

การศึกษาการทำนายค่าความเสถียรของวัคซีนเอ็มอาร์เอ็นเอสำหรับโรคโควิด 19

ด้วยวิธีการเรียนรู้ของเครื่องจักร

Predicting Stable COVID-19 mRNA Vaccine by Machine Learning

ณัฐนรี พอสุงเนิน¹ และผศ.ดร.จันตรี ผลประเสริฐ¹

คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพมหานคร

nutnaree.poo@g.swu.ac.th

บทคัดย่อ

ในปัจจุบันวัคซีนชนิดเอ็มอาร์เอ็นเอเป็นหนึ่งในเทคโนโลยีที่ถูกนำมาใช้ในการป้องกันการแพร่ระบาดของโรคโควิด 19 โดยวัคซีนชนิดนี้มีประสิทธิภาพในการป้องกันโรคได้สูงถึง 95% อย่างไรก็ตามการนำวัคซีนชนิดเอ็มอาร์เอ็นเอไปใช้งานยังติดปัญหาอยู่หลายประการ อาทิเช่น ความเสถียรของวัคซีนที่อุณหภูมิห้อง การขนส่ง เป็นต้น งานวิจัยนี้นำเสนอการเรียนรู้ของเครื่องจักรทั้งหมด 4 แบบ คือ XGboost, Random Forest, Catboost และ LightGBM ในการทำนายความเสถียรของการย่อยสลายของลำดับอาร์เอ็นเอ โดยจากผลการทดลองพบว่า โมเดลการเรียนรู้ของเครื่องจักรชนิด LightGBM เป็นตัวทำนายที่ดีที่สุดโดยมีค่า Mean Column-wise Root Mean Squared (MCRMSE) เท่ากับ 0.31265 โดยผลลัพธ์ที่ได้ชี้ให้เห็นความเป็นไปได้ในการใช้วิธีการดังกล่าวในการพัฒนาวัคซีนเอ็มอาร์เอ็นเอในอนาคต

คำสำคัญ: โรคโควิด-19, วัคซีนเอ็มอาร์เอ็นเอ, การเรียนรู้ของเครื่องจักร, การย่อยสลาย, การค้นพบยา

Abstract

At present, mRNA vaccine shows promising results in late-stage clinical trials to combat the COVID-19 pandemic. However, there are still many challenges surrounding the vaccine such as their formulation and stability at room temperature that need to be addressed. In this paper, we propose using machine learning methods such as XGboost, Random Forest, Catboost, and LightGBM to predict the stability of mRNA sequence degradation. Using OpenVaccine dataset [3], LightGBM model yields the best performance with the Mean Column-wise Root Mean Squared Error (MCRMSE) equal to 0.31265. Future works will focus on using other type of machine learning models such as deep learning to investigate its impact on predicting the stability of mRNA sequence degradation.

Keywords: Covid-19, mRNA Vaccine, Machine Learning, degradation, Drug Discovery

1. บทนำ

โรคโควิด 19 เกิดจากไวรัสโคโรนา 2019-nCoV โดยส่งผลกระทบต่อทั่วโลก องค์การอนามัยโลกได้ประกาศให้การระบาดนี้เป็นภาวะฉุกเฉินทางสาธารณสุขระหว่างประเทศยาด้านไวรัสจำเพาะแต่กำลังมีการวิจัยอยู่ขณะนี้ ปัจจุบันมีการผลิตวัคซีนเอ็มอาร์เอ็นเอสำหรับโรคโควิด 19 มีบริษัทผู้ผลิตยาหลายบริษัท เช่น โมเดอร์นา (Moderna), ไฟเซอร์ (Pfizer), ซิโนแวค (Sinovac), แอสตราเซนนิกา

