การพัฒนาและการเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการคัดกรองผู้ป่วยกลุ่มที่สุ่มเสี่ยงเป็นโรคไต ด้วยเทคนิคการทำเหมืองข้อมูล

Development and Performance Comparison of Models for Screening Randomly At-Risk Patients with Kidney Disease Using Data Mining Techniques

อนุพงศ์ สุขประเสริฐ[1], อันดี้ เอเว่นส์[2], อริสรา สัตย์ซื่อ[3], อุไรวรรณ ทิจันธุง[4], กาญจนา หินเธาว์[5] และ วราวุฒิ นาคบุญนำ[6*]

อาจารย์ประจำสาขาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม[1,]

นิสิตสาขาวิชาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม[2,3,4]

อาจารย์ประจำสาขาการจัดการ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม[5,6*]

E-Mail: warawut.n@acc.msu.ac.th[*]

(Received: November 19, 2023; Revised: May 23, 2024; Accepted: June 05, 2024)

บทคัดย่อ

โรคไตเป็นปัญหาสาธารณสุขที่สำคัญระดับโลก ส่งผลกระทบต่อคุณภาพชีวิตของผู้ป่วยและก่อให้เกิดภาระค่าใช้จ่ายในการรักษาที่สูง โดยเฉพาะอย่างยิ่งในกลุ่มประเทศรายได้ต่ำและปานกลาง ข้อมูลจากการศึกษาพบว่าความชุกของโรคไตเรื้อรังระยะที่ 3-5 ในประเทศไทยอยู่ที่ร้อยละ 12.4 ซึ่งใกล้เคียงกับค่าเฉลี่ยของโลก แต่มีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง การคัดกรองโรคไตตั้งแต่ระยะเริ่มแรกจึงมีความสำคัญอย่างยิ่งในการชะลอความเสื่อมของไตและลดอัตราการเกิดภาวะแทรกซ้อน งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาตัวแบบสำหรับคัดกรองกลุ่มเสี่ยงโรคไตโดยใช้เทคนิคการทำเหมืองข้อมูล และเปรียบเทียบประสิทธิภาพของอัลกอริทึมการจำแนกประเภทข้อมูล 3 วิธี ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree: C45) ต้นไม้ป่าสุ่ม (Random Forest) และเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor: KNN) โดยใช้ชุดข้อมูลจำนวน 1,500 ระเบียนจากฐานข้อมูลสาธารณะ ซึ่งประกอบด้วยข้อมูลทั้งกลุ่มเสี่ยงและไม่เสี่ยงต่อการเป็นโรคไต ผลการศึกษาพบว่า เทคนิค KNN ที่มีค่า k เท่ากับ 11 ให้ประสิทธิภาพสูงสุด โดยมีค่าความถูกต้องเท่ากับ 89.20% และค่าประสิทธิภาพโดยรวมเท่ากับ 17.35% ซึ่งแสดงให้เห็นถึงศักยภาพในการนำไปประยุกต์ใช้สำหรับคัดกรองกลุ่มเสี่ยงได้อย่างแม่นยำ ผลลัพธ์ที่ได้จากงานวิจัยนี้มีประโยชน์ต่อการพัฒนาระบบสนับสนุนการตัดสินใจทางคลินิก เพื่อช่วยบุคลากรทางการแพทย์ในการระบุกลุ่มเสี่ยงและวางแผนการดูแลรักษาได้อย่างเหมาะสม อันจะนำไปสู่การลดภาระค่าใช้จ่ายในการรักษาและพัฒนาคุณภาพชีวิตของผู้ป่วยโรคไตในระยะยาว

คำสำคัญ: โรคไต, เหมืองข้อมูล, ตัดแบบคัดกรองโรค, เทคนิคเพื่อนบ้านใกล้ที่สุด (K-nearest neighbors: KNN))

ABSTRACT

Kidney disease is a major global public health problem that greatly affects patients' quality of life and imposes high treatment costs, especially in low- and middle-income countries. Studies have shown that the prevalence of chronic kidney disease stages 3-5 in Thailand is 12.4%, which is close to the global average but has been increasing continuously. Early screening for kidney disease is crucial in slowing down kidney deterioration and reducing complication rates. This study

aimed to develop a model for screening individuals at risk of kidney disease using data mining techniques and to compare the performance of three data classification algorithms: Decision Tree, Random Forest, and K-Nearest Neighbors (KNN). A dataset of 1,500 records from a public database, including both at-risk and not-at-risk individuals for kidney disease, was used. The results showed that the KNN technique with a k value of 11 yielded the highest performance, with an accuracy of 89.20% and an overall F-measure of 17.35%, demonstrating its potential for accurate risk group screening. The findings of this research are beneficial for developing clinical decision support systems to assist healthcare professionals in identifying at-risk individuals and planning appropriate care, which can lead to reduced treatment costs and improved long-term quality of life for kidney disease patients.

