



Advancing skin toxicity predictions with HaCaT cells and machine learning



จัดทำโดย นศ.ก. เคลิมเดช ฤณุวัฒน์
อาจารย์ที่ปรึกษา ผศ.ดร.ธราพงษ์ ศรีสุขุม
คณะเภสัชศาสตร์ มหาวิทยาลัยขอนแก่น



หลักการและเหตุผล

ความเป็นพิษทางผิวน้ำ คือ ผลเสียที่เกิดจากการได้รับสารผ่านทางผิวน้ำ ทั้งเกิดแค่เฉพาะที่หรือทั่วระบบไปเลวเวียนโลหิตในมนุษย์หรือสัตว์ ฉะนั้น จึงมีความจำเป็นที่จะต้องทดสอบความเป็นพิษของสารเคมีก่อนที่จะสามารถนำไปใช้ได้

ในปัจจุบัน การทดสอบความเป็นพิษทางผิวน้ำของสารเคมีสามารถทำการทดสอบได้โดยหลายวิธี และหนึ่งในนั้นคือการใช้แบบจำลองทางคอมพิวเตอร์

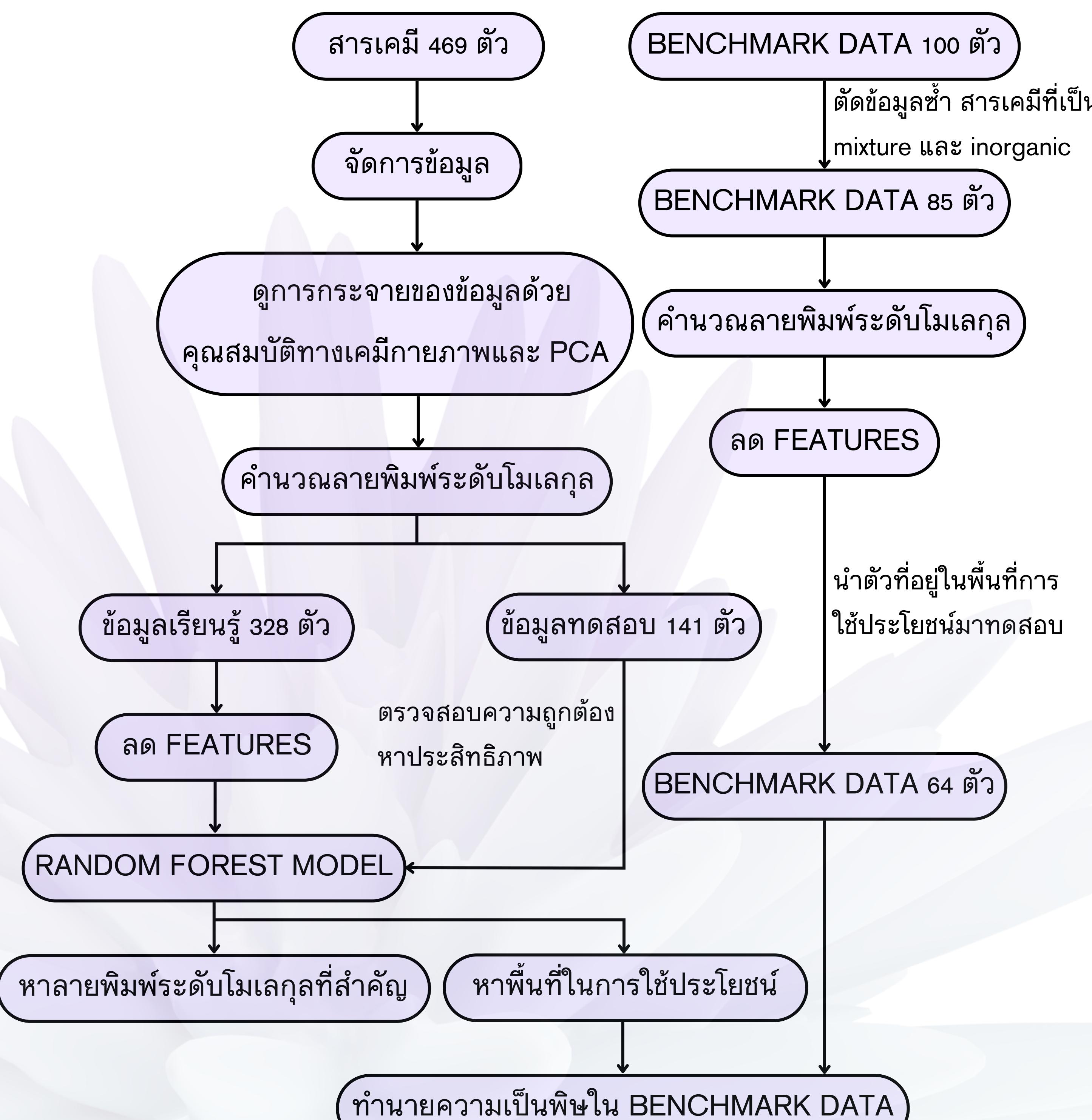
HaCaT cell เป็นเซลล์ไลน์เคราติโนไซด์อมตะของมนุษย์ (Immortalized human keratinocyte cell line) ชนิด aneuploid ซึ่งมีการพัฒนาให้เป็นเซลล์เคราติโนไซด์ของมนุษย์ในหลอดทดลอง เนื่องจากมีคุณสมบัติเพิ่มจำนวนมากและอัตราการตายต่ำในหลอดทดลอง จึงใช้ในงานวิจัยทางวิทยาศาสตร์อย่างแพร่หลาย