(Astra Zeneca) เป็นต้น ได้ผลิตวัคซีนออกมาได้ผลในการป้องกันไวรัสประมาณ 50-100% และคงสภาพได้เมื่อเก็บในอุณหภูมิติดลบไปจนถึง 8 องศาเซลเซียส เนื่องจากวัคซีนเอ็มอาร์เอ็นเอ(mRNA)เป็นหนึ่งในตัวเลือกของการทำวัคซีนที่เร็วและมีประสิทธิภาพสำหรับโรคโควิด แต่จากข้อมูลการเก็บรักษา มีข้อจำกัดการออกแบบโมเลกุลเอ็มอาร์เอ็นเอเนื่องจากยังไม่สามารถรักษาความเสถียรจากคุณสมบัติการย่อยสลายตามธรรมชาติได้

ดังนั้นเมื่อกลางปีพ.ศ.2563 ชุมชน Eterna นำโดย ศาสตราจารย์ Rhiju Das นักชีวเคมีเชิงคำนวณที่ Stanford's School of Medicine และ Eterna เป็นแพลตฟอร์มวิดีโอเกมออนไลน์ รวบรวมนักวิทยาศาสตร์และนักเล่นเกมเพื่อไขปริศนาคัดค้นยาและจากความรู้ด้านวิทยาศาสตร์ ข้อมูลของชุมชน Kaggle ทาง Eterna จึงมอบข้อมูลสำหรับการแข่งขันเพื่อพัฒนาแบบจำลองสำหรับคู่อัตราการย่อยสลายของอาร์เอ็นเอโดยจะทำนายอัตราการย่อยสลายที่เป็นไปได้ของแต่ละฐานของโมเลกุลอาร์เอ็นเอซึ่งได้รับการฝึกฝนจากชุดข้อมูล

งานวิจัยนี้ผู้วิจัยจึงศึกษาการสร้างแบบจำลองการทำนายค่าความเสถียรของวัคซีนเอ็มอาร์เอ็นเอสำหรับโรคโควิด 19 และเปรียบเทียบการทำนายความเสถียรของโมเดลด้วยวิธีการเรียนรู้ของเครื่องจักรใช้เป็นแนวทางในการพัฒนาการวิเคราะห์วัคซีนเอ็มอาร์เอ็นเอต่อไป

2. งานวิจัยที่เกี่ยวข้อง

วัคซีนเอ็มอาร์เอ็นเอเป็นวัคซีนที่มีความยืดหยุ่นสูงสำหรับการผลิตและการใช้งานดีกว่าแบบดั้งเดิมเนื่องจากมีความสามารถสูงสามารถพัฒนาอย่างรวดเร็วและใช้ต้นทุนต่ำ [2] Edison Ong et al (2020) [1] ได้พัฒนาวัคซีนที่มีประสิทธิภาพ และปลอดภัยสำหรับโรคติดเชื้อที่เกิดจากไวรัสโคโรนาพวกเขาใช้ Vaxign reverse Vaccinology 19 และอัลกอริทึมของการเรียนรู้ของเครื่องจักรในการจำแนกรวมถึงการถดถอยโลจิสติก (Logistic regression), เครื่องเวกเตอร์สนับสนุน (Support vector machine), การหาเพื่อนบ้านใกล้เคียง (K-Nearest Neighbor), Random forest และ xgboost ประสิทธิภาพที่ดีที่สุดคือ xgboost งานวิจัยใช้การเรียนรู้ของเครื่องเพื่อค้นหาโปรตีนที่จะชกนภูมิคุ้มกันแต่ไม่ได้ศึกษาความเสถียรของวัคซีน ต่อมา Ankit Singhal [4] เสนองานวิจัยโดยใช้การเรียนรู้เชิงลึกเป็นวิธีการทำนายความเสถียรและการเกิดปฏิกิริยาและความเสี่ยงต่อการย่อยสลายของลำดับอาร์เอ็นเอซึ่งค่อนข้างมีความซับซ้อน ดังนั้นงานวิจัยฉบับนี้ได้สนใจแนวคิดในการทำโมเดลอย่างง่ายและได้ประสิทธิภาพใกล้เคียงกัน

3. วิธีดำเนินการวิจัย

3.1 ชุดของข้อมูล

ข้อมูลโครงการ Open Vaccine: COVID-19 mRNA Vaccine Degradation Prediction [3] จาก Kaggle โดยความร่วมมือของStanford's School of Medicine และ Eterna โดยจากข้อมูลจะแบ่งเป็น