Keywords: kidney disease, data mining, disease screening model, K-nearest neighbors (KNN)

บทนำ

โรคไตเป็นปัญหาสุขภาพที่สำคัญและมีผลกระทบอย่างกว้างขวางทั่วโลก โดยส่งผลต่อคุณภาพชีวิตของผู้ป่วยอย่างมาก ไตมีบทบาทสำคัญในการควบคุมสมดุลของของเหลวในร่างกาย การกำจัดของเสีย และการรักษาระดับความดันโลหิต ในสหรัฐอเมริกา โรคไตเป็นสาเหตุสำคัญของการเสียชีวิต โดยมีผู้ป่วยโรคไตประมาณ 37 ล้านคน ซึ่งส่วนใหญ่ไม่ได้รับการวินิจฉัยอย่างถูกต้อง นอกจากนี้ ประมาณร้อยละ 40 ของผู้ที่มีการทำงานของไตลดลงอย่างมากแต่ไม่ได้รับการฟอกไต ไม่ทราบว่าตนเองเป็นโรคไตเรื้อรัง ทุก ๆ 24 ชั่วโมงจะมีผู้ป่วยรายใหม่ที่ต้องเริ่มรับการรักษาด้วยการฟอกไต โดยโรคเบาหวานและความดันโลหิตสูงเป็นสาเหตุหลักของโรคไตถึง 3 ใน 4 ของผู้ป่วยรายใหม่ ในปี 2019 ผู้ป่วยโรคไตในสหรัฐอเมริกามีค่าใช้จ่ายในการรักษาสูงถึง 87.2 พันล้านเหรียญ [1] ข้อมูลจาก Global Burden of Disease Study 2019 ซึ่งเผยแพร่ในวารสาร The Lancet ชี้ให้เห็นว่า โรคไตเรื้อรังเป็นหนึ่งในสาเหตุหลักของการสูญเสียปีสุขภาวะ (Disability-Adjusted Life Years: DALYs) และมีแนวโน้มของภาระโรคที่เพิ่มขึ้นอย่างต่อเนื่องในช่วง 30 ปีที่ผ่านมา โดยอัตราการสูญเสีย DALYs จากโรคไตเรื้อรังเพิ่มขึ้นถึงร้อยละ 58.8 ระหว่างปี 2533 ถึงปี 2562 [2] สำหรับสถานการณ์โรคไตในประเทศไทย ข้อมูลจากการศึกษาของสมาคมโรคไตแห่งประเทศไทยพบว่า คนไทยป่วยเป็นโรคไตเรื้อรังประมาณร้อยละ 17.6 หรือราว 8 ล้านคน โดย 80,000 คนเป็นโรคไตระยะสุดท้าย และมีแนวโน้มเพิ่มขึ้นทุกปี [3] และความชุกเฉลี่ยของโรคไตเรื้อรังระยะที่ 3-5 ในภูมิภาคเอเชียอยู่ที่ร้อยละ 11.2 (95% CI; 9.3–13.2%) จากการศึกษาในประชากร 14 ประเทศ รวมถึงไทย พบว่าประเทศไทยมีความชุกของโรคอยู่ที่ร้อยละ 12.4 ซึ่งใกล้เคียงกับค่าเฉลี่ยความชุกของโรคไตเรื้อรังระยะที่ 3-5 ในระดับโลกในปี ค.ศ. 2015 ที่ร้อยละ 10.6 ในขณะที่หลายประเทศในเอเชีย เช่น ฟิลิปปินส์ บังคลาเทศ ศรีลังกา ปากีสถาน อิหร่าน และมองโกเลีย มีความชุกของโรคไตสูงกว่าค่าเฉลี่ยของโลก [4] โรคไตส่งผลกระทบต่อการสร้างฮอร์โมนที่ควบคุมความดันโลหิตและการสร้างเม็ดเลือดแดงในร่างกาย ซึ่งสาเหตุสำคัญของโรคไตคือ โรคเบาหวานและความดันโลหิตสูง ในการวิจัยนี้ ผู้วิจัยได้รวบรวมข้อมูลเกี่ยวกับภาวะโรคไตและผลกระทบต่อสุขภาพในประชากรกลุ่มเสี่ยง จำนวน

