

プレスリリース

2023年12月5日

報道関係者各位

国立大学法人宮崎大学
株式会社NTT データ
ファイザー株式会社

日本初、複数医療機関の電子カルテデータに適用可能な 薬物治療効果判定 AI モデルを構築

～電子カルテデータから薬物治療効果を抽出するモデルにより、さらに迅速なエビデンス提供へ～

国立大学法人宮崎大学(宮崎県宮崎市、学長:鮫島 浩、以下、宮崎大学)、株式会社NTTデータ(本社:東京都江東区、代表取締役社長:佐々木 裕、以下、NTTデータ)およびファイザー株式会社(本社:東京都渋谷区、代表取締役社長:原田 明久、以下、ファイザー)はこの度、複数医療機関の電子カルテデータに適用可能な、肺がん患者さんの薬物治療効果を判定する AI モデルを構築しました。本モデルは、電子カルテの非構造化データを自然言語処理し、薬物治療効果を抽出するものです。

今回の研究では、宮崎大学の電子カルテデータをベースに、大規模言語モデル BERT¹を用いてモデルを構築したうえで、6つの医療機関の電子カルテデータへの適用可能性と実用性を検証しました。その結果、複数医療機関の電子カルテデータに適用可能であることや、本モデルで抽出した薬物治療効果から算出した臨床研究の評価項目は、人が抽出した結果と同様の傾向を示すことを確認しました。

【背景】

電子カルテデータは、薬剤の治療効果や安全性などの臨床アウトカム取得の可能性のあるものとして、活用が期待されています。一方で、臨床アウトカムに関わる多くのデータは、経過記録や画像検査レポート等のテキストに含まれており、これらの非構造化データを構造化しなければ解析は困難です。

このような課題から、三者は非構造化データを用いた臨床アウトカムの評価手法の確立に向けて、2020年から共同研究²を進めてきました。その研究では、宮崎大学の電子カルテデータを用いて、肺がん患者さんを対象に薬物治療効果等のアウトカムを医師が評価しました。そして、その結果を自然言語処理したところ、治療効果の評価において、文章の肯定形・否定形や、文脈の情報が重要であることが判明しました。

今回の研究では、文脈情報が考慮可能な大規模言語モデル BERT を用いて薬物治療効果を判定するモデルを構築し、対象を単独医療機関から6つの医療機関に広げ、次世代医療基盤法に基づく認定匿名加工医療情報作成事業者である、一般社団法人ライフデータイニシアティブ(以下、LDI)が保有する多施設電子カルテデータベースに適用させることで、迅速かつ大規模データに基づくリアルワールドエビデンス³の創出を目指しました。研究の概要は以下の通りです。

【今回の研究の概要】

研究目的

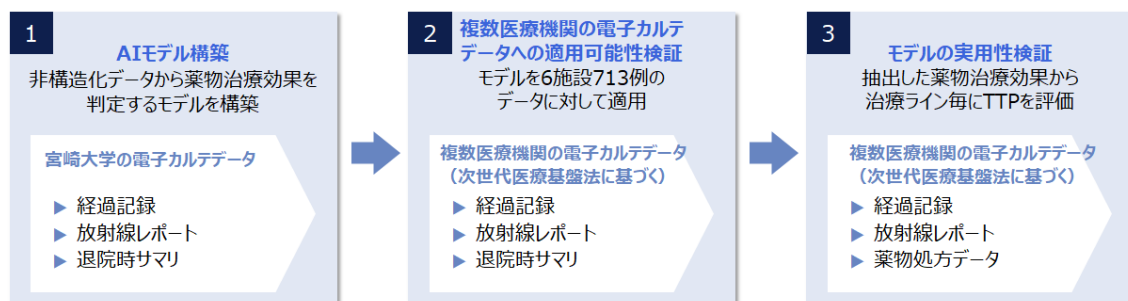
- 宮崎大学の電子カルテデータと、LDIで有する複数の医療機関の電子カルテデータベースをそれぞれ利用して、肺がん患者を対象に薬物治療効果を判定するモデルを構築し、非構造化データからの臨床アウトカム抽出方法を検討する。
- 抽出した薬物治療効果から、臨床研究で用いられる評価項目である疾患進行までの時間(Time-to-Progression disease、以下、TTP)を治療ラインごとに評価し、本手法の実用性を検討する。

プレスリリース

- 宮崎大学倫理審査委員会にて2021年9月7日に承認された。また、LDIの利用目的等審査委員会にて2021年10月28日に承認された。

研究方法

1. 宮崎大学医学部附属病院に通院または入院した、肺がん患者31例の電子カルテデータを対象に医師が評価した学習データを作成し、BERTを用いた薬物治療効果判定モデルの構築手法を検討する。ドメイン特化BERT構築フレームワーク⁴を適用して事前学習を行い、がん治療ドメインに特化したモデルを構築したうえでファインチューニングを行う。
2. 上記1で構築したモデルについて、次世代医療基盤法に基づき得られた6つの医療機関の肺がん患者713例の電子カルテデータに対して適用し、複数医療機関の電子カルテデータへの適用可能性を検証する。
3. 抽出した薬物治療効果から、臨床研究で用いられる評価項目であるTTPを評価。人が抽出した結果とモデルが推定した結果を比較評価し、本手法の実用性を確認する。



