

2023 年度年次報告書

パンデミックに対してレジリエントな社会・技術基盤の構築

2022 年度採択研究代表者

清水 秀幸

東京医科歯科大学 M&D データ科学センター
教授

微生物学と AI の統合知による次世代の感染症制圧

研究成果の概要

2年次目の研究項目は「研究項目 1. 病原体ゲノム配列情報から治療薬を迅速に探索し構造最適化を行う人工知能の開発」については微生物 CPI データを学習する高精度人工知能の開発と大規模機能アッセイデータを学習する人工知能の開発、「研究項目 2. 抗菌活性ペプチドを高精度で予測する人工知能の確立」については1年次目で修得・整形済みの抗菌活性ペプチドの機能予測 AI の開発であった。

さきがけ採択以前から取り組んできたアミノ酸配列のみから阻害剤を探索する AI である LIGHHOUSE が査読付き学術誌に掲載された (iScience 2022) が、哺乳類データを学習しているため、微生物に特化した chemical protein interaction (CPI) データに特化した改良版のフレームワークを作ることで来るパンデミックにレジリエントな社会の実現に貢献したい。そのために世界保健機関 WHO が規定する臨床医学的に重要な細菌 20 種類ほどの CPI データを収集、解析し、またそのデータに適した AI の構造を設計することで、活性値を相関係数 0.8 台と非常によく予測できる AI モデルの開発に成功した。

2つ目のプロジェクトとして掲げている抗菌活性ペプチドの高精度予測については、公開されている ADAM、APD、CAMP、LAMP 等のデータベースからデータを取得し1年次目に AI の学習に適する形への変換や整形等を行っていた。2年次では、自然界にある大量のタンパク質配列を学習した言語モデルを微生物の持つアミノ酸配列で調整 (ドメイン適応) させた上で、その配列に抗菌活性が期待されるかどうかを判定する AI モデルの開発に成功した。同じ目的意識を持つ研究者らによって複数の AI モデルが直近でいくつか報告されているが、我々のモデルはそれら既存の手法と比べ最も性能が高いことがコンピューター上での検証から分かった。

2年次までに基盤となる AI モデルの開発がおおよそ完了し、3年次・4年次でのさらなるブラッシュアップや実験検証につなげていく。