บทความวิจัย

การจำแนกประเภทหลายหมวดหมู่ด้วยการเรียนรู้ของเครื่องเพื่อพยากรณ์ระดับน้ำตาลสะสมในเลือดของผู้ป่วยเบาหวานชนิดที่ 2

ณพัชร โพธิ์ศรี* และ วฤษาย์ ร่มสายหยุด
แขนงวิชาเทคโนโลยีดิจิทัล สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08 4153 4625 อีเมล: naphatposri@gmail.com DOI: 10.14416/j.kmutnb.2024.11.007
รับเมื่อ 25 เมษายน 2567 แก้ไขเมื่อ 26 สิงหาคม 2567 ตอบรับเมื่อ 12 กันยายน 2567 เผยแพร่ออนไลน์ 27 พฤศจิกายน 2567

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อจำแนกกลุ่มระดับน้ำตาลสะสมในเลือดและการควบคุมโรคของผู้ป่วยเบาหวานชนิดที่ 2 ด้วยแบบจำลองการเรียนรู้ของเครื่อง โดยรวบรวมข้อมูลเวชระเบียนจำนวน 28,431 เวชระเบียน 37 คุณลักษณะ นำมาผ่านกระบวนการจัดเตรียมข้อมูล การคัดเลือกคุณลักษณะด้วยวิธีการลดคุณลักษณะโดยการวัดค่าสารสนเทศ (Information Gain Feature Reduction; IG) การกำจัดคุณลักษณะแบบวนซ้ำ (Recursive Feature Elimination; RFE) และการวัดความสำคัญของคุณลักษณะด้วยวิธีป่าสุ่ม (Random Forest Importance; RFI) การจัดการกับข้อมูลไม่สมดุลด้วยเทคนิค Synthetic Minority Over-sampling Technique; SMOTE (SM) เทคนิค Borderline-SMOTE (BM) และเทคนิค Adaptive Synthetic (ADA) และสร้างแบบจำลองด้วย 5 อัลกอริทึม การจำแนกประเภทหลายหมวดหมู่ ผลการวิจัยพบว่าแบบจำลอง IG + SM ร่วมกับอัลกอริทึมป่าสุ่ม (Random Forest) มีประสิทธิภาพสูงสุด โดยมีร้อยละความถูกต้อง ความแม่นยำ ความครบถ้วน และค่า F1 Score เท่ากับ 81.89 82.13 81.89 81.83 ตามลำดับ จากผลการวิจัยสามารถนำแบบจำลองไปประยุกต์ใช้เพื่อประกอบการตัดสินใจส่งตรวจติดตามระดับน้ำตาลสะสมในเลือดเพิ่มเติมจากปกติและช่วยเพิ่มประสิทธิภาพในการควบคุมโรคเบาหวานในเขตพื้นที่รับผิดชอบของโรงพยาบาล

คำสำคัญ: การเรียนรู้ของเครื่อง การจำแนกประเภทหลายหมวดหมู่ อัลกอริทึมป่าสุ่ม โรคเบาหวานชนิดที่ 2 ระดับน้ำตาลสะสม

การอ้างอิงบทความ: ณพัชร โพธิ์ศรี และ วฤษาย์ ร่มสายหยุด, "การจำแนกประเภทหลายหมวดหมู่ด้วยการเรียนรู้ของเครื่องเพื่อพยากรณ์ระดับน้ำตาลสะสมในเลือดของผู้ป่วยเบาหวานชนิดที่ 2," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 35, ฉบับที่ 4, หน้า 1–14, เลขที่บทความ 254-097602, ต.ค.–ธ.ค. 2568.

Research Article

Machine Learning-Based Multiclass Classification for Predicting the Cumulative Blood Sugar Levels in Type 2 Diabetes Patient

Naphat Posri* and Walisa Romsaiyud
Digital Technology, School of Science and Technology, Sukhothai Thammathirat Open University, Nonthaburi, Thailand

* Corresponding Author, Tel. 08 4153 4625, E-mail: naphatposri@gmail.com
DOI: 10.14416/j.kmutnb.2024.11.007
Received 25 April 2024; Revised 26 August 2024; Accepted 12 September 2024; Published online: 27 November 2024

Abstract

This research aims to classify cumulative blood sugar levels (Hemoglobin A1c) and assess disease control in type 2 diabetes patients using machine learning models. A total of 28,431 medical records with 37 attributes were collected and processed through data preparation. Feature selection was performed using Information Gain (IG), Recursive Feature Elimination (RFE), and Random Forest Importance (RFI) methods. Imbalanced data was addressed using the Synthetic Minority Over-sampling Technique; SMOTE (SM), Borderline-SMOTE (BM), and Adaptive Synthetic (ADA) techniques. Models were developed using five multiclass classification algorithms. The results demonstrated that the IG + SM model, combined with the Random Forest algorithm, yielded the highest performance, with an accuracy, precision, recall, and F1 score of 81.89%, 82.13%, 81.89%, and 81.83% respectively. These findings can be applied to support decision-making for additional cumulative blood sugar level testing beyond routine practices and to enhance the efficiency of diabetes control within the hospital's service area.

Keywords: Machine Learning, Multiclass Classification, Random Forest, Type 2 Diabetes, Hemoglobin A1c

Please cite this article as: N. Posri and W. Romsaiyud, "Machine learning-based multiclass classification for predicting the cumulative blood sugar levels in type 2 diabetes patient," *The Journal of KMUTNB*, vol. 35, no. 4, pp. 1–14, ID. 254-097602, Oct.–Dec. 2025 (in Thai).

1. บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนโยบายไทยแลนด์ 4.0 เป็นกลไกสำคัญในการนำประเทศให้ก้าวไปสู่การเป็นประเทศที่มีความมั่งคั่งและยั่งยืน ทำให้เกิดการส่งเสริมการนำเทคโนโลยีดิจิทัลมาประยุกต์ในทางการแพทย์มากขึ้น [1] สอดคล้องกับกระทรวงสาธารณสุขได้ผลักดันให้เกิดการพัฒนาบริการแพทย์และสาธารณสุขไปสู่ยุคดิจิทัลผ่านนโยบายโรงพยาบาลอัจฉริยะ (Smart Hospital) ที่สนับสนุนการนำเทคโนโลยีดิจิทัลมาใช้ในระบบงานต่าง ๆ ของโรงพยาบาล [2] ตัวอย่างการนำเทคโนโลยีมาใช้งาน เช่น การจำแนกภาพเอกซเรย์ทรวงอกด้วยโครงข่ายประสาทเทียมในการวินิจฉัยโรคโควิด-19 [3] การจำแนกความเสี่ยงการเป็นโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล [4] เป็นต้น

จากการศึกษาพบว่า การนำเทคโนโลยีดิจิทัลมาประยุกต์ใช้ในทางการแพทย์มีหลากหลายแนวทาง แต่หนึ่งในแนวทางที่สำคัญ คือ การเรียนรู้ของเครื่องแบบจำแนกประเภทมาประยุกต์ใช้งานด้านสุขภาพ เพื่อทำนายภาวะสุขภาพในด้านต่าง ๆ เช่น การพยากรณ์การควบคุมโรคเบาหวาน โดยสร้างแบบจำลองการจำแนกประเภท เพื่อพยากรณ์ระดับน้ำตาลสะสมที่แบ่งออกเป็น 2 กลุ่มผลลัพธ์ ได้แก่ กลุ่มที่ควบคุมไม่ได้ และกลุ่มปกติ [5] การพยากรณ์ความเสี่ยงของการเกิดภาวะแทรกซ้อนจากโรคเบาหวาน โดยสร้างแบบจำลองการจำแนกประเภท เพื่อคาดการณ์การเกิดขึ้นหรือไม่เกิดภาวะแทรกซ้อนในผู้ป่วยที่รักษาไม่ต่อเนื่อง [6] เป็นต้น