1. train.json - ข้อมูลการฝึกอบรมการฝึกอบรมมีค่าของเอาท์พุทที่เกิดขึ้นจริง (Ground Truth) หลายค่า ใช้คะแนนเฉพาะ reactivity, deg_Mg_pH10 และ deg_Mg_50C
2. test.json - ชุดทดสอบโดยไม่มีคอลัมน์ใด ๆ ที่เกี่ยวข้องกับค่าของเอาท์พุทที่เกิดขึ้นจริง

ตารางที่ 1 รายละเอียดคอลัมน์ของชุดข้อมูลทดสอบ

คอลัมน์	รายละเอียดคอลัมน์
id	รหัสสำหรับแต่ละตัวอย่าง
seq_scored	ค่าจำนวนเต็มแสดงถึงจำนวนตำแหน่งที่ใช้ในการให้คะแนนด้วยค่าที่ทำนาย
seq_length	ความยาวของลำดับ
sequence	ลำดับ RNA การรวมกันของ A, G, U และ C สำหรับแต่ละตัวอย่าง
Structure	อาร์เรย์ของ () และ . อักษรที่อธิบายว่าฐานควรวางจับคู่หรือไม่จับคู่ ฐานที่จับคู่แสดงโดยวงเล็บเปิดและปิดเช่น (....)
reactivity	ค่าความสามารถในการเกิดปฏิกิริยาสำหรับ 68 ฐานแรกตามลำดับและใช้เพื่อกำหนดโครงสร้างรองที่เป็นไปได้ของตัวอย่างอาร์เอ็นเอ
deg_pH10	ค่าความสามารถในการเกิดปฏิกิริยาสำหรับ 68 ฐานแรกตามลำดับและใช้เพื่อกำหนดความเป็นไปได้ที่จะเกิดการย่อยสลายหลังจากการบ่มโดยไม่มีแมกนีเซียมที่ pH สูง (pH 10)
deg_Mg_pH10	ค่าความสามารถในการเกิดปฏิกิริยาสำหรับ 68 ฐานแรกตามลำดับและใช้เพื่อกำหนดโอกาสในการย่อยสลายหลังจากบ่มด้วยแมกนีเซียมใน pH สูง (pH 10)
deg_50C	ค่าความสามารถในการเกิดปฏิกิริยาสำหรับ 68 ฐานแรกตามลำดับและใช้เพื่อกำหนดความเป็นไปได้ที่จะเกิดการย่อยสลายหลังจากการบ่มโดยไม่มีแมกนีเซียมที่อุณหภูมิสูง (50 องศาเซลเซียส)
deg_Mg_50C	ค่าความสามารถในการเกิดปฏิกิริยาสำหรับ 68 ฐานแรกตามลำดับและใช้เพื่อกำหนดความเป็นไปได้ของการย่อยสลายหลังจากบ่มด้วยแมกนีเซียมที่อุณหภูมิสูง (50 องศาเซลเซียส)
error	ข้อผิดพลาดจากการคำนวณในการทดลองที่ได้รับในการทำปฏิกิริยาและคอลัมน์ deg_*
Predict_loop_type	อธิบายบริบทโครงสร้าง (เรียกอีกอย่างว่า "loop type") ของแต่ละอักษรตามลำดับประเภทที่ที่กำหนดโดย bpRNA จากโครงสร้าง Vienna RNAfold 2 จาก bpRNA_documentation: S: จับคู่ "Stem" M: Multiloop I: Internal loop B: Bulge H: Hairpin loop E: dangling End X: eXternal loop
signal_to_noise	ข้อผิดพลาดทางสถิติในค่าการวัดที่มากกว่า 68 เบส ต้องมากกว่า 1.0 (ไม่มีสัญญาณรบกวน ค่า min_value ต้องมากกว่า -0.5 เนื่องจากข้อมูลถูกสร้างขึ้นจากผลการทดลองในห้องปฏิบัติการจึงต้องอธิบายว่าค่าดังกล่าวได้ผ่านขั้นตอนการควบคุมคุณภาพเพื่อให้แน่ใจว่าไม่ได้ใช้ข้อมูลที่ผิดพลาด signal_to_noise ค่า
SN_filter	สัญญาณรบกวนในทั้ง 5 เpsilon ต้องมากกว่า 1.0 (ไม่มีสัญญาณรบกวนมิได้เพื่อตรวจสอบขั้นตอนการควบคุมคุณภาพของข้อมูล