1,500 ระเบียน เพื่อนำมาวิเคราะห์และพัฒนาตัวแบบสำหรับคัดกรองและพยากรณ์กลุ่มเสี่ยงต่อการเป็นโรคไต โดยเปรียบเทียบประสิทธิภาพของเทคนิคการทำเหมืองข้อมูลต่าง ๆ ผลลัพธ์ที่ได้จากงานวิจัยนี้ สามารถใช้เป็นเครื่องมือช่วยแพทย์และบุคลากรทางการแพทย์ ในการระบุกลุ่มเสี่ยงและวางแผนการคัดกรอง ติดตาม และรักษาโรคไตได้อย่างมีประสิทธิภาพ อันจะนำไปสู่การลดภาระค่าใช้จ่ายในการรักษา และการพัฒนาคุณภาพชีวิตของผู้ป่วยโรคไตในระยะยาว

ในปัจจุบันความก้าวหน้าของเทคโนโลยีสารสนเทศและปัญญาประดิษฐ์ในปัจจุบัน ทำให้มีการนำเทคนิคการทำเหมืองข้อมูล (data mining) มาประยุกต์ใช้ในการวิเคราะห์ข้อมูลทางการแพทย์ขนาดใหญ่ เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ ซึ่งอาจนำไปสู่การสร้างตัวแบบในการทำนายหรือคัดกรองโรคต่าง ๆ รวมถึงโรคไตได้อย่างมีประสิทธิภาพ [5-7] การนำเทคนิคการทำเหมืองข้อมูลมาใช้ในการวิเคราะห์ข้อมูลทางการแพทย์จึงมีความสำคัญอย่างยิ่ง เพราะจะช่วยให้สามารถค้นพบองค์ความรู้ใหม่ ที่แฝงอยู่ในชุดข้อมูลขนาดใหญ่ ซึ่งอาจเป็นประโยชน์ต่อการพัฒนาตัวแบบในการคัดกรองและพยากรณ์โรคไต รวมถึงโรคเรื้อรังอื่น ได้อย่างแม่นยำและมีประสิทธิภาพมากขึ้น นอกจากนี้ ผลลัพธ์ที่ได้จากการทำเหมืองข้อมูลยังสามารถใช้เป็นข้อมูลสนับสนุนในการวางแผนเชิงนโยบายด้านสาธารณสุข เพื่อกำหนดมาตรการในการป้องกันและควบคุมโรคไตในระดับประชากรได้อย่างเหมาะสมอีกด้วย ด้วยเหตุนี้ ผู้วิจัยจึงเล็งเห็นถึงความสำคัญในการศึกษาและพัฒนาตัวแบบทำนายกลุ่มเสี่ยงโรคไตโดยเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลต่าง ๆ เพื่อนำไปประยุกต์ใช้ประโยชน์ในทางคลินิกต่อไป

1. วัตถุประสงค์การวิจัย

1.1 เพื่อพัฒนาตัวแบบสำหรับคัดกรองกลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไตและกลุ่มที่ไม่มีความเสี่ยงต่อการเป็นโรคไต

1.2 เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมต่าง ๆ ในการพยากรณ์ผู้ป่วยกลุ่มที่สุ่มเสี่ยงเป็นโรคไต

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 การทำเหมืองข้อมูล (Data Mining) การทำเหมืองข้อมูลเป็นกระบวนการค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลขนาดใหญ่ โดยใช้วิธีการทางสถิติ การเรียนรู้ของเครื่อง และเทคนิคทางคณิตศาสตร์ เพื่อสกัดองค์ความรู้ที่เป็นประโยชน์สำหรับการตัดสินใจและแก้ไขปัญหา [8]

