

การทำนายความเสี่ยงการเกิดภาวะพร่องไทรอยด์ด้วยวิธีการเรียนรู้ ของเครื่องพร้อมการเลือกคุณลักษณะที่จำเป็น

Predicting Hypothyroidism Risk using Machine Learning with Essential Feature Selection

จักรพรรดิ โพธิพิภฏกุล^{1*}, ภัสร์ณลีน ธนาธีระไพศาล² และรัชดาพร คณาวงษ์³
Chakkraphat Photiphiphattarakul^{1*}, Phasnalini Tanateeraphaisarn², and Ratchadaporn Kanawong³

Received : May 8, 2026; Revised : June 23, 2026; Accepted : August 26, 2026

บทคัดย่อ

โรคภาวะพร่องไทรอยด์เป็นปัญหาสุขภาพที่พบได้บ่อย อาการในระยะแรกมักไม่ชัดเจนทำให้การวินิจฉัยล่าช้า งานวิจัยนี้พัฒนาโมเดลการเรียนรู้ของเครื่องเพื่อทำนายภาวะพร่องไทรอยด์โดยใช้ข้อมูลทุติยภูมิ (Secondary Data) จากผู้ป่วย 3,772 ราย จากคลัง UCI Machine Learning Repository ปัญหาหลักคือความไม่สมดุลของข้อมูล (Class Imbalance อัตราส่วน 24:1 ระหว่างกลุ่มปกติ 3,621 ราย และกลุ่มป่วย 151 ราย) และข้อมูลขาดหาย ผู้วิจัยใช้ Median Imputation ในการจัดการข้อมูลที่ขาดหาย และใช้เทคนิค SMOTE เพื่อปรับสมดุลข้อมูล ร่วมกับการคัดเลือกคุณลักษณะด้วย Random Forest Feature Importance ซึ่งสามารถลดจำนวนตัวแปรจาก 23 ตัว เหลือเพียง 8 ตัวแปรสำคัญที่ระดับเกณฑ์ 0.01 จากการเปรียบเทียบประสิทธิภาพของ SVM, Decision Tree, Random Forest และ MLP ด้วยตัวบ่งชี้ Accuracy, Precision, Recall และ F1-Score พบว่า Decision Tree ที่ผ่านการปรับพารามิเตอร์ด้วย GridSearchCV ให้ประสิทธิภาพสูงสุดด้วยค่าความแม่นยำ (Accuracy) ร้อยละ 99.9 ค่า Recall ร้อยละ 99.9 และค่า F1-Score ร้อยละ 99.6 (หรือในรูปทศนิยม 0.993, 0.999 และ 0.996 ตามลำดับ) จากการทดสอบกับ Test Set ซึ่งผ่านการตรวจสอบความเสถียรด้วย 5-Fold Cross-Validation โมเดลนี้มีความโปร่งใสสูงและง่ายต่อการตีความ ซึ่งเหมาะสมต่อการนำไปใช้เป็นระบบสนับสนุนการตัดสินใจทางการแพทย์

คำสำคัญ : ต้นไม้ตัดสินใจ การคัดเลือกคุณลักษณะ ภาวะพร่องไทรอยด์ การเรียนรู้ของเครื่อง แรนดอมฟอเรสต์

^{1*} จักรพรรดิ โพธิพิภฏกุล สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

^{1*} Chakkaphat Photiphiphathkul Department of Computer Science, Faculty of Science, Silpakorn University.

² ภัสร์ณลีน ธนาธีระไพศาล สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

² Phasnalini Tanateeraphaisarn Department of Computer Science, Faculty of Science, Silpakorn University.

³ รัชดาพร คณาวงษ์ อาจารย์ประจำภาควิชา สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

³ Ratchadaporn Kanawong Lecturer at the Department of Computer Science, Faculty of Science, Silpakorn University

Abstract

Hypothyroidism is a prevalent health condition with subtle early symptoms, leading to delayed diagnosis. This research develops a machine learning predictive model using secondary data of 3,772 patients from the UCI Machine Learning Repository. Challenges including class imbalance (24:1 ratio; 3,621 normal vs. 151 hypothyroid cases) and missing values were addressed using Median Imputation and SMOTE. We applied Random Forest-based Feature Importance for dimensionality reduction, systematically reducing features from 23 to 8 essential variables (TSH, FTI, T3, TT4, T4U, thyroid surgery history, thyroxine usage, and age) with a threshold of 0.01. Comparing SVM, Decision Tree Random Forest and MLP using Accuracy, Precision, Recall, and F1-Score, the Decision Tree optimized with GridSearchCV achieved the highest performance with 99.9% Accuracy, 99.9% Recall, and 99.6% F1-Score (decimal values: 0.993, 0.999, 0.996) on the Test Set, verified by 5-Fold Cross-Validation. The model offers high interpretability, making it suitable for clinical decision support.

Keywords : Decision Tree Feature Selection Hypothyroidism Machine Learning Random Forest

1. บทนำ

โรคพร่องไทรอยด์ (Hypothyroidism) เป็นภาวะที่ต่อมไทรอยด์ผลิตฮอร์โมนไทรอยด์ได้ไม่เพียงพอต่อความต้องการของร่างกาย ส่งผลให้กระบวนการเมแทบอลิซึมทำงานช้าลง ผู้ป่วยมักมีอาการอ่อนเพลีย น้ำหนักเพิ่มขึ้น ทนความเย็นไม่ได้ ผิวกแห้ง ผมร่วง ท้องผูก และในรายที่รุนแรงอาจส่งผลต่อการทำงานของหัวใจ และระบบประสาท (Biondi & Cooper, 2008)

