

# SINETを介したデータベース基盤とHPC基盤の連携による医療画像解析基盤実現に関する研究

## Research on realization of medical image analysis platform by linking database platform and HPC platform via SINET

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：佐藤 真一・合田 憲人・大江 和一・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 (奈良先端科学技術大学院大学)

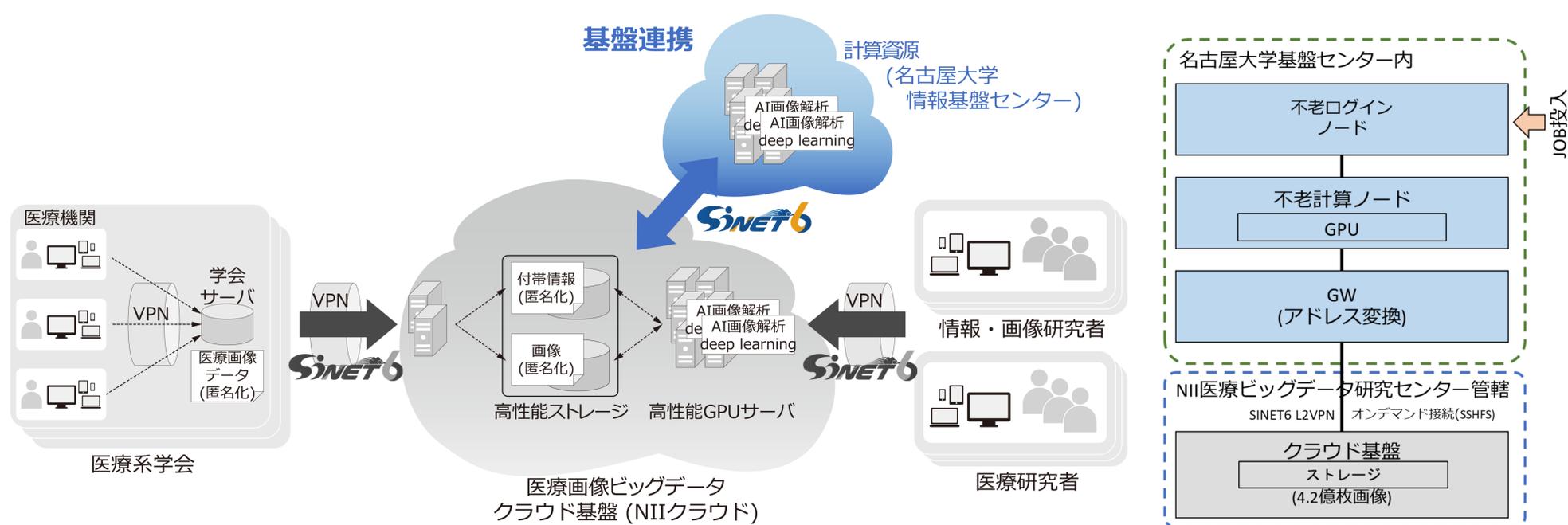
### 研究の背景と目的

画像診断支援AIを実現する学習データには多様性が求められ、異なった撮影装置・撮影条件でのデータが必要となる。NIIは医療画像の提供元を単一の医療機関ではなく、その診療科の複数の医療機関をまとめる医療系学会に求め、全国の医療機関から悉皆的に多彩な医療画像を収集し、データベース上に格納する仕組みを構築した。2017年11月の運用開始以来、クラウド基盤はデータ収集を順調に進め、2023年3月末時点で6つの医療系学会を通じて約4億2千万枚の医療画像を蓄積している。

このデータベースには小規模であるがGPUサーバが接続され、NIIを含む国内14の情報系研究チームが累計39の医療AI研究開発タスクに取り組んでいる。しかし、データ・タスク

の増加に伴って計算資源に対する需要は急激に増大し、マルチモダリティ解析や時系列解析といったタスクの高度化に伴うデータベース処理能力への要求に対して、単一拠点の計算・データベース資源だけで対応することは困難となりつつある。

そこで、データを蓄積する拠点とHPC資源を提供する情報基盤センターとをSINETを介して接続することで分散計算基盤を構成することは、有効な解決策と考えられる。しかし、医療AI研究が求めるセキュリティと性能を確保するための分散計算基盤を実現する方式は確立されていない。本研究では、医療AI研究開発のための分散計算基盤型の医療画像ビッグデータクラウド基盤を構築することを目的とする。2023年度は、SINET6のL2VPNを活用して拠点の異なるデータベース基盤ならびにHPC基盤を安全に接続する分散計算基盤を構築する。



### 研究計画

#### 研究項目1：NIIクラウド基盤とスーパーコンピュータとの基盤連携方式検証

名古屋大学情報基盤センタースーパーコンピュータ「不老」もしくはデータプラットフォーム「mdx」をSINET6のL2VPNで接続(基盤連携)し、2022年度に構築した予備実験環境における結果から導いた条件を用いて、NIIクラウド基盤とスーパーコンピュータとの基盤連携方式を検証する。スーパーコンピュータのノード上で仮想環境を起動し、その環境内からsshfs等を利用してクラウド基盤のストレージボリュームをユーザレベルでマウントする。機械学習計算は仮想環境の中でマウントされたディレクトリ先にあるデータを読みこみながら行い、クラウド基盤にある学習データ(画像等)や中間結果が連携先のスーパーコンピュータに残されることがないようにする。

#### 研究項目2：基盤連携におけるアーキテクチャ最適条件検証

本提案課題における基盤連携方式では、機械学習による学習や推論といった計算の実行中において、学習データや一時的なファイル等がネットワーク経由で送受信される。そこで、ネットワークオーバーヘッド等がスーパーコンピュータでの計算パフォーマンスに与える影響について評価し、最適な通信条件等について検討する。複数のAIモデルの並列処理でのパフォーマンスの調査などは研究項目3と連動して実施する。

#### 研究項目3：基盤連携を利用した医療画像AIモデル構築

クラウド基盤に格納されたデータと連携先のスーパーコンピュータを用いて、医療画像AIの開発を進める。クラウド基盤に蓄積されているCT像を対象として、COVID-19肺炎の画像解析、筋骨格系の画像解析、腹部臓器の形状解析など、超大規模なデータベースを活用して医療画像AIモデル構築を進める。