

代表者 **中津井 雅彦** 教授

メンバー 平野 靖 准教授  
 安部 武志 助教  
 浅井 義之 教授

## 業績例

中津井雅彦, 平野靖, 河村千乃, 安部武志, 浅井義之, 永野浩昭, 星井嘉信, 石田博, 電子カルテ自然文からの疾患レジストリデータ抽出における大規模言語モデルの有効性評価, 第44回医療情報学連合大会 (第25回日本医療情報学会学術大会), 2024年11月23日

## 本研究により解決される課題

これまで利用が容易ではなかった電子カルテの自然文に記載されている情報を、疾患レジストリへの登録サポートなどに活用できる。

## 技術の内容

電子カルテ自然文について、大規模言語モデルを用いて情報抽出を行う。大規模言語モデルは事前学習され、学内の計算機で実行可能な規模のものを使用する。情報抽出に有効な専門知識をモデルに入力し、情報抽出の精度向上を実現する。

## 技術の独自性・優位性

膨大な知識を事前学習した最新の大規模自然言語モデル (Large Language Model, LLM) を活用することにより、従来の機械学習モデルと異なり、大量の教師データを人手で用意する必要がない。外部サービスを使用せず、大学内の計算機を用いて処理を行うことで、電子カルテに記載された機微な自然文を学外に送信することなくセキュアな環境で解析できる。

## 現在得られているデータの概要

疾患レジストリへの登録サポートを想定して、乳癌・胃癌・大腸癌を対象として、手術記録、退院時サマリ・病理組織診・細胞診からの疾患レジストリ登録項目の自動抽出を試行し、特定の項目においてF値が0.89-0.95 (乳癌)、0.81-0.94 (胃癌)、0.74-0.92 (大腸癌) となり、項目は限られるものの、高精度での情報抽出を実現した。

