

令和4年厚生労働科学研究費 補助金 政策科学総合研究事業
(臨床研究等ICT基盤構築・人工知能実装研究事業)
総括研究年度終了報告書

大規模データの利活用研究の加速のための研究

研究代表者 木村 映善 (愛媛大学医学部 教授)

研究要旨

認定事業者を利用した研究を加速する施策の提言につなげることをめざし、認定事業者の潜在的な利用者に関する意識調査、海外の医療情報の収集・活用に関する調査、匿名加工医療情報を用いた AI 研究、次世代医療基盤法第 27 条下にかかるデータ融通の検討を実施した。

研究分担者

荒木 賢二 国立大学法人宮崎大学・医学部附属
病院・研究員
黒田 知宏 国立大学法人京都大学・大学院医学
研究科・教授
水島 洋 国立保健医療科学院・研究情報支援
研究センター・主任研究官
星 佳芳 国立保健医療科学院・研究情報支援
研究センター・センター長
渋谷 哲朗 国立大学法人東京大学・医科学研究
所・教授
佐々木 香織 北海道公立大学法人札幌医科大学・
医療人育成センター・教授
伊藤 伸介 中央大学・経済学部・教授
長島 公之 一般財団法人 日本医師会医療情報
管理機構・理事

B. 研究方法

本研究班は、2年目及び延長した3年目の活動として、以下に述べる分野の取組を行った。

(1) 認定事業者を活用するAI研究・制度環境に関するアンケートの実施

認定事業者の潜在的な利用者である地方自治体、学術団体を選定し、Webアンケートシステムにて実施した。

(2) 海外の医療情報の収集・分析にかかる制度・事業の調査

2022年9月に渡航し、英国で医療情報の収集、リネー、研究者のデータ利活用に供している組織として、NHS Digital (England); Scottish Safe Havens; Public Health Scotland; University of the West England; Imperial College London; Small Area Health Statistics Unit; Clinical Practice Research Datalinkを訪問し、聞き取り調査を実施した。またその聞き取り調査及び提供された資料をベースに文献調査を実施した。

(3) AI研究

前年度に構築した、LDI内に設置した機械学習環境を用いてLDIの匿名加工医療情報を分析した。J-

MIMOから提供された匿名加工医療情報については東京大学医科学研究所のデータセンター内において解析を実行した。

(4) 次世代医療基盤法第25条下にかかるデータ融通の検討

定例会議、エンジニアWG、契約WG、審査WGを延べ30回開催し、次世代医療基盤法第25条下におけるデータ交換にかかる技術的課題、複数の認定事業者間にまたがる契約、審査の課題について検討し、最終提案にむけて資料を取りまとめた。

(倫理面への配慮)

本研究は、認定事業者より提供された匿名加工医療情報を利活用した研究や、個人情報収集しないアンケートを実施しており、倫理委員会の付議対象ではない。LDIについては安全なデータセンター内で分析し、匿名加工医療情報を機械学習しその結果のみをLDIの確認のもとセキュリティ便で研究者に送付、確認する体制を構築し、情報セキュリティを確保している。J-MIMO提供の匿名加工医療情報の解析については東京大学において高度に管理されたスパコンおよび、研究者のみがアクセスできる、施錠された部屋に設置されたPCのみを用いて行った。

C. 研究結果

(1) 認定事業者を活用するAI研究・制度環境に関するアンケートの結果

アンケート依頼を送付した地方公共団体の部署は539部署となった。それらの部署からの総回答数461件中有効回答233件(有効回答率51.5%)であり、アンケート回答率は43.2%であった。31学術団体の会員からは総回答数713件があり、そのうち有効回答として474件(有効回答率66.5%)が得られた。全体的に次世代医療基盤法・認定事業者に対する認知度は低かった。主成分分析、順序カテゴリーカル解析を通し

