

การทำนายภาวะเลือดเป็นกรดจากแลคติกที่สัมพันธ์กับการใช้ยาเมทฟอร์มิน ด้วยเทคนิคการเรียนรู้ของเครื่อง: กรณีศึกษาโรงพยาบาลสุรินทร์

ศุภกร อังโชคอำนวย¹, นัทที พรประภา^{2,4}, ลาวัลย์ ศรีธาพุทธร^{3,4}

¹กลุ่มงานเภสัชกรรม โรงพยาบาลสุรินทร์

²สาขาวิชาบริบาลเภสัชกรรม คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์

³สาขาวิชาการแพทย์และสารสนเทศศาสตร์ทางสุขภาพ คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์

⁴กลุ่มวิจัยปัญญาประดิษฐ์และเมตาโบลอมิกส์ คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์

บทคัดย่อ

วัตถุประสงค์: พัฒนาแบบจำลองการทำนายภาวะเลือดเป็นกรดจากแลคติกที่สัมพันธ์กับการใช้ยาเมทฟอร์มิน (metformin-associated lactic acidosis: MALA) และเปรียบเทียบผลการทำนายภาวะ MALA จากแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องแต่ละประเภท **วิธีการ:** การศึกษาเก็บข้อมูลแบบย้อนหลังจากเวชระเบียนอิเล็กทรอนิกส์ของผู้ป่วยเบาหวานอายุ 18 ปีขึ้นไปที่ได้รับยาเมทฟอร์มินของโรงพยาบาลสุรินทร์ ระหว่าง มกราคม 2560 ถึง ธันวาคม 2564 ปัจจัยที่นำมาสร้างแบบจำลองคือ ค่าดัชนีมวลกาย ขนาดยาเมทฟอร์มินที่ได้รับต่อวัน โรคอ้วน การได้ NSAIDs ใน 6 เดือน การได้รับยากลุ่ม thiazides การได้รับยากลุ่ม beta blockers การได้รับยากลุ่ม statins ระดับการทำงานของไต ภาวะไขมันในเลือดผิดปกติ และระยะเวลาที่ได้รับการวินิจฉัยโรคเบาหวาน การศึกษาใช้เทคนิคการเรียนรู้ของเครื่องชนิดที่มีผู้สอนแบบแบ่งแยกประเภท เทคนิคการเรียนรู้ของเครื่องที่ใช้ได้แก่ Logistic Regression (LR), Multi-layer Perceptron (MLP), Support Vector Machine (SVM) และ Random Forest (RF) **ผลการวิจัย:** ผู้ป่วยเบาหวานที่ได้รับยาเมทฟอร์มิน 8,505 ราย แบ่งเป็นผู้ป่วยเบาหวานปกติ 8,387 ราย และผู้ป่วยเบาหวานที่เกิดภาวะ MALA 118 ราย แบบจำลองจากเทคนิค RF และ MLP มีความถูกต้องมากกว่า LR อย่างมีนัยสำคัญ เมื่อทดสอบพบแบบจำลองจาก RF ให้ค่า sensitivity 93.30%, specificity 93.57%, accuracy 95.26%, และ AUROC 0.992 **สรุป:** แบบจำลองการทำนายภาวะ MALA ด้วยเทคนิคการเรียนรู้ของเครื่องโดยเทคนิค RF ให้แบบจำลองที่มีประสิทธิภาพสูงสุด

คำสำคัญ: เมทฟอร์มิน ภาวะเลือดเป็นกรด โรคเบาหวาน เทคนิคการเรียนรู้ของเครื่อง

รับต้นฉบับ: 24 มี.ค. 2566, ได้รับบทความฉบับปรับปรุง: 25 เม.ย. 2566, รับลงตีพิมพ์: 28 เม.ย. 2566

ผู้ประสานงานบทความ: ลาวัลย์ ศรีธาพุทธร สาขาวิชาการแพทย์และสารสนเทศศาสตร์ทางสุขภาพ คณะเภสัชศาสตร์มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์ เลขที่ 6 ถนนราชมรรคาใน ตำบลพระปฐมเจดีย์ อำเภอเมือง จังหวัดนครปฐม 73000 E-mail: SRATTAPHUT_L@su.ac.th

Prediction of Metformin-Associated Lactic Acidosis Using Machine Learning Techniques : A Case Study of Surin Hospital

Supakorn Ungchok-amnuai¹, Nattee Phornprapha^{2,4}, Lawan Sratthaphut^{3,4}

¹Department of Pharmacy, Surin Hospital

²Department of Pharmaceutical Care, Faculty of Pharmacy,
Silpakorn University, Sanam Chandra Palace Campus

³Department of Biomedicine and Health informatics, Faculty of Pharmacy,
Silpakorn University, Sanam Chandra Palace Campus

⁴Artificial Intelligence and Metabolomics Research Group, Faculty of Pharmacy,
Silpakorn University, Sanam Chandra Palace Campus

Abstract

Objective: To develop a predictive model for metformin-associated lactic acidosis (MALA) and compare the prediction of MALA among various models using different machine learning techniques. **Methods:** The study retrospectively collected data from electronic medical records of diabetic patients aged 18 years and over who received metformin at Surin Hospital between January 2017 and December 2021. The factors used to build the model were body mass index, metformin daily dose, obesity, NSAIDs use within the past 6 months, thiazides use, use of beta blockers, use of statins, renal function, dyslipidemia and duration of being diagnosed with diabetes. The study used a supervised machine learning technique with classification algorithm. Machine learning techniques used were Logistic Regression (LR), Multi-layer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest (RF). **Results:** There were 8,505 diabetic patients receiving metformin including 8,387 of those without MALA and 118 with MALA. RF and MLP models were significantly more accurate than LR. The RF model showed sensitivity 93.30%, specificity 93.57%, accuracy 95.26%, and AUROC of 0.992. **Conclusion:** The MALA prediction model using RF machine learning technique was the most efficient model.

Keywords: metformin, lactic acidosis, diabetes, machine learning techniques

บทนำ

ปัจจุบันแนวทางการรักษาเบาหวานทั้งในไทยและต่างประเทศ แนะนำการใช้ยาเมทฟอร์มิน (metformin) เป็นยาชนิดแรกในการรักษาผู้ป่วยในกรณีที่ไม่มีความผิดปกติของตับไต เนื่องจากสามารถลดระดับน้ำตาลในเลือดได้ดีและราคาไม่แพง (1, 2) เมทฟอร์มินเป็นยาเบาหวานในกลุ่ม biguanides ที่ลดระดับน้ำตาลในเลือดจากกลไกการยับยั้งกระบวนการสร้างกลูโคสที่ตับ อาการข้างเคียงที่พบจากการใช้ยาเมทฟอร์มิน คือ คลื่นไส้ อาเจียน ท้องเสีย และภาวะเลือดเป็นกรดจากแลคติก ซึ่งพบได้น้อยแต่มีความรุนแรงถึงชีวิต ข้อห้ามในการใช้ยาเมทฟอร์มิน คือ ผู้ที่มีค่าการทำงานของไตน้อยกว่า 30 มล./นาที/1.73 ตร.ม. เพื่อป้องกันการเกิดอาการดังกล่าว (2) การเพิ่มขึ้นของระดับแลคเตทในเลือดนั้น สามารถเกิดได้จากหลายสาเหตุ คือภาวะฉุกเฉินต่าง ๆ เช่น ภาวะช็อค หัวใจหยุดเต้นเฉียบพลัน การขาดออกซิเจนของอวัยวะในร่างกาย โรคประจำตัว เช่น โรคตับ มะเร็งบางชนิด รวมถึง ยาหรือสารพิษ เช่น เมทฟอร์มิน ยาในกลุ่ม beta blocker ไซยาไนด์ คาร์บอนมอนอกไซด์ (3)