องค์การอนามัยโลก (WHO) รายงานว่าโรคต่อมไทรอยด์เป็นหนึ่งในโรคต่อมไร้ท่อที่พบบ่อยที่สุดในโลก โดยประมาณการว่าประชากรโลกกว่า 750 ล้านคนมีภาวะความผิดปกติของต่อมไทรอยด์ในรูปแบบใดรูปแบบหนึ่ง และในจำนวนนี้มีผู้ป่วยโรคพร่องไทรอยด์ราว 200 ล้านคนทั่วโลก (Chaker et al., 2017) สำหรับในประเทศไทย กรมการแพทย์ กระทรวงสาธารณสุข รายงานว่าโรคต่อมไทรอยด์มีแนวโน้มพบเพิ่มขึ้นอย่างต่อเนื่อง โดยเฉพาะในกลุ่มผู้หญิงวัยกลางคนและสูงอายุ (กรมการแพทย์ กระทรวงสาธารณสุข, 2564) ในการวินิจฉัยโรคพร่องไทรอยด์ แพทย์จำเป็นต้องอาศัยข้อมูลทางคลินิกหลายประการ ได้แก่ ผลการตรวจระดับฮอร์โมน TSH (Thyroid-Stimulating Hormone), T3 (Triiodothyronine) และ T4 (Thyroxine) ในเลือด รวมถึงอาการทางคลินิก ประวัติผู้ป่วย และผลการตรวจอื่น ๆ ในเชิงทฤษฎีการแพทย์ ระดับฮอร์โมน TSH ถือเป็นตัวบ่งชี้หลักที่มีความไวและความจำเพาะสูงที่สุดในการวินิจฉัยภาวะพร่องไทรอยด์ โดย TSH ที่สูงผิดปกติร่วมกับระดับ T3 และ T4 ที่ต่ำลงเป็นเกณฑ์มาตรฐานสากลในการยืนยันการวินิจฉัย (Garber et al., 2012) ตัวแปรเหล่านี้จึงมีความสำคัญอย่างยิ่งในการนำมาใช้เป็นคุณลักษณะ (Feature) สำหรับการสร้างแบบจำลอง AI ข้อมูลเหล่านี้ถูกจัดเก็บในระบบฐานข้อมูลทางการแพทย์ ซึ่งเป็นโอกาสให้นาเทคโนโลยีปัญญาประดิษฐ์ (AI) และการเรียนรู้ของเครื่อง (ML) มาช่วยสนับสนุนการวินิจฉัยของแพทย์ให้มีความแม่นยำและรวดเร็วยิ่งขึ้น (Chen et al., 2017) อย่างไรก็ตาม ปัญหาสำคัญที่พบในชุดข้อมูลทางการแพทย์สำหรับโรคพร่องไทรอยด์คือความไม่สมดุลของข้อมูล (Class Imbalance) (He & Garcia, 2009) ส่งผลให้แบบจำลองการเรียนรู้ของเครื่องที่สร้างขึ้นมีความลำเอียง (Bias) ในการทำนาย ปัญหานี้มีความสำคัญอย่างยิ่งในบริบทการแพทย์ เนื่องจากการพลาดการวินิจฉัย

ผู้ป่วย (False Negative) อาจส่งผลเสียต่อสุขภาพของผู้ป่วยได้อย่างรุนแรง (Canaris et al., 2000) นอกจากนี้ยังพบปัญหาความซับซ้อนของตัวแปรจำนวนมาก (High Dimensionality) ซึ่งทำให้โมเดลเกิดปัญหา Overfitting (Guyon & Elisseeff, 2003)

งานวิจัยส่วนใหญ่ที่ผ่านมายังขาดการบูรณาการระหว่างการแก้ปัญหาค่าความไม่สมดุลของข้อมูล (Class Imbalance) และการคัดเลือกคุณลักษณะ (Feature Selection) อย่างเป็นระบบในบริบทของภาวะพร่องไทรอยด์โดยตรง งานวิจัยนี้จึงมุ่งศึกษาการนำเทคนิค SMOTE (Chawla, et al., 2002) มาใช้ในการแก้ปัญหาค่าความไม่สมดุลของข้อมูล ร่วมกับการเลือกคุณลักษณะด้วย Random Forest Feature Importance (Breiman, 2001) เพื่อสร้างแบบจำลองการเรียนรู้ของเครื่องที่มีประสิทธิภาพสูงในการทำนายโรคพร่องไทรอยด์ (Guyon & Elisseeff, 2003)

2. วัตถุประสงค์การวิจัย

2.1 เพื่อพัฒนาแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) สำหรับการทำนายภาวะพร่องไทรอยด์ โดยใช้ชุดข้อมูลทุติยภูมิจากคลัง UCI Machine Learning Repository

2.2 เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึม SVM, Decision Tree, Random Forest และ MLP ด้วยตัวบ่งชี้ Accuracy, Precision, Recall และ F1-Score โดยเน้นเป็นพิเศษที่ค่า Recall และ F1-Score อันเป็นตัวชี้วัดสำคัญสำหรับข้อมูลที่มี Class Imbalance

2.3 เพื่อประเมินประสิทธิภาพของเทคนิค SMOTE และการคัดเลือกคุณลักษณะด้วย Random Forest Feature Importance ในการแก้ปัญหาค่าความไม่สมดุลของข้อมูลและลดมิติตัวแปร

2.4 เพื่อประเมินผลกระทบของการคัดเลือกคุณลักษณะต่อประสิทธิภาพของแบบจำลองที่ดีที่สุด และวิเคราะห์ความสำคัญทางคลินิกของตัวแปรที่ถูกคัดเลือก

3. สมมุติฐานของการวิจัย

3.1 การใช้เทคนิค SMOTE ร่วมกับการเลือกคุณลักษณะ (Feature Selection) จะช่วยเพิ่มประสิทธิภาพการทำนายภาวะพร่องไทรอยด์ในแง่ของ Recall และ F1-Score ได้สูงกว่าโมเดลที่ไม่ใช้เทคนิคดังกล่าวอย่างมีนัยสำคัญ

3.2 อัลกอริทึมประเภท Tree-based (Decision Tree และ Random Forest) จะให้ค่าความแม่นยำและค่าความระลึก (Recall) สูงกว่าโมเดลอื่น (SVM และ MLP) ในชุดข้อมูลที่มีความไม่สมดุล

4. การทบทวนวรรณกรรม

4.1 ภาวะพร่องไทรอยด์

โรคพร่องไทรอยด์ (Hypothyroidism) คือภาวะที่ต่อมไทรอยด์ซึ่งอยู่บริเวณคอด้านหน้าไม่สามารถผลิตฮอร์โมนไทรอยด์ได้เพียงพอ ได้แก่ ฮอร์โมน Thyroxine (T4) และ Triiodothyronine (T3) ซึ่งมีบทบาทสำคัญในการควบคุมกระบวนการเมแทบอลิซึมของร่างกาย การเต้นของหัวใจ อุณหภูมิร่างกาย และการทำงานของระบบประสาท Biondi, A., & Cooper, D. S. (2008) สาเหตุที่พบบ่อยที่สุดคือโรค Hashimoto's Thyroiditis ซึ่งเป็นโรคภูมิคุ้มกันตนเอง (Autoimmune Disease) (Ganie & Malik, 2022) สาเหตุอื่น ๆ ได้แก่ การรักษาโรคไทรอยด์เป็นพิษด้วยไอโอดีนกัมมันตรังสี การผ่าตัดต่อมไทรอยด์ การขาดสารไอโอดีน (Chaker et al., 2017)

และโรคต่อมไธสมอง การวินิจฉัยอาศัยการตรวจระดับ TSH ในเลือดเป็นหลัก โดยค่า TSH ที่สูงกว่าปกติ (>4.5 mIU/L) ร่วมกับ Free T4 ที่ต่ำกว่าปกติ บ่งชี้ถึงภาวะพร่องไทรอยด์ Garber et al., (2012)

4.2 อัลกอริทึมการเรียนรู้ของเครื่องที่ใช้ในงานวิจัย

งานวิจัยนี้เลือกใช้อัลกอริทึม 4 ประเภท ซึ่งมีหลักการทำงานดังนี้

Decision Tree (Quinlan, 1986) เป็นโมเดลที่สร้างโครงสร้างต้นไม้โดยแบ่งข้อมูลด้วยเกณฑ์ที่ลด Impurity (เช่น Gini Impurity หรือ Information Gain) ในแต่ละโหนด มีข้อดีคือตีความได้ง่าย โปร่งใส และเหมาะกับข้อมูลทางการแพทย์ที่แพทย์ต้องการทราบเหตุผลการตัดสินใจ