สถานการณ์โรคเบาหวานในประเทศไทย มีผู้ป่วยเบาหวานประมาณ 5.2 ล้านคน หรือ คิดเป็นอัตราส่วน 1 ใน 11 คน ของประชากรไทยที่อายุ 15 ปี ขึ้นไป กำลังป่วยด้วยโรคเบาหวาน [7] จากข้อมูลสถิติปีงบประมาณ พ.ศ. 2564–2566 พบว่า แนวโน้มผู้ป่วยโรคเบาหวานรายใหม่มีจำนวนเพิ่มขึ้น จากอัตราต่อแสน (100,000) ของประชากร 522.91 561.93 และ 606.10 ตามลำดับ [8] และในปีงบประมาณ พ.ศ. 2566 มีผู้ป่วยรายใหม่เพิ่มขึ้น 3 แสนคน ซึ่งเพิ่มขึ้น 2 เท่า จากปีงบประมาณ พ.ศ. 2565 มีอัตราการเสียชีวิตจากโรคเบาหวาน มากถึง 200 รายต่อวัน [9] สาเหตุส่วนหนึ่งที่ทำให้เกิดการสูญเสียและทุพพลภาพในผู้ป่วยเบาหวานเกิดจากการควบคุมระดับน้ำตาลในเลือดไม่ได้ จนทำให้เกิดภาวะแทรกซ้อนของโรค

โรคเบาหวานแบ่งออกเป็น 6 ชนิด โดยโรคเบาหวานชนิดที่ 2 (Type 2 Diabetes) เป็นชนิดที่พบได้บ่อยที่สุด พบประมาณร้อยละ 95 ของผู้ป่วยเบาหวานทั้งหมด เกิดจากภาวะดื้อต่ออินซูลิน (Insulin Resistance) ร่วมกับความบกพร่องในการผลิตอินซูลินที่เหมาะสม [10] ผู้ป่วยเบาหวานต้องได้รับการดูแลรักษาอย่างต่อเนื่อง และต้องมีการตรวจประเมินภายหลังการรักษา โดยใช้ระดับน้ำตาลในเลือดหลังอดอาหาร และระดับน้ำตาลสะสมในเลือด เป็นเกณฑ์ประเมินการควบคุมโรคของผู้ป่วย เป้าหมายของการควบคุมระดับน้ำตาลในเลือดของผู้ป่วยเบาหวานที่ไม่ได้ตั้งครรภ์แบ่งได้ 3 ระดับ ได้แก่ ควบคุมเข้มงวดมาก ควบคุมเข้มงวด และควบคุมไม่เข้มงวด [11]

จากปัญหาในทางปฏิบัติ การตรวจระดับน้ำตาลในเลือดของผู้ป่วยเบาหวานควรตรวจทั้งระดับน้ำตาลหลังอดอาหารและระดับน้ำตาลสะสม แต่ด้วยข้อจำกัดหลายประการจึงไม่สามารถตรวจระดับน้ำตาลสะสมได้ทุก 3–6 เดือน ตามแนวทางการรักษา โดยเฉลี่ยผู้ป่วยเบาหวานได้รับการตรวจระดับน้ำตาลสะสมประมาณ 1 ครั้งต่อคนต่อปี ทำให้ในการตรวจรักษาแต่ละครั้ง จึงตรวจเพียงระดับน้ำตาลหลังอดอาหาร ซึ่งในบางครั้งค่าที่ตรวจวัดได้อาจไม่สอดคล้องกับระดับน้ำตาลสะสม เนื่องจากเป็นเพียงการตรวจวัดค่าระดับน้ำตาลในเลือดเฉพาะวันที่ตรวจเลือดเท่านั้น ไม่สามารถอ้างอิงการควบคุมโรคย้อนหลัง 3 เดือนได้เหมือนกับระดับน้ำตาลสะสม และจากการศึกษาในต่างประเทศพบว่า จำนวนครั้งในการตรวจระดับน้ำตาลสะสมในเลือดที่มากกว่าแนวทางเวชปฏิบัติมีความสัมพันธ์กับผลลัพธ์ในการควบคุมโรคเบาหวานได้ดีขึ้น [12], [13]

ดังนั้น งานวิจัยนี้จึงประยุกต์ใช้เทคโนโลยีดิจิทัลกับการบริการทางการแพทย์ สร้างแบบจำลองพยากรณ์กลุ่มระดับน้ำตาลสะสมในเลือดของผู้ป่วยเบาหวานชนิดที่ 2 ด้วย

แบบจำลองการเรียนรู้ของเครื่อง เพื่อนำผลลัพธ์มาประกอบการตัดสินใจส่งตรวจตามระดับน้ำตาลสะสมในเลือดเพิ่มเติมจากปกติ ซึ่งผลจากการตรวจระดับน้ำตาลสะสมช่วยให้แพทย์สามารถปรับการรักษาเพื่อควบคุมโรคเบาหวานได้ดีขึ้น อันจะส่งผลต่อโอกาสการเกิดภาวะแทรกซ้อนจากโรคเบาหวานลดลง โดยมีวัตถุประสงค์เพื่อจำแนกกลุ่มระดับน้ำตาลสะสมในเลือด และการควบคุมโรคของผู้ป่วยเบาหวานชนิดที่ 2 ด้วยแบบจำลองการเรียนรู้ของเครื่อง

1.2 นิยามศัพท์เฉพาะ

1.2.1 ระดับน้ำตาลในเลือด

ระดับน้ำตาลหลังอดอาหาร หมายถึง ระดับน้ำตาลในเลือดหลังอดอาหารอย่างน้อย 8 ชั่วโมง (Fasting Blood Sugar; FBS) และระดับน้ำตาลสะสม (Hemoglobin A1c; HbA1c) หมายถึง ระดับน้ำตาลเฉลี่ยสะสมในเลือดเกิดจากน้ำตาลในเลือดจับกับฮีโมโกลบิน ทั้งระดับน้ำตาลหลังอดอาหารและระดับน้ำตาลสะสม สามารถนำมาใช้ในการตรวจคัดกรองโรค วินิจฉัยโรค และติดตามการรักษา [10]

1.2.2 การจำแนกประเภทหลายหมวดหมู่

หรือ Multiclass Classification เป็นส่วนของการเรียนรู้ของเครื่อง ประเภทการเรียนรู้แบบมีผู้สอน มีวัตถุประสงค์เพื่อจำแนกหมวดหมู่ของข้อมูล [14] จากบริบทของปัญหาวิจัยแบ่งกลุ่มผลลัพธ์การพยากรณ์ตามระดับการควบคุมโรคเบาหวาน เป็น 3 กลุ่ม ได้แก่

