

令和 4 年度研究開発成果概要書

採 択 番 号 222A03
 研究開発課題名 ウイルス等感染症対策に資する情報通信技術の研究開発
 課題 A ウイルス等感染症により発生するパンデミック対策に資する ICT
 副 題 COVID-19 肺炎の CT 画像を AI 解析するためのプラットフォーム開発と実証展開

(1) 研究開発の目的

COVID-19 肺炎典型度の判定において 90%以上 (目標値) の識別性能を有する診断補助 AI を研究開発する。さらに、本診断補助 AI を臨床現場において実証実験するための準備を整える。

(2) 研究開発期間

令和 3 年度から令和 4 年度 (2 年間)

(3) 受託者

大学共同利用機関法人情報・システム研究機構 <代表研究者> (以下、NII)
 学校法人順天堂
 国立大学法人東海国立大学機構

(4) 研究開発予算 (契約額)

令和 3 年度から令和 4 年度までの総額 40 百万円 (令和 4 年度 20 百万円)
 ※百万円未満切り上げ

(5) 研究開発項目と担当

研究開発項目 1 データベース基盤とネットワークを活用した AI 学習基盤

- 1-1. データベース基盤整備 (NII)
- 1-2. ネットワークを活用した COVID-19 画像解析 (NII)

研究開発項目 2 COVID-19肺炎のCT画像収集と臨床的な評価

- 2-1. データアノテーション (順天堂大学)
- 2-2. 実証実験用ソフトウェアを用いた評価実験 (順天堂大学)

研究開発項目 3 AIアルゴリズム開発と実証用ソフト実現

- 3-1. COVID-19診断補助AI開発 (名古屋大学)
- 3-2. 実証実験用ソフトウェア実装 (名古屋大学)

(6) 特許出願、外部発表等

		累計 (件)	当該年度 (件)
特許出願	国内出願	0	0
	外国出願	0	0
外部発表等	研究論文	0	0
	その他研究発表	20	10
	標準化提案・採択	0	0
	プレスリリース・報道	1	1
	展示会	0	0
	受賞・表彰	0	0

(7) 具体的な実施内容と最終成果

研究開発項目1 データベース基盤とネットワークを活用した AI 学習基盤

- COVID-19 肺炎にかかる学習用データセットの作成
- 大規模データベース (NII) と遠隔地にある計算資源 (名古屋大学) との基盤連携実験

CT 画像による COVID-19 肺炎の典型度判定を行う AI の研究開発に必要な十分な数の学習データを得るために、日々収集される CT 画像から適切なデータを抽出するワークフローを構築した。J-MID より NII へ送られてくる CT 画像は全身を対象としたものであり、その中には頭頸部のみ、あるいは腹部のみといった、肺野が含まれないものも多い。したがって、まず肺野が含まれる CT 画像のシリーズへ絞り込む必要がある。学習用のデータには、放射線学会が指定した基準で付与されたアノテーションが必須である。クラウド基盤上には多施設から送られてくるアノテーション情報を画像や所見文などとは別に収集する仕組みを持っており、今回これを COVID-19 肺炎の判定用に利用した。その結果、毎日十数万枚送られてくる CT 画像から、合計で 2,164 例の CT 画像シリーズを得て開発用のデータセットを作成した。後述するサーベイランスシステム用のデータについては、所見文から肺に関する情報を検索し、これを利用することで肺野を含む CT 画像シリーズを効率よく抽出した。

