

深層学習による子宮肉腫鑑別診断の検討

豊原佑典¹⁾、曾根献文¹⁾、黒川遼²⁾、野田勝彦³⁾、吉田要³⁾、谷本早紀¹⁾、
加藤浩介¹⁾、田口歩¹⁾、古澤啓子⁴⁾、塚崎雄大⁵⁾、谷川道洋¹⁾、鶴賀哲史¹⁾、
八杉利治⁴⁾、武知公博⁵⁾、大須賀穰¹⁾

1) 東京大学医学部附属病院 女性外科

2) 東京大学医学部附属病院 放射線科

3) サイオテクノロジー株式会社

4) 東京都立駒込病院 産婦人科

5) 公立昭和病院 産婦人科

JSAWI 第22回シンポジウム 利益相反状態の開示

筆頭演者氏名：豊原 佑典

所 属：東京大学医学部附属病院

演題名：深層学習による子宮肉腫鑑別診断の検討

演題発表内容に関連して、開示すべき利益相反状態はありません。

- 子宮肉腫は10万人あたり0.7人程度と発症頻度は稀であるが、再発も多く、5年生存率は2-6割程度と予後不良の疾患である。
- オカルト腫瘍の頻度は0.3-0.5%程度だが、腫瘍のみの摘出術では子宮肉腫の場合に腹腔内播種が懸念され、子宮筋腫との術前画像鑑別が重要である。
- 人工知能(AI)は医療において画像解析など様々な分野で発展を遂げており、子宮肉腫・子宮筋腫の鑑別をAIを用いた解析により、診断精度の向上を目指すことを本研究の目的とする。

子宮肉腫・子宮筋腫 一般的な鑑別方法

	子宮筋腫	子宮肉腫
年齢	多様	比較的閉経後女性に多い
腫瘍径	多様	比較的大きいことが多い
辺縁	整	不整
T1WI信号	低～中等度信号 出血や脂肪で高信号	不均一で低信号 壊死組織で高信号
T2WI信号	均一な低信号 変性などで高信号	中等度～高信号
ADC値 (10 ⁻³ mm ² /s)	0.88-1.40	0.79-1.17
造影MRIの造影効果	病理組織により多様 例：ヒアリン変性では造影増強効果なし 富細胞性では強い造影増強効果 等	早期の不均一な造影増強効果 中心部の造影増強効果なし
SUVmax (PET)	<7.5	>7.5

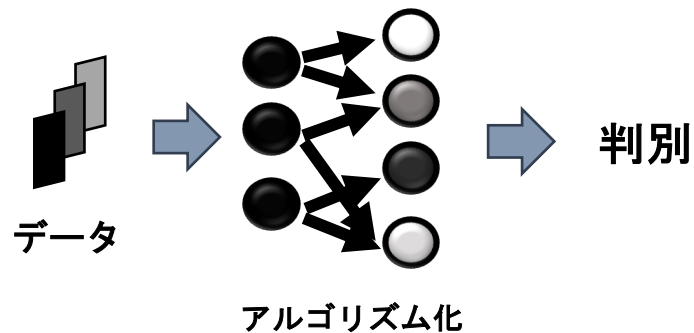
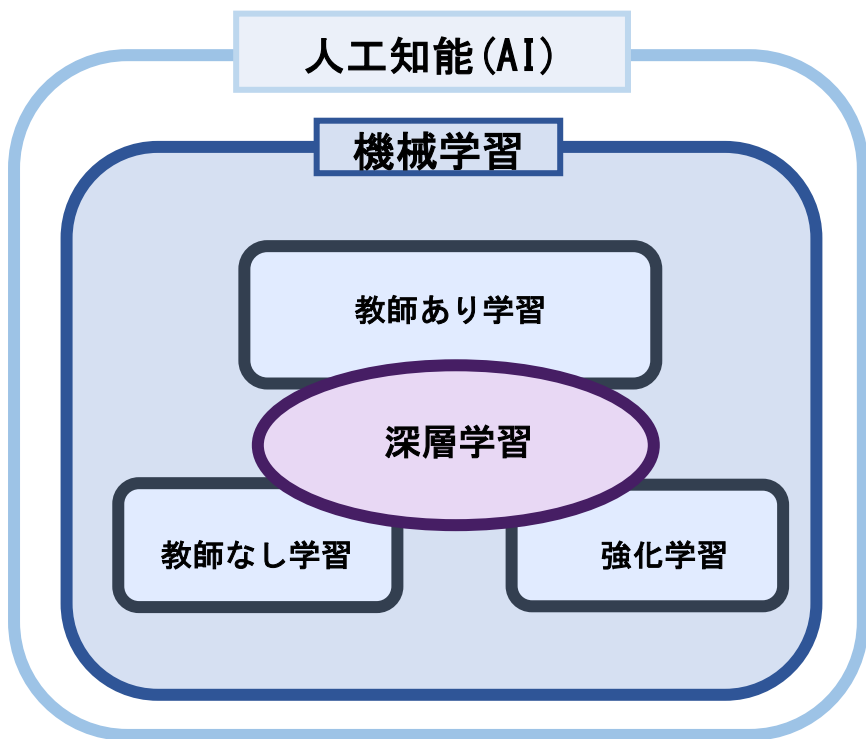
Sun, *Diagnostic and Interventional Imaging*, 2019