Random Forest (Breiman, 2001) เป็นการรวม Decision Trees หลายต้น (Ensemble Method) ที่สร้างจากการสุ่มตัวอย่างข้อมูล (Bootstrap Sampling) และสุ่มคุณลักษณะ ผลการทำนายได้จากการโหวตเสียงข้างมาก ทำให้มี Variance ต่ำกว่า Decision Tree เดียว นอกจากนี้ Random Forest ยังให้ Feature Importance Score ซึ่งวัดจากการลด Impurity เฉลี่ยเมื่อนำตัวแปรแต่ละตัวมาแบ่งข้อมูล ทำให้เหมาะสมอย่างยิ่งสำหรับการคัดเลือกคุณลักษณะ (Feature Selection) Guyon & Elisseeff, 2003)

Cortes & Vapnik (1995) กล่าวว่า Support Vector Machine (SVM) เป็นโมเดลที่หาเส้นแบ่ง (Hyperplane) ที่มีระยะห่างจากจุดข้อมูลที่ใกล้ที่สุดของแต่ละคลาสมากที่สุด (Maximum Margin) มีประสิทธิภาพดีในข้อมูลที่มีมิติสูง แต่ใช้เวลาประมวลผลนานและตีความยาก

Multi-Layer Perceptron (MLP) เป็นโครงข่ายประสาทเทียมที่มีหลายชั้น (Hidden Layers) เหมาะกับปัญหาที่มีความสัมพันธ์ซับซ้อน แต่ต้องการข้อมูลจำนวนมากและขาดความโปร่งใสในการตีความ

Technique (2002) กล่าวว่า SMOTE (Synthetic Minority Over-Sampling เป็นเทคนิคสร้างข้อมูลสังเคราะห์สำหรับคลาสส่วนน้อย โดยการนำตัวอย่างในคลาสส่วนน้อยมาสร้างจุดข้อมูลใหม่ตาม k-Nearest Neighbors แทนที่จะทำ Simple Oversampling ซึ่งช่วยลดปัญหา Overfitting ที่เกิดจากการทำซ้ำข้อมูลเดิม

4.3 งานวิจัยที่เกี่ยวข้อง

Chaganti et al. (2022) ได้ศึกษา การเปรียบเทียบ Random Forest, Decision Tree และ Naive Bayes สำหรับการทำนายโรคพร่องไทรอยด์ พบว่า Random Forest ให้ค่า Accuracy สูงสุดที่ 0.987 อย่างไรก็ตาม งานวิจัยนี้ไม่ได้แก้ปัญหา Class Imbalance ทำให้ค่า Recall ยังอยู่ในระดับต่ำ

Ganie & Malik (2022) ได้ศึกษา SVM ร่วมกับ Chi-Square Feature Selection สำหรับโรคไทรอยด์ พบว่าการเลือกคุณลักษณะช่วยเพิ่มประสิทธิภาพ แต่ยังมีข้อจำกัดด้านความสมดุลของข้อมูล

Kaur & Kumari (2022) ได้เสนอการใช้ SMOTE ร่วมกับ k-NN สำหรับโรคเบาหวาน ซึ่งแสดงให้เห็นว่า SMOTE ช่วยปรับปรุงค่า F1-Score และ AUC-ROC ได้อย่างมีนัยสำคัญในชุดข้อมูลที่ไม่สมดุล แนวทางดังกล่าวสามารถประยุกต์ใช้กับโรคพร่องไทรอยด์ได้

Abiyev & Ma'aitah (2018) นำเสนอ Fuzzy Neural Network สำหรับการวินิจฉัยโรคไทรอยด์ซึ่งมีการตีความผลได้ดีกว่า Neural Network แบบดั้งเดิม แต่มีความซับซ้อนในการปรับแต่งพารามิเตอร์สูง

Temurtas et al., (2009) ได้ใช้ Neural Network ในการวินิจฉัยโรคไทรอยด์โดยได้ค่าความแม่นยำสูง แต่แบบจำลองมีความซับซ้อนสูงและขาดการตีความ (Interpretability) ซึ่งเป็นข้อจำกัดสำคัญในการนำไปใช้ทางคลินิก

จากการทบทวนวรรณกรรมพบว่า มีข้อจำกัดสำคัญสองประการ ได้แก่ งานที่แก้ปัญหา Class Imbalance มักไม่มีการคัดเลือกคุณลักษณะอย่างเป็นระบบ และงานที่ทำ Feature Selection มักไม่ได้แก้ปัญหา Class Imbalance ส่งผลให้ค่า Recall ต่ำ งานวิจัยนี้จึงเข้ามาเติมช่องว่างดังกล่าวโดย บูรณาการทั้ง SMOTE และ Random Forest Feature Importance เข้าด้วยกันในบริบทของภาวะพร่องไทรอยด์โดยตรง

5. ข้อมูลในการวิจัย

งานวิจัยนี้ใช้ข้อมูลทุติยภูมิ (Secondary Data) จากชุดข้อมูล Hypothyroid Disease Dataset ซึ่งรวบรวมจากผู้ป่วยที่เข้ารับการตรวจที่โรงพยาบาล Garvan Institute of Medical Research ประเทศออสเตรเลีย โดยได้รับการเผยแพร่ผ่านคลัง UCI Machine Learning Repository และสามารถดาวน์โหลดได้จากแพลตฟอร์ม Kaggle (Bidyapati. (n.d.) เหตุผลที่เลือกใช้ชุดข้อมูลนี้ได้แก่ (1) เป็นชุดข้อมูลมาตรฐานที่ถูกอ้างอิงอย่างแพร่หลายในงานวิจัยด้าน ML สำหรับโรคต่อมไทรอยด์ (Chaganti et al., 2022; Ganie & Malik, 2022) (2) ประกอบด้วยตัวแปรทางคลินิกครบถ้วนทั้งผลการตรวจเลือดและประวัติผู้ป่วย และแสดงปัญหา Class Imbalance ที่ชัดเจน ซึ่งเหมาะสำหรับการศึกษาเทคนิค SMOTE (Canaris et al., 2000)

ชุดข้อมูลประกอบด้วยตัวอย่างทั้งสิ้น 3,772 รายการ แบ่งเป็นผู้ป่วยโรคพร่องไทรอยด์ (Class P: Positive/Hypothyroid) จำนวน 151 ราย (ร้อยละ 4.0) และผู้ที่ไม่เป็นโรค (Class N: Negative) จำนวน 3,621 ราย (ร้อยละ 96.0) แสดงให้เห็นอัตราส่วนความไม่สมดุลอยู่ที่ประมาณ 24:1 คุณลักษณะในชุดข้อมูลมีทั้งหมด 30 ตัวแปร ได้แก่ ข้อมูลประชากร (อายุ เพศ) ประวัติทางการแพทย์ (การตั้งครรภ์ การใช้ยา Thyroxine การรักษาด้วยไอโอดีนกัมมันตรังสี) และผลการตรวจทางห้องปฏิบัติการ (TSH, T3, TT4, T4U, FTI)

