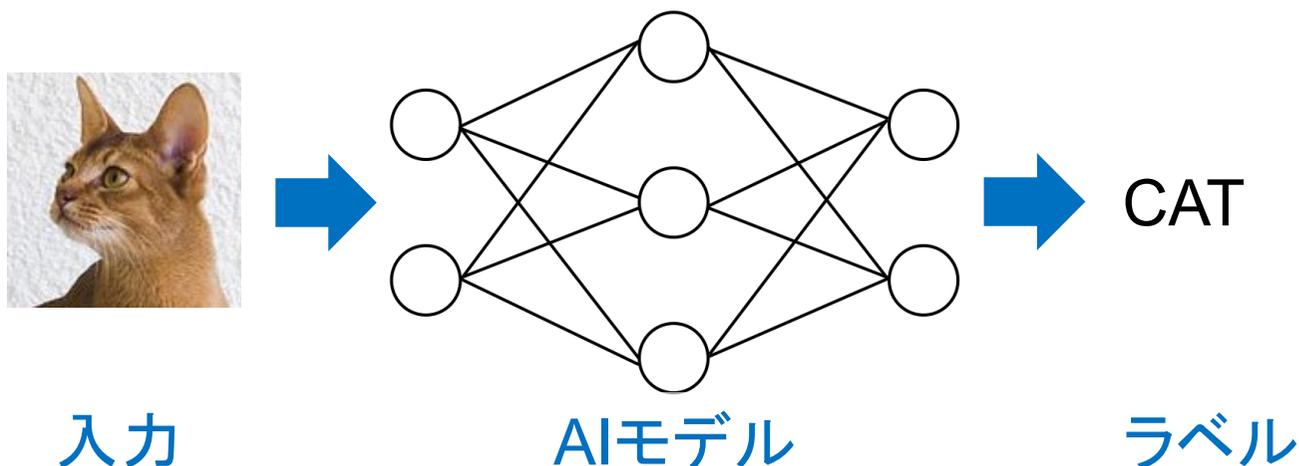
現在のところ、複雑な問題へのAI応用は進んでいない

Sala, E., Ursprung, S. 2020 Radiol: artific Intell.

理由：現在のAIデザインが複雑な問題と相性が悪いから？

# 従来のAIデザインの問題点

従来のAIデザイン（分類問題）

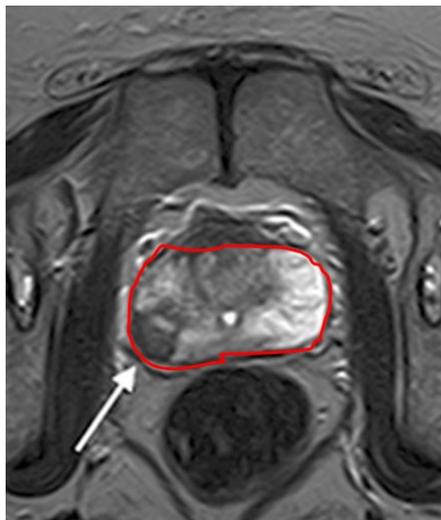


- 入力に対し、正解が存在するという前提で学習
- 未知の入力に対して正解の予測を出力する、という利用法
- 多量の網羅的なデータを用いて学習する必要性

このそれぞれに「複雑な問題」との相性の悪さがある

# 正解が存在するという前提で学習

…の問題点



[https://www.teramoto.or.jp/teramoto\\_hp/kousin/sinryou/gazoushindan/case/case206/index.html](https://www.teramoto.or.jp/teramoto_hp/kousin/sinryou/gazoushindan/case/case206/index.html)

現在のAI開発の多くは医師の合議の判断を正解とみなすが…

- 本来は正解が定義しにくい(できない)場合も多い
- 医師ごとの考え方のバリエーションが反映されない
  - 合議によるラベルは必ずしも適切ではない  
(全会一致のラベルと、意見が割れたラベルが同じ扱い)

# 入力に対して正解を予測させる利用法

...の問題点

予測された診断をどのように利用すればいいのかが難しい

- ・ **実は自分の想像と違う答えが出力されても困る**
  - AIの説明性は充分でないので、根拠が不明確
  - 「正解」なのか「異なる意見」かが区別できない
  - 最終責任は基本的に人間
- ・ **見落としを防ぐために、感度を非常に設定せざるを得ない**
  - 医師の作業効率が低下（結果的に使われない）