1) กลุ่มควบคุมโรคดี (ระดับน้ำตาลสะสม น้อยกว่าร้อยละ 7)

2) กลุ่มควบคุมโรคปานกลาง (ระดับน้ำตาลสะสมร้อยละ 7 ถึงน้อยกว่าร้อยละ 8)

3) กลุ่มควบคุมโรคไม่ดี (ระดับน้ำตาลสะสม มากกว่าเท่ากับร้อยละ 8)

2. วัสดุ อุปกรณ์และวิธีการวิจัย

วิธีการดำเนินการวิจัยในครั้งนี้ได้นำกระบวนการเรียนรู้ของเครื่องมาประยุกต์ใช้ ตามขั้นตอนที่แสดงในรูปที่ 1 โดยรายละเอียดของกระบวนการ มีดังต่อไปนี้

2.1 การเก็บรวบรวมข้อมูล

รวบรวมข้อมูล ผู้ป่วยเบาหวานชนิดที่ 2 เกี่ยวกับลักษณะหรือปัจจัยที่ส่งผลต่อระดับน้ำตาลสะสมในเลือดของผู้ป่วยเบาหวาน แบ่งออกเป็น 3 กลุ่มข้อมูล ได้แก่ 1) กลุ่มข้อมูลพื้นฐานทั่วไป 2) กลุ่มข้อมูลผลตรวจทางห้องปฏิบัติการ และ 3) กลุ่มข้อมูลโรคเบาหวาน

โดยนำข้อมูลจากฐานข้อมูลเวชระเบียนอิเล็กทรอนิกส์ของโรงพยาบาลชุมชนแห่งหนึ่ง ในรูปแบบไฟล์ตารางข้อมูล ครอบคลุมตั้งแต่ 1 มกราคม 2560 ถึง 1 ตุลาคม 2566 การวิจัยครั้งนี้ได้ผ่านการพิจารณาและรับรองจากคณะกรรมการจริยธรรมวิจัยในมนุษย์ จังหวัดนครสวรรค์ หมายเลขรับรอง NSWPHOEC-035/66 เมื่อวันที่ 27 ตุลาคม 2566

2.2 การทำวิศวกรรมคุณลักษณะและติดป้ายกำกับข้อมูล

ตรวจสอบและวิเคราะห์รูปแบบข้อมูลสูญหาย แทนค่าข้อมูลสูญหายด้วยวิธีแทนค่าด้วยหลายค่า (Multiple Imputation; MI) [15] จัดการกับข้อมูลที่มีค่าผิดปกติด้วยค่าพิสัยระหว่างควอไทล์ (Interquartile Range; IQR) และปรับข้อมูลให้สมมาตรด้วยวิธี Min-max Normalization

2.3 การคัดเลือกคุณลักษณะ

กำหนดจำนวนข้อมูลคุณลักษณะที่จะนำไปสร้างแบบจำลองด้วยวิธี Recursive Feature Elimination with Cross Validation (RFECV) โดยพิจารณาจำนวนชุดข้อมูลคุณลักษณะที่เหมาะสม จากค่าเฉลี่ยความถูกต้องของแบบจำลองและคัดเลือกคุณลักษณะจาก 3 วิธี ดังนี้ [16]

1) Information Gain Feature Reduction; IG

2) Recursive Feature Elimination; RFE

3) Random Forest Importance; RFI

2.4 การจัดการข้อมูลแบบไม่สมดุล

จัดการกับข้อมูลไม่สมดุลระหว่างคลาสผลลัพธ์ ด้วยวิธีการสุ่มเพิ่มข้อมูลเนื่องจากเป็นวิธีที่เสถียร [17] และป้องกันการสูญเสียข้อมูลที่สำคัญ [18] จาก 3 วิธี ดังนี้ 1) SMOTE (SM) ใช้หลักการการสุ่มตัวอย่างเพิ่มขึ้น

ในข้อมูลกลุ่มน้อย โดยนำข้อมูลกลุ่มน้อยมาพิจารณาจนครบทุกตัว ด้วยหลักการของ K-Nearest Neighbor; KNN 2) Borderline-SMOTE (BM) เป็นการปรับปรุงจากวิธี SM เพื่อแก้ไขปัญหากรณีข้อมูลกลุ่มน้อยมีความผิดปกติ โดยนำมาสร้างเป็นแนวขอบเขตข้อมูล และ 3) ADASYN (ADA) ใช้การแจกแจงแบบถ่วงน้ำหนักของข้อมูลกลุ่มน้อย ในการสร้างข้อมูลขึ้นมาใหม่

2.5 การแบ่งชุดข้อมูล

กำหนดการสุ่มแยกตัวอย่างข้อมูล ตามอัตราส่วน 80 : 20 สำหรับข้อมูลฝึกสอน และข้อมูลทดสอบ ตามลำดับ โดยพิจารณาจากจำนวนคุณลักษณะในการสร้างแบบจำลอง [19]

2.6 การสร้างแบบจำลอง

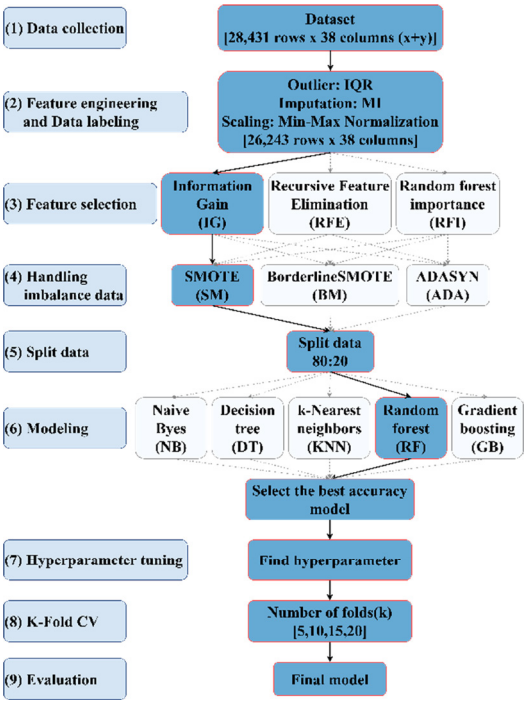
นำข้อมูลที่ผ่านการจัดเตรียมมารวมกับอัลกอริทึม และทำการฝึกสอนแบบจำลอง เพื่อให้เกิดการเรียนรู้จำรูปแบบจากชุดข้อมูล โดยใช้อัลกอริทึมการจำแนกประเภทแบบ Multi-class Classification ชนิดที่ไม่ต้องอาศัยกระบวนอื่นเสริมในการจำแนกประเภท (Native Multiclass Classification) เนื่องจากสามารถแปลความหมายผลลัพธ์ที่ได้จากแบบจำลองและประสิทธิภาพการจำแนกได้โดยตรงซึ่งได้แก่ นาอีเบย์ (Naïve Bayes; NB) ต้นไม้ตัดสินใจ (Decision Trees; DT) เพื่อนบ้านใกล้ที่สุด (KNN) ป่าสุ่ม (Random Forest; RF) และต้นไม้เร่งความเร็วแบบไล่ระดับ (Gradient Boosting; GB) [20]

2.7 การปรับแต่งไฮเปอร์พารามิเตอร์

คัดเลือกแบบจำลองที่มีประสิทธิภาพ ค่าความถูกต้องสูงสุด เพื่อนำมาปรับแต่งค่าพารามิเตอร์

