

TSUBAME 共同利用 令和7年度 学術利用 成果報告書

利用課題名 医薬プロセス最適化と新薬開発を効率化・加速する製剤処方設計 AI の開発
英文: Development of AI for drug formulation design to streamline and accelerate pharmaceutical process optimization and new drug development

利用課題責任者
Yasushi Okuno

所属
Department of Biomedical Data Intelligence,
Graduate School of Medicine, Kyoto University
<http://clinfo.med.kyoto-u.ac.jp/>

邦文抄録(300 字程度)

本研究では、創薬プロセスの最適化と新薬開発を効率化・加速を目指し、AI および機械学習を活用した創薬支援モデルの構築を試みた。TSUBAMEを活用し、がん遺伝子検査データに基づく治療奏功予測、グラフ対照学習による遺伝子制御ネットワーク解析、低分子・ペプチドにおけるタンパク質-リガンド間相互作用解析、Transformer をベースとした構造化医療データの基盤モデル構築、および CT/MRI 画像を用いた医療画像解析を実施した。これらの取り組みにより、創薬プロセスの各段階における多様なデータを活用した AI・機械学習モデルの有用性が示唆され、今後の創薬予測モデルの基盤となり得る知見が得られた。

英文抄録(100 words 程度)

This study aimed to streamline and accelerate pharmaceutical process optimization and drug discovery by constructing support models leveraging AI and machine learning. Leveraging TSUBAME, we conducted treatment response prediction based on cancer genomic data, gene regulatory network analysis via graph contrastive learning, protein-ligand interaction analysis for small molecules and peptides, the development of foundation models for structured medical data based on Transformer architectures, and medical image analysis using CT/MRI. The results suggest the utility of AI and machine learning models that leverage diverse data across various stages of the drug discovery process, providing insights that serve as a foundation for future predictive modeling in drug discovery.

Keywords: 5つ程度

Drug Discovery,
Pharmaceutical Formulation Design
Graph Neural Network,
Transformer,
Foundation model,

背景と目的

今日の創薬研究において、医薬プロセスの最適化と新薬開発の加速は最優先課題となっている。特に、製剤処方設計などの工程は、タンパク質の構造、ゲノム情

報、臨床データ、さらには画像データといった創薬プロセスの各段階における多様なデータを統合的に解析し、最適な組み合わせを導き出す高度な知能化が求められている。

本プロジェクトでは、分子レベルから臨床段階に至る多段階のデータを活用し、AI・機械学習を用いた高精度な製剤処方設計 AI および創薬支援モデルの構築を目的とした。TSUBAME を用いた大規模計算資源を活用することで、従来手法では困難であった大規模かつ高次元のデータ統合と高速な解析を実現し、医薬プロセスの効率化と新薬開発の抜本的な加速に資する基盤を構築した。

概要

本課題では、以下の 7 つの研究を実施した。

①大規模がん遺伝子検査データに基づく治療奏効予測モデル構築と関連因子探索

Center for Cancer Genomics and Advanced Therapeutics (C-CAT) が保有する大規模ながん遺伝子検査パネルデータを用い、ゲノムや患者情報の組み合わせから、多剤併用にも適用可能な薬物治療効果予測手法の開発を行った。

②ノックダウン細胞株の遺伝子制御ネットワークを用いたグラフニューラルネットワークの転移学習と薬剤応答推定

グラフニューラルネットワークを用い、生物学的妥当性と長期的な治療効果の予測を両立したがん患者の仮想的な遺伝子ノックダウンモデルを構築した。

③遺伝子制御ネットワークに対するグラフ対照学習

遺伝子制御ネットワークに対する新たなグラフ対照学習手法を提案した。Node と Augmentation の同時分布に対する KL divergence を用いた確率モデルとしての定式化を行い、摂動に関して教師が存在する場合の数理的な拡張を与えた。

④AlphaFold3 を用いたタンパク質-低分子リガンド相互作用予測

タンパク質-低分子リガンドの相互作用を評価するため、AlphaFold3 (AF3) による複合体立体構造予測を行い、その複合体構造の相互作用評価や特徴量の解析を行った。

⑤タンパク質-ペプチド複合体のドッキング評価

タンパク質-ペプチド複合体の相互作用を評価するため、ドッキングにより多数のデコイ構造を生成し、計算結果の妥当性の検証のため、相互作用の可視化作業を行った。

⑥構造化医療データを用いた基盤モデルの構築

日本全国約 230 万人の構造化医療データを用い、診断・薬剤・日齢・性別を統合した Transformer ベースの基盤モデルを構築した。

⑦CT 画像を用いた物体検出モデルの構築

腸閉塞をきたした症例の CT 画像を用いて、水平断スライスに矩形アノテーションを施し、物体検出モデルの学習/推論を行った。

結果および考察

C-CAT の大規模データを用いた解析では、ゲノムと患者情報の統合により、多剤併用にも対応可能な治療効果予測手法を構築した。TSUBAME を活用することで、大規模な遺伝子変異特徴量の作成が実現した。

ノックダウン細胞株の解析では、Graph Neural Network を用いて生物学的妥当性を持つ仮想モデルを構築し、長期的な治療効果を予測するための基盤を確立した。

グラフ対照学習では、KL ダイバージェンスを用いた確率モデルにより、実験的・人工的摂動分布を潜在空間で整合させ、妥当性と推論精度を両立する新たな数理的拡張を実現した。

タンパク質-低分子リガンド複合体の予測では、TSUBAMEを用いた大規模な AF3 の推論が可能になり、これにより網羅的なタンパク質-低分子リガンド間相互作用の評価が可能になった。

タンパク質-ペプチド複合体の評価では、ドッキングによるデコイ生成と TSUBAME での可視化を通じ、計算結果の妥当性を確認した。これにより、精密な相互作用評価スキームを整備した。

構造化医療データの基盤モデル構築では、230 万人規模のデータを用いて複数サイズの Transformer を事前学習し、疾患・薬剤開始予測における予測性能と計算資源効率のトレードオフを明らかにした。

CT 画像を用いた物体検出では、矩形アノテーションに基づく学習により腸閉塞病変の検出を試行し、将来的な自動診断支援モデル構築に向けた成果を得た。

まとめ、今後の課題

本研究により、ゲノム、臨床、画像、および分子ネットワークといった医薬開発の各段階における多様なデータを統合した創薬支援 AI の構築可能性が示さ

れた。特に、AlphaFold3 やグラフニューラルネットワーク、Transformer 基盤モデルの活用は、製剤処方設計や治療効果予測の精度を大幅に向上させ、医薬プロセスの効率化に寄与することが期待される。

今後の課題として、構築した AI・機械学習モデルの汎化性能のさらなる検証に加え、予測結果を実際の製剤工程や臨床現場へとフィードバックするための最適化アルゴリズムの実装が挙げられる。TSUBAME の計算資源を継続的に活用し、より大規模なデータセットへの対応と推論の高速化を進めることで、新薬開発のさらなる加速を目指す。