※表は上記より抜粋

信号/造影増強効果は正常子宮筋層と比較した際の性状

子宮肉腫・子宮筋腫の鑑別方法の検討は
多くの研究評価がなされているが、確立した診断方法に至っていない。

- 人工知能とは コンピューターを用いて脳の活動を再現（諸説あり）



Ex) 赤りんごと青りんごの区別

機械学習



“色”に着目するように教える



学習を積んで判別できるようになる

深層学習



“自分”でそれぞれの“違い”を考える



学習を積んで判別できるようになる

機械学習を用いた子宮肉腫・子宮筋腫の鑑別の先行研究

	著者	発表年	雑誌	画像種類	肉腫	筋腫	精度	学習方法
1	Malek et al.	2019	European journal of radiology	造影MRI	9	33	accuracy 91% AUC 0.972	機械学習
2	Xie et al.	2019	European journal of radiology	単純MRI	29	49	accuracy 73.9% AUC 0.83	機械学習
3	Nakagawa et al.	2019	Clinical Radiology	単純MRI PET	11	56	AUC 0.92	機械学習
4	Lakhman et al.	2018	European radiology	造影MRI	19	22	accuracy 75%	機械学習
5	Gerges L et al.	2018	American Journal of Roentgenology	造影MRI	17	51	sensitivity 88.2% specificity 78.4%	機械学習
6	Wang et al.	2020	European Radiology	単純MRI	53	84	AUC 0.91	機械学習
7	Malek et al.	2020	Scientific Reports	造影MRI	21	84	Accuracy 96.2% Sensitivity 100% Specificity 95%	機械学習

課題：深層学習での先行検討はなく、既報の症例数も少ない。

対象：子宮癌肉腫を除く子宮肉腫症例

(leiomyosarcoma, endometrial stromal sarcoma, adenosarcoma, STUMP, undifferentiated uterine sarcomaなどの組織型を含める)

対象の条件：2008年～2020年までの東京大学医学部附属病院、東京都立駒込病院、公立昭和病院の症例
手術および生検で組織学的に上記の診断が下されている子宮原発肉腫の症例
手術前にMRIが撮像されている症例（フィルム画像は含まない）
転移巣の有無は問わない（今回は局所の腫瘍成分の鑑別を目的とするため）
対象画像の増量のため、複数回撮像されてMRI画像も含めた。

対照画像：子宮筋腫の症例

上記同様組織学的に証明され、MRIが撮像されている症例
明らかな卵巣腫瘍のない症例
手術前1年以内かつ直近のGnRH療法がない症例

対象に含めた画像シリーズ：

T1WIaxi=T1強調軸位断像, T1WIsag=T1強調矢状断像, fsT1WIaxi=脂肪抑制T1強調軸位断像, fsT1WIsag=脂肪抑制T1強調矢状断像,
T2WIaxi=T2強調軸位断像, T2WIsag=T2強調矢状断像, T2WICor=T2強調冠状断像, fsT2WIaxi=脂肪抑制T2強調軸位断像,
DWI/ADC=拡散強調像/ADC map, dynamic axi=ダイナミック造影軸位断像, dynamic sag=ダイナミック造影矢状断像,
fsT1CEaxi=脂肪抑制造影T1強調軸位断像, fsT1CEsag=脂肪抑制造影T1強調矢状断像, fsT1CEcor=脂肪抑制造影T1強調冠状断像