2.2 การจำแนกประเภทข้อมูล (Data Classification) การจำแนกประเภทข้อมูลเป็นการสร้างตัวแบบเพื่อจัดกลุ่มข้อมูลโดยใช้คุณลักษณะของข้อมูล ซึ่งมีจุดมุ่งหมายเพื่อทำนายกลุ่มของข้อมูลใหม่ที่ยังไม่เคยพบมาก่อน [9] การจำแนกประเภทที่ถูกต้องและแม่นยำเป็นสิ่งสำคัญในการวิเคราะห์ข้อมูลและการสร้างตัวแบบที่น่าเชื่อถือสำหรับใช้ในการตัดสินใจและการพยากรณ์

2.3 การวัดประสิทธิภาพการจำแนกประเภทข้อมูล (Classification Performance Evaluation) การวัดประสิทธิภาพของตัวแบบการจำแนกประเภทเป็นขั้นตอนสำคัญในการประเมินความถูกต้องและความน่าเชื่อถือของตัวแบบ โดยใช้ตัวชี้วัดต่าง ๆ เช่น ค่าความถูกต้อง (accuracy) ค่าความระลึก (recall) ค่าความแม่นยำ (precision)

และค่า F1 [10] ค่าเหล่านี้ช่วยให้สามารถเปรียบเทียบและเลือกตัวแบบที่เหมาะสมสำหรับแก้ปัญหาการจำแนกประเภทได้

Sanmarchi et al. [11] ได้ทบทวนวรรณกรรมเกี่ยวกับการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (machine learning) ในการวินิจฉัยโรคไตเรื้อรัง (CKD) โดยมีวัตถุประสงค์เพื่อสรุปและวิเคราะห์ระเบียบวิธีวิจัยและผลการศึกษาที่ผ่านมา ผลการทบทวนพบว่า เทคนิคที่ได้รับความนิยมและมีประสิทธิภาพสูงในการจำแนกผู้ป่วย CKD ได้แก่ Decision Tree, Random Forest, Naive Bayes, Support Vector Machine และ k-Nearest Neighbor นอกจากนี้ ยังพบว่าการเลือกใช้คุณลักษณะ (feature selection) ที่เหมาะสมมีความสำคัญต่อความถูกต้องในการวินิจฉัย โดย Random Forest สามารถให้ค่าความถูกต้อง ในการจำแนกได้สูงสุดถึง 99.8% อย่างไรก็ตาม ข้อจำกัดที่พบในงานวิจัยที่ผ่านมา ได้แก่ ขนาดของกลุ่มตัวอย่างที่ค่อนข้างน้อย และการขาดการตรวจสอบความทั่วไปของตัวแบบ (model generalization) บนประชากรที่หลากหลาย การศึกษานี้ให้ข้อมูลที่เป็นประโยชน์เกี่ยวกับระเบียบวิธีและผลการวิจัยของการใช้ machine learning ในการวินิจฉัย CKD ซึ่งสอดคล้องกับผลการทบทวนวรรณกรรมโดยรวมที่พบว่า เทคนิคเหล่านี้มีศักยภาพสูงในการตรวจจับโรคไตได้อย่างแม่นยำ แต่ยังจำเป็นต้องมีการศึกษาเพิ่มเติมในกลุ่มประชากรขนาดใหญ่และหลากหลายมากขึ้น เพื่อพัฒนาตัวแบบที่มีความน่าเชื่อถือและสามารถนำไปใช้งานได้จริงในทางคลินิก

Polat et al. [12] ได้ศึกษาการใช้เทคนิคการทำเหมืองข้อมูลเพื่อวินิจฉัยโรคไตเรื้อรัง โดยเปรียบเทียบประสิทธิภาพของอัลกอริทึม SVM และ ANN (Artificial Neural Network) ผลการศึกษาพบว่า SVM ให้ค่าความถูกต้องที่ 98.5% ในขณะที่ ANN อยู่ที่ 95%

นอกจากนี้ Salekin and Stankovic [13] ได้ทบทวนวรรณกรรมงานวิจัยที่เกี่ยวข้องกับการใช้เทคนิคการทำเหมืองข้อมูลและ machine learning ในการตรวจจับโรคไตในช่วงแรก ผลการทบทวนพบว่า เทคนิคเหล่านี้มีศักยภาพสูงในการวินิจฉัยโรคไตตั้งแต่ระยะเริ่มแรก โดยอัลกอริทึมที่นิยมใช้และมีประสิทธิภาพดี ได้แก่ ANN, Decision Tree, SVM, k-NN, Naive Bayes และ ensemble methods