	TSH	FTI	TT4	on thyroxine	T3	T4U	age	thyroid surgery	
0	1.30	109.0	125.0		0	2.5	1.14	34	0
1	4.10	107.0	102.0		0	2.0	0.98	15	0
2	0.98	120.0	109.0		0	2.0	0.91	40	0
3	0.16	107.0	175.0		1	1.9	0.98	67	0
4	0.72	70.0	61.0		0	1.2	0.87	67	0
...	...	...	...		...	...	...	...	...
3767	1.40	107.0	103.0		0	2.0	0.98	22	0
3768	1.00	114.0	124.0		0	2.1	1.08	64	0
3769	5.10	105.0	112.0		0	1.8	1.07	71	0
3770	0.70	87.0	82.0		0	2.0	0.94	69	0
3771	1.00	92.0	99.0		0	2.2	1.07	60	0

3621 rows x 8 columns

ภาพที่ 1 ตัวอย่างข้อมูลที่ใช้ในการวิจัย

6. วิธีดำเนินงานวิจัย

งานวิจัยนี้ใช้ข้อมูลทุติยภูมิจากฐานข้อมูลสาธารณะ (Public Database) โดยไม่มีขั้นตอนเก็บข้อมูลภาคสนาม ผู้วิจัยดำเนินการทดลองด้วยภาษา Python เวอร์ชัน 3.9 โดยใช้ไลบรารีหลักดังนี้: Scikit-learn เวอร์ชัน 1.2.2 สำหรับการสร้างและประเมินโมเดล ML และ Imbalanced-learn เวอร์ชัน 0.10.1 สำหรับ SMOTE สำหรับการคัดเลือกคุณลักษณะ (Feature Selection) ได้กำหนดเกณฑ์ (Threshold) ไว้ที่ 0.01 ซึ่งทำให้สามารถคัดเลือก 8 ตัวแปรสำคัญที่มีนัยสำคัญต่อการทำนาย ได้แก่

1. TSH, FTI, TT4, T3, T4U (ข้อมูลประเภทตัวเลข - Continuous)
2. On Thyroxine, Thyroid Surgery (ข้อมูลประเภทหมวดหมู่ - Binary)
3. Age (ข้อมูลประเภทตัวเลข)

6.1 การเตรียมข้อมูล

6.1.1 การจัดการข้อมูลที่ขาดหาย

ข้อมูลทั้งหมดที่มีค่า '?' ถูกแปลงเป็น NaN จากนั้นลบคอลัมน์ที่ไม่มีผลต่อการทำนาย ทำให้ข้อมูลลดลงจาก 3,772 แถว 30 คอลัมน์ เหลือ 3,621 แถว 23 คอลัมน์ โดยยังคงเก็บข้อมูลสำคัญทางการแพทย์ไว้ เช่น อายุและเพศ เนื่องจากเป็นปัจจัยเสี่ยงสำคัญของโรคไทรอยด์

6.1.2 การเติมค่าที่ขาดหายและการแปลงข้อมูล

แปลงค่าในคอลัมน์ตัวเลขจาก Object เป็น Numeric จากนั้นเติมค่า NaN ด้วยค่า Median ของคอลัมน์นั้น ๆ เนื่องจากค่า Median มีความทนทานต่อ Outlier มากกว่าค่า Mean สำหรับข้อมูลหมวดหมู่ เช่น 'yes'/'no' หรือ 'M'/'F' ทำการแปลงเป็นตัวเลขโดยใช้ Label Encoding เช่น 'yes' = 1, 'no' = 0

6.2 การแบ่งชุดข้อมูล

แบ่งข้อมูลออกเป็น Training Set 80% (2,896 ตัวอย่าง) และ Test Set 20% (725 ตัวอย่าง) โดยกำหนด random_state = 42 เพื่อให้ผลลัพธ์สามารถทำซ้ำได้ (Reproducibility) และใช้เป็นมาตรฐานในการเปรียบเทียบโมเดลต่าง ๆ

6.3 การตรวจสอบและแก้ไขปัญหาความไม่สมดุลของข้อมูล

จากการตรวจสอบพบว่าข้อมูล Training Set มีความไม่สมดุล (Class Imbalance) อย่างชัดเจน โดยกลุ่มปกติ (Class N: Negative) มีจำนวนตัวอย่างประมาณ 2,775 ตัว ขณะที่กลุ่มป่วย (Class P: Hypothyroid) มีเพียงประมาณ 121 ตัว คิดเป็นอัตราส่วนประมาณ 23 :1 จึงใช้เทคนิค SMOTE (Chawla et al., 2002) เพื่อสร้างข้อมูลสังเคราะห์สำหรับกลุ่มป่วย (Class P) ให้มีจำนวนเท่ากับกลุ่มปกติ (Class N) เพื่อให้โมเดลสามารถเรียนรู้ทั้งสองกลุ่มได้อย่างสมดุล

6.4 การเลือกคุณลักษณะที่สำคัญ

ใช้โมเดล Random Forest Classifier (Breiman, 2001) ในการวัด Feature Importance ซึ่งคำนวณจากการลด Impurity เฉลี่ย (Mean Decrease in Impurity) ในแต่ละโหนดของต้นไม้ทุกต้นในป่า โดยกำหนด threshold ที่ 0.01 ผลลัพธ์พบว่าจากทั้งหมด 23 คุณลักษณะ มีเพียง 8 คุณลักษณะที่ผ่านเกณฑ์ (ดังแสดงในภาพที่ 2) การลดจำนวนคุณลักษณะช่วยลดความซับซ้อนของโมเดลและเพิ่มความเร็วในการประมวลผล (Guyon & Elisseeff, 2003)

TSH	0.608859
FTI	0.097133
TT4	0.096453
on thyroxine	0.062679
T3	0.035840
T4U	0.031191
age	0.026725
thyroid surgery	0.010472

ภาพที่ 2 ค่าความสำคัญจาก Random Forest

6.5 การสร้างและทดสอบโมเดลเบื้องต้น

ทำการสร้างและทดสอบโมเดล 4 ประเภท ในหลายสถานการณ์เพื่อเปรียบเทียบประสิทธิภาพ ได้แก่ (1) โมเดลพื้นฐาน (2) SMOTE (3) Feature Selection และ (4) Feature Selection + SMOTE

6.6 การปรับแต่ง Hyperparameters ด้วย GridSearchCV

จากผลการทดลองเบื้องต้น พบว่าการใช้ Feature Selection ให้ผลลัพธ์ที่ดีกว่า PCA อย่างมีนัยสำคัญ จึงเลือกใช้วิธีนี้ในการปรับแต่งเพิ่มเติม โดยใช้ GridSearchCV ร่วมกับ Cross-Validation 5 Folds เพื่อค้นหา Hyperparameters ที่เหมาะสมที่สุด โดยมีช่วงค่าพารามิเตอร์ (Search Space) ดังนี้