対象画像：子宮肉腫 62例 （のべ73回。複数回撮影のMRI画像を含む）
子宮筋腫 200例

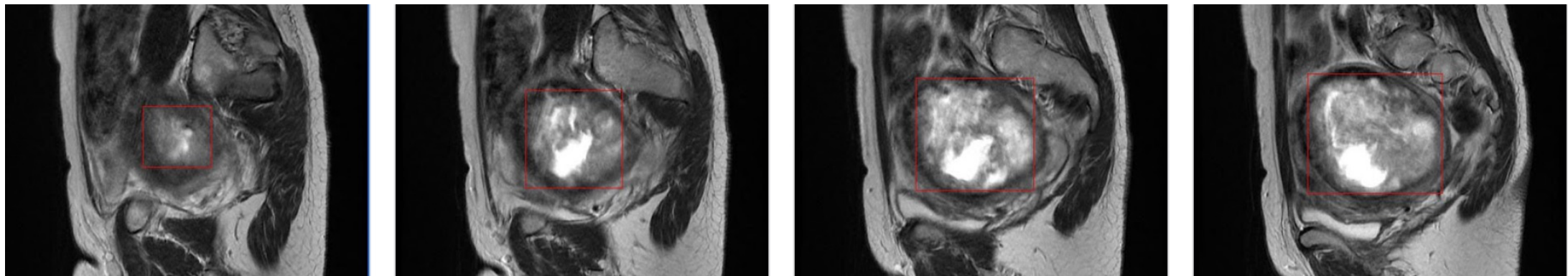
画像データの最適化：

- ・造影画像は遅延相で統一。
- ・複数の拡散強調画像はb値の高い画像を採用 (b値800-1500)。

画像調整：

下記のように肉腫画像には肉腫部位周辺を赤四角でアノテーションし、
筋腫・肉腫ともに、腫瘍の含まれるスライスと含まれないスライスに分離した。
(肉腫部位に範囲指定に対しては放射線診断専門医と協議を行った上で範囲を指定した。)