ชุดข้อมูลประกอบด้วยจำนวน 6034 ลำดับ อาร์เอ็นเอ โดยชุดการฝึกอบรมและตรวจสอบความถูกต้อง

ประกอบด้วย 3029 ของลำดับเหล่านี้ โดยมีความยาว 107 ฐานนิวคลีโอไทด์ ชุดข้อมูลการทดสอบประกอบด้วย ลำดับ 3005 ที่มีความยาว 130 ฐานนิวคลีโอไทด์

3.2 สถิติที่ใช้ในการวิจัย

โดยปกติเราสามารถคำนวณ RMSE เพื่อให้ได้เมตริกการประเมินตัวเลขเดียวสำหรับการคาดการณ์ของเรา แต่ถ้าเราคาดการณ์หลายค่าพร้อมกัน เช่น จำเป็นต้องทำนายอัตราการย่อยสลายภายใต้เงื่อนไขที่หลากหลาย เราจะได้รับหลายค่า RMSE ที่แตกต่างกันหนึ่งค่าสำหรับแต่ละคอลัมน์ MCRMSE เป็นเพียงค่าเฉลี่ยของค่า RMSE ทั้งหมดสำหรับแต่ละคอลัมน์ดังสมการตามด้านล่าง

$$MCRMSE = \frac{1}{N_t} \sum_{j=1}^{N_t} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{ij} - \hat{Y}_{ij})^2}$$

โดย MCRMSE คือ ค่าเฉลี่ยของค่า RMSE

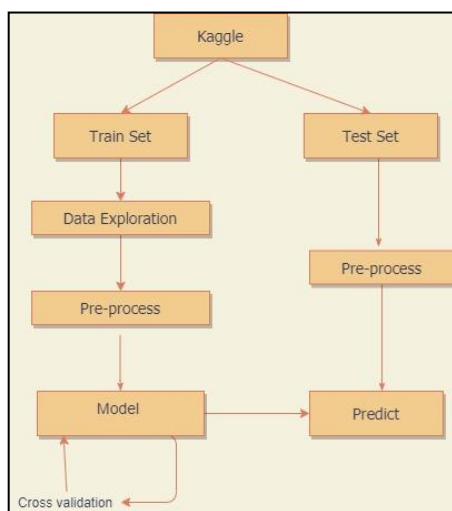
N_t คือ จำนวนคอลัมน์ scored ground truth target

n คือ จำนวนของชุดทดสอบ

y คือ ค่าจริง \hat{y} คือค่าที่ทำนายได้

3.3 ขั้นตอนการดำเนินการวิจัย

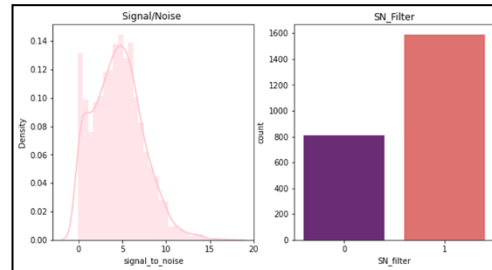
การดำเนินการวิจัยเป็นไปตามแผนผังจากรูป โดยนำข้อมูลจาก Kaggle ทั้งชุดฝึกฝนและชุดทดสอบมาเข้ากระบวนการการสำรวจข้อมูลและเตรียมข้อมูลก่อนจะนำเข้าโมเดลและนำไปคาดการณ์ค่าอัตราการย่อยสลาย



