

データ基盤とHPC基盤連携を利用した連合学習の検証実験

Verification Experiment of Federated Learning Using Data Infrastructure and HPC Infrastructure Integration

研究代表者：村尾 晃平 (国立情報学研究所) / 副代表者：森 健策 (名古屋大学)
共同研究者：合田 憲人・佐藤 真一・二宮 洋一郎 (国立情報学研究所) / 明石 敏昭 (順天堂大学) / 大竹 義人・Soufi Mazen (奈良先端科学技術大学院大学)

研究の背景と目的

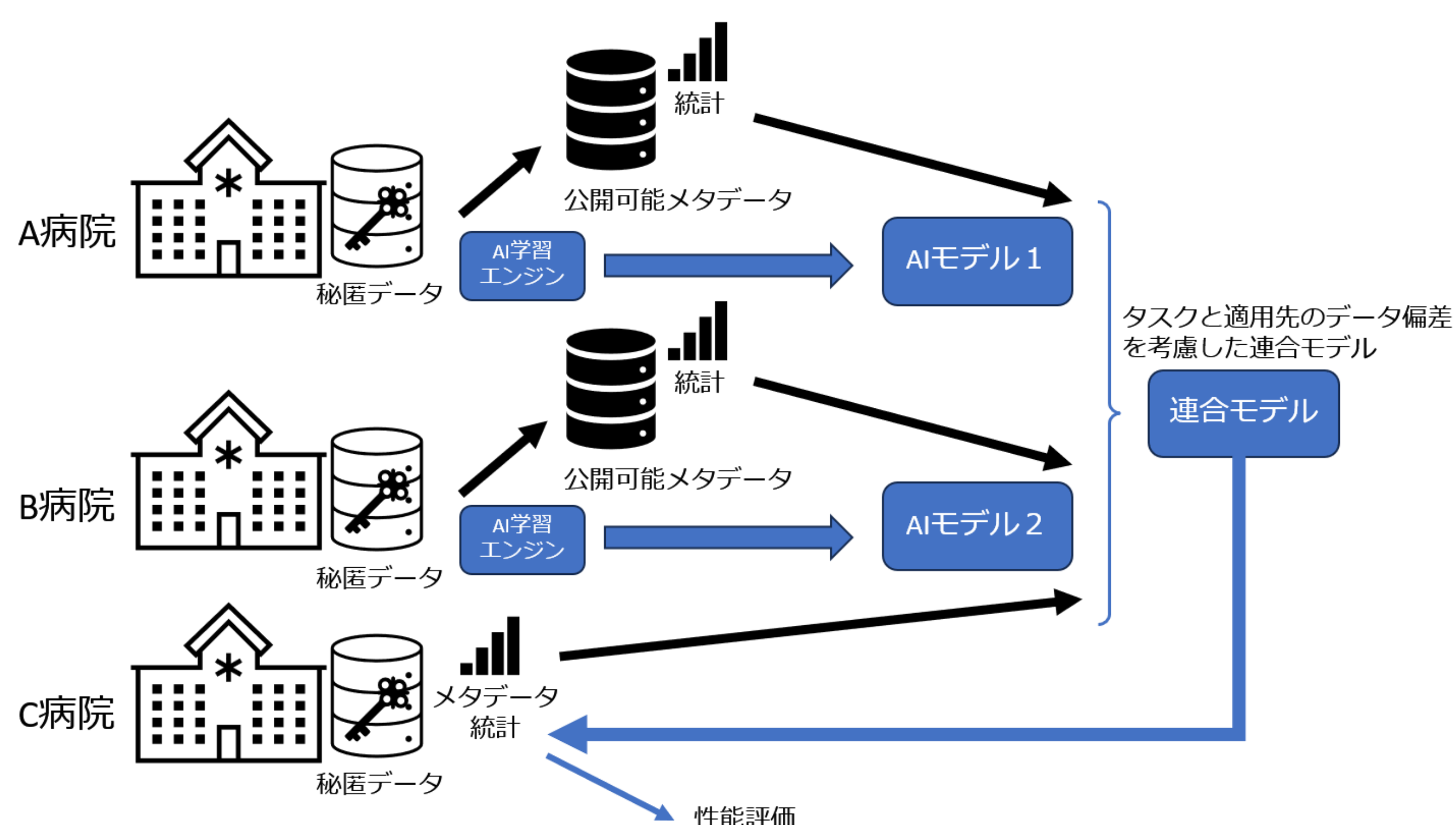
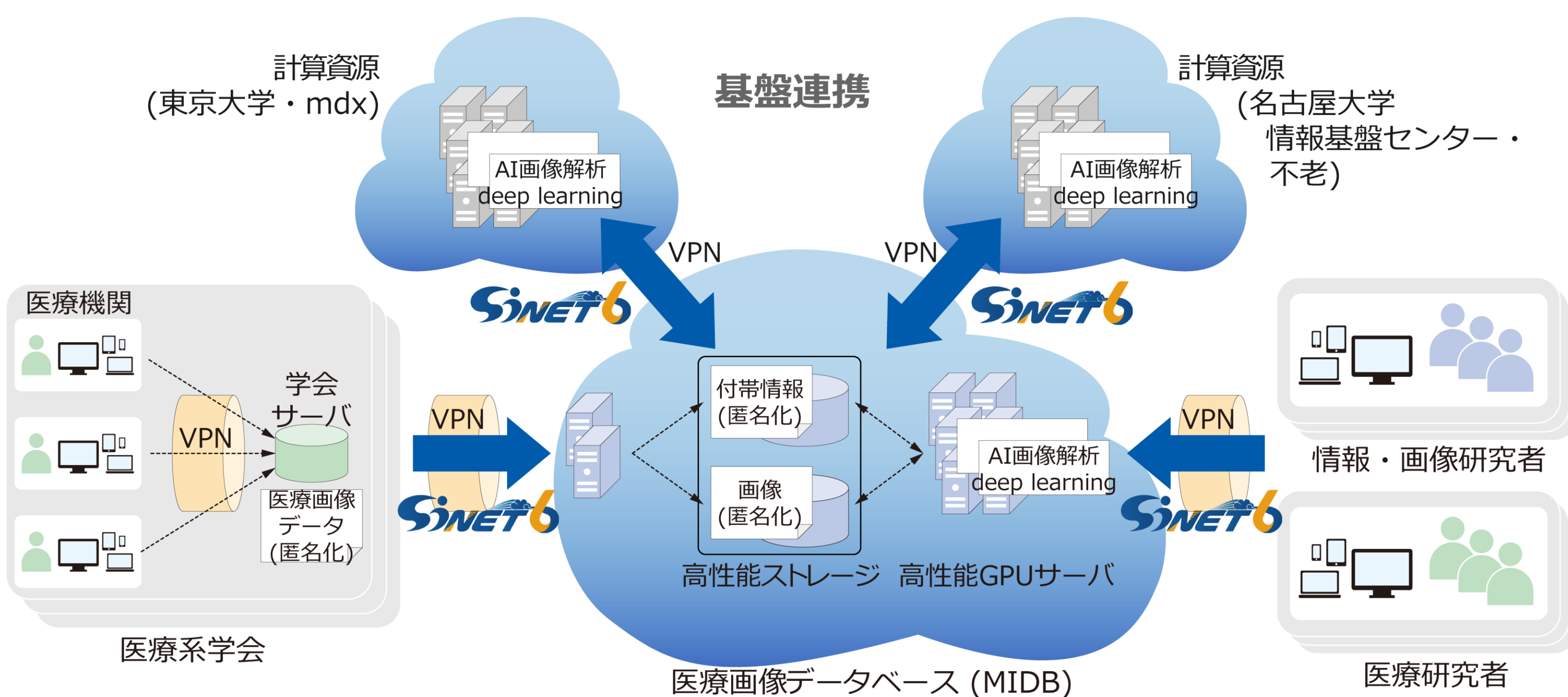
医療データのように機微なデータは、匿名化のレベルに応じて公開レベルも異なる。そのような状況で機械学習を進める際、複数のデータ提供元のデータを必ずしも一か所に集めて学習させることができず、各々のデータ提供元でデータを秘匿してモデルを作成し、複数の施設からのモデルを集約するなどの対策が必要となる（連合学習）。