2.4 ปัจจัยเสี่ยงและสาเหตุของโรคไตเรื้อรัง

โรคไตเรื้อรังเป็นภาวะที่การทำงานของไตเสื่อมลงอย่างค่อยเป็นค่อยไป ซึ่งอาจเกิดจากหลายสาเหตุ โดยปัจจัยเสี่ยงที่สำคัญของโรคไตเรื้อรัง ได้แก่ โรคเบาหวาน ความดันโลหิตสูง โรคหัวใจและหลอดเลือด ภาวะไขมันในเลือดสูง การติดเชื้อทางเดินปัสสาวะ นิ่วในไต โรคไตอักเสบ และพันธุกรรม เป็นต้น [14-15] นอกจากนี้ ภาวะอ้วนและการสูบบุหรี่ก็เป็นปัจจัยเสี่ยงสำคัญที่ทำให้ไตเสื่อมเร็วขึ้น โดยการศึกษาของ Kovesdy et al. [16] พบว่า ผู้ที่มีภาวะอ้วนและมีปัจจัยเสี่ยงอื่นร่วมด้วย เช่น เบาหวานหรือความดันโลหิตสูง จะมีความเสี่ยงต่อการเกิดโรคไตเรื้อรังเพิ่มขึ้นเป็น 7-8 เท่า ส่วนการสูบบุหรี่นั้น Xia et al. [17] พบว่าผู้สูบบุหรี่มีอัตราการลดลงของอัตราการกรองของไต (eGFR) เร็วกว่าผู้ที่ไม่สูบถึงร้อยละ 30 ในระยะเวลา 5 ปี ชี้ให้เห็นว่าการสูบบุหรี่เป็นปัจจัยที่ส่งเสริมให้ไตเสื่อมเร็วขึ้น ดังนั้น การทำความเข้าใจเกี่ยวกับปัจจัยเสี่ยงและสาเหตุของโรคไต จึงมีความสำคัญต่อการพัฒนาตัวแบบเพื่อใช้ในการคัดกรองและพยากรณ์โรคไตเรื้อรัง โดยคุณลักษณะหรือตัวแปรที่จะนำมาใช้ในการสร้างตัวแบบนั้น ควรครอบคลุมปัจจัยเสี่ยงต่าง ๆ ที่กล่าวมาข้างต้น เพื่อให้ได้ตัวแบบที่มีประสิทธิภาพสูงในการทำนายความเสี่ยงของการเกิดโรคไตเรื้อรังได้อย่างแม่นยำ

จากการทบทวนวรรณกรรม แสดงให้เห็นว่าเทคนิคการทำเหมืองข้อมูลและการเรียนรู้ของเครื่องสามารถนำมาประยุกต์ใช้ในการคัดกรองและพยากรณ์ผู้ป่วยโรคไตได้อย่างมีประสิทธิภาพ โดยพบว่าอัลกอริทึมต่าง ๆ มีจุดแข็งที่แตกต่างกันในการจำแนกกลุ่มผู้ป่วยโรคไตจากข้อมูลทางการแพทย์ ซึ่งการเปรียบเทียบประสิทธิภาพของตัวแบบเหล่านี้จะช่วยให้สามารถเลือกใช้เทคนิคที่เหมาะสมกับลักษณะของปัญหาและชุดข้อมูลได้

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

เทคนิคที่ใช้ในการวิเคราะห์ คือ เทคนิคการทำเหมืองข้อมูล (Data Mining) โดยใช้เทคนิคการจำแนกประเภทข้อมูล (Data Classification) ด้วยวิธีการแบ่งการพยากรณ์เปรียบเทียบประสิทธิภาพ Classificationแบ่งออกเป็น 3 เทคนิค ได้แก่ เทคนิค Decision tree (C45), เทคนิค K-Nearest Neighbors, เทคนิค Random Forest โดยใช้โปรแกรม RapidMiner Studio มาใช้สำหรับการเตรียมข้อมูลก่อนการ วิเคราะห์และการพยากรณ์