NII が保有する医療画像大規模データベースと名古屋大学情報基盤センターのスーパーコンピュータを接続し、COVID-19 画像解析を実施するための方式を検討し、実証実験環境を設計、試行した。実験に際して、セキュリティを確保しつつ計算性能を担保することを考慮し、起こりうる障害に対する耐性や反応についても考慮した。まずセキュリティを確保するため、NII 柏分館に設置した試行実験用のストレージサーバと名古屋大学に設置した GPU サーバとを学術ネットワーク SINET6 L2VPN を用いて接続した。ただし、名古屋大学の基盤センター側では多数の GPU サーバの中から基盤連携用のサーバを切り出すことを考慮する必要があるため、GPU サーバの IP アドレスを該当 SINET 上の IP アドレスに変換するプロキシサーバを導入した。接続方式としては、元データを GPU サーバに残さないことを前提に、NFS や SSHFS、その他のオープンソースのツールを検討した。メンテナンス性の観点で NFS と SSHFS に絞られたが、セキュリティの観点で NFS はユーザの ID やグループ名などを継承する必要があること、I/O 性能の観点で SSHFS は NFS よりも write 性能で優れている場合があることがわかり、SSHFS を採用する方針となった。実際に NII 柏分館に置いたデータに 300km 以上遠隔地の名古屋大学の GPU サーバから機械学習の I/O を動作させ、速度性能に問題の無いことを確認できた (回線速度 10Gbps : 逐次処理で約 0.8Gbps、3 並列処理で約 1.6Gbps の速度)。SSHFS でも接続オプションの選択でセキュリティの問題が発生しないようノウハウを蓄積した。起こりうる障害については、特に機械学習中のネットワーク切断やストレージサーバのダウンについて検証を行い、セキュリティ上やメンテナンス上の問題が発生しないことを確認した。以上の基盤連携の設計・検証により、次年度実運用に入れる目途が立った。

研究開発項目2 COVID-19 肺炎の CT 画像収集と臨床的な評価

- 2,164 例の放射線診断専門医の読影による RSNA 分類とアノテーション付与
- 順天堂大学附属順天堂医院で実証実験用ソフトウェアの適用
- サーベイランスシステムのための AI 陽性判定症例数の経時的な収集

COVID-19 感染患者の胸部 CT を放射線診断専門医の提言に従って、(1) COVID-19 肺炎に典型的な所見を有するもの、(2) 典型的とは言えない非特異的な所見で不確定なもの、(3) 非典型的なもの、(4) 肺炎の所見がないもの、の 4 つに分類した。2020 年 7 月から 2021 年 6 月までに撮影された CT の内、COVID-19 感染が PCR 検査で証明された症例を抽出し、それぞれの症例を放射線診断専門医が読影して (1) - (4) に分類したその結果を EIRL アノテーションソフトでアノテーションを作成、合計 2,164 例の COVID-19 肺炎に関連するアノテーションが付与された画像データを J-MID オンプレミスサーバを経由して NII クラウド基盤に送信した。

順天堂大学附属順天堂医院で撮影した胸部単純 CT を実証実験用ソフトウェアである診断補助 AI で検証するために、順天堂の医療情報システムと連携したシステムの構築を検討し、順天堂医院で利用している AI プラットフォームへの適応が可能であることを確認した。プラットフォームでは 1 分毎に PACS に Q/R をかけ、自動的に胸部 CT 画像を抽出し、AI の検証を行っている。医療機器ではないため PACS に結果を返すことはできないが、リモートデスクトップを利用して読影端末上からプラットフォーム上の AI 判定結果「COVID-19 Probability」を参照することが可能である。1 日毎の結果をリスト化して CSV に書き出し、順天堂医院でのサーベイランスも行った。また、ある一日に撮影された胸部 CT (90 件、有病率 0.03) の AI 診断結果を放射線診断専門医による RSNA 分類や PCR 結果も踏まえた臨床的な COVID-19 肺炎の有無をもとに精度の検証を行った。RSNA 分類を基準にしたところ感度は 0.65、特異度は 0.74 であり、臨床診断では感度 1.00、特異度 0.68 となった。COVID-19 肺炎の臨床診断に対する感度が高いものの、特異度はやや低いため、国内における比較的低い有病率の下で発生頻度の軽微な変化を捉えることに関しては難しいが、常時モニタリングすることで肺炎の発生状況のトレンドを知ることはできると推測される。