アノテーション事例



判別方法：「良性=筋腫」、「悪性=肉腫」の二分類で評価した。

深層学習の学習・評価方法：

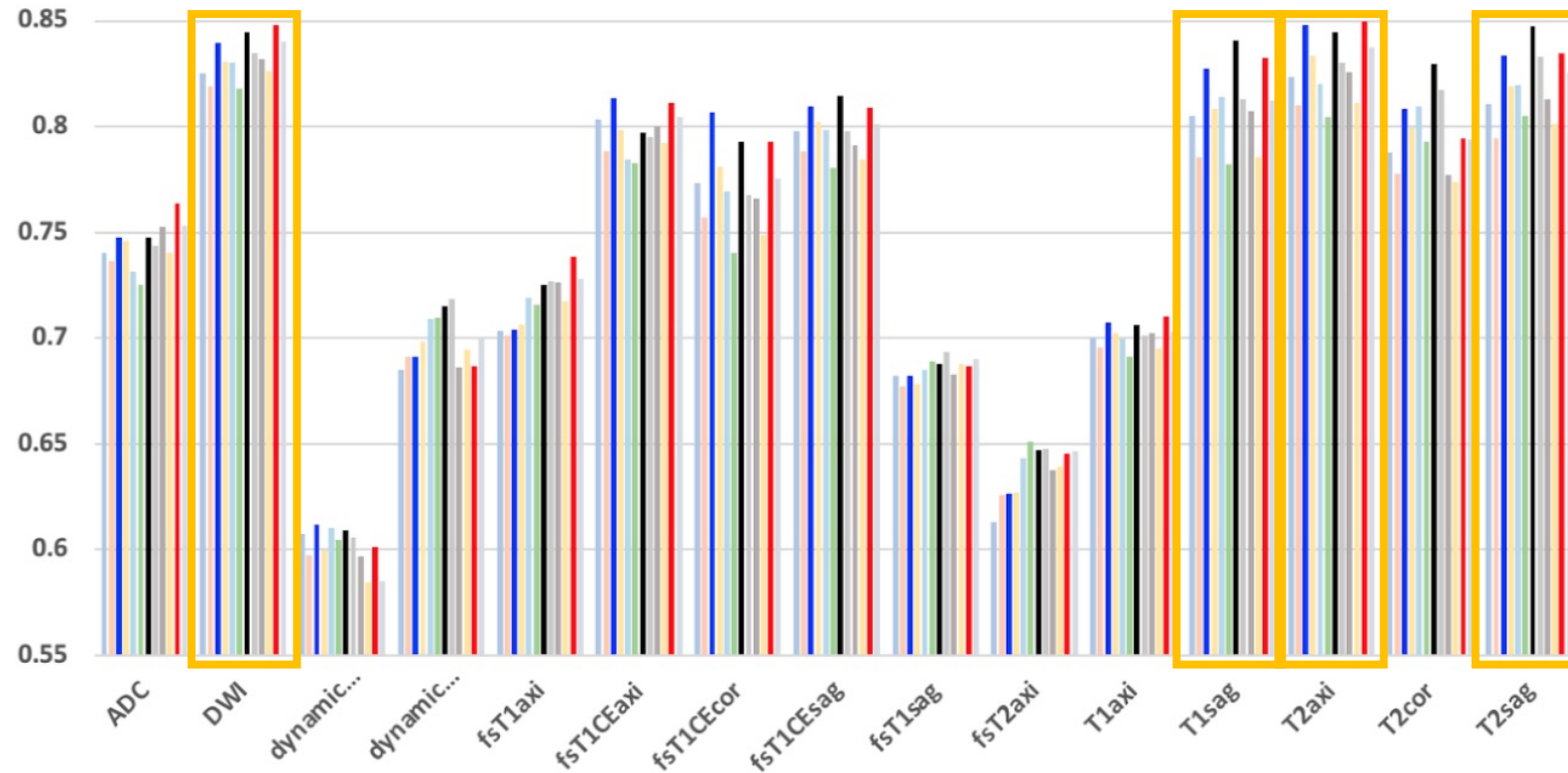
- ・ MobileNetV2をニューラルネットワークとして使用した。
- ・ 症例の画像を6分割し、学習セット:評価セット=5:1に分割した。
- ・ クロスバリデーション法を用いて上記を6モデル作成し、学習・評価を繰り返した。
- ・ 画像の回転や反転などにより、オーグメンテーション（学習用画像数の増加）を図った。
- ・ これを24セット繰り返した。

解析方法：・ アンサンブル法評価（1スライスごとに6モデルで予測）、
マルチ法評価（各症例のシーケンス単位で6モデルで予測した結果の平均（合計でも可）で判定）
など評価方法の比較を行った。
・ 各症例のシリーズ単位の判定には、全スライスの判定結果の割合を用いた。
・ さらにシリーズを組み合わせて最良の成績を得る組み合わせを検討した。

評価方法：診断精度を感度・特異度の平均値として評価した。

結果 (スライス単位の診断精度)