# 多量の網羅的なデータを用いる必要性

...の問題点

- **合議でラベルをつけるには莫大なコストがかかる**
  - 一例あたりに医師が要する時間が長い
  - 症例が簡単に集まらない
- **意見が割れるような課題では、合議のメンバー選定が難しい**

## 要するに

- ・ 入力に対し、正解が存在するという前提で学習
- ・ 未知の入力に対して正解の予測を出力する、という利用法
- ・ 多量の網羅的なデータを用いて学習する必要性

**既存のAIデザインでは、複雑な問題へのAI応用は進まない**



**あらゆる入力に対して一つの正解を前提とするデータ収集法、AI設計・利用法を離れなければならない**



**ご紹介する内容**

# 新技術の特徴・従来技術との違い

## 新技術を構成する2つの要素

### 1. インターフェースデザイン

バーチャルカンファレンス

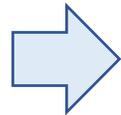
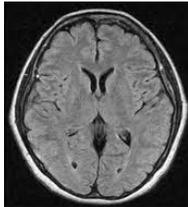
### 2. 医師の考えのバリエーションを織り込むAIデザイン

1. を実現するための技術的な工夫

# 新規1：バーチャルカンファレンス

## 従来型AI診断

患者画像

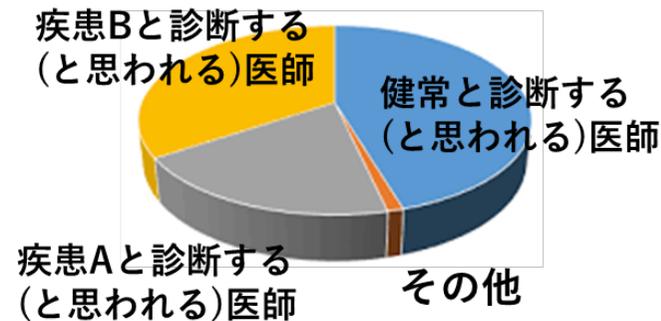


【診断レポート】

健常 : 70 %  
疾患A : 20 %  
疾患B : 10 %

## 提案するAI診断

医師が100人いたら、判断はどう分かれる？  
(100人の医師のAIでカンファ)



権威の大先生の診断  
予測は別枠？

+



大先生は疾患●と診断  
するでしょう

判断が難しい問題を、医師はカンファレンスで解決

- ・ 専門家同士が意見を交換し、満場一致は求めない
- ・ 判断の最終責任者は主治医

↓ 応用

AIに複数の医師がどのように判断するか予測させ、人間が最終判定  
(現状ではどちらにしても最終責任は人間)

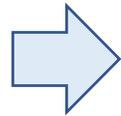
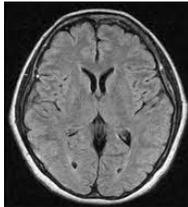


あたかもカンファ  
(合議)のよう？

# 新規1：バーチャルカンファレンス

## 従来型AI診断

患者画像



【診断レポート】

健常 : 70 %  
疾患A : 20 %  
疾患B : 10 %

唯一の正解  
(存在するか不明)

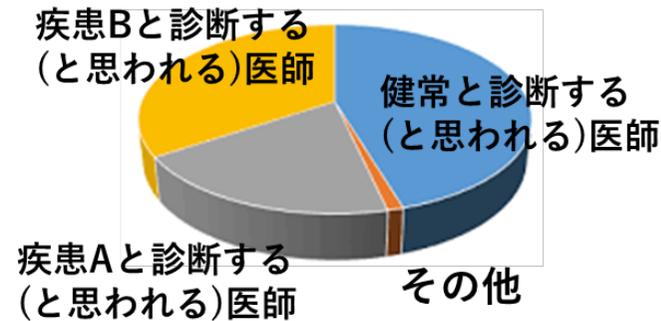
反映できない

見落とし防止

普通

## 提案するAI診断

医師が100人いたら、判断はどう分かれる？  
(100人の医師のAIでカンファ)



権威の大先生の診断  
予測は別枠？

+



大先生は疾患●と診断  
するでしょう

各医師の診断  
(必ず存在)