ภาวะเลือดเป็นกรดจากแลคติกที่สัมพันธ์กับการใช้ยาเมทฟอร์มิน (metformin-associated lactic acidosis: MALA) คือ ภาวะเลือดเป็นกรดจากแลคติกที่เชื่อว่ามีสาเหตุจากการใช้ยาเมทฟอร์มิน โดยระดับแลคเตทที่เพิ่มสูงขึ้นในเลือดเกิดจาก กลไกการออกฤทธิ์ของยาที่ยับยั้งกระบวนการสร้างกลูโคสและมีความเสี่ยงสูงขึ้นในผู้ป่วยที่มีภาวะไตบกพร่องเนื่องจากการกำจัดยาออกน้อยลง อุบัติการณ์ของการเกิดภาวะดังกล่าวในประเทศสหรัฐอเมริกาเท่ากับ 3-10 รายต่อ 1 แสนประชากร (4) ปัจจุบันยังมีความขัดแย้งของข้อมูลจากการศึกษาต่าง ๆ ว่าปัจจัยใดที่มีผลต่อการเกิดภาวะ MALA รวมถึงยังไม่มีเครื่องมือที่นำมาใช้ทำนายโอกาสในการเกิดภาวะดังกล่าว งานวิจัยนี้จึงนำข้อมูลเกี่ยวกับปัจจัยต่าง ๆ มาสร้างแบบจำลองด้วยเพื่อทำนายโอกาสในการเกิดภาวะ MALA เพื่อป้องกันและเฝ้าระวังอันตรายที่จะเกิดขึ้นกับผู้ป่วยเบาหวานต่อไป

วิธีการวิจัย

การศึกษานี้เป็นการเก็บข้อมูลเชิงสังเกตแบบย้อนหลัง ซึ่งผ่านการพิจารณาจากคณะกรรมการพิจารณาโครงการวิจัยมหาวิทยาลัยศิลปากร (เลขที่โครงการ REC 65.0525-081-4507) และจากคณะกรรมการจริยธรรมการวิจัยของโรงพยาบาลสุรินทร์ (หนังสือรับรองเลขที่ 21/2565)

การเก็บข้อมูล

การศึกษาเก็บข้อมูลจากเวชระเบียนอิเล็กทรอนิกส์จากโปรแกรม HosXP ของผู้ป่วยเบาหวานที่อายุ 18 ปีขึ้นไป ที่ได้รับยาเมทฟอร์มินไม่เกิน 2,550 มก./วัน จากโรงพยาบาลสุรินทร์ ตั้งแต่ปี พ.ศ. 2560 – 2564 ที่แผนกผู้ป่วยนอกและอุบัติเหตุฉุกเฉินของ โรงพยาบาลสุรินทร์ โดยศึกษาในผู้ป่วยทุกรายที่ผ่านเกณฑ์คัดเข้าดังกล่าว

ข้อมูลด้านคุณลักษณะประชากรและสังคมที่เก็บได้แก่ เพศ (ชาย/หญิง) ดัชนีมวลกาย และอายุ (ปี) ทั้งนี้ในกลุ่มผู้ป่วยเบาหวานที่เกิดภาวะ MALA นับอายุตั้งแต่วันเกิดถึงวันที่ได้รับการวินิจฉัย ส่วนกลุ่มผู้ป่วยเบาหวานปกติ นับอายุตั้งแต่วันเกิดถึงวันที่มารับการรักษาครั้งสุดท้าย

ข้อมูลด้านคลินิกที่ศึกษาได้แก่ ขนาดยาเมทฟอร์มินที่ได้รับต่อวัน (มก.) ระยะเวลาตั้งแต่ได้รับการวินิจฉัยโรคเบาหวาน (ปี) การควบคุมระดับน้ำตาลในเลือด (ควบคุมได้/ควบคุมไม่ได้) โดยพิจารณาจากค่าระดับน้ำตาลสะสมล่าสุด (HbA1C<7%) หากไม่มีค่าดังกล่าว พิจารณาจากค่าระดับน้ำตาลในเลือดขณะอาหาร (fasting blood sugar: FBS< 130 มก./ดล) ระดับการทำงานของไตโดยแบ่งตามค่าอัตราการกรองของไต (eGFR) เป็น 6 ระดับตามตารางที่ 1 และ 2 โรคปอดอุดกั้นเรื้อรัง (chronic obstructive pulmonary disease: COPD) (มี/ไม่มี) โรคหัวใจล้มเหลว (มี/ไม่มี) โรคตับ (มี/ไม่มี) ได้แก่ โรคตับแข็ง โรคตับจากสุรา (cirrhosis; K7x) หรือโรคมะเร็งตับ (liver cancer; C22x, C787) ภาวะไขมันผิดปกติ (มี/ไม่มี) โรคความดันโลหิตสูง (มี/ไม่มี) ยาชนิดอื่นที่ได้รับรวม (ได้รับ/ไม่ได้รับ) ดังนี้ กลุ่มยาลดความดันโลหิต angiotensin converting enzyme

ตารางที่ 1. การคำนวณค่า eGFR ด้วยสูตร Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) (5)

เพศ	ระดับ serum creatinine (mg/dL)	สูตรที่ใช้
ชาย	น้อยกว่าหรือเท่ากับ 0.7	$GFR = 144 \times (SCr/0.7)^{-0.329} \times (0.993)^{\text{อายุ}}$
	มากกว่า 0.7	$GFR = 144 \times (SCr/0.7)^{-1.209} \times (0.993)^{\text{อายุ}}$
หญิง	น้อยกว่าหรือเท่ากับ 0.9	$GFR = 141 \times (SCr/0.7)^{-0.411} \times (0.993)^{\text{อายุ}}$
	มากกว่า 0.9	$GFR = 141 \times (SCr/0.7)^{-1.209} \times (0.993)^{\text{อายุ}}$

ตารางที่ 2. การแบ่งระยะของโรคไตเรื้อรังตามเกณฑ์ของอัตราการกรองของไตตามข้อแนะนำเวชปฏิบัติการดูแลผู้ป่วยโรคไตเรื้อรังก่อนการบำบัดทดแทนไต พ.ศ. 2565 (5)

ระยะของโรคไตเรื้อรัง	อัตราการกรองของไต (มิลลิลิตร/นาที่/1.73 ตารางเมตร)
ระยะที่ 1	มากกว่าหรือเท่ากับ 90
ระยะที่ 2	60 - 89
ระยะที่ 3a	45 - 59
ระยะที่ 3b	30 - 44
ระยะที่ 4	15 - 29
ระยะที่ 5	น้อยกว่า 15

inhibitors (ACE inhibitors), angiotensin receptor blockers (ARBs), thiazide diuretics, loop diuretics, beta blockers, calcium channel blockers (CCBs), ยาลดระดับไขมันในเลือด statins หรือ fibrates และการได้รับยาในกลุ่ม non-steroidal anti-inflammatory drugs (NSAIDs) (ได้รับ/ไม่ได้รับ) โดยเป็นข้อมูลย้อนหลัง 6 เดือน