2. กลุ่มเป้าหมาย

ข้อมูลที่ใช้ในการวิจัยนี้เป็นข้อมูลกลุ่มเสี่ยงเป็นโรคไต ปี 2015 เป็นเวลา 1 ปี จำนวน 1500 ระเบียน จากเว็บไซต์ https://data.world/ [18] ซึ่งเป็นการเก็บข้อมูลกลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไตและกลุ่มที่ไม่มีความเสี่ยงต่อการเป็นโรคไต

3. วิธีดำเนินงานวิจัย

ในการพยากรณ์ผู้ป่วยกลุ่มที่สุ่มเสี่ยงเป็นโรคไต โดยดำเนินการตามกระบวนการ CRISP-DM (Cross-Industry Standard Process for Data Mining: CRISP-DM) [19] ซึ่งประกอบด้วย 6 ขั้นตอนหลัก ได้แก่ 1) การทำความเข้าใจธุรกิจ (Business Understanding) โดยศึกษาและทำความเข้าใจปัญหาเกี่ยวกับการคัดกรองผู้ป่วยกลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไต และกำหนดวัตถุประสงค์ของการวิจัย 2) การทำความเข้าใจข้อมูล (Data Understanding) โดยรวบรวมข้อมูลกลุ่มเสี่ยงเป็นโรคไตจากแหล่งข้อมูลที่น่าเชื่อถือ (https://data.world/) และทำความเข้าใจโครงสร้างและคุณลักษณะของชุดข้อมูล 3) การเตรียมข้อมูล (Data Preparation) โดยคัดเลือกและทำความสะอาดข้อมูลที่เกี่ยวข้องกับการวิจัย และแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์ 4) การสร้างแบบจำลอง (Modeling) โดยใช้เทคนิคการจำแนกประเภทข้อมูล 3 วิธี ได้แก่ ต้นไม้ตัดสินใจ ต้นไม้ป่าสุ่ม และเพื่อนบ้านใกล้ที่สุด และแบ่งข้อมูลเป็นชุดสำหรับสร้างแบบจำลอง (Training set) และชุดทดสอบประสิทธิภาพ (Testing set) ด้วยวิธี 10-fold cross-validation 5) การประเมินผล (Evaluation) โดยประเมินประสิทธิภาพของแบบจำลองที่สร้างขึ้นด้วยเกณฑ์วัดต่าง ๆ เช่น ค่าความถูกต้อง ค่าประสิทธิภาพโดยรวม ค่าความไว และค่าจำเพาะ และเปรียบเทียบประสิทธิภาพของอัลกอริทึมต่าง ๆ เพื่อหาตัวแบบที่เหมาะสมที่สุดสำหรับการคัดกรองโรคไต และ 6) การนำไปใช้งาน (Deployment) โดยนำผลการวิจัยและตัวแบบที่ได้ไปประยุกต์ใช้ในการคัดกรองและพยากรณ์โรคไตในทางปฏิบัติ รวมถึงเผยแพร่ผลการวิจัยเพื่อเป็นแนวทางในการพัฒนาระบบสนับสนุนการตัดสินใจทางการแพทย์ต่อไป การนำเสนอวิธีดำเนินการวิจัยข้างต้นแสดงให้เห็นถึงกรอบแนวคิดและกระบวนการทำงานอย่างเป็น

ระบบ ซึ่งสอดคล้องกับวัตถุประสงค์ของการวิจัยที่ต้องการพัฒนาและเปรียบเทียบประสิทธิภาพของตัวแบบในการคัดกรองผู้ป่วยกลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไต โดยใช้วิธีการเปรียบเทียบประสิทธิภาพกระบวนการดังภาพที่ 1

ภาพที่ 1 กรอบแนวคิดในการทำวิจัย

4. กระบวนการมาตรฐานในการทำเหมืองข้อมูล

วิธีดำเนินงานวิจัยสำหรับคนที่มีความสุ่มเสี่ยงที่จะเป็นโรคไตและกลุ่มที่มีไม่มีความเสี่ยงที่จะเป็นโรคไต โดยจะอ้างอิงข้อมูลแบบ (CRISP-DM) โดยมีทั้งหมด 6 ขั้นตอนประกอบไปด้วย