診断精度



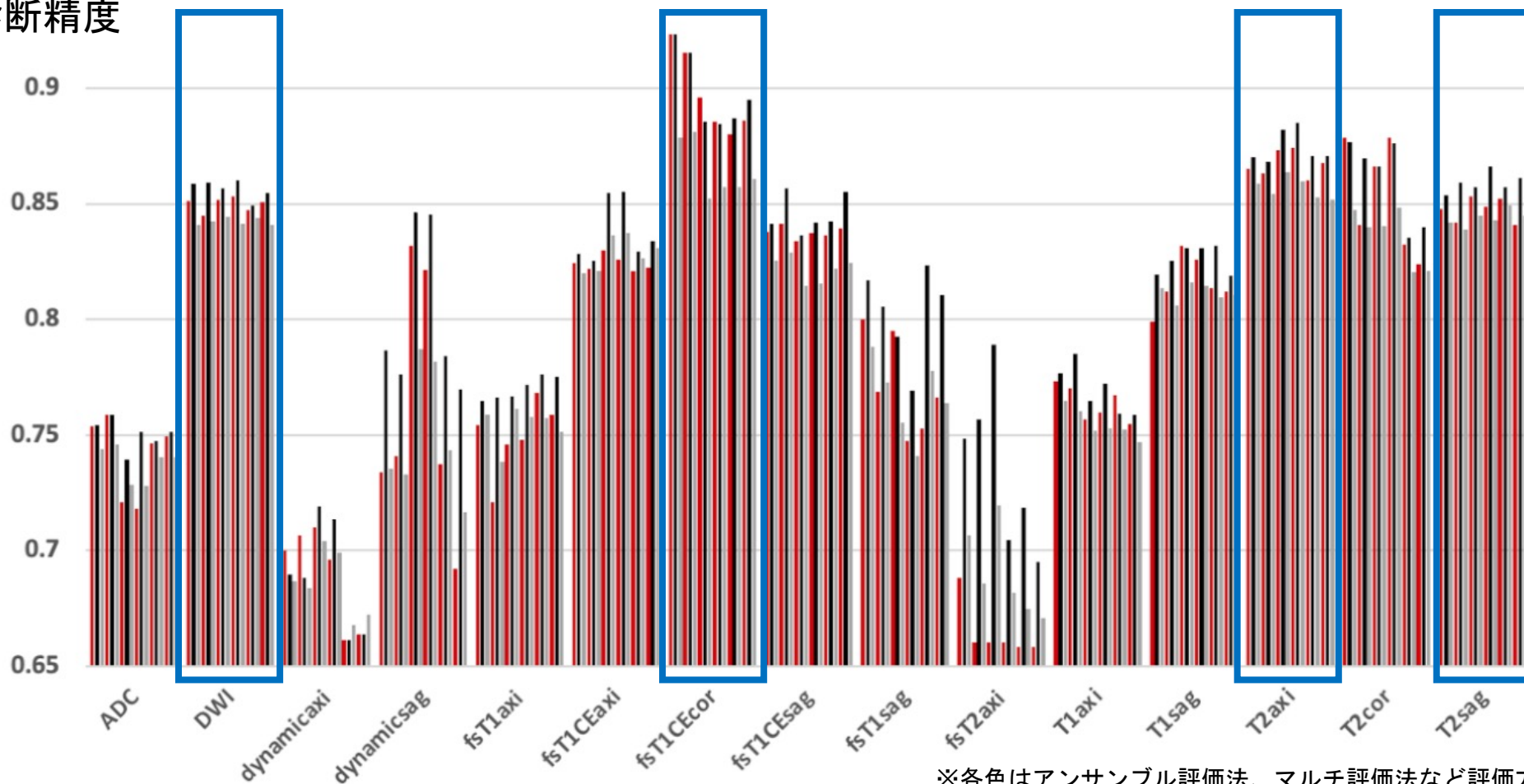
※各色はアンサンブル評価法、マルチ評価法など評価方法による違いを表す。
ここでは詳細を割愛。

成績上位にはDWI (拡散強調像)、T1sag (T1強調矢状断像)、T2axi (T2強調軸位断像)、T2sag (T2強調矢状断像) が挙げられた。

※診断精度 = 感度・特異度の平均

結果 (シーケンスごとと症例単位の診断精度)

診断精度



※各色はアンサンブル評価法、マルチ評価法など評価方法による違いを表す。
ここでは詳細を割愛。

※診断精度 = 感度・特異度の平均

成績上位にはDWI (拡散強調像)、fsT1Ecor (脂肪抑制造影T1強調冠状断像)、T2axi (T2強調軸位断像)、T2sag (T2強調矢状断像)が挙げられた。

