

受賞報告・研究紹介:

日本バイオインフォマティクス生命医薬情報学連合大会 ポスター賞受賞

宮田 龍太 (みやた りょうた) /琉球大学工学部エネルギー環境工学コース・助教
〒903-0213 沖縄県中頭郡西原町千原 1 番地 E-mail: miyata26@tec.u-ryukyu.ac.jp

キーワード：バイオインフォマティクス, 機械学習, 遺伝子選択問題

※本研究成果は、2020年9月1日(火)～3日(木)にオンライン開催された「2020年日本バイオインフォマティクス学会年会 第9回生命医薬情報学連合大会」にて筆頭著者である修士2年の藤澤孝太くんによって発表され、後日、ポスター賞が贈呈されました (<https://www.jsbi.org/news/annual-meetings/20200914-nenkai/>)。昨年度の優秀口頭発表賞に続く藤澤くんの受賞は新聞記事でも取り上げて頂きました (<https://ryukyushimpo.jp/news/entry-1208930.html> <https://www.okinawatimes.co.jp/articles/-/653081>)。

1. はじめに

病気に関連する遺伝子群をなるべく少数で特定することが目的である遺伝子選択問題の解決策として、昨年度は機械学習とエンリッチメント解析(生物学的知見)によるハイブリッドアプローチを提案しました (<https://www.tec.u-ryukyu.ac.jp/wp-content/uploads/2020/03/3e008f6cbe7ae1779d17342832019667.pdf>)。この手法は候補となる遺伝子数が数千個までなら非常に有効ですが、最新の次世代シーケンサーで出力されるような数万個のオーダーでは十分に候補を絞り切ることが難しいケースもありました。

そこで図1のように、「教師なし機械学習」と「教師あり機械学習」を組み合わせることで疾患関連遺伝子群を強力に絞り込む手法を開発し、肺腺がん細胞のデータセットを使って有効性を確認しました。

2. 方法と結果

具体的には表1に示す通り、3つの独立したデータセットに対して提案手法でそれぞれ遺伝子選択を行い、教師なし学習・教師あり学習を単体で適用した場合とで選択遺伝子数を比較しました。その結果、いずれの場合も提案手法が最も候補を限定できていることがわかりました。さらに図2の通り、提案手法で選択された遺伝子のみから腫瘍細胞か通常細胞かが高精度で判別できることも明らかになりました。

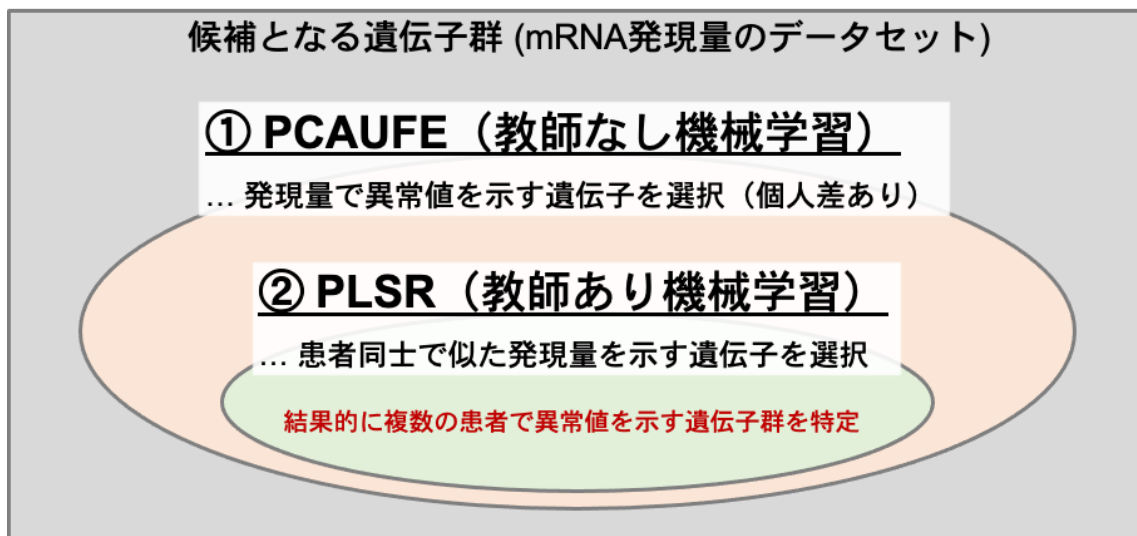


図1: 本研究で提案するアプローチの概念.