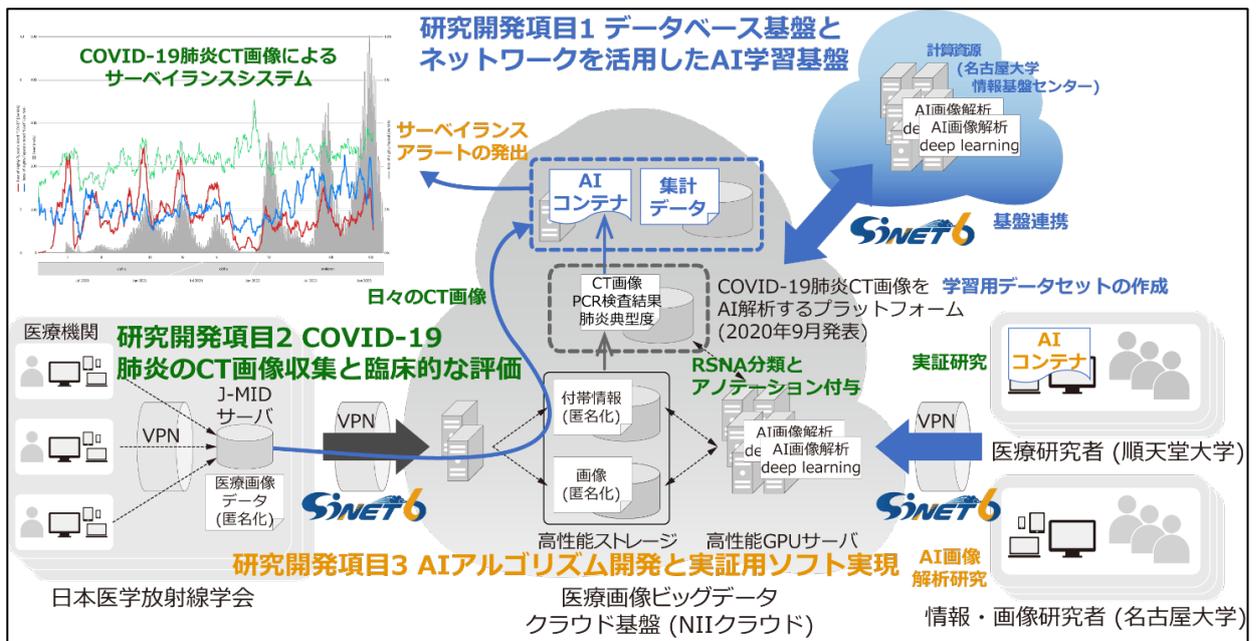
COVID-19 肺炎の発生状況をモニタリングするサーベイランスシステムについては、NII クラウド基盤に登録されたデータを解析している。順天堂大学では J-MID に蓄積された 140 万件を超えるレポートをデータベース化することで、機械的に突合できるような環境を整備し正確な AI 陽性判定症例数の経時的な変化を求めることを可能にした。

研究開発項目 3 AI アルゴリズム開発と実証用ソフト実現

- CT 画像から自動識別を行う COVID-19 診断補助 AI の改善、約 92% の分類精度を実現
- 診断補助 AI を用いた実証実験用ソフトウェアを実装
- 継続的に AI の自動判別結果を得ることでサーベイランスシステムの運用