6.1 การทำความเข้าใจเกี่ยวกับธุรกิจ (Business Understanding) บริบทกลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไตและกลุ่มที่ไม่มีความเสี่ยงต่อการเป็นโรคไต ส่วนสำคัญในการวิเคราะห์ข้อมูลลดกลุ่มที่ไม่มีความเสี่ยงต่อการเป็นโรคไต งานวิจัยศึกษาเพื่อวิเคราะห์ข้อมูลการพยากรณ์ เป็นหรือไม่เป็น

6.2 การทำความเข้าใจเกี่ยวกับข้อมูล (Data Understanding) งานวิจัยนี้ใช้ชุดข้อมูลจริงเกี่ยวกับกลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไต โดยใช้ข้อมูลอ้างอิงจากปี 2015 เป็นข้อมูลชุดจริงที่อ้างอิงมาจากเว็บไซต์ www.data.world.com โดยรวบรวมข้อมูลไว้ในรูปแบบไฟล์ Excel เพื่อนำข้อมูลไปทำการวิเคราะห์ต่อไป

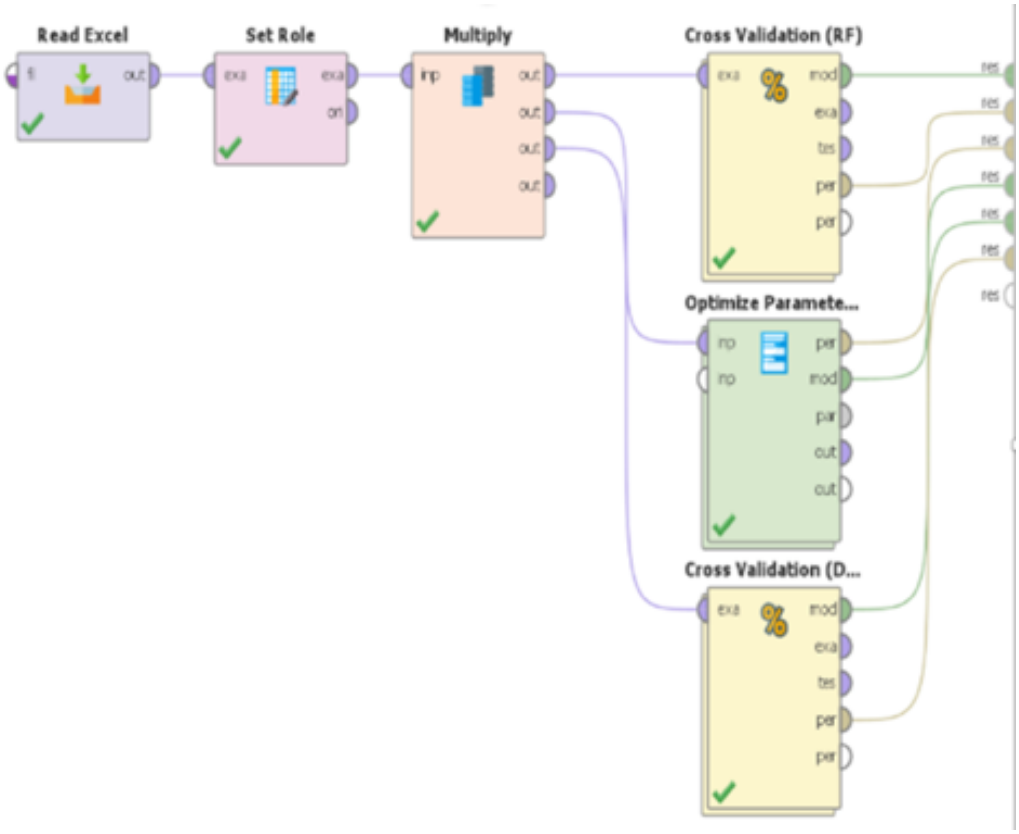
6.3 การเตรียมข้อมูล (Data Preparation) เป็นการเตรียมข้อมูลก่อนไปทำการวิเคราะห์ในงานวิจัยมีทั้งหมด 3 ขั้นตอน

6.3.1 การคัดเลือกข้อมูล (Data Selection) ผู้วิจัยได้ตัดแอตทริบิวต์ที่ไม่สามารถอธิบายค่าข้อมูลอื่น ๆ ได้แก่ แอตทริบิวต์สถานการณ์ดำรงชีวิต (Living Situation) แอตริบิวต์สถานะการจ้างงาน