ภาพที่ 3 ขั้นตอนการดำเนินงานของงานวิจัย

กระบวนการการสำรวจข้อมูลค่า Summary แต่ละคอลัมน์ และตรวจสอบรายละเอียดของข้อมูล

ตรวจสอบ Signal to noise และ SN filter และตรวจสอบค่าเฉลี่ยการย่อยสลายของเป้าหมาย ทำการนำข้อมูลคอลัมน์ที่ไม่ได้ใช้งานออก



ภาพที่ 4 ค่าSignal to noise และ SN filter เป็นค่าที่ใช้ตรวจสอบคุณภาพของข้อมูลเนื่องจากเป็นข้อมูลที่ได้จากห้องปฏิบัติการ

นำข้อมูลที่ได้จากกระบวนการเตรียมข้อมูลมาใช้ทดสอบโมเดลโดยงานวิจัยนี้ใช้โมเดลเปรียบเทียบกันทั้งหมด 4 โมเดล คือ Xgboost ใช้เป็น Baseline, Random Forest, Catboost และ LightGBM

3.3.1 XGBoost

เป็นโมเดลที่นำเอา Decision Tree มาต่อกัน โดยที่แต่ละ decision tree จะเรียนรู้จาก error ของ tree ก่อนหน้า เมื่อมีการเรียนรู้ของ tree ต่อเนื่องกันจนมีความลึกมากพอและโมเดลจะหยุดเรียนรู้เมื่อไม่เหลือรูปแบบของ error จาก tree ก่อนหน้าให้เรียนรู้แล้ว

3.3.2 RandomForest

เป็นโมเดลที่นำ Decision Tree หลาย ๆ tree มา Train ร่วมกัน (ตั้งแต่ 10 ต้น ถึงมากกว่า 1000 ต้น) โดยที่แต่ละ tree จะได้รับ feature และ data เป็น subset ของ feature และ data ทั้งหมดแบบสุ่มตอนคาดการณ์แต่ละ Decision Tree ทำการคาดการณ์ของตัวเองและเลือกผล final prediction

3.3.3 Catboost

เป็นโมเดลการจัดการสำหรับคุณสมบัติตามหมวดหมู่ได้รับความนิยมเมื่อเทียบกับอัลกอริธึม Gradient boosting อื่นๆเนื่องจากคุณสมบัติการใช้

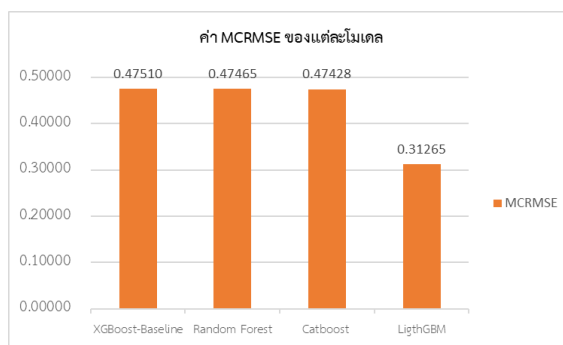
Oblivious Trees หรือ Symmetric Trees เพื่อให้โมเดลทำงานได้เร็วขึ้น

ในระดับหนึ่ง ทำให้โมเดลมีประสิทธิภาพที่เทียบเคียงกับการใช้วิธีการเรียนรู้เชิงลึกที่มีความซับซ้อนกว่า

3.3.4 Light GBM

เป็นเฟรมเวิร์ก Gradient Boosting Framework ที่รวดเร็ว, ประสิทธิภาพสูงโดยใช้อัลกอริธึมแผนผังและงานการเรียนรู้ของเครื่องอื่น ๆ อีกมากมายใช้วิธีการแยกความลึกของ Trees

ในการตั้งค่าพารามิเตอร์ใช้เทคนิค 5-Fold cross-validation และทำการปรับโมเดลเพื่อหาค่าพารามิเตอร์ที่เหมาะสมด้วยการใช้ GridSearchCV