て、認定事業者の認知度の向上と認定事業者の利活用を推進するという2つの施策を推進する必要性が見いだされた(木村, 2023a)。

(2) 海外の医療情報の収集・分析にかかる制度・事業

英国において、匿名加工情報を研究者に提供する代わりに、データ加工、分析環境も内包したTrusted Research Environment (TRE) というコンセプトを推進しており、TREによってはリモートアクセスも許可していることが確認された。このTREの運用にあたって英国のOffice for National Statisticsが考案したFive Safe Models、すなわち”安全な利用者”、”安全なプロジェクト”、”安全な施設/環境”、”安全なデータ”、”安全な分析結果”という5つの安全管理対策の軸にそった事業展開は、わが国でも認定事業者にTREのコンセプトを導入し事業拡大する際の参考になると思われる。次に、データのリンケージにあたって、直接識別子による名寄せとデータそのもののリンクの役務をIndexerとLinkerという独立した組織に担わせることで個人情報の露出を最低限におさえた秘匿性の高い運用を実現していることが確認された。本邦が学ぶべき点として、TRE、Five Safe Models、IndexerとLinkerの分離、利用者への教育体制の拡充が確認された。これらの知見をまとめて1編の査読付き論文として発表した(伊藤 2024)。

(3) AI研究

a) シーケンス解析

オーダーリング情報を対象にシーケンシャルパターンマイニングを実施、頻出医療指示パターン(SPM)の抽出を行う。そして、これらの最長共通部分を導出すると典型的な医療指示パターンを導出できるとしている。複数の医療機関の頻出医療指示パターンを導出し比較することで、医療機関を越えて共通する指示パターンとバリエーションを抽出できることが確認された(Le, 2023)。

b) 敗血症の予測モデル

入院後の敗血症モデルを予測するための機械学習モデルを開発し、これまでに開発したモデルよりも高い指標(AUC: 0.769、F 値: 0.528、適合率: 0.536、再現率: 0.523)を実現し、多施設から提供された大規模なデータを用いることのメリットが確認された。

c) (ϵ, k) -乱択匿名化の開発

k-匿名化と差分プライバシー双方の要求を満たす新しい匿名化手法を開発し、EMアルゴリズムと組み合わせることにより高いプライバシーレベルを保ったまま、統計量を高精度に維持できる可能性を提示した(Tetsuo, 2020)(Yamamoto, 2023)。

(4) 次世代医療基盤法第25条下にかかるデータ融通の検討

審査WGでのとりまとめとして、複数の認定事業者にわたる申請は最初に申請を受理した認定事業者でワンストップ対応とし、その事業者が審査とリスク評価の審査結果を共有していく方向性を確認した。事業者間の申請書類、共有する項目について突合処理、検討を継続中である。契約WGでは、認定事業者間の医療情報の提供に関する契約書ひな型案を完成させた。エンジニアWGでは、データ授受の形式、手段、媒体、データの統合時のクレンジングの責任分解点、文字コードの扱いについて合意を形成した。これらは最終報告書の付帯資料としてまとめた。

(5) 認定事業者にかかる政策への提言

最終年度の総括として、研究班によって認定事業者の利活用を推進するために効果的な施策の提言を最終報告書の付帯資料として取りまとめた。

C、結論

この研究期間において、4種類のAIにかかる研究、各調査を終了し、4編の査読付き論文及び9編の学会発表、シンポジウムを開催した。これらの取り組みを通して、認定事業者から提供される匿名加工医療情報はAIの研究にも有用であることが確認された。

また、次世代医療基盤法第27条下の運用の方向性について整理がなされた。様々な認定事業者に医療情報が散在して蓄積されるが、標準化された医療情報を収集すること、データの保有状況について情報共有すること、相互運用性を担保すること、契約の流れを整理することで、将来においてデータを相互に融通する見通しが立てられた。

そして、認定事業者に関するアンケート調査、英国の渡航調査を通して認定事業者にかかる状況、RWDの活用状況について知見が整理され、認定事業者の利活用を推進するために必要な施策への提言が取りまとめられた。