การศึกษาจำแนกผู้ป่วยออกเป็น 2 กลุ่ม ได้แก่ กลุ่มผู้ป่วยเบาหวานที่เกิด MALA จากการวินิจฉัยของอายุรแพทย์โรคไต โดยเวชระเบียนไม่ได้มีการระบุความเสี่ยงอื่นๆ ที่นำไปสู่การเกิดภาวะเลือดเป็นกรดจากแลคติก และกลุ่มผู้ป่วยเบาหวานที่ไม่เกิด MALA ทั้งนี้ การศึกษาได้ตรวจสอบความซ้ำซ้อนของข้อมูลด้วยหมายเลขโรงพยาบาลของผู้ป่วย (Hospital Number: HN)

การวิเคราะห์ข้อมูลและคัดเลือกตัวแปร

การวิเคราะห์ปัจจัยด้านคุณลักษณะประชากรและสังคมและด้านคลินิกของผู้ป่วยใช้สถิติเชิงพรรณนา การวิเคราะห์ปัจจัยที่มีผลต่อการเกิด MALA ที่เป็นตัวแปรกลุ่มใช้ chi-square ปัจจัยดังกล่าว ได้แก่ เพศ (ชาย/ หญิง) โรคอ้วน (BMI มากกว่าหรือเท่ากับ 25) การควบคุมระดับน้ำตาลในเลือดให้ได้ตามเป้าหมาย (ได้/ไม่ได้) โรคปอดอุดกั้นเรื้อรัง (มี/ไม่ มี) โรคหัวใจล้มเหลว (มี/ไม่มี) โรคตับ (มี/ไม่มี) ภาวะไขมันในเลือดผิดปกติ (มี/ไม่มี) โรคความดันโลหิตสูง (มี/ไม่มี) การได้รับยาในกลุ่ม ACE inhibitors หรือ ARBs (ได้รับ/ไม่ได้รับ) การได้รับยาในกลุ่ม thiazide diuretics (ได้รับ/ไม่ได้รับ) การได้รับยาในกลุ่ม loop diuretics (ได้รับ/ไม่ได้รับ) การได้รับยาในกลุ่ม beta blockers (ได้รับ/ไม่ได้รับ) การได้รับยาในกลุ่ม CCBs (ได้รับ/ไม่ได้รับ) การได้รับยาในกลุ่ม statins (ได้รับ/ไม่ได้รับ) การได้รับยาในกลุ่ม fibrates

(ได้รับ/ไม่ได้รับ) และการได้รับยา NSAIDs ย้อนหลัง 6 เดือน (ได้รับ/ไม่ได้รับ) ส่วนการวิเคราะห์ปัจจัยที่มีผลต่อการเกิด MALA ที่เป็นตัวแปรเชิงปริมาณทดสอบด้วยสถิติชนิด independent T-test ปัจจัยดังกล่าว ได้แก่ อายุ (ปี) ดัชนีมวลกาย ขนาดยาเมทฟอร์มินที่ได้รับต่อวัน (มิลลิกรัม) ระยะเวลาที่ได้รับการวินิจฉัยโรคเบาหวาน (ปี) และระดับการทำงานของไต (มล./นาที่/1.73ตร.ม.) การทดสอบกำหนดค่านัยสำคัญทางสถิติที่ระดับ $P < 0.05$ หลังจากนั้นนำปัจจัยที่มีนัยสำคัญทางสถิติไปใช้ในแบบจำลองพยากรณ์

การปรับสมดุลข้อมูล

การปรับสมดุลข้อมูล 2 กลุ่ม (MALA และ non-MALA) ใช้วิธีการสังเคราะห์ข้อมูลเพิ่ม (synthetic minority oversampling techniques: SMOTE) และปรับสมดุลข้อมูลด้วยวิธีสุ่มลด (under-sampling) ด้วยโปรแกรม WEKA (เวอร์ชัน 3.9.6) การปรับสมดุลข้อมูลด้วยวิธีการสังเคราะห์ข้อมูลเพิ่ม (6) เป็นการสุ่มสังเคราะห์ข้อมูลเพิ่มขึ้นจากข้อมูลเดิมที่มีอยู่ด้วยหลักการเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors: K-NN) โดยสุ่มเลือกข้อมูลที่มีอยู่ 1 ค่า หลังจากนั้นพิจารณาข้อมูลใกล้เคียง k ค่า จากนั้นคำนวณระยะทางระหว่างค่าที่สุ่มกับค่าข้อมูลใกล้เคียงแต่ละค่า เพื่อหาค่าระยะทางที่น้อยที่สุด ระหว่างค่าที่สุ่มกับค่าข้อมูลใกล้เคียง จากนั้นสร้างข้อมูลเทียมระหว่างค่าข้อมูลที่สุ่มกับค่าข้อมูลใกล้เคียงตัวที่ให้ค่าระยะทางที่น้อยที่สุด โดยกระบวนการจะดำเนินไปเรื่อย ๆ ตามจำนวนที่ต้องการสังเคราะห์เพิ่ม การศึกษานี้ใช้ SMOTE ซึ่งเป็นการสุ่มข้อมูลที่มีเพื่อสังเคราะห์ข้อมูลเพิ่มในกลุ่มที่เกิด MALA ในอัตราการส่วนที่ทำให้ทั้ง 2 กลุ่มใกล้เคียงกันมากที่สุด (เพิ่มขึ้น 7,000%) ดังตารางที่ 3 ส่วนการปรับสมดุลข้อมูลด้วยวิธีการสุ่มลด (6) ทำได้โดยสุ่มลดข้อมูลในกลุ่มที่มีค่ามากให้ลดลง ในอัตราส่วนหรือจำนวนที่ต้องการ โดยการศึกษาในทำในกลุ่ม non-MALA โดยสุ่มลดให้ข้อมูลทั้ง 2 กลุ่มมีจำนวนเท่ากัน

ตารางที่ 3. จำนวนผู้ป่วยในกลุ่ม MALA ที่สังเคราะห์ข้อมูลเพิ่มด้วยเทคนิค SMOTE

การปรับสมดุลข้อมูล	Non-MALA	MALA	รวม
จำนวนผู้ป่วยจริง (ราย)	8,387	118	8,505
จำนวนผู้ป่วยที่ผ่าน			
การสังเคราะห์ข้อมูลเพิ่ม (ราย)	8,387	8,378*	16,765

*กำหนดให้ percentage = 7,000 และ k=5

ตารางที่ 4. จำนวนผู้ป่วยในกลุ่ม Non-MALA ที่ปรับสมดุลข้อมูลด้วยวิธีการสุ่มลด

การปรับสมดุลข้อมูล	Non-MALA	MALA	รวม
จำนวนผู้ป่วยจริง (ราย)	8,387	118	8,505
จำนวนผู้ป่วยที่ปรับสมดุล			
ข้อมูลด้วยวิธีการสุ่มลด (ราย)	118*	118	236