反映できる

判断補助

高い！



あたかもカンファ  
(合議)のよう？

予測対象

多様な考え方

役立て方

開発難度

# 開発難度が高いわけ

各医師ごとに十分量の学習データを準備するのは大変

- ・ ラベルを作るのは大変なコスト
- ・ 診療データを使ってAIを開発したい

医師の「偏り」をなくすのも難しい



AI学習には多量のデータが必要

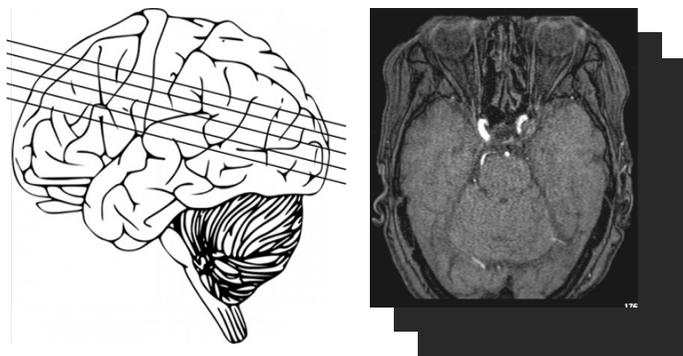


新規2：考えのバリエーションを織り込むAIデザイン

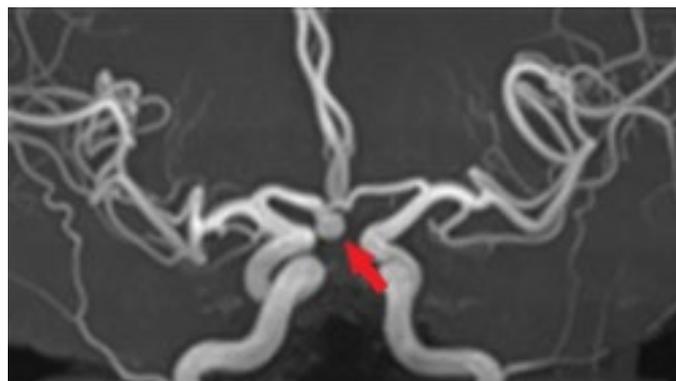
# 実験例: MRI画像から脳動脈瘤を診断

## 動脈瘤か否かを判定するAIを構築する

### MR Angiography (脳ドック)



[http://jp.vector.me/browse/379542/brain\\_01](http://jp.vector.me/browse/379542/brain_01)



<https://www.akita-noken.jp/general/sick/brain-nerve/page-2060/>

### 実験対象として選ぶ理由

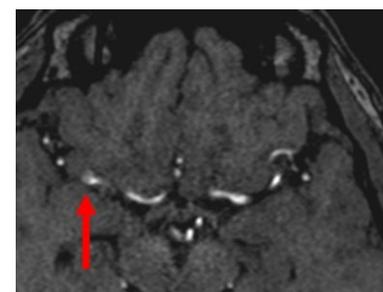
徐々に病変が増大する。微小病変も多い

- ・ 微小病変の臨床的意義は様々
- ・ 判断が難しい例が少なくない

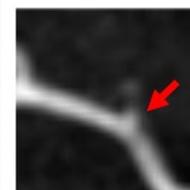
→ 正解が定義しにくい



血管の起始部が動脈瘤のように見えている



微小な動脈瘤??



# 実験データの概要

## 脳ドックデータ

4人の診断医が一次診断を行った症例を蓄積

4人の合計：3143例



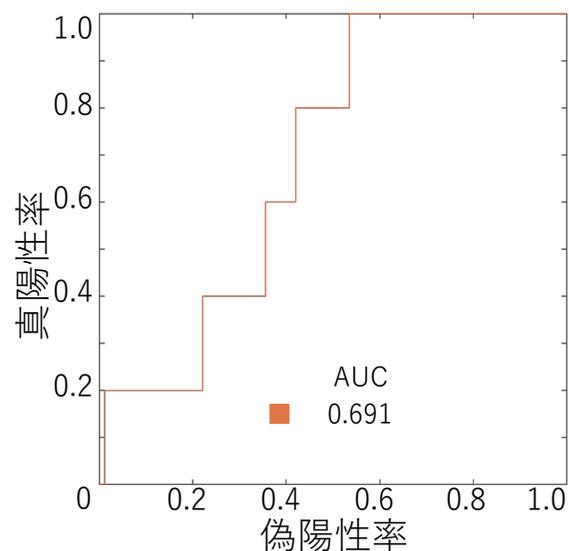
自動検出された候補領域に対し、  
病変のあり/なしを判定

	判定領域数 (/ 件)	陽性判定数 (割合)
医師1	1848 (2.5)	21 (1.1%)
医師2	2123 (2.6)	106 (5.0%)
医師3	2796 (2.6)	171 (6.1%)
医師4	1178 (2.4)	185 (15.7%)
合計	7945 (2.5)	483 (6.1%)