3. おわりに

上記で紹介したような研究分野は、創薬や遺伝子治療の足がかりとして期待されています。今後はさまざまな種類の病気に関するデータに提案手法を適用し、多面的に有用性を検証していく予定です。

表 1：肺腺がん関連遺伝子の数

肺腺がん患者のデータセット	候補となる遺伝子数	先行研究 i PLSR	先行研究 ii PCAUFE	提案手法
① GSE27262	24,442	304	8,480	69
② GSE32863	25,441	840	9,430	229
③ GSE40791	24,442	235	6,115	46

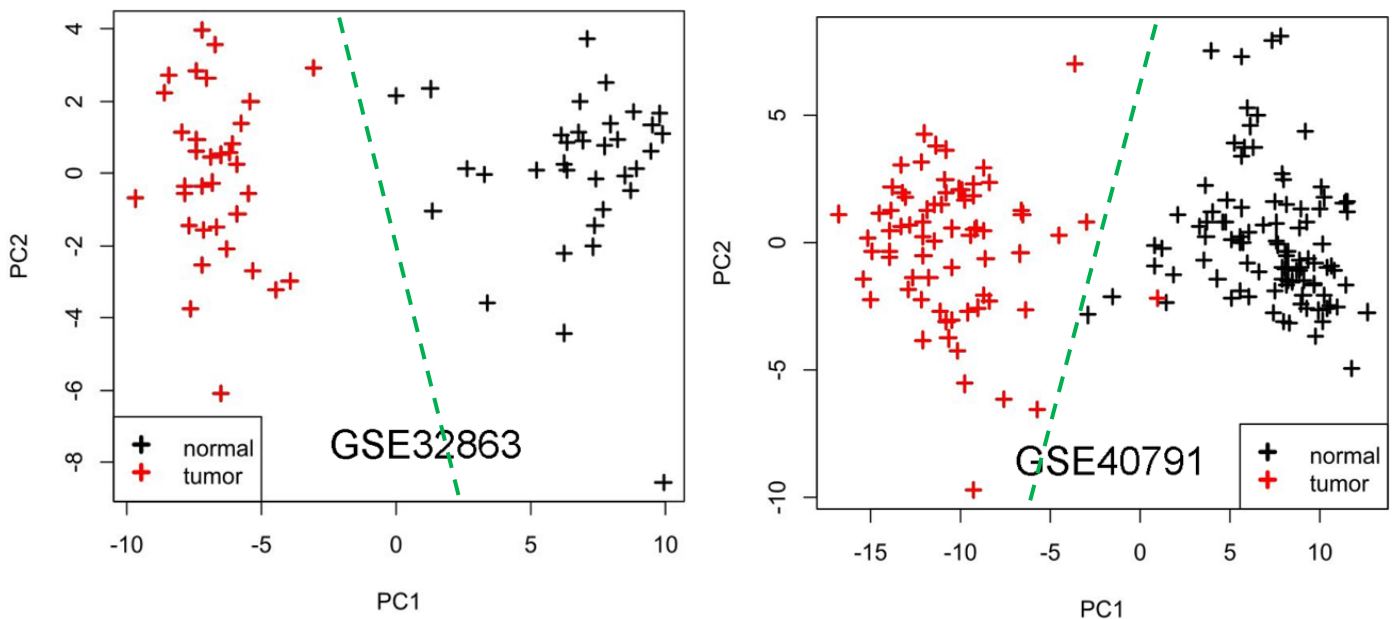


図 2: 提案手法で選択した 69 個の遺伝子 (表 1①) から肺腺がん細胞か通常細胞か判別分析した結果。



図 3：学長記者懇談会での記念撮影 (左：指導教員の宮田龍太助教、右：藤澤孝太くん)