การทดสอบความเป็นพิษในปัจจุบันสนับสนุนหลักการ 3Rs จึงนำมาสู่การพัฒนาแบบจำลองการศึกษาความล้มเหลวของโครงสร้างและความเป็นพิษทางผิวน้ำโดยใช้วิธีการเรียนรู้ของเครื่องคอมพิวเตอร์ (Machine learning) เพื่อใช้ในการคำนวณความเป็นพิษทางผิวน้ำของสารเคมีที่ต้องการทดสอบ

วัตถุประสงค์

- พัฒนาแบบจำลองการเรียนรู้ของเครื่องคอมพิวเตอร์ (Machine learning) ในการคำนวณความเป็นพิษทางผิวน้ำโดยการใช้ข้อมูลความสัมพันธ์โครงสร้างและความเป็นพิษของสารเคมี
- ตรวจทานแบบจำลองในการคำนวณความเป็นพิษทางผิวน้ำที่สร้างขึ้น โดยการใช้ข้อมูลความสัมพันธ์โครงสร้างและความเป็นพิษของสารเคมี
- หาประสิทธิภาพของแบบจำลองที่พัฒนาขึ้นด้วยการเรียนรู้ของเครื่อง โดยนำแบบจำลองไปคำนวณ Benchmark dataset

วิธีดำเนินการวิจัย

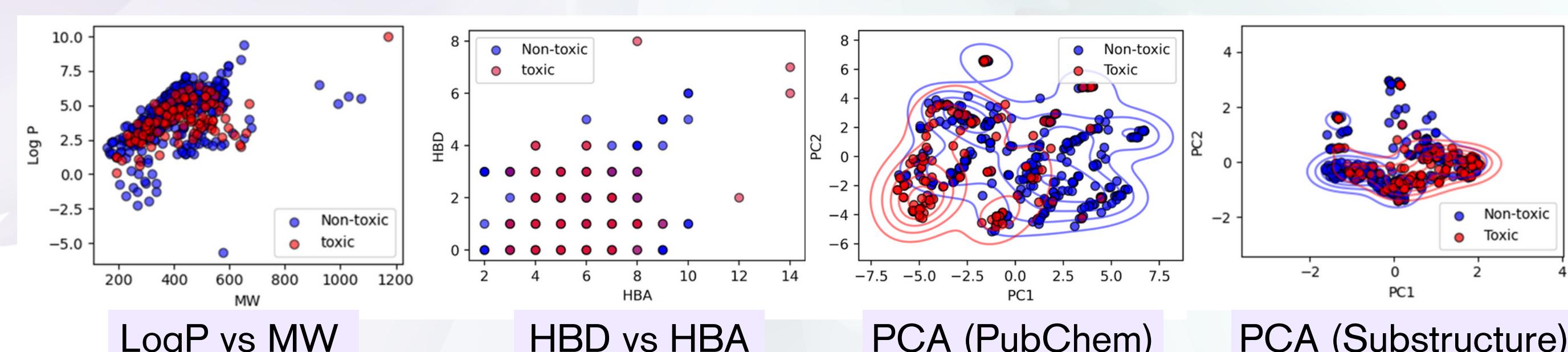


สรุปผล

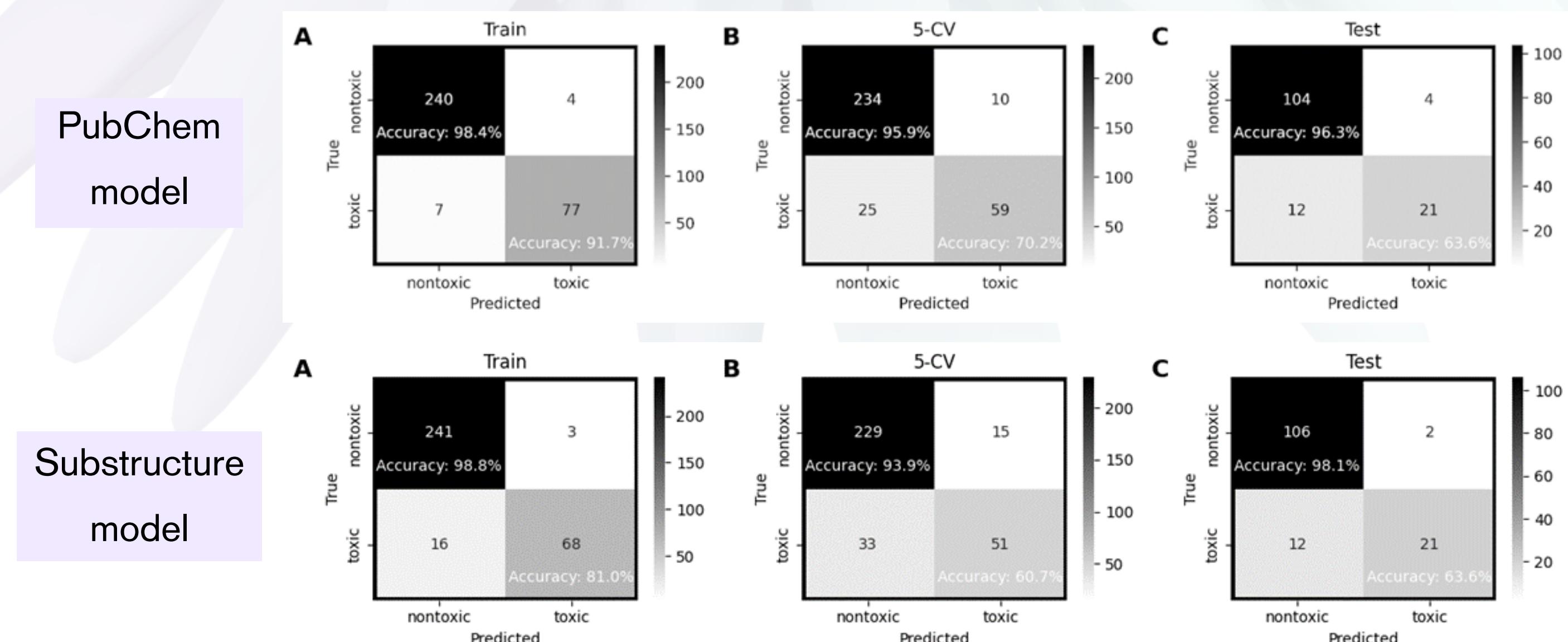
แบบจำลองทางคอมพิวเตอร์ที่ถูกสร้างโดยอัลกอริทึมแบรนดومฟอร์เรสต์มีประสิทธิภาพในการคำนวณความเป็นพิษทางผิวน้ำเพื่อประเมินการระคายเคืองผิวน้ำด้านความถูกต้อง (accuracy) และความจำเพาะ (specificity) มากกว่าการทดสอบพิษในหลอดทดลองด้วยเซลล์ EpiSkin™, SkinEthic™, LabCyt EPI-MODEL24 และ KeraSkin™ และสามารถใช้แบบจำลองนี้เพื่อกำหนด Benchmark ได้โดยมีความถูกต้องอยู่ที่ร้อยละ 72 และ 64 ด้วยการใช้แบบจำลองที่สร้างด้วยลายพิมพ์ระดับโมเลกุล PubChem และ Substructure ตามลำดับ