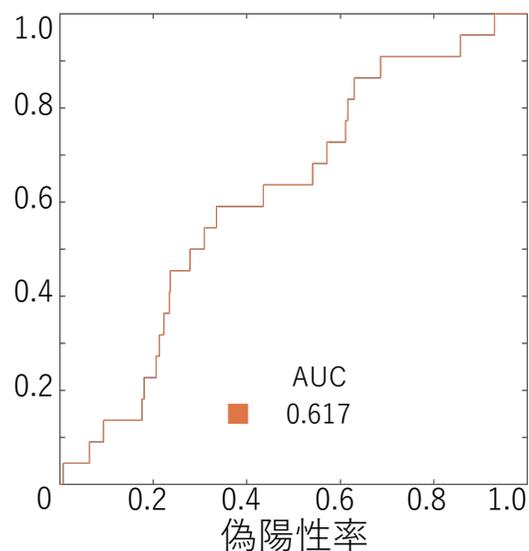
# 4名分のAIを個別に作る場合

【ROC解析】 真陽性率と偽陽性率が高いレベルでバランスするか？

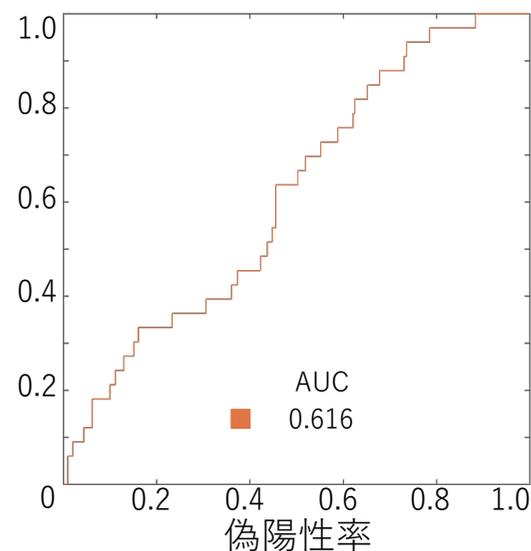
医師 1