そこで、本研究では、医療画像の観点で安全かつ効率的な連合学習の方法について研究する。医療画像の場合は、装置の撮影条件など物理的なメタ情報に加えて、対象部位の情報や年齢・性別といったメタ情報もあり、これらには公開可能な情報も多い。適用先のメタ情報も合わせてこれらを手掛かりに効率的な連合学習方法を探索するのが目的である。

我々は、2017年度から構築してきた医療画像データベース

(MIDB) と2023年度からJHPCNのプロジェクトの一環として構築してきたHPC基盤と連携したインフラ設備を有している。データ基盤としてはMIDBであるが、HPC基盤としては名古屋大学情報基盤センターの不老（現在メンテナンス中で10月から不老・弐としてサービス再開予定）および東京大学のmdx Iがある。

一方、MIDBは日本放射線学会が主導するJapan Medical Image Database (J-MID) から日々CTおよびMR画像と関連する所見データなどを悉皆的に受信しており、画像枚数は2026年6月1日現在で10億枚を越えている。その中には10施設の病院からの様々なメーカー・機種・画像再構成条件のデータが混在しており、これらを利用することで施設間ドメインシフトが生じるようなAIモデルに対し、連合学習の手法や効果を検証する計画である。



研究計画

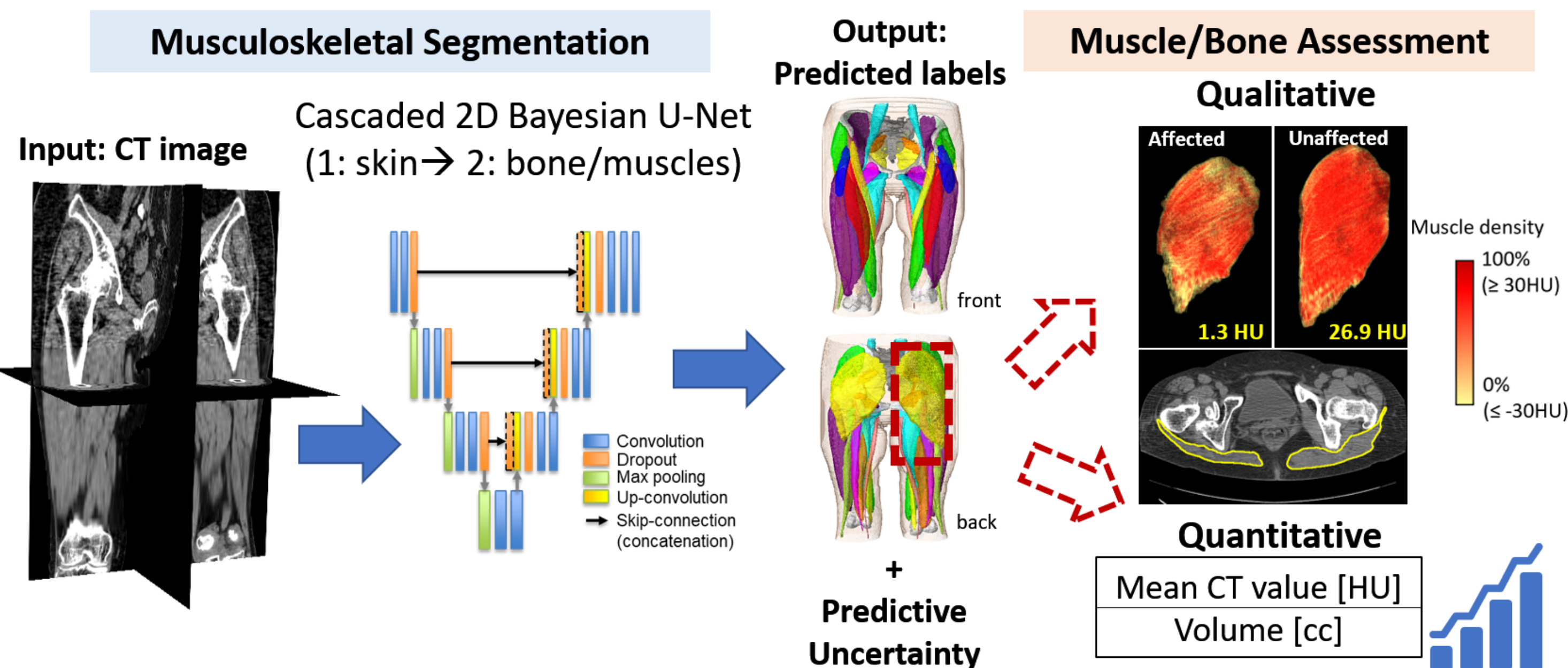
まず、MIDB で受信している CT データを施設ごと・撮影条件ごとに分類し、施設間でドメインシフトが生じる可能性の高い組合せを抽出する。次に、セグメンテーション AI が出力する 不確実性指標を用いて、その大きさを定量化する。最後に、施設ごとに独立に学習させたモデルを連合し、より多様な施設に有効な連合モデルを構築できるか検証する。