ภาพที่ 4 ค่า MCRMSE ของแต่ละโมเดล

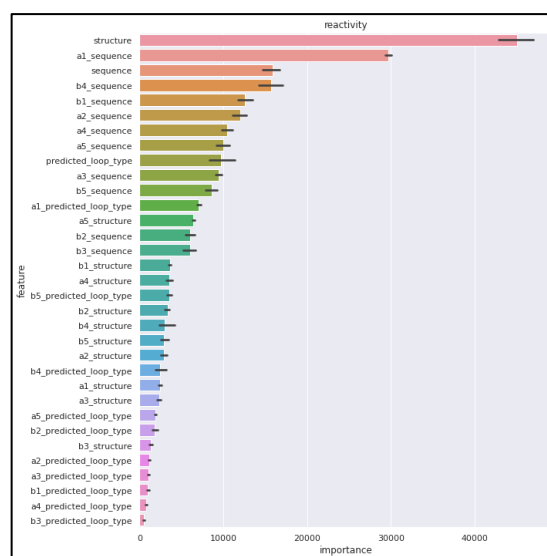
4. ผลการศึกษาและการอภิปรายผล

ผลการศึกษาจากการตั้งค่าพารามิเตอร์ต่าง ๆ ซึ่งเป็นพารามิเตอร์ที่เหมาะสม (Best Parameter) ตามตารางสามารถรายงานและอภิปรายผลได้ดังนี้

นำโมเดลที่ได้จากเทคนิค LigthGBM มาหา Feature Importance มี Feature ที่ส่งผลกับโมเดลแตกต่างกันไปตามรูป

ตารางที่ 2 พารามิเตอร์ที่เหมาะสมของแต่ละโมเดลในที่ใช้ในการเรียนรู้

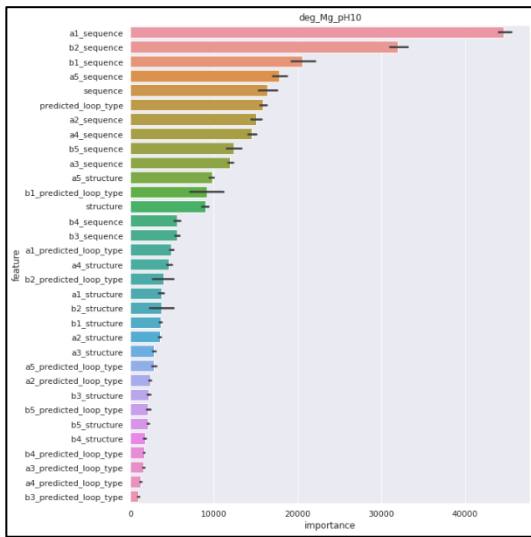
โมเดล	พารามิเตอร์
XGBoost	learning_rate: 0.1, estimator__max_depth: 5 n_estimators: 800 subsample=0.8 colsample_bytree=0.75 reg_lambda=2 reg_alpha=1
Random Forest	max_depth: 11 n_estimators: 800
Catboost	learning_rate: 0.1 max_depth: 8 n_estimators: 300
LigthGBM	learning_rate: 0.05



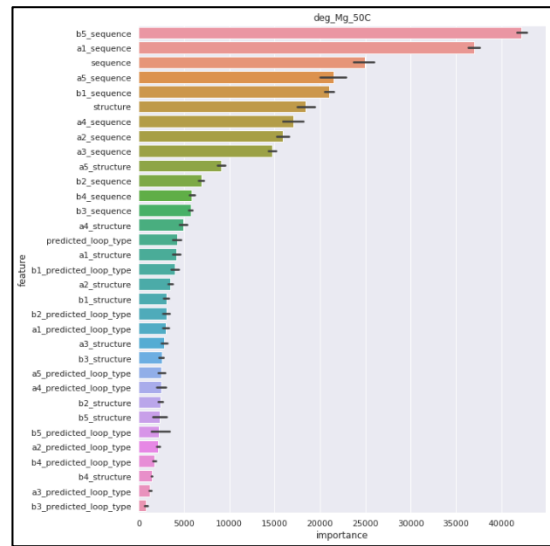
ภาพที่ 5 Feature importance ของค่า reactivity