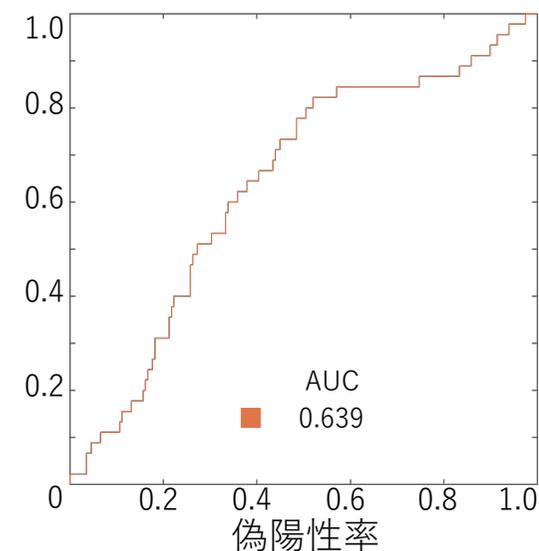
医師 2



医師 3



医師 4

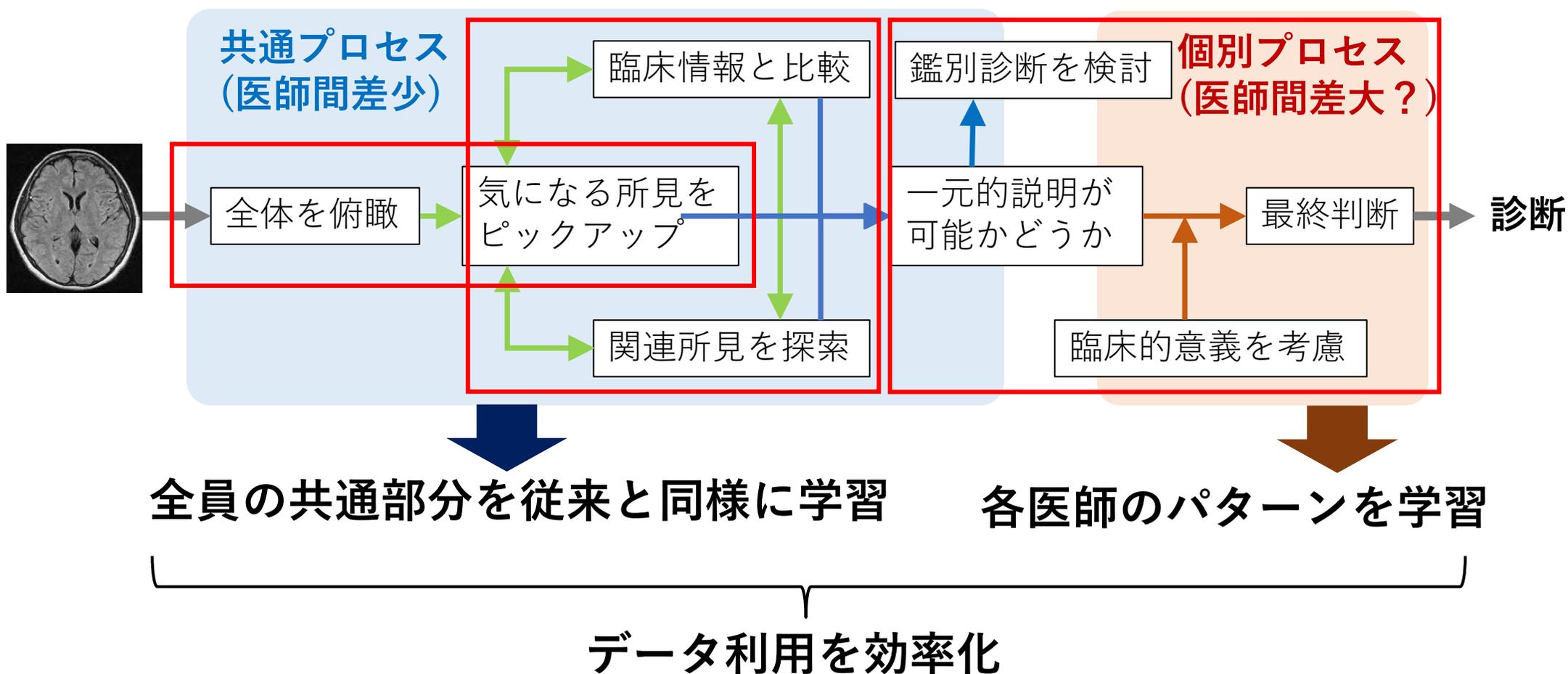


**役立ちそうなAIにはなっていない**

(データ数が十分でないため正しい学習ができていない?)