1. Decision Tree: max_depth = [3, 5, 7, 10, None], min_samples_split = [2, 5, 10], min_samples_leaf = [1, 2, 4]
2. Random Forest: n_estimators = [100, 200, 300], max_depth = [5, 10, None], max_features = ['sqrt', 'log2'], min_samples_split = [2, 5]
3. MLP: hidden_layer_sizes = [(50,), (100,)], (100,50)], activation = ['relu', 'tanh'], alpha = [0.0001, 0.001], learning_rate = ['constant', 'adaptive']

6.7 การประเมินผลและเลือกโมเดล

ประเมินด้วย Accuracy, Precision, Recall และ F1-Score โดยเน้นเป็นพิเศษที่ค่า Recall เนื่องจากในบริบทการแพทย์ การพลาดวินิจฉัยผู้ป่วย (False Negative) มีต้นทุนสูงกว่าการวินิจฉัยเกิน (False Positive)

7. ผลการทดลอง

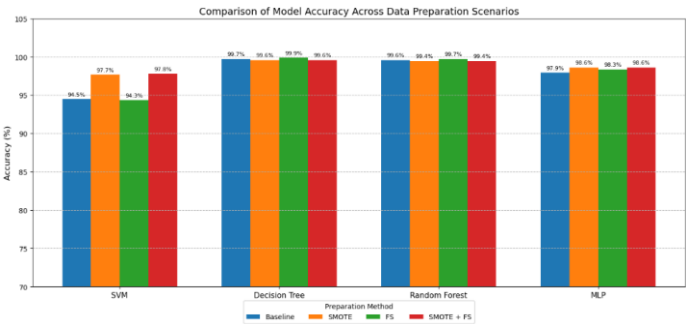
ตารางที่ 1 สรุปผลการทดลองของโมเดลที่เลือกใช้ Feature Selection และ SMOTE

Model	Precision	Recall	F1	Accuracy
SVM	0.939	0.710	0.776	0.944
SVM + SMOTE	0.903	0.973	0.934	0.976
SVM + Feature Selection	0.924	0.709	0.773	0.943
SVM + SMOTE + Feature Selection	0.902	0.987	0.939	0.977
Decision Tree	0.991	0.991	0.991	0.997
Decision Tree + SMOTE	0.978	0.997	0.987	0.995
Decision Tree + Feature Selection	0.993	0.999	0.996	0.999
Decision Tree + SMOTE + Feature Selection	0.978	0.997	0.980	0.987
Random Forest	0.984	0.990	0.987	0.995
andom Forest + SMOTE	0.971	0.997	0.983	0.994
Random Forest + Feature Selection	0.985	0.998	0.991	0.997
Random Forest + SMOTE + Feature Selection	0.971	0.997	0.983	0.994

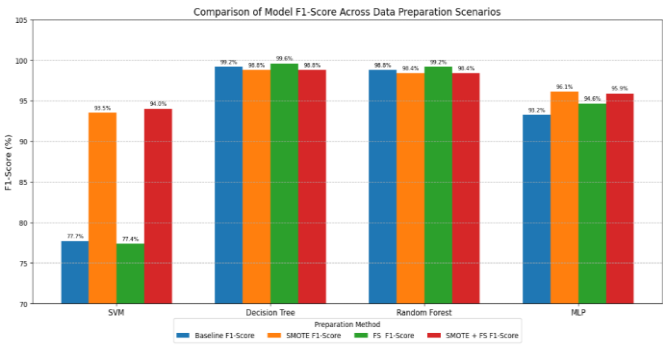
ตารางที่ 1 (ต่อ)

Model	Precision	Recall	F1	Accuracy
MLP	0.972	0.900	0.932	0.979
MLP + SMOTE	0.934	0.992	0.961	0.986
MLP + Feature Selection	0.982	0.915	0.946	0.983
MLP + SMOTE + Feature Selection	0.952	0.965	0.958	0.986

จากตารางที่ 1 พบว่าโมเดล Decision Tree + Feature Selection ให้ประสิทธิภาพสูงสุดด้วยค่า Precision = 0.993, Recall = 0.999, F1-Score = 0.996 และ Accuracy = 0.999 (ร้อยละ 99.3, 99.9, 99.6 และ 99.9 ตามลำดับ) รองลงมาคือ Random Forest + Feature Selection ที่มีค่า Accuracy = 0.997 ส่วน SVM (Baseline) ให้ประสิทธิภาพต่ำสุดโดยเฉพาะค่า Recall = 0.710 ซึ่งบ่งชี้ถึงข้อจำกัดในการตรวจจับผู้ป่วยที่มีภาวะโรคจริง ทั้งนี้บทคัดย่อและสรุปผลอ้างอิงค่าจากกรณี Decision Tree + Feature Selection ซึ่งเป็นโมเดลที่ดีที่สุด (Accuracy 0.999 หรือ 99.9 %) สอดคล้องกับข้อมูลในตารางที่ 1 อย่างครบถ้วน

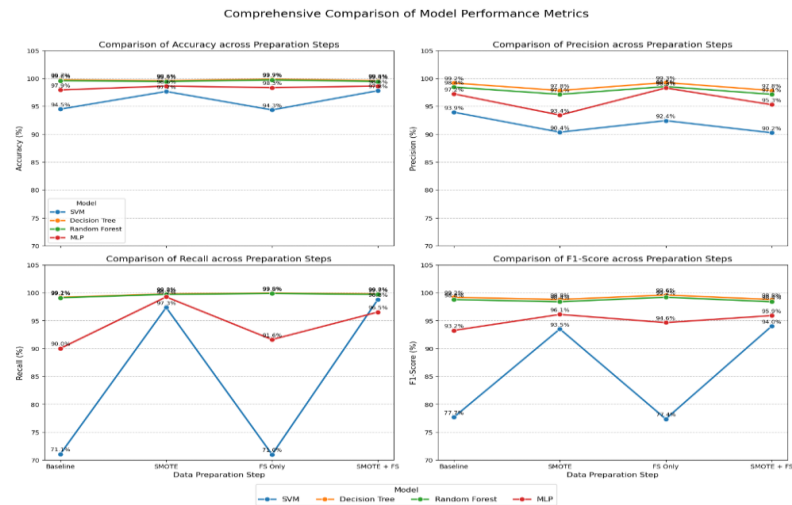


ภาพที่ 3 Bar Plot แสดงค่า Accuracy ของทุกโมเดล



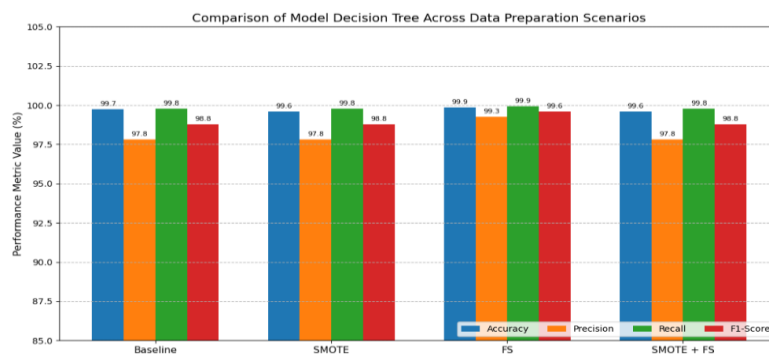
ภาพที่ 4 Bar Plot แสดงค่า F1-Score ของทุกโมเดล