*distributionSpread = 1.0

หรือ อัตราส่วนของจำนวนตัวอย่างระหว่างทั้ง 2 กลุ่มเท่ากับ 1:1 ดังตารางที่ 4

การพัฒนาแบบจำลอง

การศึกษาแบ่งข้อมูลที่ผ่านมาการปรับสมดุลเป็น 2 ชุดด้วยวิธี random sampling ได้แก่ ชุดฝึกสอน (training set) จำนวนร้อยละ 80 สำหรับใช้พัฒนาแบบจำลอง และชุดทดสอบ (test set) จำนวนร้อยละ 20 สำหรับประเมินแบบจำลอง โดยข้อมูลชุด SMOTE มีข้อมูลจำนวนทั้งสิ้น 16,765 ราย แบ่งเป็นชุดฝึกสอน 13,142 ราย (ร้อยละ 80) และชุดทดสอบ 3,353 ราย (ร้อยละ 20) ข้อมูลชุด under-sampling มีข้อมูลจำนวนทั้งสิ้น 236 ราย แบ่งเป็น ชุดฝึกสอน 189 ราย (ร้อยละ 80) และชุดทดสอบ 47 ราย (ร้อยละ 20)

การศึกษาค่า hyperparameter ที่เหมาะสมของแบบจำลองที่สร้างด้วยเทคนิค Logistic Regression (LR), Multi-layer Perceptron (MLP), Support Vector Machines (SVM), Random Forest (RF) ด้วยวิธี cross-validation คือ การแบ่งจำนวนข้อมูลออกเป็น k ชุดเท่า ๆ กัน จากนั้นนำข้อมูล $k-1$ ชุดสำหรับสร้างแบบจำลอง และข้อมูล 1 ชุดสำหรับทดสอบ จากนั้นสลับกันทดสอบไปจนครบจำนวน k รอบ (7) จากนั้นหาค่า hyperparameter ที่เหมาะสม ค่า hyperparameter คือ ค่าพารามิเตอร์ที่มีความจำเป็นในขั้นตอนการสร้างแบบจำลองของเทคนิคการเรียนรู้ของเครื่อง โดยเป็นขั้นตอนที่มีผลต่อประสิทธิภาพของแบบจำลอง โดยการหาค่า hyperparameter ที่เหมาะสม (hyperparameter tuning) มีความแตกต่างไปตามแต่ละเทคนิคที่เลือกใช้ การศึกษานี้ทดลองปรับค่า hyperparameter ต่าง ๆ ที่ละขั้นตอน (random search) โดยแต่ละเทคนิคปรับค่า hyperparameter ดังตารางที่ 5 การศึกษาเปรียบเทียบค่าความอ่อนไหว ความแม่นยำ F1 score และความถูกต้อง (ความหมายอยู่ในหัวข้อถัดไป) ที่

ได้เมื่อมีการปรับค่า hyperparameter ของแต่ละเทคนิค และตัดสินใจว่า ข้อมูลชุดฝึกสอนที่มีการปรับสมดุลข้อมูลด้วย SMOTE หรือ under-sampling แบบใดมีความเหมาะสมกว่ากัน

ตารางที่ 5. การปรับค่า hyperparameter ของแต่ละเทคนิค

เทคนิค	ค่า hyperparameter
LR	-ridge: 1×10^{-8} – 10 -max iterations: (-1) – 1,000
MLP	-learning rate: 0.1 – 0.9 -momentum: 0.1 – 0.9 -hidden layers: 1 – 5 -neurons: 10 – 100 -training time: 500 – 5,000
SVM	-ประเภทของ kernel: polykernel, RBF และ normalized polykernel -regularize parameter (c): 0.1 – 20 -tolerance: 1×10^{-8} – 1×10^{-3}
RF	-number of iterations: 100 – 1,000 -number of features: 1 – 4 maximum depth: 10 – 100

ผลลัพธ์การทำนาย

การศึกษานำแบบจำลองการเรียนรู้ที่ใช้เทคนิค LR, MLP, SVM และ RF ที่มีค่า hyperparameter ที่เหมาะสม มาเปรียบเทียบด้วยโปรแกรม WEKA (weka.experiment.PairedCorrectedTTTester) และปรับค่าทดสอบเป็นชนิดแบ่งกลุ่ม (classification) เพื่อเปรียบเทียบค่าความถูกต้องในการทำนาย (percent_correct) โปรแกรมจะเปรียบเทียบค่าความถูกต้องที่ได้ของแต่ละเทคนิคกับเทคนิคอ้างอิง (LR) ด้วยสถิติชนิด paired T-test การประเมินแบบจำลองที่ได้จากเทคนิคที่ให้ผลเหมาะสมที่สุดใช้ข้อมูลชุดทดสอบและหาค่า sensitivity, specificity, accuracy, AUROC ดังแสดงในตารางที่ 6 โดยค่า recall (ค่าความอ่อนไหว) เป็นการวัดความถูกต้องของแบบจำลองหาได้จาก $A / (A+C)$ ค่า precision (ค่าความแม่นยำ) เป็นการวัดความแม่นยำของแบบจำลองหาได้จาก $A / (A+B)$ ค่า F1 score คือ ค่าเฉลี่ยของค่าความอ่อนไหวและค่าความแม่นยำ หาได้จาก = $\frac{2 \times \text{ค่าความอ่อนไหว} \times \text{ค่าความแม่นยำ}}{\text{ค่าความอ่อนไหว} + \text{ค่าความแม่นยำ}}$ ส่วนค่า

ตารางที่ 6. ผลลัพธ์การทำนาย (8)

ผลการทำนาย จากแบบจำลอง	การเกิด MALA ในสภาวะจริง	
	ผู้ป่วยเกิด MALA จริง	ผู้ป่วยไม่เกิด MALA
เกิด MALA	A (ผลบวกจริง: true positive)	B (ผลบวกปลอม: false positive)
ไม่เกิด MALA	C (ผลลบปลอม: false negative)	D (ผลลบจริง: true negative)

accuracy (ค่าความถูกต้อง) คือ ค่าความน่าจะเป็นที่ได้ผลการทำนายที่ถูกต้อง หาได้จาก $(A+D) / (A+B+C+D)$ ค่า sensitivity (ค่าความไว) คือ ค่าความน่าจะเป็นที่แบบจำลองทำนายว่าผู้ป่วยเกิด MALA เมื่อผู้ป่วยรายนั้นเกิด MALA ขึ้นจริง หาได้จาก $A / (A+C)$ ค่า specificity (ค่าความจำเพาะ) คือ ค่าความน่าจะเป็นที่แบบจำลองทำนายว่าผู้ป่วยไม่เกิด MALA เมื่อผู้ป่วยรายนั้นไม่เกิดภาวะนั้นจริง หาได้จาก $D / (B+D)$ ส่วนโค้ง ROC หรือ receiver operating characteristics เป็นกราฟที่แสดงความสัมพันธ์ระหว่างค่าความไว (แกน X) และ ค่า 1-ค่าความจำเพาะ (แกน Y) พื้นที่ใต้กราฟ (area under the ROC curve: AUROC) ดังกล่าวใช้บอกค่าประสิทธิภาพของแบบจำลอง โดยค่าที่ดีที่สุดมีค่าเท่ากับ 1 หรือเข้าใกล้ 1

ผลการวิจัย

คุณลักษณะประชากรและสังคม

ผู้ป่วยเบาหวานที่ได้รับยาการรักษาด้วยยาเมทฟอร์มินที่ผ่านเกณฑ์คัดเข้าทั้งหมด 8,505 ราย แบ่งเป็น