CT 画像からの自動識別を行う COVID-19 診断補助 AI の性能向上を図った。AI の構成要素である「(1) 肺野・異常陰影セグメンテーション AI」、画像所見に基づいて COVID-19 症例の可能性を判定する「(2) 判別 AI」それぞれの向上を行った。「(1) 肺野・異常陰影セグメンテーション AI」では、肺の炎症発生の平均的な位置を表す Average Template を導入し、Average Template を用いた炎症領域セグメンテーション手法を開発した。この手法は学習データが少ない場合においてもセグメンテーション実施を高める効果がある。「(2) 判別 AI」では、3D Convolutional Neural Network (3D CNN) を用いた自動判別に対し、近年注目される新しい画像分類モデルである Vision Transformer や、Vision Transformer と類似した構造を持つ MLP-Mixer を併用して判別精度の向上を図った。実験の結果、3D CNN のみを用いる場合と比較して、Vision Transformer や MLP-Mixer を併用することで自動判別精度の向上が確認された。MLP-Mixer を使用した場合に、3D CNN のみと比べて約 5% の判別精度向上が見られた。さらに、3D CNN と attention 機構を併用した自動判別手法を開発し、約 92% の自動判別性能を得た。

診断補助 AI を用いた実証実験用ソフトウェアを継続的に適用し、データベース基盤に順次格納される最新の CT 画像に対して AI の自動判別を行った。これにより、近年の新たな感染傾向を反映した統計的データを得た。また、この成果を順天堂大学における CT 撮影装置と連動した診断補助 AI として実装し、運用可能であることを確認した。



(8) 研究開発成果の展開・普及等に向けた計画・展望

1. 計画

直近の計画では、クラウド基盤を社会インフラとして継続するために、利用料金の徴収を含めた財政強化を図る。医師の診断に影響を与える医療 AI は PMDA 認証を得る必要がある。費用と時間がかかる医療 AI そのものの社会実装は直近の計画とはせず、将来目指すべき展望とする。また、サーベイランスシステムの稼働は継続して行い、結果を公表することで社会還元とする。以下に今後の計画と展望を具体的に述べる。

・医療 AI 研究開発プラットフォームとしてクラウド基盤を維持

日々悉皆的に医療画像を収集し、データベース化しているクラウド基盤は、COVID-19 pandemic において必要な学習データを迅速に準備して研究開発に取り掛かることができることを実証した。このような研究開発プラットフォームは他になく、社会インフラとして維持整備することを目指す。国に働き掛けると同時に、利用者から相応の利用料金を負担してもらうなどクラウド基盤事業化の仕組みを作る。事業化については、研究所そのものが料金を徴収することはできず、外郭団体の設立が必要となるなど、関連機関での検討が必要である。

利用者からみて魅力的なプラットフォームとするために、時系列データやマルチモーダルデータの拡充を図ると同時に、付帯情報、特に所見文の構造化を進める。

・成果の公表や公開

サーベイランスシステムを継続して稼働し、結果をウェブページなどで公表する。また、本課題で研究開発した肺 CT 画像解析に関するモデル (AI) をオープンアクセスとして公開する。基盤連携のプロトコルをドキュメント化し、SINET を利用して遠隔資源を連携して利用する際の標準的な手法を公開する。

2. 展望

・クラウド基盤

医療 AI 研究開発を支援する医療画像データベースとして、産学を問わず広く研究者に利用されるプラットフォームとなる。企業の利用には医療機器・ソフトウェア開発製造と製薬を想定し、学術では画像解析・自然言語解析のみならず、医学分野からの利用も想定している。特に、医学分野では、大学医学部に所属する医師だけでなく、市中病院の勤務医や開業医の研究参加が増加することを見込んでいる。

- 医療 AI

COVID-19 肺炎 CT 画像サーベイランスの実行を継続し、その結果をウェブ公開する。また、SARS-CoV-2 だけでなく、他のウイルス性肺炎を併せて検出する AI も稼働し、季節性を含む流行性の呼吸器感染症に対するアラートシステムを整備する。サーベイランスに使う医療 AI を院内システムに組み込んで診断補助とするために PMDA 認証などの取得を目指す。あるいは、既存の PACS 開発企業と連携して、PACS へのオプション組み込みを目指す。

- 基盤連携

物理的に遠隔地に分散した医療画像データベースと計算資源を連携するプロトコルとして、本プロジェクトで確立した手法が標準として利用される。

以上