結果（シーケンス組み合わせの診断精度）

	成績上位10番以内のシーケンス組合せ	感度	特異度	診断精度
1	DWI+T2axi+T2sag	0.8970	0.9076	0.9023
2	DWI+T1axi+T1sag+T2axi+T2cor+T2sag+dynamiacaxi+dynamicsag+fsT1CEaxi+fsT1CEcor+fsT1CEsag+fsT1sag+fsT2axi	0.9198	0.8842	0.9020
3	DWI+T1axi+T1sag+T2axi+T2cor+T2sag+dynamiacaxi+dynamicsag+fsT1CEaxi+fsT1CEcor+fsT1axi+fsT1sag+fsT2axi	0.9264	0.8768	0.9016
4	DWI+T1axi+T1sag+T2axi+T2cor+T2sag+dynamiacaxi+dynamicsag+fsT1CEcor+fsT1CEsag+fsT1axi+fsT1sag+fsT2axi	0.9343	0.8688	0.9015
5	ADC+DWI+T1sag+T2axi+T2cor+T2sag+dynamiacaxi+dynamicsag+fsT1CEcor+fsT1CEsag+fsT1axi+fsT1sag+fsT2axi	0.9118	0.8908	0.9013
6	DWI+T1axi+T1sag+T2axi+T2cor+T2sag+dynamicsag+fsT1CEaxi+fsT1CEcor+fsT1CEsag+fsT1axi+fsT1sag+fsT2axi	0.9289	0.8724	0.9006
7	DWI+T1sag+T2axi+T2cor+T2sag+dynamiacaxi+dynamicsag+fsT1CEaxi+fsT1CEcor+fsT1CEsag+fsT1axi+fsT1sag+fsT2axi	0.9166	0.8842	0.9004
8	DWI+T1axi+T1sag+T2axi+T2cor+T2sag+dynamiacaxi+dynamicsag+fsT1CEaxi+fsT1CEcor+fsT1CEsag+fsT1axi+fsT2axi	0.9216	0.8790	0.9003
9	DWI+T1axi+T1sag+T2axi+T2cor+T2sag+dynamiacaxi+dynamicsag+fsT1CEaxi+fsT1CEcor+fsT1CEsag+fsT1axi+fsT1sag	0.9286	0.8717	0.9002
10	DWI+T1axi+T2axi+T2cor+T2sag+dynamiacaxi+dynamicsag+fsT1CEaxi+fsT1CEcor+fsT1CEsag+fsT1axi+fsT1sag+fsT2axi	0.9305	0.8688	0.8996

シーケンスを組み合わせることで診断精度は0.9程度まで向上した。
DWI（拡散強調像）， T2axi（T2強調軸位断像）， T2sag（T2強調矢状断像）は
上位10番以内のすべての組み合わせに含まれていた。

- 深層学習による子宮肉腫・子宮筋腫の鑑別において、画像シリーズを組み合わせることにより、診断精度は9割程度に向上した。
- DWI、T2WIの重要性が高い結果であった。
中でもDWI, T2WI axi, T2WI sagのみの組み合わせが最も特異度が高かった理由として、典型的な筋腫はT2WIで強い低信号を示すので、T2WIだけで（T1WIは含まず）判断する方が肉腫と筋腫とのoverlapが少なかった可能性は考慮される。
- しかし、DWI、T2WIを含む症例数が相対的に多数であったため、精度が上がりやすかった可能性はある（下図通り）。

シーケンス	T1axi	T1sag	fsT1axi	fsT1sag	T2axi	T2sag	T2cor	fsT2axi	DWI	ADC	dynamic axi	dynamic sag	fsT1CEsag	fsT1CEaxi	fsT1CEcor	全体
症例数(例)	251	103	175	142	258	259	156	86	254	240	43	61	189	161	61	263

- 希少癌の症例数確保は困難を要し、シーケンス数や撮像などのばらつきなどを補正することには限界点ではあるが症例数増量により、さらなる成績向上の余地が期待できる。

- 深層学習による子宮肉腫・子宮筋腫の鑑別研究を行い、9割前後の診断精度を得られた。
- 深層学習においては、拡散強調画像、T2強調画像が重要な因子と考えられた。
- 今後の症例数の増量による成績の向上が期待される。

本発表の機会を与えてくださりました

杏林大学医学部産科婦人科学教室 教授 小林陽一先生に

深謝いたします。

また、本研究においてご協力いただきました

- ・ 東京都立駒込病院 八杉利治先生 古澤啓子先生 川田淑子先生
- ・ 公立昭和病院 武知公博先生 塚崎雄大先生

多大なご尽力に深く感謝申し上げます。