กลุ่ม non-MALA เท่ากับ 8,387 ราย (ร้อยละ 98.61) และกลุ่ม MALA เท่ากับ 118 ราย (ร้อยละ 1.39) ประชากรส่วนใหญ่เป็นเพศหญิง โดยแบ่งเป็นเพศหญิง 4,884 (ร้อยละ 58.23) และ 75 ราย (ร้อยละ 63.56) ในกลุ่ม non-MALA และกลุ่ม MALA ตามลำดับ ค่ามัธยฐานอายุเท่ากับ 60 ปีในกลุ่ม non-MALA และ เท่ากับ 61 ปีในกลุ่ม MALA (พิสัย 18-95 ปี) ค่ามัธยฐานดัชนีมวลกาย เท่ากับ 24.80 กิโลกรัมต่อตารางเมตร ในกลุ่ม non-MALA และ เท่ากับ 20.30 กิโลกรัมต่อตารางเมตรในกลุ่ม MALA ทั้งนี้ ผู้ป่วยในกลุ่ม non-MALA ร้อยละ 51.26 และกลุ่ม MALA ร้อยละ 82.20 มีระดับดัชนีมวลกายน้อยกว่า 25.0 กิโลกรัมต่อตารางเมตร ดังตารางที่ 7

ข้อมูลด้านคลินิก

จากตารางที่ 8 ค่ามัธยฐานของขนาดยาเมทฟอร์มินที่ได้รับต่อวันเท่ากับ 1,000 มิลลิกรัมในกลุ่ม non-MALA และ 2,000 มิลลิกรัมในกลุ่ม MALA (พิสัย 250-2,550 มิลลิกรัมต่อวัน) ค่ามัธยฐานของระยะเวลาที่ได้รับการวินิจฉัยโรคเบาหวานเท่ากับ 5 และ 4 ปีในกลุ่ม non-MALA และ MALA ตามลำดับ (พิสัย 1-44 ปี) ผู้ป่วยเบาหวานส่วนใหญ่ไม่สามารถควบคุมระดับน้ำตาลในเลือดได้ตามเป้าหมาย เท่ากับ 4,931 ราย (ร้อยละ 58.79) และ 73 ราย (ร้อยละ 61.86) ในกลุ่ม non-MALA และ MALA ตามลำดับ

มัธยฐานระดับการทำงานของไตที่ได้คำนวณจากสูตร CKD-EPI เท่ากับ 89 และ 73 มล./นาที/1.73 ตร.ม. ในกลุ่ม non-MALA และ MALA ตามลำดับ จากค่าอัตราการกรองของไตพบว่า กลุ่ม non-MALA อยู่ในระยะที่ 1 เท่ากับ

ตารางที่ 7. ข้อมูลคุณลักษณะประชากรและสังคม (N=8,505)

คุณลักษณะประชากรและสังคม	Non-MALA	MALA
จำนวน (ร้อยละ)	8,387 (98.61)	118 (1.39)
เพศ, จำนวน (ร้อยละ)		
หญิง	4,884 (58.23)	75 (63.56)
ชาย	3,503 (41.77)	43 (36.44)
อายุ, มัธยฐาน (ปี)	60	61
ดัชนีมวลกาย, ค่าเฉลี่ย±SD (กิโลกรัมต่อตารางเมตร)	25.31±5.16	21.08±4.23
ดัชนีมวลกาย, มัธยฐาน (กิโลกรัมต่อตารางเมตร)	24.80	20.30
< 25.0, จำนวน (ร้อยละ)	4,299 (51.26)	97 (82.20)
≥ 25.0, จำนวน (ร้อยละ)	4,088 (48.74)	21 (17.80)

ตารางที่ 8. ข้อมูลด้านคลินิก

ข้อมูลด้านคลินิก	Non-MALA	MALA
จำนวน (ร้อยละ)	8,387 (98.61)	118 (1.39)
ขนาดยาเมทฟอร์มินที่ได้รับ (มก./วัน, มิลลิกรัม)	1,000	2,000
ขนาดยาเมทฟอร์มินที่ได้รับ, ค่าเฉลี่ย±SD (มก./วัน)	1,355.58±606.88	1,843.64±518.20
ระยะเวลาที่ได้รับการวินิจฉัยเบาหวาน (มก./วัน, ปี)	5	4
ระยะเวลาที่ได้รับการวินิจฉัยเบาหวาน, ค่าเฉลี่ย±SD (ปี)	6.50±6.14	4.76±4.00
การควบคุมระดับน้ำตาลตามเป้าหมาย, จำนวน (ร้อยละ)		
ตามเป้าหมาย	3,456 (41.21)	45 (38.14)
ไม่ได้ตามเป้าหมาย	4,931 (58.79)	73 (61.86)
ระดับการทำงานของไตจากสูตร CKD-EPI (มก./นาที่/1.73ตร.ม.)	89	73
ระดับการทำงานของไตจากสูตร CKD-EPI, ค่าเฉลี่ย±SD (มก./นาที่/1.73ตร.ม.)	84.92±23.25	73.72±22.10
ระดับการทำงานของไต จำแนกตามระยะ, จำนวน (ร้อยละ)		
ระยะที่ 1	4,065 (48.47)	29 (24.58)
ระยะที่ 2	2,877 (34.30)	57 (48.31)
ระยะที่ 3a	1,007 (12.00)	22 (18.64)
ระยะที่ 3b	409 (4.88)	0 (0.00)
ระยะที่ 4	26 (0.31)	10 (8.47)
ระยะที่ 5	3 (0.04)	0 (0.00)
โรคร่วม, จำนวน (ร้อยละ)		
โรคความดันโลหิตสูง	4,660 (53.18)	74 (62.71)
ภาวะไขมันผิดปกติ	4,171 (49.73)	24 (20.34)
โรคตับ (ตับแข็ง มะเร็งตับ)	218 (2.60)	5 (4.24)
โรคหัวใจล้มเหลว	97 (1.16)	4 (3.39)
โรคปอดอุดกั้นเรื้อรัง	37 (0.44)	1 (0.85)
ยาชนิดอื่นที่ได้รับร่วม, จำนวน (ร้อยละ)		
statins	6,622 (78.96)	70 (59.32)
ACE inhibitors หรือ ARBs	3,992 (47.60)	63 (53.39)
calcium channel blockers	2,499 (29.80)	30 (25.42)
beta blockers	1,025 (12.22)	7 (5.93)
thiazide diuretics	319 (3.80)	9 (7.63)
fibrates	298 (3.55)	4 (3.39)
loop diuretics	162 (1.93)	2 (1.69)
NSAIDs ในเวลาย้อนหลัง 6 เดือน	1,912 (22.80)	40 (33.90)

3,065 ราย (ร้อยละ 48.47) และระยะที่ 2 เท่ากับ 2,877 ราย (ร้อยละ 34.30) กลุ่ม MALA อยู่ในระยะที่ 2 เท่ากับ 57 ราย (ร้อยละ 48.31) และระยะที่ 1 เท่ากับ 29 ราย (ร้อยละ 24.58) กลุ่ม non-MALA มีโรคร่วมที่ได้รับการวินิจฉัยโรค คือ ความดันโลหิตสูงเท่ากับ 4,660 ราย (ร้อยละ 53.18)