ผลการวิจัย

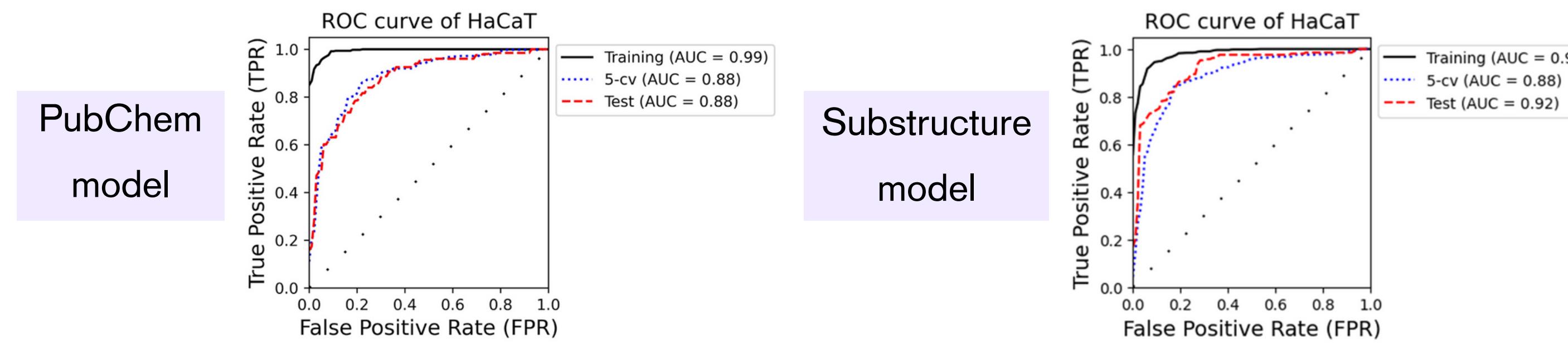
การกระจายข้อมูล



Confusion matrix



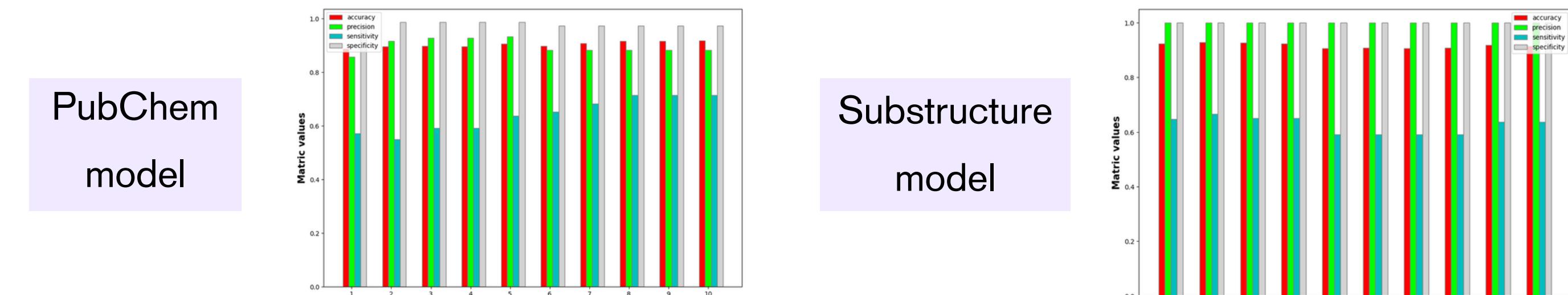
Receiver Operating Characteristics (ROC) Curve