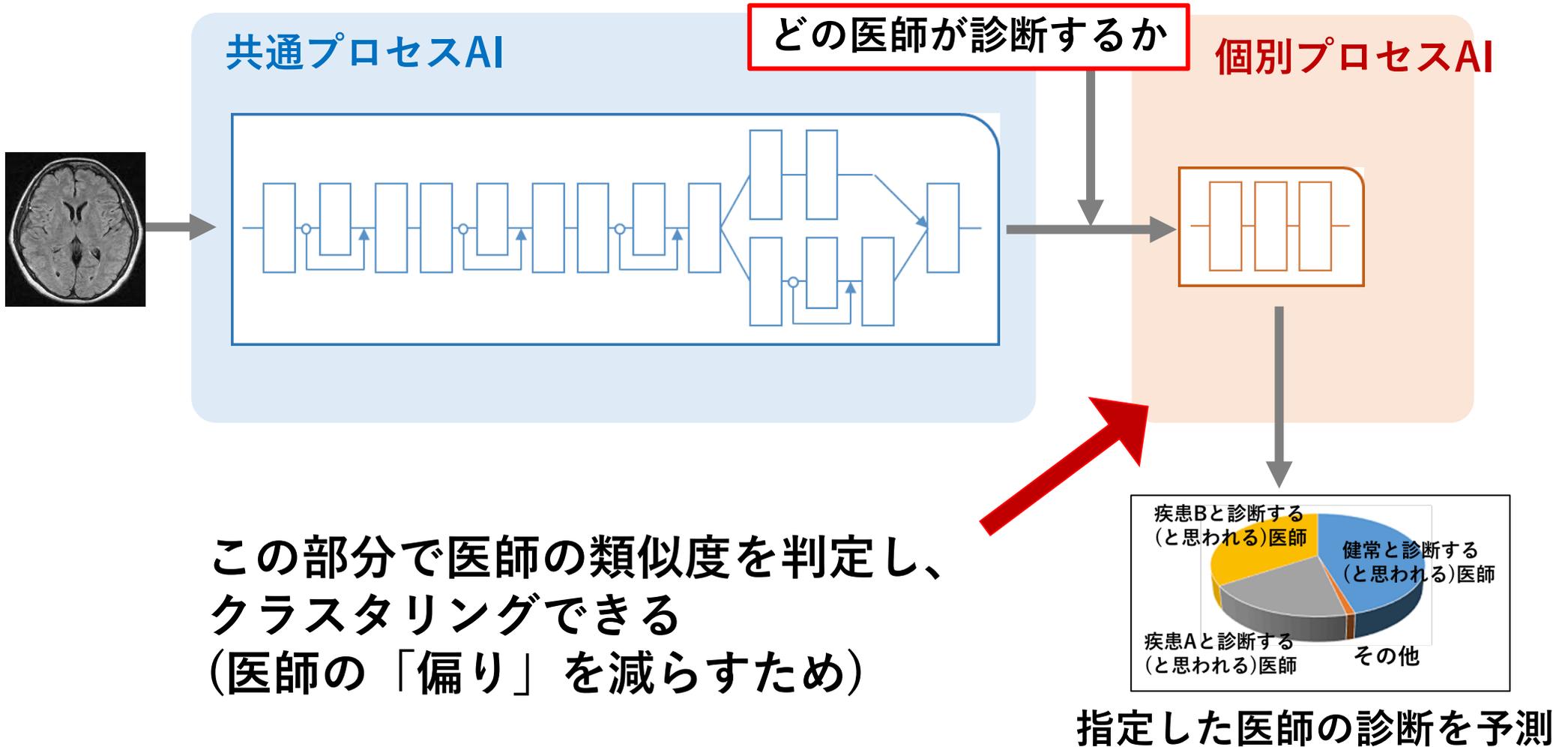
# 新規2：考えのバリエーションを織り込むAI

- ・ **4人分の判断を予測するAIを一つ作る**
- ・ 医師の思考プロセスには共通部分も多いことを利用



# 新規2：考えのバリエーションを織り込むAI

- ・ **4人分の判断を予測するAIを一つ作る**
- ・ 医師の思考プロセスには共通部分も多いことを利用



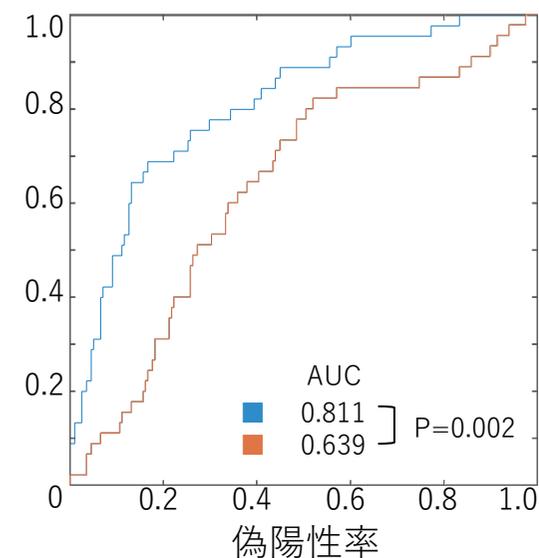
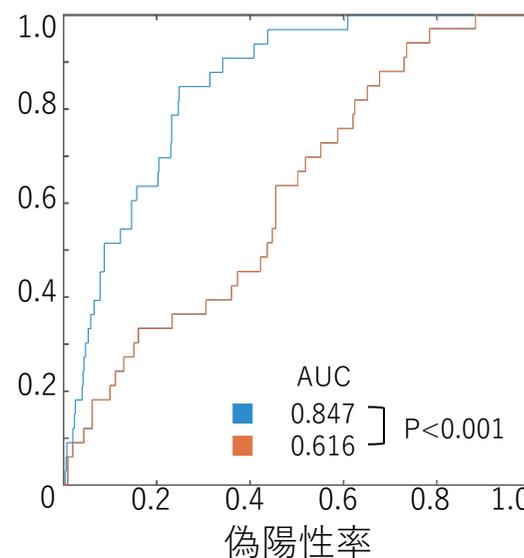