จากภาพที่ 3 และ 4 แสดง Bar Plot เปรียบเทียบค่า Accuracy และ F1-Score ของทุกโมเดลในทุกสถานการณ์ ซึ่งยืนยันว่า Decision Tree และ Random Forest ให้ผลสูงกว่า SVM และ MLP อย่างชัดเจน โดยเฉพาะในสถานการณ์ที่ใช้ Feature Selection



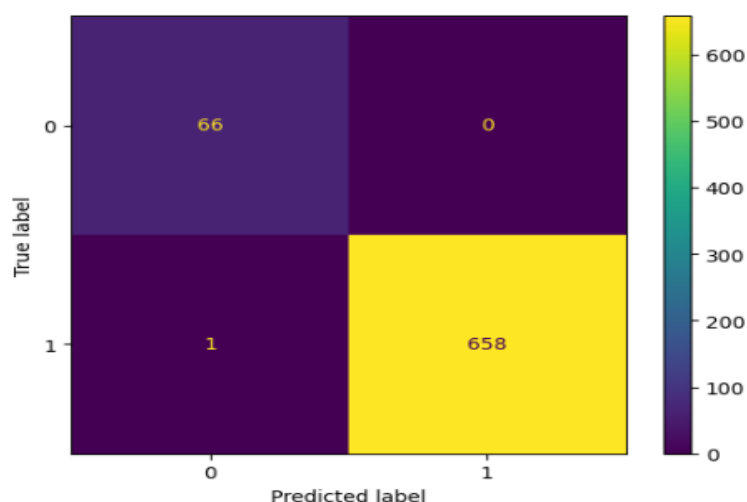
ภาพที่ 5 Line Graph แสดงผลลัพธ์ของทุกโมเดล

จากภาพที่ 5 จะเห็นว่าโมเดล Decision Tree นั้นให้ผลลัพธ์ที่ดีที่สุด โดยให้ผลลัพธ์ที่ดีกว่าโมเดล Support Vector Machine และ Multi-Layer Perceptron อย่างชัดเจน และใกล้เคียงกับโมเดล Random Forest เป็นอย่างมาก



ภาพที่ 6 Bar Graph แสดงผลลัพธ์ของโมเดล Decision Tree

จากภาพที่ 6 แสดง Bar Graph เปรียบเทียบ Decision Tree ในสถานการณ์ต่างๆ ยืนยันว่าการทำ Feature Selection ให้ผลลัพธ์ที่ดีที่สุด โดยได้ค่า Accuracy = 0.999, Precision = 0.993, Recall = 0.999 และ F1-Score = 0.996



ภาพที่ 7 Confusion Matrix ของโมเดล Decision Tree (Feature Selection)

จากภาพที่ 7 แสดง Confusion Matrix ของโมเดล Decision Tree + Feature Selection บน Test Set (725 ตัวอย่าง) โดยแกน True label แสดงคลาสจริงและแกน Predicted label แสดงคลาสที่โมเดลทำนาย ผลลัพธ์มีดังนี้: True Negative (TN) = 66 หมายถึงผู้ที่ไม่เป็นโรคและโมเดลทำนายถูกต้องว่าไม่เป็นโรค, False Positive (FP) = 0 หมายถึงผู้ที่ไม่เป็นโรคแต่โมเดลทำนายผิดว่าเป็นโรค, False Negative (FN) = 1 หมายถึงผู้ที่ เป็นโรคแต่โมเดลพลาดการวินิจฉัย และ True Positive (TP) = 658 หมายถึงผู้ที่ เป็นโรคพร้อม ไทรอยด์จริงและแบบจำลองสามารถทำนายระบุตัวตนได้อย่างถูกต้องว่าเป็นโรค ซึ่งแสดงให้เห็นว่าโมเดลพลาด การวินิจฉัยผู้ป่วยเพียง 1 รายจากทั้งหมด อันเป็นผลลัพธ์ที่ดีมากในทางคลินิก

8. อภิปรายผลการวิจัย

8.1 ประสิทธิภาพการคัดเลือกคุณลักษณะ (Feature Selection)

การใช้ Random Forest Feature Importance (Breiman, 2001) ช่วยลดตัวแปรจาก 23 เหลือ 8 ตัวแปรหลัก ได้แก่ TSH, FTI, TT4, T3, T4U, ประวัติการผ่าตัดไทรอยด์ การใช้ยาไทรอยด์ และอายุ การตั้งเกณฑ์ Threshold ที่ 0.01 เป็นจุดที่เหมาะสมเนื่องจากช่วยตัดตัวแปรที่มีความสำคัญต่ำซึ่งอาจเป็นสัญญาณรบกวน (Noise) ออกไป ทำให้โมเดลสามารถโฟกัสไปที่ตัวแปรที่มีนัยสำคัญทางคลินิกได้อย่างแม่นยำ (Guyon & Elisseeff, 2003)

8.2 ความสำคัญทางคลินิกของตัวแปรที่คัดเลือก

ตัวแปรทั้ง 8 ที่คัดเลือกได้มีความสอดคล้องกับมาตรฐานการวินิจฉัยทางการแพทย์ในปัจจุบันอย่างสูง (Garber et al., 2012; Jonklaas et al., 2014) ดังนี้

1. TSH เป็นตัวบ่งชี้หลักและมี Feature Importance สูงสุด (0.609) ซึ่งสอดคล้องกับแนวทางเวชปฏิบัติสากลที่ใช้ TSH เป็น First-line Test ในการวินิจฉัยภาวะพร่องไทรอยด์ (Garber et al., 2012)
2. FTI และ TT4 สะท้อนระดับฮอร์โมนไทรอยด์ที่ใช้ยืนยันการวินิจฉัยร่วมกับ Biondi & Cooper (2008)
3. T3 และ T4U ใช้ประเมินความรุนแรงและแยกประเภทของภาวะพร่องไทรอยด์ (Garber et al., 2012)

4. ประวัติการผ่าตัดไทรอยด์ เป็นสาเหตุโดยตรงของภาวะพร่องไทรอยด์ภายหลัง (Post-Surgical Hypothyroidism) (Jonklaas et al., 2014)

5. การใช้ยาไทรอกซีน บ่งชี้ว่าผู้ป่วยได้รับการรักษาอยู่แล้ว ซึ่งเป็นข้อมูลสำคัญทางคลินิก