2.8 การทำ K-fold Cross-validation (KFCV)

เพื่อเพิ่มความน่าเชื่อถือของการประมาณค่าในแบบจำลองโดยพิจารณาจำนวน k ที่ 5 10 15 20 กับค่าความถูกต้อง และนำเสนอค่าความถูกต้องสูงสุด [21], [22]



รูปที่ 1 ขั้นตอนการทดลอง

2.9 การประเมินประสิทธิภาพแบบจำลอง

ด้วย 4 เมตริกหลัก แสดงค่าความถูกต้อง ค่าความแม่นยำ ค่าความครบถ้วน และค่า F1 Score

3. ผลการวิจัย

3.1 แบบจำลองการเรียนรู้ของเครื่อง

ข้อมูลที่ได้จากฐานข้อมูลเวชระเบียน ประกอบด้วยจำนวนข้อมูลทั้งหมด 28,431 แถว 37 คุณลักษณะแบ่งออกเป็น 3 กลุ่มข้อมูล ได้แก่ กลุ่มข้อมูลทั่วไป กลุ่มข้อมูลผลตรวจทางห้องปฏิบัติการ และกลุ่มข้อมูลโรคเบาหวาน (ตารางที่ 1)

ภายหลังการจัดการข้อมูลสูญหาย การจัดการกับข้อมูลที่มีค่าผิดปกติ และการปรับข้อมูลให้สมมาตร ส่งผลให้ได้ชุดข้อมูลจำนวน 26,243 แถว ซึ่งลดลงร้อยละ 7.69 จากจำนวนเวชระเบียนเริ่มต้น โดยชุดข้อมูลประกอบด้วย 37 คอลัมน์คุณลักษณะ และ 1 คอลัมน์ผลลัพธ์ ทั้งนี้ การติดป้ายกำกับข้อมูลผลลัพธ์ (Label) ได้ดำเนินการตั้งแต่ขั้นตอนการรวบรวมข้อมูลจากฐานข้อมูลเวชระเบียน (ตารางที่ 2)

ตารางที่ 1 ข้อมูลและคุณลักษณะ

No	Feature	Type	Description
1	1_age_year	N	Age (year)
2	1_sex	C	Sex
3	1_bw	N	Body weight (kg)
4	1_ht	N	Height (cm)
5	1_bmi	N	Body mass index
6	1_sbp	N	Systolic pressure
7	1_dbp	N	Diastolic pressure
8	1_waist	N	Waist (cm)
9	1_smoking	C	Smoking
10	1_drinking	C	Alcohol drinking
11	1_ud_ht	C	Hypertension
12	1_ud_dlp	C	Dyslipidemia
13	1_ud_heart	C	Heart disease
14	2_sugar	N	Fasting blood level
15	2_lipid_chol	N	Lipid: Cholesterol
16	2_lipid_tg	N	Lipid: Triglyceride
17	2_lipid_hdl	N	Lipid: HDL
18	2_lipid_ldl	N	Lipid: LDL
19	2_kidney_cr	N	Kidney: Creatinine
20	2_kidney_egfr	N	Kidney: eGFR
21	2_kidney_stage	N	Kidney: stage
22	2_blood_hb	N	Hemoglobin
23	2_blood_hct	N	Hematocrit
24	2_blood_mcv	N	MCV
25	2_blood_mchc	N	MCHC
26	2_latest_hba1c	N	Latest HbA1c
27	2_previous_fbs_120day	N	Mean FBS 120 days
28	3_dm_duration_year	N	Duration of diabetes
29	3_dm_fam_hx	C	Family diabetes
30	3_dm_rx_nph_day	N	Amount of NPH
31	3_dm_rx_mixtard_day	N	Amount of Mixtard
32	3_dm_rx_gpz_day	N	Amount of Glipizide
33	3_dm_rx_mfm_day	N	Amount of Metformin

ตารางที่ 1 ข้อมูลและคุณลักษณะ (ต่อ)

No	Feature	Type	Description
34	3_dm_rx_pio_day	N	Amount of Pioglitazone
35	3_dm_rx_folic_day	N	Amount of Folic
36	3_dm_rx_ff_day	N	Amount of Ferrous
37	3_dm_rx_asa_day	N	Amount of Aspirin

หมายเหตุ: N = Numerical, C = Categorical

ตารางที่ 2 ป้ายกำกับข้อมูลผลลัพธ์

Label	Description	HbA1c Level
1	Good control	<7%
2	Moderate control	7 – <8%
3	Poor control	≥8%

การคัดเลือกจำนวนคุณลักษณะด้วยวิธี RFECV เพื่อหาจำนวนชุดข้อมูลคุณลักษณะที่เหมาะสมในการสร้างแบบจำลอง (รูปที่ 2) เมื่อพิจารณาจากค่าเฉลี่ยความถูกต้องที่เริ่มคงที่พบว่า จำนวนคุณลักษณะที่เหมาะสมที่นำไปสร้างแบบจำลองในขั้นตอนต่อไป คือ 15 คุณลักษณะ และดำเนินการคัดเลือกคุณลักษณะด้วยวิธี IG, RFE และ RFI (ตารางที่ 3)

ตารางที่ 3 การคัดเลือกคุณลักษณะแต่ละวิธี

IG	RFE	RFI
1_age_year	1_age_year	1_age_year
1_bmi	1_bmi	1_bmi
2_blood_mchc	2_blood_mchc	2_blood_mchc
2_lastest_hba1c	2_lastest_hba1c	2_lastest_hba1c
2_lipid_chol	2_lipid_chol	2_lipid_chol
2_lipid_ldl	2_lipid_ldl	2_lipid_ldl
2_lipid_tg	2_lipid_tg	2_lipid_tg
2_previous_fbs_120day	2_previous_fbs_120day	2_previous_fbs_120day
2_sugar	2_sugar	2_sugar

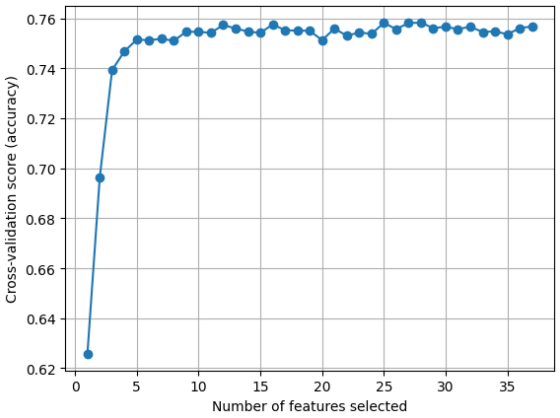
ตารางที่ 3 การคัดเลือกคุณลักษณะแต่ละวิธี (ต่อ)

IG	RFE	RFI
3_dm_duration_year	3_dm_duration_year	3_dm_duration_year
2_blood_hct	1_bw	1_bw
2_blood_mcv	1_dbp	1_dbp
2_lipid_hdl	2_blood_hb	1_sbp
3_dm_rx_mfm_day	2_kidney_cr	2_kidney_cr
3_dm_rx_mixtard_day	2_kidney_egfr	2_kidney_egfr