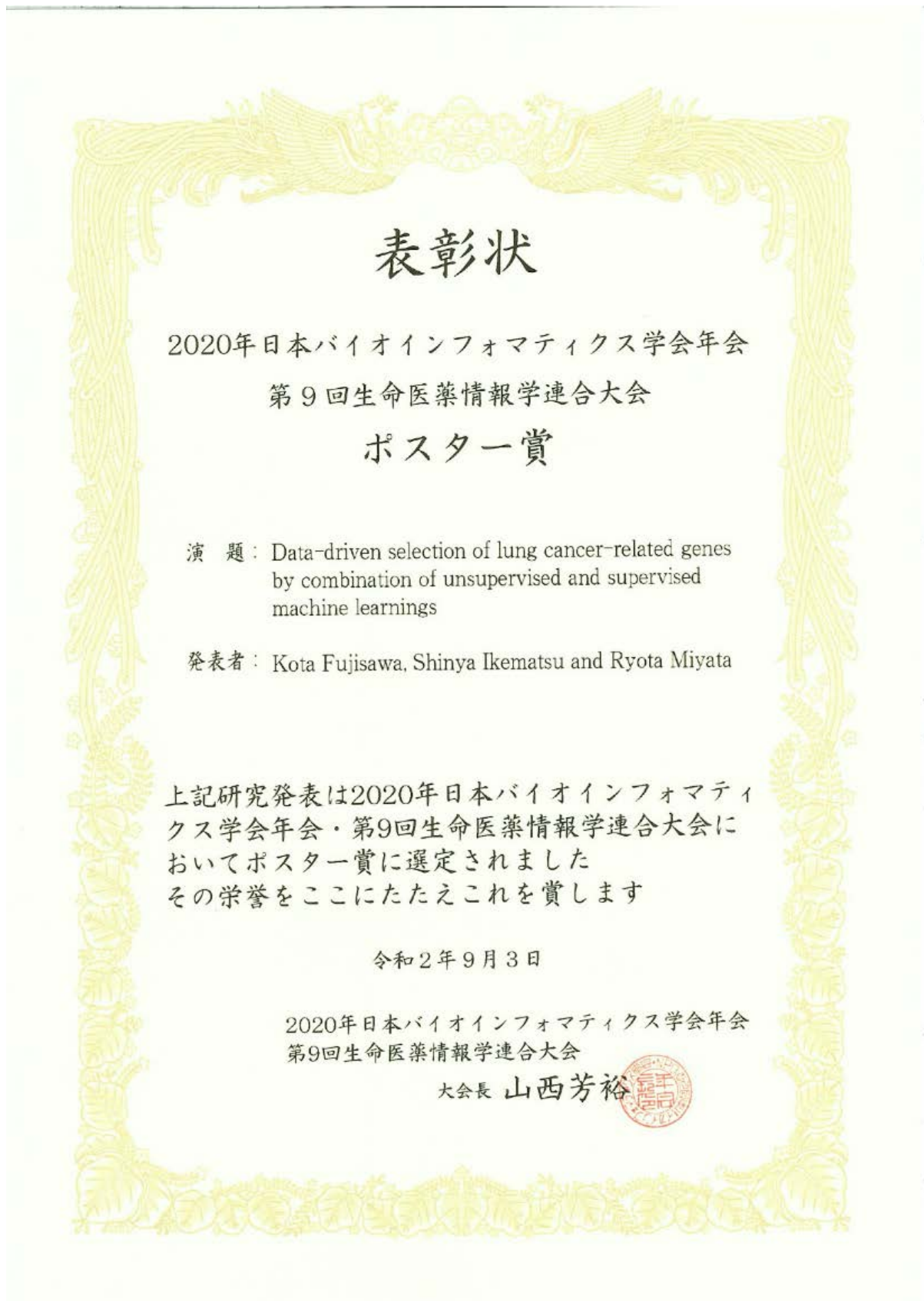


図4：藤澤くんが獲得した賞状