จาก Baseline ได้ค่า MCRMSE 0.4751 นั้น เมื่อนำเทคนิค Random Forest และ Catboost มาใช้งานและหาพารามิเตอร์ที่เหมาะสมพบว่าค่า MCRMSE ตีกว่าเล็กน้อยคือ 0.47465 และ 0.47428 แต่เมื่อใช้ LigthGBM พบว่า ค่าตีกว่าและแตกต่างจาก baseline ค่อนข้างเยอะคือ 0.31265 ซึ่งอาจเป็นผลมาจาก LightGBM เป็นเทคนิคที่เหมาะสมกับข้อมูลที่ใหญ่ประมาณหนึ่งโดย Trees ของมันเติบโตไปทางแนวตั้ง และลดการสูญเสียอัลกอริธึมที่สำคัญได้

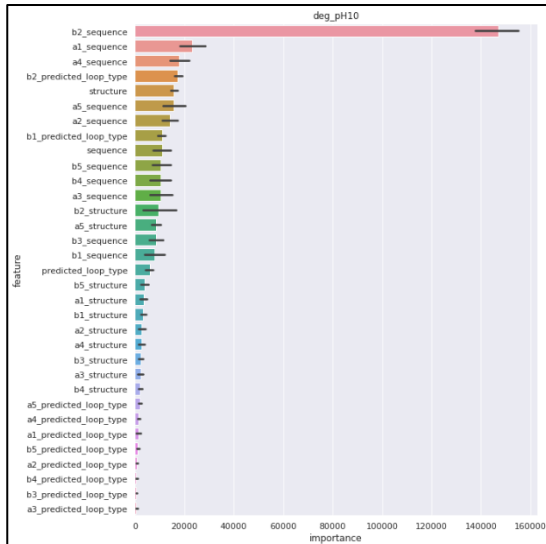
การประชุมวิชาการเสนอผลงานวิจัยระดับชาติ ด้านวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏจันทรเกษม ครั้งที่ 4
วันที่ 22 พฤษภาคม 2564 ณ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏจันทรเกษม



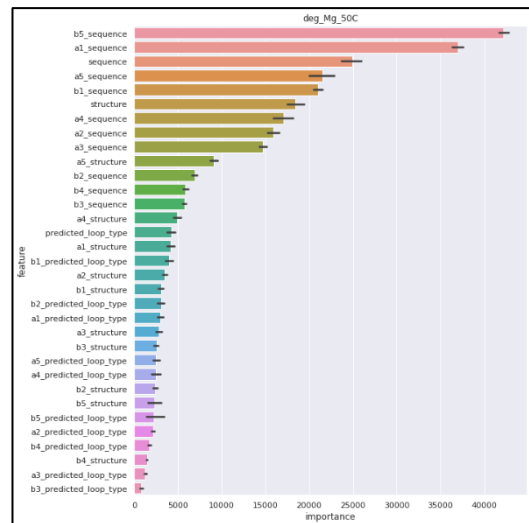
ภาพที่ 6 Feature importance ของค่า deg_Mg_pH10



ภาพที่ 8 Feature importance ของค่า deg_Mg_50C



ภาพที่ 7 Feature importance ของค่า deg_pH10



ภาพที่ 9 Feature importance ของค่า deg_50C

แต่ละเป้าหมายมีตำแหน่งของอาร์เอ็นเอในการย่อยสลายที่แตกต่างกันไป ตามตารางคือ Feature importance ที่สำคัญที่สุด ที่น่าสนใจคือ pH10 เมื่อมีแมกนีเซียมมาเกี่ยวข้องตำแหน่งในการย่อยสลายจะเปลี่ยน แต่ที่อุณหภูมิ 50 องศาเซลเซียสยังคงเป็นตำแหน่งในการย่อยสลายเดิม

ตารางที่ 3 Feature Importance ที่สำคัญที่สุดของแต่ละเป้าหมาย

Target	Feature importance
reactivity	structure
deg_Mg_pH10	a1_sequence
deg_Mg_50C	b5_sequence
deg_pH10	b2_sequence
deg_50C	b5_sequence