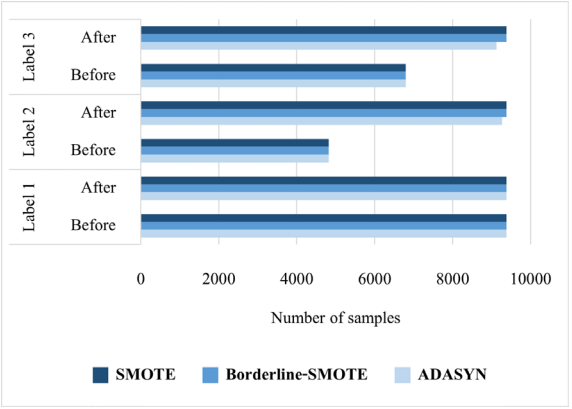
การคัดเลือกคุณลักษณะด้วยวิธี IG, REF และ RFI พบว่า มีคุณลักษณะที่ถูกเลือกซ้ำจากทั้ง 3 วิธี จำนวน 10 คุณลักษณะ จากชุดคุณลักษณะทั้งหมด 15 คุณลักษณะ แสดงให้เห็นถึงความสอดคล้องของชุดคุณลักษณะที่ถูกคัดเลือก โดยคุณลักษณะระดับน้ำตาลสะสมครั้งล่าสุด (2_lastest_hba1c) เป็นคุณลักษณะที่มีค่าสารสนเทศ (Information Gain) และค่าความสำคัญ (Importance) สูงสุด ในแต่ละวิธีของการคัดเลือกคุณลักษณะ

การจัดการกับข้อมูลไม่สมดุลด้วยวิธี SM, BM และ ADA ซึ่งเป็นการสังเคราะห์ข้อมูลตัวอย่างเพิ่ม (รูปที่ 3) พบว่า ก่อนสังเคราะห์ข้อมูลเพิ่มจำนวนตัวอย่างแบ่งตามป้ายกำกับของผลลัพธ์ (Label) ป้ายกำกับ 1 2 และ 3 มีจำนวนตัวอย่างเท่ากับ 9,381 4,819 และ 6,794 ตามลำดับ ภายหลังสังเคราะห์ข้อมูลเพิ่ม วิธี SM และ BM ทำการสังเคราะห์ข้อมูลตัวอย่างเพิ่มขึ้น เท่ากับจำนวนข้อมูลย่อยที่มากที่สุดในชุดข้อมูลตัวอย่าง ซึ่งในที่นี้ คือ ข้อมูลย่อยที่มีป้ายกำกับ “1” (ควบคุมโรคได้ดี)

การแบ่งส่วนชุดข้อมูลสำหรับการฝึกสอน และทดสอบแบบจำลองใช้อัตราส่วน 80 : 20 ตามลำดับ สร้างแบบจำลองโดยใช้อัลกอริทึม การจำแนกประเภท 5 อัลกอริทึม ร่วมกับการเตรียมข้อมูลที่ผ่านการคัดเลือกคุณลักษณะ 3 วิธี และการจัดการข้อมูลไม่สมดุล 3 วิธี ส่งผลให้มีการสร้างแบบจำลองทั้งหมด 45 แบบจำลอง พบว่า แบบจำลองที่มีค่าความถูกต้องสูงสุด คือ แบบจำลอง IG + SM + RF และนำมาปรับแต่ง



รูปที่ 2 ค่าเฉลี่ยความถูกต้องและจำนวนคุณลักษณะ



รูปที่ 3 การสังเคราะห์ตัวอย่างเพิ่ม

ค่าไฮเปอร์พารามิเตอร์ด้วยวิธี GridSearchCV ส่งผลให้ค่าเฉลี่ยความถูกต้องเพิ่มขึ้น (ตารางที่ 4)

ตารางที่ 4 การปรับแต่งค่าไฮเปอร์พารามิเตอร์

IG + SM + RF	Before Tuning	After Tuning
Accuracy (%)	75.48	75.73

จากนั้น นำแบบจำลองมาเพิ่มความน่าเชื่อถือของการประมาณค่าในแบบจำลองด้วย KFCV (ตารางที่ 5) พบว่าค่า $k = 15$ เหมาะสมกับข้อมูลชุดดังกล่าว ในการเรียนรู้จำรูปแบบข้อมูล ทำให้แบบจำลองมีค่าความถูกต้องสูงสุด

และจากการทดลองพบว่า เมื่อค่า k สูงขึ้น $(k \geq 20)$ เวลาประมวลผลจะนานขึ้น แต่ค่าความถูกต้องไม่เพิ่มขึ้นจากเดิม

ตารางที่ 5 การทำ K-fold Cross-validation

K-fold CV (k)	5	10	15	20
Accuracy (%)	81.54	81.70	**81.89**	81.83

3.2 ผลการประเมินประสิทธิภาพแบบจำลอง

ภายหลังการปรับแต่งไฮเปอร์พารามิเตอร์ และ KFCV $(k = 15)$ พบว่า แบบจำลอง IG + SM + RF มีประสิทธิภาพสูงสุด โดยมีค่าความถูกต้องร้อยละ 81.89 ค่าความแม่นยำร้อยละ 82.13 ค่าความครบถ้วนร้อยละ 81.89 และค่า F1 Score ร้อยละ 81.83 (ตารางที่ 6)

ตารางที่ 6 ผลการประเมินประสิทธิภาพแบบจำลองทั้ง 45 แบบจำลอง

FS	IDH	Classification	Accuracy	Precision	Recall	F1 score
IG	SM	NB	0.6479	0.6939	0.6479	0.6583
IG	SM	DT	0.6902	0.7098	0.6902	0.6982
IG	SM	KNN	0.6304	0.6801	0.6304	0.6476
IG	**SM**	**RF**	**0.7548***	0.7643	0.7548	0.7587
IG	SM	GB	0.7413	0.7663	0.7413	0.7501
IG	BM	NB	0.6533	0.6923	0.6533	0.6628
IG	BM	DT	0.6855	0.7000	0.6855	0.6917
IG	BM	KNN	0.6213	0.6673	0.6213	0.6382
IG	BM	RF	0.7546	0.7608	0.7546	0.7566
IG	BM	GB	0.7424	0.7623	0.7424	0.7492
IG	ADA	NB	0.6512	0.6908	0.6512	0.6602
IG	ADA	DT	0.6815	0.6964	0.6815	0.6878
IG	ADA	KNN	0.6211	0.6716	0.6211	0.6390
IG	ADA	RF	0.7525	0.7599	0.7525	0.7552
IG	ADA	GB	0.7460	0.7657	0.7460	0.7527
RFE	SM	NB	0.5957	0.7033	0.5957	0.6218
RFE	SM	DT	0.6719	0.6946	0.6719	0.6810
RFE	SM	KNN	0.6119	0.6642	0.6119	0.6303
RFE	SM	RF	0.7499	0.7598	0.7499	0.7538
RFE	SM	GB	0.7398	0.7653	0.7398	0.7487
RFE	BM	NB	0.6005	0.6994	0.6005	0.6254
RFE	BM	DT	0.6765	0.6955	0.6765	0.6845
RFE	BM	KNN	0.6136	0.6598	0.6136	0.6307
RFE	BM	RF	0.7504	0.7578	0.7504	0.7528
RFE	BM	GB	0.7459	0.7656	0.7459	0.7526
RFE	ADA	NB	0.6037	0.7030	0.6037	0.6287
RFE	ADA	DT	0.6818	0.6980	0.6818	0.6887
RFE	ADA	KNN	0.6121	0.6592	0.6121	0.6291