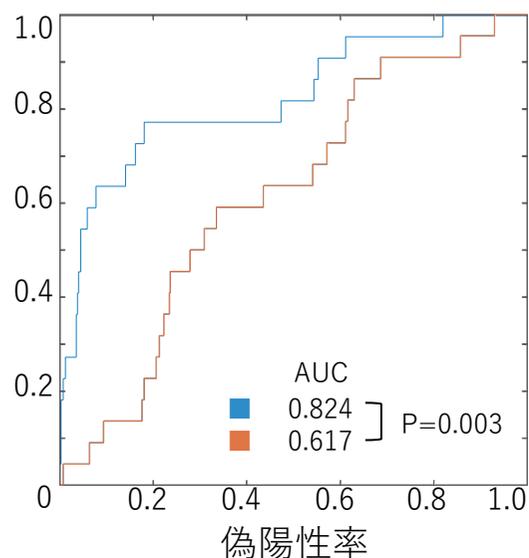
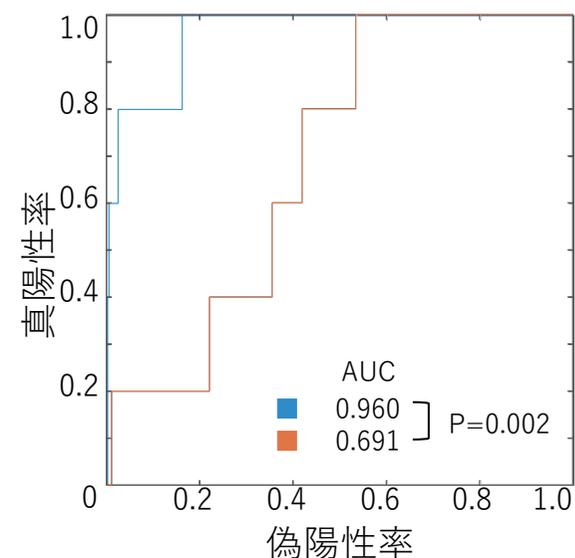
# 提案デザインによる予測精度

医師 1

医師 2

医師 3

医師 4



(統計学的比較はDeLong検定による)

**提案法により大幅に精度が向上**  
技術的な問題が改善している

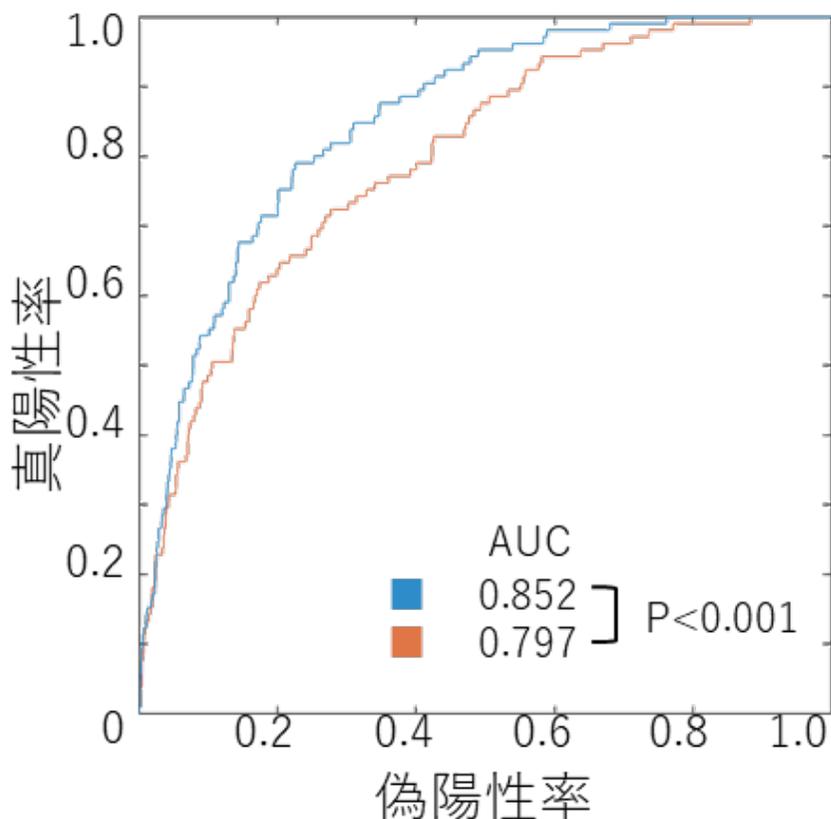
→ 医師ごとにどのように考え方が違うのか、をモデル化できたことで効率の良い学習を行うことができたか