D. 健康危険情報

本研究はリアルワールドデータ(日々の診療行為から発生したデータや、診療報酬、健診データ)の二次利用にもとづいた研究であり、侵襲性のある活動はない。また、次世代医療基盤法の認定事業者の匿名加工医療情報を用いる研究であるため、人を対象とする生命科学・医学系研究に関する倫理指針の対象外である。

認定事業者の潜在的利用者に対する認定事業者の認識にかかるアンケート調査についても、代表研究者が所属する愛媛大学において指针对象外でありIRBの付議対象とはならないという判断を受けている。

E. 研究発表

1. 論文・国際会議論文

1. Tetsuo Shibuya. Privacy preserving technologies toward safe medical science. The 47th Annual Meeting of the Japanese Society for Investigative Dermatology (JSID), December 2-4. 2022.
2. Le Hieu Hanh, 松尾 亮輔, 山崎 友義, 荒木 賢二, 横田 治夫. シーケンスバリエーションの比較と電子カルテの分析への応用. 日本データベース学会データドリブンスタディーズ論文誌. 2023;1-5: 1-8.
3. Yamamoto A, Kimura E, Shibuya T, editors. (ϵ, k) -Randomized Anonymization: ϵ -Differentially Private Data Sharing with k-Anonymity. , HEALTHINF, Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies;1: 287-97, 2023.
4. 木村 映善, 高田 春樹, 荒木 賢二, 黒田 知宏,

星 佳芳, 佐々木 香織, 伊藤 伸介. 次世代医療基盤法の認定事業者の認知状況, 活用への意識に関する社会調査. 保健医療科学 2023;72(5):454-63.
5. 伊藤 伸介, 木村 映善, 佐々木 香織. イギリスにおける大規模データの二次利用の新たな展開 — スコットランドの事例を中心に —. 統計研究彙報. 2024;81:89-108.

2. 学会発表

1. Emily Jefferson. The UK' s progress to wards enabling secure, researcher access to s ensitive health data at a UK population scal e. 第43回医療情報学連合大会. 2023:120.
2. Le Hieu Hanh, 松尾 亮輔, 山崎 友義, 横田 治夫. 千年カルテの匿名加工医療情報を利用した多医療機関の電子カルテに対するシーケンス解析. 医療情報学43(Suppl). 2023:309-14.
3. 三浦 堯之, 紀伊 真昇, 市川 敦謙, 山本 充子, 木村 映善. 合成データ生成技術の医療データへの適用と課題. 第27回日本医療情報学会春季学術大会. 2023:68-9.
4. 三浦 堯之, 紀伊 真昇, 市川 敦謙, 山本 充子, 木村 映善. 実医療データを用いた合成データ生成技術の有用性検証の課題. 医療情報学43(Suppl). 2023:pp. 315-20.
5. 佐々木 香織. e-Health, big-data時代の生政治 : Big-Brotherイメージの超克. 立命館生存学研究. 2023;7:13-7.
6. 佐々木 香織, 木村 映善, 伊藤 伸介. Scottish Safe HavenにおけるTREとFive Safes Model から学ぶ -改正次世代医療基盤法に基づく「仮名加工医療情報」の利活用に向けて-. 医療情報学43(Suppl). 2023:pp. 325-8.
7. 山本 章人, 木村 映善, 渋谷 哲朗. 医療情報における差分プライバシーなk匿名化について. 医療情報学43(Suppl). 2023:pp. 306-8.
8. 木村 映善. Scottish Safe HavenにおけるFive Safesの解釈と運用. 2023年度統計関連学会連合大会. 2023:3CPM02-1.
9. 松田 敦義, 池田 遼太郎, 荒木 賢二, 松尾 亮輔. 次世代医療基盤法に基づく匿名加工医療情報を利用した入院後敗血症の予測モデルの 開発における課題と対策の検討. 医療情報学43(Suppl). 2023:321-4.

F. 知的財産権の出願・登録状況
該当なし