ภาวะไขมันผิดปกติเท่ากับ 4,171 ราย (ร้อยละ 49.73) และโรคตับเท่ากับ 218 ราย (ร้อยละ 2.60) ส่วนกลุ่ม MALA ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูงเท่ากับ 74 ราย (ร้อยละ 62.71) ภาวะไขมันผิดปกติเท่ากับ 24 ราย (ร้อยละ 20.34) และโรคตับเท่ากับ 5 ราย (ร้อยละ 4.24)

ยาชนิดอื่นที่ได้รับร่วมกับการรับยาเมทฟอร์มมีนในกลุ่ม non-MALA คือ ยาลดระดับไขมันชนิด statins (6,622 รายหรือร้อยละ 78.96) ยากลุ่มลดระดับความดันโลหิตชนิด ACEI หรือ ARBs (3,992 รายหรือร้อยละ 47.60) ยาลดระดับความดันโลหิตชนิด CCBs (2,499 รายหรือร้อยละ 29.80) และการได้รับยาในกลุ่ม NSAIDs ในเวลาอย่างน้อย 6 เดือน (1,912 รายหรือร้อยละ 22.80) ส่วนกลุ่ม MALA มีผู้ป่วยได้รับยาลดระดับไขมันชนิด statins 70 ราย (ร้อยละ 59.32%) ยากลุ่มลดระดับความดันโลหิตชนิด ACE inhibitors หรือ ARBs 63 ราย (ร้อยละ 53.39) ยาลดระดับความดันโลหิตชนิด CCBs 30 ราย (ร้อยละ 25.42) และ NSAIDs ใน 6 เดือนอย่างน้อย 40 ราย (ร้อยละ 33.90) ดังตารางที่ 8

ปัจจัยที่สัมพันธ์กับ MALA

การวิเคราะห์ด้วย chi-square หรือ independent sample t test พบ 10 ปัจจัยที่มีความสัมพันธ์กับการเกิดภาวะ MALA อย่างมีนัยสำคัญทางสถิติสำหรับ ได้แก่ ดัชนีมวลกาย (ตัวแปรเชิงปริมาณในหน่วยกิโลกรัมต่อตารางเมตร) (P<0.001) โรคอ้วน (ดัชนีมวลกายมากกว่าหรือเท่ากับ 25.0 กิโลกรัมต่อตารางเมตร) (P<0.001) ขนาดยาเมทฟอร์มมีนที่ได้รับต่อวันในหน่วยมิลลิกรัม (P<0.001) การได้รับยา NSAIDs อย่างน้อย 6 เดือน (P<0.001) การได้รับยากลุ่ม thiazide diuretics (P<0.001) การได้รับยาในกลุ่ม beta blockers (P<0.001) การได้รับยากลุ่ม statins (P<0.001) ระดับการทำงานของไต (eGFR) (มล./นาที่/1.73 ตร.ม.) (P<0.001) ภาวะไขมันในเลือดผิดปกติ (P<0.001) และจำนวนปีที่ได้รับการวินิจฉัยโรคเบาหวาน (P=0.001) ปัจจัยทั้งสิบถูกนำมาใช้สร้างแบบจำลองพยากรณ์ต่อไป

การสร้างแบบจำลองด้วย hyperparameter ด้วยเทคนิคต่าง ๆ

จากตารางที่ 9 การเปรียบเทียบค่าความอ่อนไหว ความแม่นยำ F 1 score และความถูกต้องด้วย hyperparameter ต่าง ๆ พบว่า ข้อมูลชุดฝึกสอนที่มีการปรับสมดุลข้อมูลด้วย SMOTE ให้ค่าที่เหมาะสมกว่าชุดข้อมูลฝึกสอน

ตารางที่ 9. การหาค่า hyperparameter ของเทคนิค LR, MLP, SVM และ RF

เทคนิค	S/U*	ค่าความอ่อนไหว (ร้อยละ)	ค่าความแม่นยำ (ร้อยละ)	F1 score	ค่าความถูกต้อง (ร้อยละ)
LR					
- ridge = 1×10^{-8}	S	89.00	84.60	0.867	86.41
- max iterations = -1					
LR					
- ridge = 10	U	80.00	81.70	0.809	80.95
- max iterations = 10					
MLP					
- learning rate = 0.1, momentum = 0.2	S	97.40	89.70	0.934	93.13
- hidden layers = 1, neurons = 10					
- training time = 5,000					
MLP					
- learning rate = 0.3, momentum = 0.2	U	78.90	70.80	0.746	73.02
- hidden layers = 2, neurons = 10					
- training time = 500					
SVM					
- polykernel	S	89.50	84.00	0.869	86.50
- regularization parameter (c) = 20.0					
- tolerance = 0.001					

ตารางที่ 9. การหาค่า hyperparameter ของเทคนิค LR, MLP, SVM และ RF (ต่อ)

เทคนิค	S/U*	ค่าความอ่อนไหว (ร้อยละ)	ค่าความแม่นยำ (ร้อยละ)	F1 score	ค่าความถูกต้อง (ร้อยละ)
SVM					
- polykernel	U	81.10	81.90	0.815	81.48
- regularization parameter (c) = 1.0					
- tolerance = 0.001					
RF					
- number of iterations = 100	S	98.00	96.70	0.974	97.36
- number of features = 1					
- maxdepth = 50					
RF					
- number of iterations = 100	U	80.00	73.10	0.764	75.13
- number of features = 1					
- maxdepth = 1					

*S = SMOTE, U = under-sampling, cross-validation = 10

ที่มีการปรับสมดุลด้วย under-sampling จึงนำข้อมูลชุดฝึกสอนที่มีการปรับสมดุลด้วยวิธี SMOTE มาทดสอบด้วยโปรแกรม WEKA (weka.experiment.PairedCorrectedT Tester) โดยใช้ค่า hyperparameter ที่เหมาะสมของแต่ละเทคนิค และปรับค่าทดสอบเป็นชนิดแบ่งกลุ่ม (classification) เพื่อเปรียบเทียบค่าความถูกต้องในการทำนาย (percent_correct) โปรแกรมจะเปรียบเทียบค่าความถูกต้องที่ได้ของแต่ละเทคนิค โดยเป็นการเปรียบเทียบเทคนิคใด ๆ กับเทคนิคอ้างอิง (LR) ที่ละ 1 เทคนิคด้วยสถิติชนิด paired T-test พบว่าเทคนิค MLP และ RF ให้ค่า percent_correct

ที่สูงกว่าเทคนิค LR อย่างมีนัยสำคัญ (P=0.05) จึงนำเทคนิคดังกล่าวมาใช้ทดสอบกับชุดข้อมูลทดสอบ (training set) เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองต่อไป ดังตารางที่ 10

การประเมินแบบจำลอง

ข้อมูลชุดทดสอบของข้อมูลที่ปรับสมดุลข้อมูลด้วย SMOTE ถูกใช้ทดสอบแบบจำลองที่ประกอบด้วยปัจจัย ได้แก่ ดัชนีมวลกาย โรคอ้วน ขนาดยาเมทฟอร์มินที่ได้รับ ต่อวัน การได้รับยา NSAIDs ย้อนหลัง 6 เดือน การได้รับยา