ตารางที่ 6 ผลการประเมินประสิทธิภาพแบบจำลองทั้ง 45 แบบจำลอง (ต่อ)

FS	IDH	Classification	Accuracy	Precision	Recall	F1 score
RFE	ADA	RF	0.7500	0.7586	0.7500	0.7529
RFE	ADA	GB	0.7455	0.7646	0.7455	0.7520
RFI	SM	NB	0.6013	0.7082	0.6013	0.6275
RFI	SM	DT	0.6843	0.7004	0.6843	0.6910
RFI	SM	KNN	0.6159	0.6687	0.6159	0.6337
RFI	SM	RF	0.7516	0.7619	0.7516	0.7556
RFI	SM	GB	0.7380	0.7637	0.7380	0.7471
RFI	BM	NB	0.6064	0.7050	0.6064	0.6315
RFI	BM	DT	0.6860	0.7022	0.6860	0.6928
RFI	BM	KNN	0.6062	0.6560	0.6062	0.6239
RFI	BM	RF	0.7535	0.7622	0.7535	0.7566
RFI	BM	GB	0.7422	0.7621	0.7422	0.7491
RFI	ADA	NB	0.6091	0.7050	0.6091	0.6340
RFI	ADA	DT	0.6759	0.6934	0.6759	0.6831
RFI	ADA	KNN	0.6129	0.6589	0.6129	0.6290
RFI	ADA	RF	0.7514	0.7589	0.7514	0.7540
RFI	ADA	GB	0.7400	0.7608	0.7400	0.7469

หมายเหตุ: * The best accuracy model

4. อภิปรายและสรุปผล

ผลการประเมินประสิทธิภาพของแบบจำลอง การพยากรณ์ระดับน้ำตาลสะสมในเลือดสำหรับผู้ป่วยเบาหวานชนิดที่ 2 แสดงให้เห็นว่าแบบจำลอง IG + SM + RF มีประสิทธิภาพสูงสุด โดยมีค่าความถูกต้องร้อยละ 81.89 ค่าความแม่นยำร้อยละ 82.13 ค่าความครบถ้วนร้อยละ 81.89 และค่า F1 Score ร้อยละ 81.83 ผลลัพธ์ดังกล่าว สอดคล้องกับงานวิจัยก่อนหน้าที่ศึกษาในลักษณะเดียวกัน [23], [24] แต่เนื่องจากความแตกต่างของชุดข้อมูลที่ใช้ในการศึกษาที่มีความเฉพาะต่อลักษณะประชากร อาจไม่สามารถนำผลการวิจัยมาเปรียบเทียบกันได้โดยตรง

อย่างไรก็ตาม แบบจำลองที่พัฒนาขึ้นสามารถนำไปประยุกต์ใช้ในบริบทของพื้นที่ศึกษา เพื่อคาดการณ์ระดับน้ำตาลสะสมในเลือดของผู้ป่วยเบาหวานชนิดที่ 2 ที่เข้ารับการรักษา ผลการพยากรณ์สามารถใช้ประกอบการตัดสินใจในการส่งตรวจระดับน้ำตาลสะสมเพิ่มเติม ซึ่งช่วยเพิ่มประสิทธิภาพในการควบคุมโรคเบาหวาน ในเขตพื้นที่รับผิดชอบของโรงพยาบาล

ข้อจำกัดของการศึกษาครั้งนี้ คือ ข้อมูลที่นำมาใช้ในการพัฒนาแบบจำลอง ส่วนใหญ่เป็นข้อมูลเชิงโครงสร้างจากฐานข้อมูลเชิงสัมพันธ์ในขณะที่กระบวนการตัดสินใจทางการแพทย์ยังคงพึ่งพาข้อมูลการซักประวัติและการตรวจร่างกาย ซึ่งมักอยู่ในรูปแบบข้อมูลไม่มีโครงสร้าง การนำเทคนิคการประมวลผลภาษาธรรมชาติมาประยุกต์ใช้ร่วมกับการเรียนรู้ของเครื่อง อาจช่วยเพิ่มประสิทธิภาพของแบบจำลองได้

นอกจากนี้ การพัฒนาของเทคโนโลยีฐานข้อมูลสุขภาพขนาดใหญ่ ที่สามารถเชื่อมโยงข้อมูลจากหลายแหล่ง ส่งผลให้จำนวนคุณลักษณะในชุดข้อมูลเพิ่มขึ้นอย่างมาก ดังนั้น การเรียนรู้เชิงลึก (Deep Learning) จึงมีบทบาทสำคัญในการพัฒนาแบบจำลองให้มีประสิทธิภาพสูงขึ้น จากความสามารถในการค้นพบรูปแบบที่ซับซ้อนซึ่งแฝงอยู่ในชุดข้อมูลขนาดใหญ่

เอกสารอ้างอิง

[1] P. Srisodsaluk, B. Chitpakdee, and S. Chanseang, "Innovation for the Care of Older Persons in Thailand during the Thailand 4.0 Era," *UBRU Journal for Public Health Research*, vol. 9, no. 2, pp. 47–54, 2020 (in Thai).

[2] Health Administration Division. (2024). *Smart Hospital*. [Online] (in Thai). Available: https://it-phdb.moph.go.th/.

[3] N. Sriwiboon, "Improvement the performance of the Chest X-ray image classification with convolutional neural network model by using image augmentations technique for COVID-19 diagnosis," *The Journal of KMUTNB*, vol. 31, no. 1, pp. 109–117, 2020 (in Thai).

[4] N. Nonsiri, R. Manassila, and K. Somkanta, "Data classifying to diagnose diabetes risk using data mining techniques," *The Journal of KMUTNB*, vol. 33, no. 2, pp. 538–547, 2022 (in Thai).

[5] Y. L. Cheng, Y. R. Wu, K. Der Lin, C. H. R. Lin, and I. M. Lin, "Using machine learning for the risk factors classification of glycemic control in Type 2 diabetes mellitus," *Healthcare (Basel)*, vol. 11, no. 8, pp. 1141, 2023.

[6] Y. Fan, E. Long, L. Cai, Q. Cao, X. Wu, and R. Tong, "Machine learning approaches to predict risks of diabetic complications and poor glycemic control in nonadherent Type 2 diabetes," *Frontiers in Pharmacology*, vol. 12, pp. 665951, 2021.

[7] Hfocus. (2023, November). *5.2 million Thais diagnosed with diabetes, with over 20 million people suffering from obesity*. [Online] (in Thai). Available: https://www.hfocus.org/.

[8] Health data center. (2024, January). *HDC - Dashboard*. [Online] (in Thai). Available: https://hdcservice.moph.go.th.

[9] Hfocus. (2023, November). *As of the year 2023, there are a total of 3.3 million Thais diagnosed with diabetes*. [Online] (in Thai). Available: https://www.hfocus.org/.

[10] Diabetes Association of Thailand, *Clinical Practice Guideline for Diabetes 2023*. Bangkok: Srimuang Printing, 2023 (in Thai).

[11] *Handbook of integrated, people-centered health services in new normal diabetic & hypertensive clinic for healthcare workers*, Division of medical technical and academic affairs Ministry of Public Health, Nonthaburi, 2020 (in Thai).