# 本当に医師ごとの判断は違うのか

- 本当に医師の考え方に差はあるのか？
- 先の実験の差は単に学習データの量の差なのではないか？



提案モデルと比較モデルの両方に、全医師のデータを入力した場合

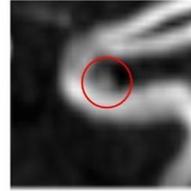


学習データが同一でもモデル間に有意差

→ 医師間差が存在し、提案モデルはそれを含めて学習できている

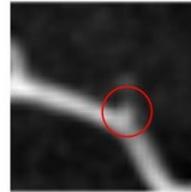
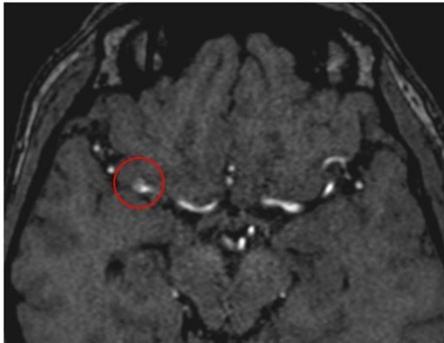
# 出力例

症例1: 48歳, 男性



	AI予測	実際の診断
医師1	Negative	-
医師2	Negative	-
医師3	Positive	Positive
医師4	Positive	-

症例2: 46歳, 男性



	AI予測	実際の診断
医師1	Negative	-
医師2	Negative	Negative
医師3	Negative	-
医師4	Positive	-

- ・ 判断に悩みうる状況と優勢な解を提示
- ・ 医師の数が増えても同様の手法が適用できると思われる (4人→100人)

→ バーチャルカンファの実現へ

## 新規2：まとめ

### 各医師ごとに十分量の学習データを準備するのは大変

- ・ 提案法により、比較的少量のデータで個人パターンを学習可能
- ・ 臨床データをそのまま学習データとして使用可能

### 医師の「偏り」をなくすのも難しい

- ・ 医師をクラスタリング、類似した意見を統合する

# 従来のAI開発と比較した利点と課題

## 「難しい問題」におけるAI活用が可能に

- ・最終判断はあくまでも医師  
バーチャルカンファで最終判断のために必要な情報を提示

## 開発コストの圧縮

- ・臨床データをそのまま使えることは非常に大きなメリット  
(AI開発の少なくとも半分はデータ準備)

## 残された課題

- ・作業時間がどう変化するかなどの臨床研究
- ・医師の偏りをなくすためのクラスタリングが可能か実証

# 想定される用途

- **医療画像診断**  
→ 放射線領域 (CT, MRI, PETなど)、病理学、その他
- **“高度な専門技術による判定を補助するAI”**  
→ 専門家が判断に迷うのは医学だけではない？
- **“データに既知のバイアス要因がある課題”へのAI応用**  
→ データごとに特徴がある場合へのAI応用  
(例) 測定装置によるデータの誤差など

# 企業様への期待

スペシャリストが欲しがる機能に特化したAIを、  
現実的なデータから開発したい

- ・ 診断医の診断効率を向上
- ・ 専門医が少ない病院での診断力向上
- ・ 医療AI開発コストの大幅な削減

## 医療用AIを開発されている企業様

- ・ 共同研究によりデータ収集、AI学習、インターフェース開発

## AI開発用にデータ収集を行っている企業様

- ・ 特定の分野に関するAI開発、インターフェース開発

## その他、ご興味を持ってくださった企業様

- ・ 医療以外への応用など。ご相談ください。

# 本技術に関する知的財産権

発明の名称： 画像診断装置、画像診断方法、画像診断プログラム、及び、記録媒体

出願番号： 2022-027278

出願人： 量子科学技術研究開発機構

発明者： 立花 泰彦

# お問い合わせ先

**量子科学技術研究開発機構**

**イノベーションセンターまでお願いいたします**

TEL: 043-206-3027

FAX: 043-206-4061

email: [chizai@qst.go.jp](mailto:chizai@qst.go.jp)