ตารางที่ 10. ค่า percent_correct ของแต่ละเทคนิคเมื่อเปรียบเทียบด้วยสถิติ paired T-test (two tailed)

วิธีการจัดการ ข้อมูลที่ไม่สมดุล	ชุดข้อมูล	ค่าความถูกต้องของการทำนาย			
		LR ^a	MLP	SVM	RF
SMOTE	ชุดข้อมูลเรียนรู้ (training data)	86.49	93.19 ^V	86.47	97.37 ^V
	ชุดข้อมูลมาตรฐาน [0,1] (normalization [0,1])	86.49	93.19 ^V	86.47	97.36 ^V
	ชุดข้อมูลมาตรฐาน [-1,1] (normalization [-1,1])	86.49	93.20 ^V	86.47	97.37 ^V
	ชุดข้อมูลมาตรฐาน Z (standardization)	86.49	93.20 ^V	84.68	97.35 ^V

LR=Logistic Regression, MLP=Multi-layer perceptron, SVM=Support Vector Machine, RF=Random Forest

^a = เทคนิคอ้างอิงสำหรับเปรียบเทียบ, ^V = มากกว่าอย่างมีนัยสำคัญ (P=0.05)

ตารางที่ 11. ค่าความไว ค่าความจำเพาะ ค่าความถูกต้อง และ AUROC ที่ได้จากการทดสอบแบบจำลอง

จำนวนข้อมูล	เทคนิค*	ค่าความไว (%)	ค่าความจำเพาะ (%)	ค่าความถูกต้อง (%)	AUROC
3,353	MLP	90.90	94.99	92.84	0.965
3,353	RF	93.30	93.57	95.26	0.992

* MLP=Multi-layer Perceptron, RF=Random Forest

กลุ่ม thiazide diuretics การได้รับยาในกลุ่ม beta blockers การได้รับยากกลุ่ม statins ระดับการทำงานของไต ภาวะไขมันในเลือดผิดปกติ และระยะเวลาที่ได้รับการวินิจฉัยโรคเบาหวาน เทคนิค RF ให้ค่าความไว 93.30% ค่าความจำเพาะ 93.57% ค่าความถูกต้อง 95.26% และ AUROC 0.992 ซึ่งเหมาะสมกว่าค่าที่ได้จากเทคนิค MLP ดังตารางที่ 11

การอภิปรายและสรุปผล

การศึกษานี้ใช้ 10 ปัจจัย ได้แก่ ดัชนีมวลกาย โรคอ้วน ขนาดยาเมทฟอร์มินที่ได้รับต่อวัน การได้รับยา NSAIDs ย้อนหลัง 6 เดือน การได้รับยากกลุ่ม thiazide diuretics การได้รับยาในกลุ่ม beta blockers การได้รับยากกลุ่ม statins ระดับการทำงานของไต ภาวะไขมันในเลือดผิดปกติ และระยะเวลาที่ได้รับการวินิจฉัยโรคเบาหวาน เพื่อสร้างแบบจำลองพยากรณ์การเกิดภาวะ MALA โดยใช้วิธีการแก้ไขความไม่สมดุลของข้อมูลด้วย SMOTE พบว่าเทคนิค MLP และ RF ให้ค่าความถูกต้องในการทำนายสูงกว่าเทคนิคอื่นอย่างมีนัยสำคัญ เมื่อทดสอบด้วยข้อมูลชุดทดสอบ เทคนิค RF ให้ค่าการทำนายจำแนกผู้ป่วยเบาหวานที่เกิด MALA ออกจากผู้ป่วยเบาหวานปกติได้เหมาะสมกว่าเทคนิค MLP เมื่อพิจารณาจากค่าความไวและ AUROC

ขนาดยาเมทฟอร์มินที่ได้รับต่อวันเป็นหนึ่งในปัจจัยที่นำมาใช้สร้างแบบจำลองพยากรณ์ในการศึกษานี้ ซึ่ง สอดคล้องกับการศึกษาของ Eppenga และคณะ (9) ที่พบว่า การได้รับเมทฟอร์มินขนาดสูง (มากกว่า 2,000 มิลลิกรัมต่อวัน) มีความเสี่ยงต่อการเกิด MALA มากกว่ากลุ่มที่ได้รับขนาดต่ำกว่า ปัจจัยเกี่ยวกับระดับการทำงานของไตในแบบจำลอง สอดคล้องกับหลายการศึกษาที่พบว่าระดับการทำงานของไตที่ลดลงส่งผลต่อความเสี่ยงในการเกิด MALA เนื่องจากเมทฟอร์มินที่เพิ่มสูงในร่างกาย (3,

10, 11) โดยตามหลักเภสัชจลนศาสตร์ ยาเมทฟอร์มินถูกกำจัดออกจากร่างกายผ่านทางไต (4)

การศึกษานี้พบว่า โรคประจำตัวได้แก่ โรคปอดอุดกั้นเรื้อรัง โรคหัวใจล้มเหลว และโรคตับ ไม่มีความสัมพันธ์กับการเกิดภาวะ MALA แต่กลุ่ม MALA มีสัดส่วนผู้ที่มีโรคประจำตัวดังกล่าวสูงกว่า แต่จากการศึกษาก่อนหน้าพบว่าการมีโรคประจำตัวดังกล่าวเพิ่มความเสี่ยงในการเกิด MALA (3, 10)

การจัดการความไม่สมดุลของข้อมูลระหว่างกลุ่มด้วยเทคนิค SMOTE พบว่า ชุดข้อมูลที่ได้ให้แบบจำลองพยากรณ์ที่มีความเหมาะสมมากกว่าการใช้เทคนิค under-sampling ผลการศึกษานี้ยังเป็นเช่นเดียวกับการศึกษาอื่นๆ ที่ใช้ SMOTE ในการจัดการความไม่สมดุลของข้อมูล (6,12)

การศึกษาในอดีตจำนวนมากพบว่า การใช้แบบจำลองพยากรณ์จากเทคนิคการเรียนรู้ของเครื่องชนิดต่าง ๆ มีประสิทธิภาพที่ดี เช่น แบบจำลองเพื่อทำนายการเกิดภาวะไตวายเรื้อรังในชาวไทยที่พบว่า เทคนิค MLP และ RF มีประสิทธิภาพสูงกว่าเมื่อเทียบกับ LR (13) หรือแบบจำลองเพื่อทำนายการพัฒนาของภาวะไตวายระยะสุดท้ายจากผู้ป่วยไตวายระยะเริ่มต้นด้วยเทคนิค RF (14) รวมถึงการศึกษาในสหรัฐอเมริกาที่ใช้เทคนิค RF และ SVM เพื่อทำนายอาการข้างเคียงจากการใช้ยาเมทฟอร์มินได้อย่างมีประสิทธิภาพ (15) ซึ่งสอดคล้องกับการศึกษาที่ใช้เทคนิค RF ทำนายการเกิด MALA ได้อย่างมีประสิทธิภาพ