[12] C. Imai, L. Li, R. A. Hardie, and A. Georgiou, "Adherence to guideline-recommended HbA1c testing frequency and better outcomes in patients with type 2 diabetes: A 5-year retrospective cohort study in Australian general practice," *BMJ quality & safety*, vol. 30, no. 9, pp. 706–714, 2021.

[13] T. Weiss, A. Edwards, D. Lautsch, S. Rajpathak, and K. Snow, "Hemoglobin A1C testing frequency among patients with type 2 diabetes within a US payer system: A retrospective observational study," *Current medical research and opinion*, vol. 37, no. 11, pp. 1859–1866, 2021.

[14] Achieve.Plus. (2020, August). *The 4 important classifications in Supervised Learning*. [Online] (in Thai). Available: https://medium.com.

[15] J. H. Lee and J. C. Huber, "Evaluation of multiple imputation with large proportions of missing data: How Much Is Too Much?," *Iranian*

Journal of Public Health, vol. 50, no. 7, pp. 1372–1380, 2021.

[16] W. Romsaiyud, *Machine Learning for Predictive Data Analytics and Applications*. 1st ed. Nonthaburi: Sukhothai Thammathirat Open University, 2023 (in Thai).

[17] M. Khushi, K. Shaukat, T. M. Alam, I. A. Hameed, S. Uddin, S. Luo, X. Yang, and M. C. Reyes, "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data," *IEEE Access*, vol. 9, pp. 109960–109975, 2021.

[18] T. Pasangthien and B. Yimwadsana, "Rebalancing clinical data with probabilistic random oversampling," *Journal of the Thai Medical Informatics Association*, vol. 8, no. 2, pp. 68–72, 2022 (in Thai).

[19] V. R. Joseph, "Optimal ratio for data splitting," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 15, no. 4, pp. 531–538, 2022.

[20] ProjectPro. (2024, April). *How to Solve a Multi Class Classification Problem with Python?*. [Online]. Available: https://www.projectpro.io

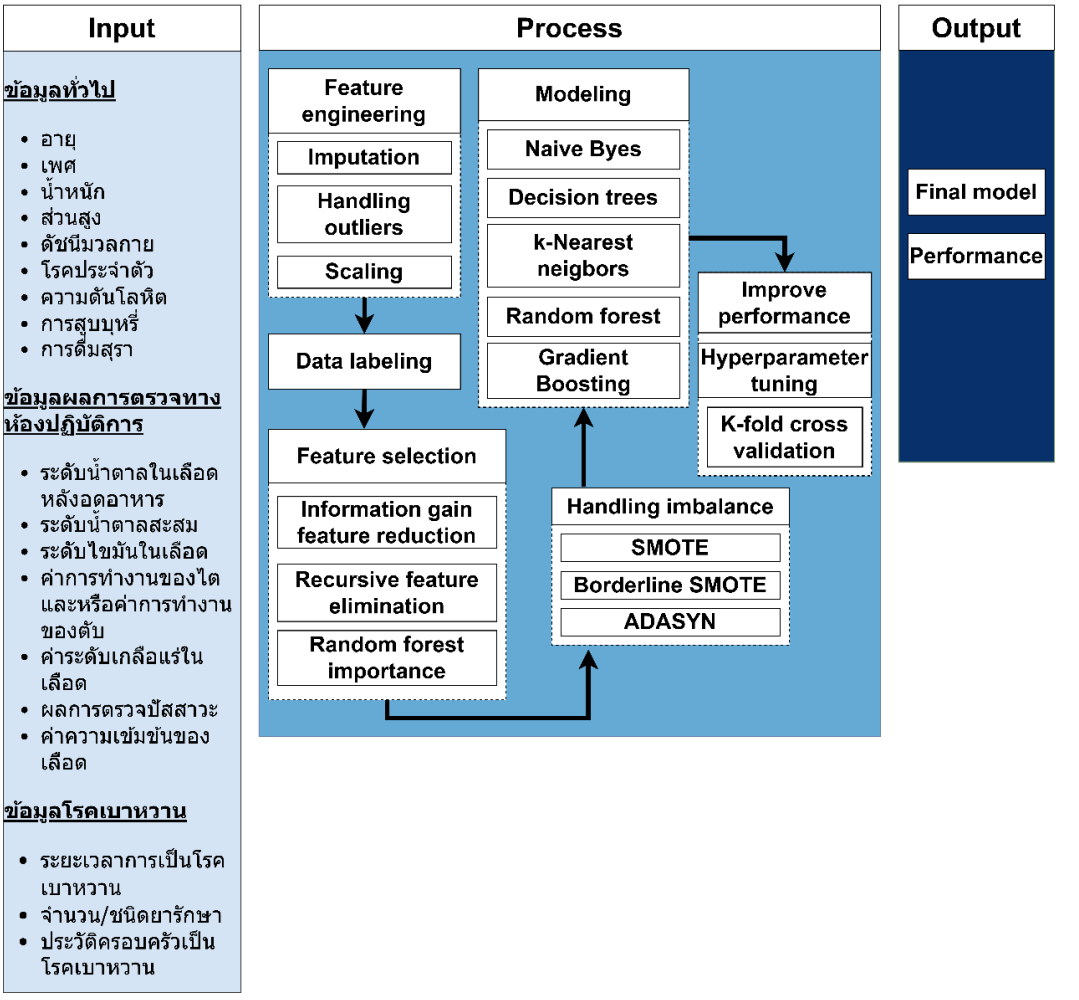
[21] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?," *Computational Statistics*, vol. 36, no. 3, pp. 2009–2031, 2021.

[22] K. Mnich, A. Polewko-Klim, A. K. Golinska, W. Lesinski, and W. R. Rudnicki, "Super learning with repeated cross validation," in *2020 International Conference on Data Mining Workshops (ICDMW) IEEE*, 2020, pp. 629–635.

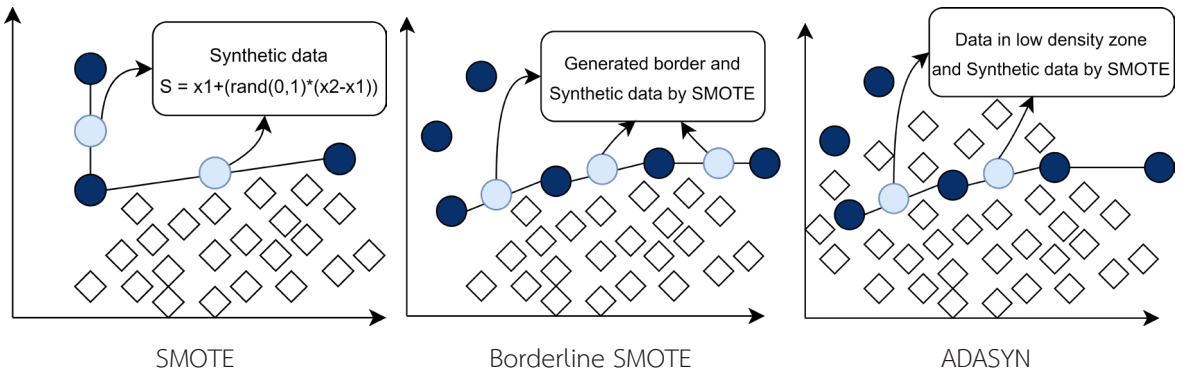
[23] M. A. Sahid, M. U. H. Babar, and M. P. Uddin, "Predictive modeling of multi-class diabetes mellitus using machine learning and filtering iraqi diabetes data dynamics," *PloS one*, vol. 19, no. 5, pp. e0300785, 2024.