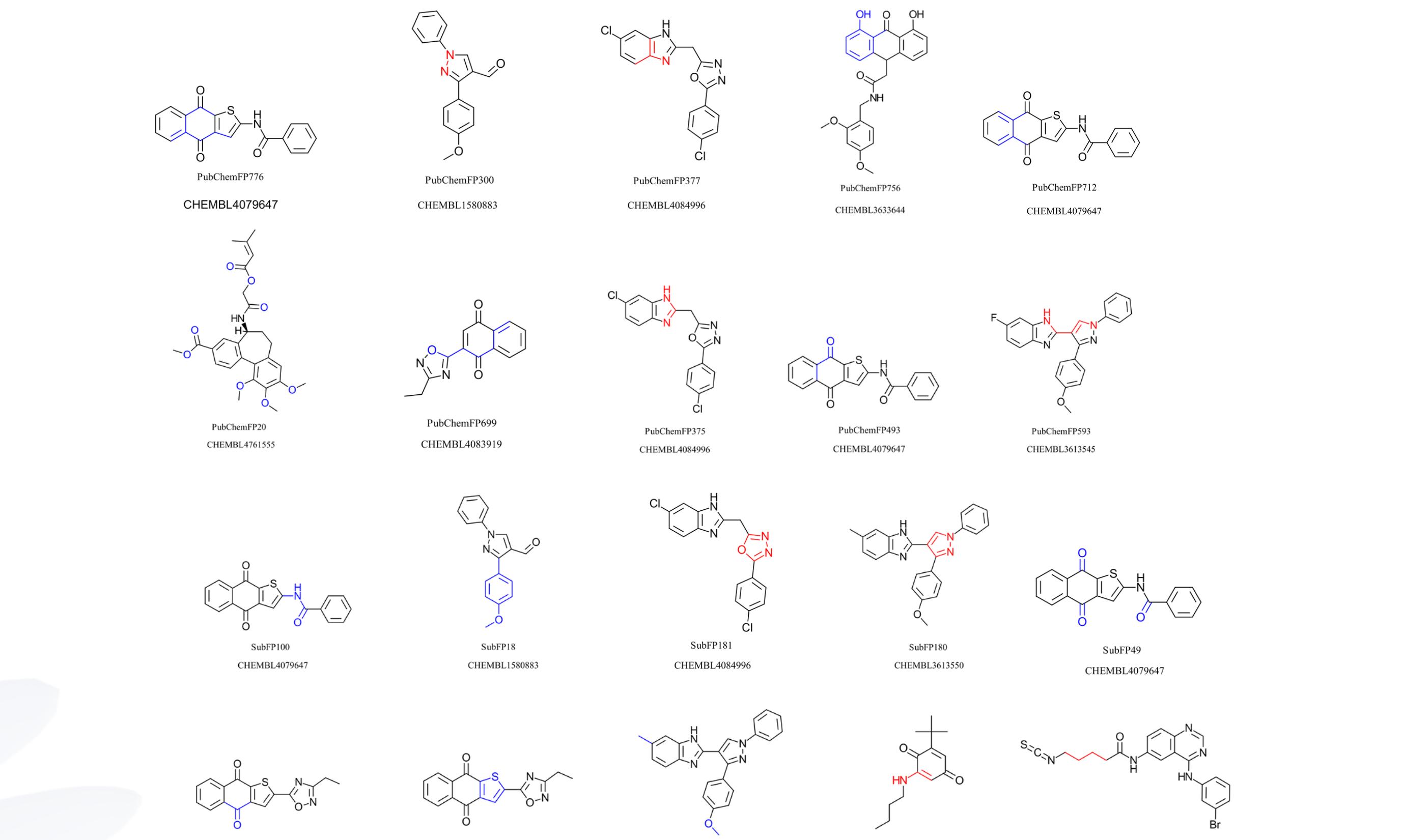
Accuracy, Precision, Sensitivity and Specificity

Random Forest model	accuracy			precision			sensitivity			specificity		
	train	cv	test	train	cv	test	train	cv	test	train	cv	test
PubChem	0.97	0.89	0.89	0.95	0.85	0.84	0.92	0.70	0.64	0.98	0.96	0.96
Substructure	0.94	0.85	0.90	0.96	0.77	0.91	0.81	0.61	0.64	0.99	0.94	0.98

Applicability domain



ลายพิมพ์ระดับโมเลกุลที่สำคัญ



ประสิทธิภาพของแบบจำลองทางคอมพิวเตอร์