(1) データ整備状況

MIDBで 2018年 3月から 2025年 12月までに受信した CT データを整理する。送信元施設は増減を経て累計10施設である。このうち、AI 解析に利用可能なデータを選別する。AI 解析に利用可能なデータとは、体軸に垂直な断面で 32枚以上のスライスを含み、3次元データとして扱えるものと定義する。撮影条件については、ベンダー名+モデル名+再構成フィルタ名の組合せが 1,500種類以上存在することが判明した。ベンダー名だけを見ても施設間で大きなばらつきがあり（表中の数値は検査数）、施設間ドメインシフトの要因となり得る。

(2) AIモデルについて

本研究で用いる AI モデルは、不確実性推論機能を備えた MSKSegmenter である※。連合モデルの構築方法としては、まずは単純な加重平均による統合を試行する。最終的には、正解データが存在しない状況でも、不確実性指標を活用することで施設間ドメインシフトを克服できるかを検証する。



※Soufi, M., Otake, Y., Iwasa, M. et al. Validation of musculoskeletal segmentation model with uncertainty estimation for bone and muscle assessment in hip-to-knee clinical CT images. Sci Rep 15, 125 (2025). <https://doi.org/10.1038/s41598-024-83793-7>

ベンダー名	施設A	施設B	施設C	施設D	施設E	施設F	施設G	施設H	施設I	施設J	合計
⊕ CANON	107,200	189,048	1,775,148	577,924	1,345,462	440,212	948,054	439,246	165,620	221,016	6,208,930
⊕ GE	0	381,000	0	327,474	0	334,410	219,084	47,520	20,016	0	1,329,504
⊕ PHILIPS	49,948	0	196,878	25,570	0	0	0	514	69,356	201,932	544,198
⊕ SIEMENS	0	0	0	214	0	13,284	389,384	0	11,468	0	414,350
総計	157,148	570,048	1,972,026	931,182	1,345,462	787,906	1,556,522	487,280	266,460	422,948	8,496,982

謝辞
MIDBは、戦略的イノベーション創造プログラム (SIP) 「統合型ヘルスケアシステムの構築」 JPJ012425の補助を受けて運用している。名古屋大学情報基盤センターとの連携においては山田一成氏にお世話になっている。