5. สรุปและข้อเสนอแนะ

แม้ว่าการสร้างโมเดลเอ็มอาร์เอ็นเอที่เสถียรจะยังคงเป็นเรื่องยากแต่การใช้ชุดข้อมูลของลำดับของเอ็มอาร์เอ็นเอกำลังเป็นที่นิยมมากขึ้นในการวิจัย

การใช้วิธีการเรียนรู้ของเครื่องจักรในงานวิจัยนี้มีค่า MCRMSE อยู่ประมาณ 0.31-0.48 การใช้เทคนิคดังกล่าวมีความสามารถในการเพิ่มความเร็วและประสิทธิภาพของการค้นพบวัคซีนเอ็มอาร์เอ็นเอที่เสถียรและมีผลต่อการวิจัยในสาขาอื่น ๆ ที่เกี่ยวข้องของเทคนิคนี้อาจถูกใช้เพื่อทดสอบคัดกรองเพื่อลบลำดับที่ไม่เสถียรออกไป ผู้วิจัยใช้ผลของ Public Leaderboard เมื่อนำไปเปรียบเทียบกับผลการทดลองอื่นอยู่ในระดับปานกลางซึ่งอาจจะเหมาะสมเมื่อนำไปต่อยอดด้วยการวิเคราะห์ด้านอื่น ๆ ประกอบกัน

ข้อจำกัดของวิธีนี้คือ ความยาวของลำดับของโมเดลเอ็มอาร์เอ็นเอที่ใช้ความยาวที่ใช้ในงานวิจัยนี้อยู่ระหว่าง 107-130 ฐานในขณะที่วัคซีน Covid-19 ที่แท้จริงน่าจะมี ความยาวตั้งแต่ 3,000-4,000 ฐาน แต่ผลของงานนี้แสดงให้เห็นว่าอัลกอริทึมการทำนายดังกล่าวมีความเป็นไปได้และมีศักยภาพในการประหยัดเวลาในระหว่างกระบวนการวิจัยในระหว่างการระบาดของโรค

ในระยะยาวเทคนิคดังกล่าวอาจช่วยในการวิจัยได้ดีขึ้นในการทำความเข้าใจเหตุผลเบื้องหลังความเสถียรของโมเดลเอ็มอาร์เอ็นเอบางชนิดและช่วยในการพัฒนาเทคโนโลยีที่เกี่ยวข้อง โดยหวังเป็นอย่างยิ่งว่างานนี้จะ

ประโยชน์ต่อนักวิทยาศาสตร์ข้อมูลคนอื่น ๆ ในการสร้างการคาดการณ์ที่ดีขึ้นในอนาคต

กิตติกรรมประกาศ

ขอขอบคุณการสนับสนุนทุนนำเสนอผลงานจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

เอกสารอ้างอิง

- [1] Ong, Wong, Huffman, & He, (2020) and Wurtmann & Wolin, (2009). *COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning*. doi:10.1101/2020.03.20.000141
- [2] Schlake, T., Thess, A., Fotin-Mleczek, M., & Kallen, K.-J. (2012). Developing mRNA-vaccine technologies. *RNA Biology*, 9(11), 1319-1330. doi:10.4161/rna.22269
- [3] Open Vaccine: *COVID-19 mRNA Vaccine Degradation Prediction*. <https://www.kaggle.com/c/stanford-covid-vaccine/data>.
- [4] Singhal, A. (2020). Application and Comparison of Deep Learning Methods in the Prediction of RNA Sequence Degradation and Stability. *Cornell University*.
- [5] Florindo, H. F., Kleiner, R., Vaskovich-Koubi, D., Acúrcio, R. C., Carreira, B., Yeini, E., Satchi-Fainaro, R. (2020). Immune-mediated approaches against COVID-19. *Nature Nanotechnology*, 15(8), 630-645.
- [6] Pardi, N., Hogan, M. J., Porter, F. W., และ Weissman, D. (2018). mRNA vaccines - a new era in vaccinology. *Nat Rev Drug Discov*, 17(4), 261-279.