6. อายุ เป็นปัจจัยเสี่ยงสำคัญ โดยพบความชุกของภาวะพร่องไทรอยด์สูงขึ้นในผู้สูงอายุ

8.3 การวิเคราะห์ความแม่นยำและความเสี่ยง Overfitting

เนื่องจากค่า Accuracy สูงถึง 99.9% ผู้วิจัยได้วิเคราะห์ความเสี่ยงเรื่อง Data Leakage พบว่าในชุดข้อมูลนี้ ตัวแปร TSH และ FTI มีความสัมพันธ์เชิงตัวเลขกับสถานะโรคที่เด่นชัดมากตามธรรมชาติของโรคไทรอยด์ อย่างไรก็ตาม ผู้วิจัยได้ป้องกัน Overfitting ด้วยวิธี 5-Fold Cross-Validation และพบว่าโมเดลยังคงให้ผลลัพธ์ที่เสถียรในทุกกลุ่มข้อมูลทดสอบ ซึ่งสนับสนุนความน่าเชื่อถือของผลลัพธ์ที่ได้

8.4 ผลกระทบของ SMOTE

ผลการทดลองชี้ให้เห็นว่าการใช้ SMOTE ไม่ได้ช่วยเพิ่มประสิทธิภาพอย่างมีนัยสำคัญเมื่อใช้ร่วมกับ Feature Selection เนื่องจากข้อมูลหลังการทำ Feature Selection มีโครงสร้างที่แยกออกจากกันได้ดีชัดเจน (Well-Separated Classes) การเพิ่มข้อมูลสังเคราะห์บริเวณขอบเขตการตัดสินใจ (Decision Boundary) อาจส่งผลให้โมเดลเกิดความสับสน ดังจะเห็นได้จากค่า Precision ที่ลดลงเล็กน้อยในบางกรณี ซึ่งสอดคล้องกับผลการศึกษาของ Kaur & Kumari (2022) ที่พบว่า SMOTE มีประสิทธิภาพสูงสุดในข้อมูลที่คลาสยังมีความทับซ้อนกัน

8.5 การเปรียบเทียบกับงานวิจัยอ้างอิง

เมื่อเปรียบเทียบกับงานวิจัยที่ผ่านมา พบว่าผลลัพธ์ของงานวิจัยนี้ (Accuracy = 0.999) สูงกว่า (Chaganti et al., 2022) ที่ได้ Accuracy 0.987 และสูงกว่า Ganie & Malik & Malik, (2022). ที่ใช้ SVM + Chi-Square เนื่องจากงานวิจัยนี้บูรณาการทั้ง Feature Selection และการแก้ปัญหา Class Imbalance พร้อมกัน นอกจากนี้ Decision Tree ร่วมกับ Feature Selection ที่ตรงจุดให้ผลดีกว่าการใช้ PCA เนื่องจาก PCA ทำให้ความสัมพันธ์ระหว่างค่าฮอร์โมนเสียรูปและยากต่อการตีความโดยแพทย์ (Quinlan, 1986)

8.6 ข้อจำกัดเชิงอัลกอริทึมและข้อจำกัดของงานวิจัย

แม้ Decision Tree จะให้ผลดีเยี่ยม แต่มีข้อจำกัดเรื่อง High Variance หรือความอ่อนไหวต่อข้อมูลนำเข้า หากข้อมูลชุดใหม่มีการเปลี่ยนแปลงเพียงเล็กน้อย โครงสร้างต้นไม้อาจเปลี่ยนไปทั้งหมด ผู้วิจัยจึงแนะนำว่าในกรณีที่มีข้อมูลมีสัญญาณรบกวนสูง การใช้ Random Forest (Breiman, 2001) อาจเป็นทางเลือกที่มีความทนทาน (Robustness) มากกว่าในระยะยาว นอกจากนี้ ชุดข้อมูลที่ใช้มาจากโรงพยาบาลแหล่งเดียว (Single-center Study) ในประเทศออสเตรเลีย ซึ่งอาจมีความแตกต่างจากประชากรในบริบทอื่น จึงจำเป็นต้องทำการทดสอบความสามารถในการสรุปผล (External Validation) กับข้อมูลจากแหล่งอื่น โดยเฉพาะข้อมูลจากประชากรไทย ก่อนนำไปใช้ในทางคลินิกจริง

8.7 ข้อเสนอแนะด้านการนำไปใช้งานจริง (Deployment)

เนื่องจาก Decision Tree มีความโปร่งใสและตีความได้ง่าย จึงเหมาะสำหรับการพัฒนาเป็นระบบสนับสนุนการตัดสินใจทางการแพทย์ (Clinical Decision Support System: CDSS) โดยมีแนวทางดังนี้

1. พัฒนาเป็นเว็บแอปพลิเคชันหรือแอปพลิเคชันมือถือสำหรับแพทย์ทั่วไปในการประเมินความเสี่ยงเบื้องต้น โดยใช้ 8 ตัวแปรสำคัญ (TSH, FTI, T3, TT4, T4U, ประวัติการผ่าตัด การใช้ยา และอายุ) เป็นข้อมูลนำเข้า
2. บูรณาการกับระบบ Hospital Information System (HIS) เพื่อแจ้งเตือนอัตโนมัติเมื่อผลตรวจห้องปฏิบัติการบ่งชี้ความเสี่ยง

3. นำ 8 ตัวแปรสำคัญไปจัดทำเป็นแบบฟอร์มคัดกรองเบื้องต้น (Screening Form) สำหรับใช้ในสถานพยาบาลปฐมภูมิหรือพื้นที่ห่างไกลที่ขาดแคลนผู้เชี่ยวชาญด้านต่อมไทรอยด์

9. สรุปผลการวิจัย

งานวิจัยนี้พัฒนาและเปรียบเทียบโมเดลการเรียนรู้ของเครื่อง 4 รูปแบบ ได้แก่ SVM, Decision Tree, Random Forest และ MLP สำหรับการทำนายภาวะพร่องไทรอยด์ โดยใช้ข้อมูลชุดเดียวจาก UCI Machine Learning Repository จำนวน 3,772 ราย (กลุ่มป่วย 151 ราย, กลุ่มปกติ 3,621 ราย)

ผลการวิจัยสรุปได้ดังนี้ ประการแรก Decision Tree + Feature Selection ให้ประสิทธิภาพสูงสุดในทุกมาตรวัด โดยมีค่า Precision = 0.993, Recall = 0.999, F1-Score = 0.996 และ Accuracy = 0.999 (ร้อยละ 99.3, 99.9, 99.6 และ 99.9 ตามลำดับ) ซึ่งผ่านการตรวจสอบความเสถียรด้วย 5-Fold Cross-Validation แล้ว ประการที่สอง Random Forest Feature Importance สามารถลดจำนวนตัวแปรจาก 23 เหลือ 8 ตัวแปร ซึ่งสอดคล้องกับมาตรฐานการวินิจฉัยทางการแพทย์ (Garber et al., 2012; Jonklaas et al., 2014) ประการที่สาม SMOTE ไม่ได้ช่วยเพิ่มประสิทธิภาพอย่างมีนัยสำคัญเมื่อข้อมูลมีการแยกแยะที่ชัดเจนหลัง Feature Selection เมื่อพิจารณาสมมติฐาน พบว่า สมมติฐานที่ 1 ได้รับการยืนยันบางส่วน กล่าวคือ Feature Selection ช่วยเพิ่ม Recall และ F1-Score อย่างมีนัยสำคัญ แต่ SMOTE ให้ผลปรับปรุงที่ไม่ชัดเจนเมื่อใช้ร่วมกับ Feature Selection สมมติฐานที่ 2 ได้รับการยืนยัน เนื่องจากโมเดล Tree-based ทั้ง Decision Tree และ Random Forest ให้ค่า Accuracy และ Recall สูงกว่า SVM และ MLP ในทุกสถานการณ์