研究期間

2021年9月～2022年3月

各者の役割

宮崎大学:

- 研究計画の立案
- 宮崎大学医学部附属病院に蓄積されたデータのうち、研究対象者として選択基準を満たす患者さんについて匿名化された電子カルテデータを提供
- 医療および医療情報学分野の専門家としての評価および助言

ファイザー:

- 研究計画の立案
- 研究に利用するデータの特典
- データ解析結果の評価

NTT データ:

- 研究計画の立案
- LDIの次世代医療基盤法に基づくリアルワールドデータ³から、研究対象者として選択基準を満たす患者さんについての電子カルテデータの抽出
- 宮崎大学から受領した匿名化された電子カルテデータ、LDIの電子カルテデータの解析

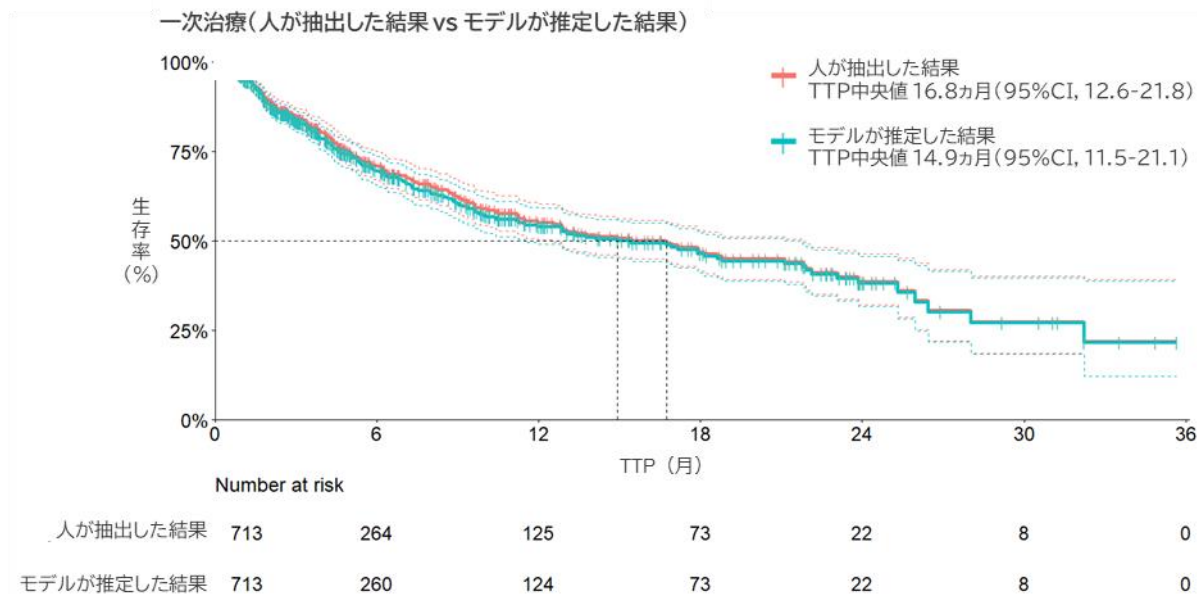
研究結果

1. ドメイン特化BERT構築フレームワークを適用して事前学習を行い、がん治療ドメインに特化したモデルを構築した。さらに宮崎大学の電子カルテデータでファインチューニングすることで

プレスリリース

薬物治療効果判定モデルを構築し、精度検証した(感度 0.63、陽性的中率 0.42、F1 スコア 0.50)。

- 上記 1 で構築したモデルを複数医療機関の電子カルテデータに対して適用した。顕著な精度の低下はみられなかった(感度 0.54、陽性的中率 0.40、F1 スコア 0.45)。
- モデルで推定した薬物治療効果と薬物処方等の構造化データを組み合わせて、TTP を治療ラインごとに推定した。人が抽出した結果とモデルが推定した結果を比較評価したところ、同様の傾向を示した(下図*)。



これらの結果から、構築した薬物治療効果判定モデルは複数医療機関の電子カルテデータに適用可能であり、また臨床研究で用いられる評価項目を評価可能であると確認した。今後、学習データ量を増やす等により、さらなる精度向上が見込まれる。

*図は以下の論文から転載

(Licensed under CC BY-NC 4.0 DEED ; <https://creativecommons.org/licenses/by-nc/4.0/>)