การศึกษานี้มีข้อดี คือ เป็นการศึกษาในกลุ่มผู้ป่วยเบาหวานชาวไทยที่เก็บข้อมูลย้อนหลังเป็นระยะเวลานาน ส่งผลให้มีข้อมูลผู้ป่วยจำนวนมาก ปัจจัยที่เลือกนำมาสร้างแบบจำลองเป็นข้อมูลทั่วไปของผู้ป่วยเบาหวานที่หาได้ไม่ยาก อย่างไรก็ตาม ภาวะ MALA เป็นอาการข้างเคียงจากยาที่พบได้น้อย จึงได้มีการปรับสมดุลข้อมูลด้วยเทคนิค SMOTE เพื่อแก้ปัญหาค่าความไม่สมดุลของข้อมูลดังกล่าว นอกจากนี้ ข้อมูลทางคลินิกบางตัวมีมากกว่า 1 ค่าหรือมี

การเปลี่ยนแปลงในช่วงระยะเวลาของการศึกษา ผู้วิจัยได้ออกแบบการศึกษาโดยใช้ข้อมูลล่าสุดที่มี ข้อควรพิจารณาอีกประการหนึ่ง คือ เนื่องจากการศึกษานี้เป็นการเก็บข้อมูลย้อนหลัง จึงพบความไม่ครบถ้วนของข้อมูล ซึ่งจะใช้วิธีการแทนค่าด้วยค่ามัธยฐานของตัวแปรนั้น ๆ

แบบจำลองพยากรณ์ที่สร้างจากเทคนิคการเรียนรู้ของเครื่องชนิด RF สามารถทำนายการเกิดภาวะ MALA ในผู้ป่วยเบาหวานได้อย่างมีประสิทธิภาพ โดยสามารถนำมาใช้เฝ้าระวังอาการข้างเคียงดังกล่าวที่อาจเกิดขึ้นเพื่อความปลอดภัยสูงสุดของผู้ป่วยเบาหวานที่ได้รับยาเมทฟอร์มิน

ข้อเสนอแนะ

การศึกษาในอนาคตควรนำแบบจำลองที่ได้จากการศึกษานี้ไปใช้ทดสอบกับผู้ป่วยเบาหวานจากโรงพยาบาลอื่น ๆ เพื่อทดสอบประสิทธิภาพ นอกจากนี้ยังควรทำการศึกษานี้ไปข้างหน้าโดยนำแบบจำลองที่ได้จากการศึกษานี้ทดสอบกับผู้ป่วยเบาหวานที่ได้รับยาเมทฟอร์มินและติดตามผลไปตามระยะเวลาที่กำหนด ทั้งนี้ผลที่ได้จากแบบจำลองเป็นเพียงส่วนหนึ่งในการช่วยตัดสินใจหรือเฝ้าติดตามอาการข้างเคียงอาจเกิดขึ้นได้จากการใช้ยา ยังคงต้องอาศัยการวินิจฉัยจากแพทย์ผู้เชี่ยวชาญเป็นผู้ตัดสินใจเกี่ยวกับการดูแลผู้ป่วย

กิตติกรรมประกาศ

การวิจัยครั้งนี้สำเร็จได้ด้วยคำแนะนำจากคณาจารย์ ขอขอบคุณเภสัชกรในกลุ่มงานเภสัชกรรมของโรงพยาบาลสุรินทร์ที่ช่วยเหลือระหว่างเก็บข้อมูล ขอขอบคุณคุณสุริศา ศรีสวัสดิ์ ที่ให้คำแนะนำและความช่วยเหลือต่าง ๆ ตลอดจนขอขอบคุณผู้ป่วยเบาหวานทุกท่านที่ผู้วิจัยนำข้อมูลมาใช้ในการศึกษาครั้งนี้ โดยงานวิจัยได้รับการสนับสนุนบางส่วนจากคณะเภสัชศาสตร์มหาวิทยาลัยศิลปากร

เอกสารอ้างอิง

1. Diabetes Association of Thailand. Clinical practice guideline for diabetes 2018 [online]. 2018 [cited Apr 14, 2023]. Available from: dmthai.org/new/index.php/sara-khwam-ru/bukhlakr-thangkar-phaethy/cpg/guideline-diabetes-care-2018

2. American Diabetes Association. 9. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes—2021. *Diabetes Care*. 2021; 44(Suppl 1): S111-S124. doi: 10.2337/dc21-S009.
3. Andersen LW, Mackenhauer J, Roberts JC, Berg KM, Cocchi MN, Donnino MW. Etiology and therapeutic approach to elevated lactate levels. *Mayo Clin Proc* 2013; 88: 1127–40.
4. Inzucchi SE, Lipska KJ, Mayo H, Bailey CJ, McGuire DK. Metformin in patients with type 2 diabetes and kidney disease: a systematic review. *JAMA* 2014; 312: 2668- 75.
5. The Nephrology Society of Thailand. Clinical practice recommendation for the evaluation and management of chronic kidney disease in adults 2022 (revised edition) [online]. 2022 [cited Apr 11, 2023]. Available from: www.nephrothai.org/wp-content/uploads/2023/03/CKD-guideline-2565-revised-edition.pdf
6. Kesornsit W, Lorchirachoonkul V, Jitthavech J. Imbalanced data problem solving in classification of diabetes patients. *KKU Research Journal*. 2018; 18: 11–21.
7. Pruneski JA, Williams RJ, Nwachukwu BU, Ramku mar PN, Kiapour AM, Martin RK, et al. The development and deployment of Machine Learning Models. *Knee Surg Sports Traumatol Arthrosc* 2022; 30: 3917–23.
8. Shouval R, Bondi O, Mishan H, Shimoni A, Unger R, Nagler A. Application of machine learning algorithms for clinical predictive modeling: A data-mining approach in SCT. *Bone Marrow Transplant*. 2013; 49: 332–7.
9. Eppenga WL, Lalmohamed A, Geerts AF, Derijks HJ, Wensing M, Egberts A, et al. Risk of lactic acidosis or elevated lactate concentrations in metformin users with renal impairment: a population-based cohort study. *Diabetes Care*. 2014; 37: 2218-24.
10. Trinkley KE, Anderson HD, Nair KV, Malone DC, Saseen JJ. Assessing the incidence of acidosis in

- patients receiving metformin with and without risk factors for lactic acidosis. *Ther Adv Chronic Dis.* 2018; 9: 179–90.
11. Kuan IH, Savage RL, Duffull SB, Walker RJ, Wright DF. The association between metformin therapy and lactic acidosis. *Drug Safety.* 2019; 42: 1449-69.
 12. Soto C, Seresangtakul P, Tangvoraphonkchai V. Model for cervical cancer result prediction using artificial neural network. *KKU Research Journal.* 2013; 13: 39–50.
 13. Thammasudjarit R, Ingsathit P, Saputro SA, Ingsathit A, Thakkinstian A. Comparison of machine learning with logistic regression for prediction of chronic kidney disease in the Thai adult population. *Ramathibodi Medical Journal.* 2021; 44: 1–12. doi.org/10.33165/rmj.2021.44.4.250334
 14. Su C-T, Chang Y-P, Ku Y-T, Lin C-M. Machine learning models for the prediction of renal failure in chronic kidney disease: A retrospective cohort study. *Diagnostics.* 2022; 12: 2454. doi: 10.3390/diagnostics12102454.
 15. Mehrpour O, Saeedi F, Hoyte C, Goss F, Shirazi FM. Utility of support vector machine and decision tree to identify the prognosis of metformin poisoning in the United States: Analysis of national poisoning data system. *BMC Pharmacol Toxicol.* 2022; 23: 49. doi: 10.1186/s40360-022-00588-0.