[24] A. Almahdawi, Z. S. Naama, and A. Al-Taie, "Diabetes prediction using machine learning," in *2022 3rd Information Technology To Enhance e-learning and Other Application (IT-ELA) IEEE*, 2022, pp. 186–190.

ภาคผนวก



รูปที่ 4 กรอบแนวคิดการวิจัย



รูปที่ 5 การจัดการข้อมูลแบบไม่สมดุล (Handling Imbalance Data)

ตารางที่ 7 รายละเอียดคุณลักษณะ

ลำดับ	คุณลักษณะ	รายละเอียด	ข้อมูลที่จัดเก็บ
1	1_age_year	อายุ (ปี)	ตัวเลข
2	1_sex	เพศ	1 = ชาย 2 = หญิง
3	1_bw	น้ำหนัก (กก.)	ตัวเลข
4	1_ht	ส่วนสูง (ซม.)	ตัวเลข
5	1_bmi	ดัชนีมวลกาย (กก./ตร.ม.)	ตัวเลข
6	1_sbp	ความดันโลหิต (ตัวบน)	ตัวเลข
7	1_dbp	ความดันโลหิต (ตัวล่าง)	ตัวเลข
8	1_waist	รอบเอว (ซม.)	ตัวเลข
9	1_smoking	ประวัติสูบบุหรี่	1 = ไม่เคยสูบ 2 = ยังสูบอยู่ หรือเลิกสูบได้ยังไม่ถึง 1 เดือน 3 = เลิกสูบแล้วอย่างน้อย 1 เดือน 4 = ไม่ได้ซักประวัติ
10	1_drinking	ประวัติดื่มสุรา	1 = ไม่ดื่ม 2 = ดื่ม 3 = เคยดื่ม แต่เลิกแล้ว
11	1_ud_ht	โรคประจำตัว ความดันโลหิตสูง	1 = มี 0 = ไม่มี
12	1_ud_dlp	โรคประจำตัว ไขมันในเลือดสูง	1 = มี 0 = ไม่มี
13	1_ud_heart	โรคประจำตัว หัวใจ	1 = มี 0 = ไม่มี
14	2_sugar	ระดับน้ำตาลอดอาหาร	ตัวเลข
15	2_lipid_chol	ไขมันคลอเลสเตอรอล	ตัวเลข
16	2_lipid_tg	ไขมันไตรกลีเซอไรด์	ตัวเลข
17	2_lipid_hdl	ไขมันดี เอชดีเอล	ตัวเลข
18	2_lipid_ldl	ไขมันเลว เอลดีเอล	ตัวเลข
19	2_kidney_cr	ค่าการทำงานของไต ครีอะตินีน	ตัวเลข
20	2_kidney_egfr	ค่าการทำงานของไต การกรอง	ตัวเลข
21	2_kidney_stage	ระดับขั้นไตเสื่อม	ตัวเลข
22	2_blood_hb	ค่าฮีโมโกลบิน (Hemoglobin)	ตัวเลข
23	2_blood_hct	ค่าฮีมาโตคริต (Hematocrit)	ตัวเลข
24	2_blood_mcv	ค่าขนาดเม็ดเลือดแดง	ตัวเลข
25	2_blood_mchc	ค่าความชีดจางเม็ดเลือดแดง	ตัวเลข
26	2_lastest_hba1c	ระดับน้ำตาลสะสมครั้งล่าสุด	ตัวเลข

ตารางที่ 7 รายละเอียดคุณลักษณะ (ต่อ)

ลำดับ	คุณลักษณะ	รายละเอียด	ข้อมูลที่จัดเก็บ
27	2_previous_fbs_120day	ระดับน้ำตาลอดอาหารเฉลี่ยในช่วง 120 วันก่อนหน้า	ตัวเลข
28	3_dm_duration_year	ระยะเวลาการเป็นโรคเบาหวาน (ปี) นับจากปีที่ถูกวินิจฉัยโรค	ตัวเลข
29	3_dm_fam_hx	ประวัติคนในครอบครัวเป็นโรคเบาหวาน	1 = มี 0 = ไม่มี
30	3_dm_rx_nph_day	ปริมาณยาฉีด NPH ต่อวัน	ตัวเลข
31	3_dm_rx_mixtard_day	ปริมาณยาฉีด Mixtard ต่อวัน	ตัวเลข
32	3_dm_rx_gpz_day	ปริมาณยากิน Glipizide ต่อวัน	ตัวเลข
33	3_dm_rx_mfm_day	ปริมาณยากิน Metformin ต่อวัน	ตัวเลข
34	3_dm_rx_pio_day	ปริมาณยากิน Pioglitazone ต่อวัน	ตัวเลข
35	3_dm_rx_folic_day	ปริมาณยากิน Folic ต่อวัน	ตัวเลข
36	3_dm_rx_ff_day	ปริมาณยากิน Ferrous ต่อวัน	ตัวเลข
37	3_dm_rx_asa_day	ปริมาณยากิน Aspirin ต่อวัน	ตัวเลข
38	4_group_hba1c	การควบคุมโรคตามระดับน้ำตาลสะสมในเลือด (ดี ปานกลาง ไม่ดี)	1 = ดี 2 = ปานกลาง 3 = ไม่ดี

ตารางที่ 8 ค่า Information Gain และ Feature Importance ของการคัดเลือกคุณลักษณะ

IG	Information Gain	RFE	Importance	RFI	Importance
1_age_year	0.0191	1_age_year	0.0379	1_age_year	0.0252
1_bmi	0.0272	1_bmi	0.0433	1_bmi	0.0269
2_blood_mchc	0.0302	2_blood_mchc	0.0408	2_blood_mchc	0.0248
2_lastest_hba1c	0.3627	2_lastest_hba1c	0.2538	2_lastest_hba1c	0.2128
2_lipid_chol	0.0250	2_lipid_chol	0.0385	2_lipid_chol	0.0262
2_lipid_ldl	0.0194	2_lipid_ldl	0.0384	2_lipid_ldl	0.0256
2_lipid_tg	0.0277	2_lipid_tg	0.0444	2_lipid_tg	0.0289
2_previous_fbs_120day	0.3201	2_previous_fbs_120day	0.1657	2_previous_fbs_120day	0.1478
2_sugar	0.1730	2_sugar	0.1037	2_sugar	0.0899
3_dm_duration_year	0.0329	3_dm_duration_year	0.0396	3_dm_duration_year	0.0261
2_blood_hct	0.0172	1_bw	0.0386	1_bw	0.0242
2_blood_mcv	0.0437	1_dbp	0.0374	1_dbp	0.0239
2_lipid_hdl	0.0288	2_blood_hb	0.0383	1_sbp	0.0244
3_dm_rx_mfm_day	0.0414	2_kidney_cr	0.0386	2_kidney_cr	0.0247
3_dm_rx_mixtard_day	0.0379	2_kidney_egfr	0.0407	2_kidney_egfr	0.0256