10. ข้อเสนอแนะในการวิจัย

10.1 ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. นำโมเดล Decision Tree + Feature Selection พัฒนาต่อเป็นแอปพลิเคชันหรือ Web Service เพื่อให้แพทย์ทั่วไปใช้ประเมินความเสี่ยงภาวะพร่องไทรอยด์เบื้องต้น โดยป้อนค่าตัวแปรทั้ง 8 ตัว ได้แก่ TSH, FTI, T3, TT4, T4U, ประวัติการผ่าตัดไทรอยด์, การใช้ยาไทรอกซีน และอายุ

2. จัดทำแบบฟอร์มคัดกรองเบื้องต้น (Screening Form) โดยใช้ 8 ตัวแปรสำคัญดังกล่าวเพื่อใช้ในสถานพยาบาลปฐมภูมิ คลินิกชุมชน หรือพื้นที่ห่างไกลที่ขาดแคลนผู้เชี่ยวชาญด้านต่อมไทรอยด์ ก่อนส่งต่อผู้ป่วยเพื่อตรวจยืนยัน

3. บูรณาการโมเดลกับระบบ Hospital Information System (HIS) เพื่อแจ้งเตือนแพทย์อัตโนมัติเมื่อผลตรวจห้องปฏิบัติการของผู้ป่วยบ่งชี้ความเสี่ยงสูง

4. ในกรณีที่ข้อมูลทางคลินิกมีสัญญาณรบกวนสูง (Noisy Data) หรือมีความหลากหลายของประชากรมาก แนะนำให้เลือกใช้ Random Forest แทน Decision Tree เนื่องจากมีความทนทาน (Robustness) มากกว่า

10.2 ข้อเสนอแนะเพื่อการวิจัยครั้งต่อไป

1. ทำการ External Validation โดยทดสอบโมเดลกับข้อมูลผู้ป่วยจากโรงพยาบาลหลายแห่ง โดยเฉพาะข้อมูลจากประชากรไทย เพื่อประเมินความสามารถในการสรุปผล (Generalizability) ก่อนนำไปใช้จริงทางคลินิก

2. ศึกษาเปรียบเทียบเทคนิค Feature Selection แบบอื่น เช่น LASSO Regularization, Recursive Feature Elimination (RFE) หรือ Mutual Information เพื่อหาแนวทางที่ให้ผลดีกว่า Random Forest Feature Importance ในบริบทของโรคพร่องไทรอยด์

3. เปรียบเทียบกับโมเดลขั้นสูงอื่น ๆ เช่น XGBoost, LightGBM หรือ CatBoost ซึ่งเป็น Gradient Boosting Ensemble ที่มีรายงานว่าให้ผลดีในข้อมูลตาราง (Tabular Data) ทางการแพทย์
4. พัฒนาอัลกอริทึมที่สามารถทำงานร่วมกับข้อมูลภาพ (Image Data) จากการอัลตราซาวด์ต่อมไทรอยด์ร่วมกับข้อมูลทางห้องปฏิบัติการ เพื่อเพิ่มความแม่นยำทางคลินิกในแบบ Multimodal Learning
5. ทดลองใช้เทคนิค Explainable AI (XAI) เช่น SHAP (SHapley Additive exPlanations) เพื่อเพิ่มความโปร่งใสในการอธิบายผลการทำนายให้แพทย์และผู้ป่วยเข้าใจได้ง่ายขึ้น

11. เอกสารอ้างอิง

- กรมการแพทย์ กระทรวงสาธารณสุข. (2564). *แนวทางเวชปฏิบัติโรคต่อมไทรอยด์*. กระทรวงสาธารณสุข ประเทศไทย.
- Abiyev, R. H., & Ma'aitah, M. K. S. (2018). Deep convolutional neural networks for chest diseases detection. *Journal of Healthcare Engineering*, 2018, Article 4168538.
- Biondi, A., & Cooper, D. S. (2008). The clinical significance of subclinical thyroid dysfunction. *Endocrine Reviews*, 29(1), 76–131.
- Breiman, L. (2001). *Random forests*. *Machine Learning*, 45(1), 5–32.
- Canaris, G. J., Manowitz, N. R., Mayor, G., & Ridgway, E. C. (2000). The Colorado thyroid disease prevalence study. *Archives of Internal Medicine*, 160(4), 526–534.
- Chaganti, S., Rustam, F., DePaula, A. C. A., Daghrery, A., Mazin, I., & Ashraf, I. (2022). Thyroid disease prediction using selective features and machine learning techniques. *Cancers*, 14(16), 3914.
- Chaker, L., Bianco, A. C., Jonklaas, J., & Peeters, R. P. (2017). Hypothyroidism. *The Lancet*, 390(10101), 1550–1562.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*, 5, 8869–8879.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Ganie, S. M., & Malik, M. B. (2022). An ensemble machine learning approach for thyroid disease detection. *International Journal of Applied Engineering Research*, 17(4), 267–273.
- Garber, J. R., Cobin, R. H., Gharib, H., Hennessey, J. V., Klein, I., Mechanick, J. I., & Woeber, K. A. (2012). Clinical practice guidelines for hypothyroidism in adults. *Thyroid*, 22(12), 1200–1235.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182.
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
- Jonklaas, J., et al. (2014). Guidelines for the treatment of hypothyroidism. *Thyroid*, 24(12), 1670–1751.
- Bidyapati. (n.d.). Hypothyroid disease dataset. Kaggle. <https://www.kaggle.com/datasets/bidyapati/hypothyroid>

- Kaur, P., & Kumari, S. (2022). Diabetes prediction using SMOTE and k-nearest neighbor algorithm. *International Journal of Advanced Computer Science and Applications*, 13(5), 100–108.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.
- Roberts, C. G. P., & Ladenson, P. W. (2004). Hypothyroidism. *The Lancet*, 363(9411), 793–803.
- Temurtas, H., Yumusak, N., & Temurtas, F. (2009). *A comparative study on diabetes disease diagnosis using neural networks*. *Expert Systems with Applications*, 36(4), 8610–8615.
- Weetman, A. P. (2021). An update on the pathogenesis of Hashimoto's thyroiditis. *Journal of Endocrinological Investigation*, 44(5), 883–890.
- World Health Organization. (2007). *Assessment of iodine deficiency disorders and monitoring their elimination: A guide for programme managers (3rd ed.)*. World Health Organization.