Araki K, Matsumoto N, Togo K, et al. Developing Artificial Intelligence Models for Extracting Oncologic Outcomes from Japanese Electronic Health Records. Adv Ther. 2023;40(3):934-950. <https://doi.org/10.1007/s12325-022-02397-7>

【今回の研究の意義と今後について】

今回の研究により、複数医療機関の電子カルテデータに適用可能な薬物治療効果判定 AI モデルを構築できることを確認しました。今後、治療効果の薬剤間の比較や肺がん以外の疾患の薬物治療効果判定等に、非構造化データの活用が広がる可能性が見込まれます。多施設の大規模電子カルテデータベースから医療に関する臨床アウトカムを効率的に収集して活用することができれば、さらなる個別化医療の進展や、適切な医薬品への早期のアクセス等、さまざまなメリットが期待されます。今後も、リアルワールドデータ利活用の有用性を高める取り組みにより、医薬品をより適切な患者さんに届け、医薬品の価値を最大化し、医療に貢献することを目指してまいります。

【結果の公表】

今回の研究結果の一部は、第 62 回日本呼吸器学会学術講演会(2022 年 4 月 22 日~24 日)、第 89 回日本呼吸器学会九州支部秋季学術講演会(2022 年 10 月 14 日~15 日)、第 42 回医

プレスリリース

療情報学連合大会(2022年11月17日～20日)、第91回日本呼吸器学会九州支部秋季学術講演会(2023年10月27日～28日)において発表されたほか、ジャーナル *Advances in Therapy* (2022年12月、<https://doi.org/10.1007/s12325-022-02397-7>)、*Health and Technology* (2023年2月、<https://doi.org/10.1007/s12553-023-00739-1>)、医療情報学(2023年10月)に論文としてそれぞれ掲載されました。

〈出典〉

- ¹ BERT(Bidirectional Encoder Representations from Transformers)とは、2018年10月にGoogleが発表した自然言語処理モデルです。自然言語処理分野のさまざまなベンチマークにおいて従来モデルの精度を上回る等、近年非常に注目されています。
<https://arxiv.org/abs/1810.04805>
- ² 日本初、電子カルテから薬物治療効果判定に有益な情報を取得
<https://www.pfizer.co.jp/pfizer/company/press/2021/2021-11-24>
<https://www.nttdata.com/global/ja/news/release/2021/112400/>
- ³ リアルワールドデータとは、医療現場での診療行為から得られる医療情報の総称です。リアルワールドエビデンスは、リアルワールドデータの解析から得られたエビデンスです。
- ⁴ ドメイン特化BERT構築フレームワークは、NTTデータが開発した、ドメイン特化BERTを自動で構築する仕組みです。ドメイン特化BERTは、業務領域(ドメイン)特有な用語や言い回しを含む文書に対して、類似表現を含む文書を大量に用意してBERTに追加学習を施すことで、専門性の高い文書でも高い精度を出せるようにしたものです。
<https://www.nttdata.com/jp/ja/news/release/2021/031600/>
<https://www.nttdata.com/jp/ja/data-insight/2021/032202/>

国立大学法人宮崎大学について

宮崎大学は、「世界を視野に 地域から始めよう」のスローガンのもと、地域との連携を密にした人材育成を行うとともに、宮崎最大の知的拠点として発展しています。生命科学、環境科学、エネルギー科学、食の科学、多領域共生分野の科学などを中心に、分野を超えた、融合的で特色ある高度な学術研究成果を世界へ発信しています。

NTT データについて

NTT データは、豊かで調和のとれた社会づくりを目指し、世界 50 カ国以上で IT サービスを提供しています。デジタル技術を活用したビジネス変革や社会課題の解決に向けて、お客さまとともに未来を見つめ、コンサルティングからシステムづくり、システムの運用に至るまで、さまざまなサービスを提供します。

ファイザーについて:患者さんの生活を大きく変えるブレイクスルーを生み出す

ファイザーはサイエンスとグローバルなリソースを活用し、人々が健康で長生きし、生活を大きく改善するための治療法をお届けしています。私たちは、革新的な医薬品やワクチンを含むヘルスケア製品の探索・開発・製造における品質・安全性・価値の基準を確立するよう努めています。ファイザーの社員は、生命や生活を脅かす疾患に対するより良い予防法や治療法を提供することで、日々、世界中の人々の健康に貢献しています。世界有数の革新的医薬品企業の責務として、信頼できる医療に誰もが容易にアクセスできるように、世界中の医療従事者、政府、地域社会と協力しています。人々の期待に応えるため、私たちは 170 年以上にわたり前進し続けてきました。詳細はホームページ、公式 SNS をご覧ください。

www.pfizer.com (米ファイザー本社)、www.pfizer.co.jp、[Instagram](#)、[X](#) (日本法人)

【本件に関するお問い合わせ先】

国立大学法人宮崎大学 企画総務部総務広報課

kouhou@of.miyazaki-u.ac.jp

電 話:0985-58-7114