

「イノベーション創発に資する人工知能基盤技術の創出と統合化」
平成 28 年度採択研究代表者

H29 年度 実績報告書

浜本 隆二

国立がん研究センター
分野長

人工知能を用いた統合的ながん医療システムの開発

§ 1. 研究実施体制

(1) 浜本グループ

①研究代表者: 浜本 隆二 (国立がん研究センター研究所 分野長)

②研究題目

・本研究の中心的役割を担い、統合的推進を図る。

(2) 岡野原グループ

①研究代表者: 岡野原 大輔 ((株) Preferred Networks 取締役副社長)

②研究題目

・深層学習技術を用いた統合的な解析を行う。特に microRNA データ解析及びクリニカルシーケ
ンスデータ解析を中心に行う。

(3) 瀬々グループ

①研究代表者: 瀬々 潤 (産業技術総合研究所人工知能研究センター 研究チーム長)

②研究題目

・機械学習全般の解析手法を用いた統合的解析を行う。特にエピジェネティクス解析及び医療画
像データ解析を中心に行う。

§ 2. 研究実施の概要

我々は AI 技術を活用した統合的ながん医療システムの開発を通して、国民の皆様にもっと最適な医療を提供することを目的に本プロジェクトを推進している。本プロジェクトは、1. AI 技術を用いたエピジェネティクスデータベース解析/臨床情報を用いた統合的解析、2. AI 技術による microRNA を対象とした早期がん診断システムの開発、3. AI 技術による高精度の画像診断システム及び自動放射線治療システムの開発、4. AI 技術を用いたクリニカルシーケンスデータ解析/Precision Medicine を指向したがん治療の最適化という、4つのテーマについて研究を推進している。平成 29 年度の成果を下記に示す。

1. AI 技術を用いたエピジェネティクスデータベース解析/臨床情報を用いた統合的解析

本年度は機械学習・深層学習技術を用いたマルチオミクス解析を行う基盤として、次世代 ChIP-seq 解析法の開発に取り組んだ。その結果、formalin-fixed, paraffin-embedded (FFPE) サンプルから効率的に ChIP-seq 解析が可能なシステムの構築に成功した。本件に関しては、ヒストン修飾 (H3K27 acetylation, H3K4 tri-methylation) の同定のみならず、転写因子 CCCTC-binding Factor (CTCF) の結合部位の同定に成功した。FFPE サンプルを用いた ChIP-seq 解析で転写因子結合部位の同定に成功した例としては世界初であり、現在論文発表準備中である (Nature Methods, *in preparation*)。本システムにおいては、ヒト型汎用ロボット LabDroid も導入し、再現性が高い高精度のデータ取得が可能となっている。引き続き臨床検体を用いた解析を進展させ、データベースを拡充していくことを予定している。

2. AI 技術による microRNA を対象とした早期がん診断システムの開発

本年度は昨年度に引き続き、免疫チェックポイント阻害剤ニボルマブ投与患者血清の収集後 microRNA 解析を行い、得られたデータを機械学習・深層学習技術を用いて解析することで、新規バイオマーカーの探索を行った。特に肺がんに関しては、数種類の microRNA を用いることで、精度高くニボルマブの感受性予測を行うことに成功した。

3. AI 技術による高精度の画像診断システム及び自動放射線治療システムの開発

内視鏡画像解析に関しては、昨年度に引き続き“深層学習技術を活用した大腸がん及び前がん病変発見のためのリアルタイム内視鏡診断サポートシステムの開発”に取り組んだ。その結果、隆起型の病変に関しては、感度 98%、特異度 99%及び正診率 98.8%と世界トップレベルの精度を誇るシステムの開発に成功した。本研究内容に関しては、現在論文発表の準備をしており (Nature Biomedical Engineering, *in preparation*) さらに、早期社会実装を目指し現在医薬品医療機器総合機構 (PMDA) と面談を進めている。

放射線画像解析に関しては、自動化された Radiomics 解析系により、単純な腫瘍サイズの変化では捉えきれない画像特徴量の経時的変化を網羅的かつ定量的に解析することが可能となった。さらに、セグメンテーションを用いた 3D-UNet の最終出力層を fine tuning することによって、脳腫瘍領域に対して選択的に精度を向上させた検出器の構築に成功した。

4. AI 技術を用いたクリニカルシーケンスデータ解析/Precision Medicine を指向したがん治療の最適化

クリニカルシーケンスを機械学習・深層学習技術を用いて解析する上で最も問題となるのは、検体の数 (N) と比してパラメーター数 (P) が多いという、新 NP 問題である。平成 28 年度は新 NP 問題を克服するアルゴリズムの開発に取り組み、TCGA の肺がんデータを用いて、扁平上皮がんと腺がんを 80%という高い正確性をもって分類することに成功した。