(Employment Status) และแอตทริบิวต์ปีที่ตรวจ (Survey Year) ซึ่งจะเหลือแอตทริบิวต์ที่ใช้ในการวิเคราะห์ข้อมูล 19 แอตทริบิวต์ จำนวน 1,500 ระเบียน

6.3.2 การแปลงรูปข้อมูล (Data Transformation) งานวิจัยเรื่องโรคไตได้ทำข้อมูลในโปรแกรม Microsoft Excel โดยจัดเก็บในรูปแบบของไฟล์ CSV เพื่อนำข้อมูลต่อไปนี้ไปวิเคราะห์ข้อมูล โดยใช้โปรแกรม RapidMiner Studio ไปวิเคราะห์ข้อมูล

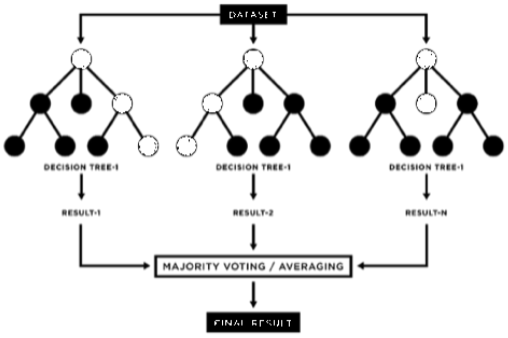
6.4 การสร้างแบบจำลอง (Modeling) ขั้นตอนนี้จะเป็นการสร้างแบบจำลองเพื่อพิจารณาว่ากลุ่มผู้ป่วยที่มีความเสี่ยงและกลุ่มที่ไม่มีความเสี่ยงต่อโรคไต โดยใช้โปรแกรม RapidMiner Studio เพื่อคาดการณ์โมเดลโดยใช้เทคนิค Cross Validation เพื่อสร้างและประเมินประสิทธิภาพของแต่ละโมเดล ดังแสดงในภาพที่ 3



ภาพที่ 2 ขั้นตอนการสร้างตัวแบบและการวัดประสิทธิภาพของตัวแบบด้วยโปรแกรม RapidMiner Studio Version 10

6.4.1 เทคนิคต้นไม้ป่าสุ่ม (Random Forest) เป็นเทคนิคหนึ่งที่ใช้สำหรับการจำแนกประเภทข้อมูล ประกอบด้วยการรวมแผนผังการตัดสินใจหลายแบบเข้าด้วยกัน เพื่อมาใช่วยในการค้นหาคำตอบซึ่งมีประสิทธิภาพมากยิ่งกว่าการใช้ตัวแบบเดียว [20] เพื่อลดความสัมพันธ์นี้ ข้อมูลตัวอย่างและคุณลักษณะจึงได้รับการ

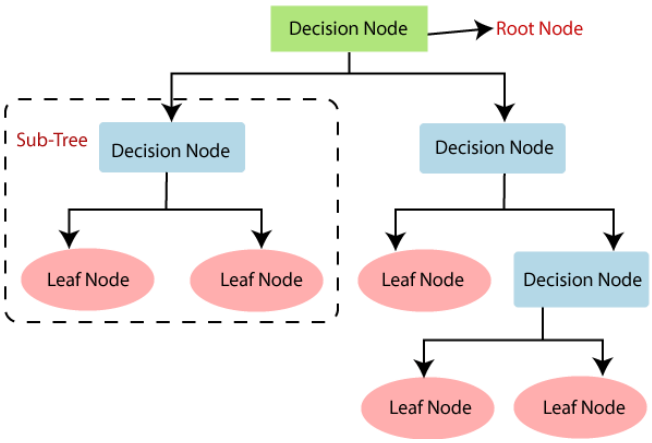
สุ่มเลือก สามารถทนต่อข้อมูลที่เสียหายและซับซ้อนได้ และเผยให้เห็นถึงความสำคัญของคุณลักษณะสำหรับการวิเคราะห์ข้อมูล ความสะดวกในการใช้งานและความยืดหยุ่นทำให้เกิดการยอมรับความสามารถในจัดการกับปัญหาการจำแนกประเภทและการถดถอยได้ [21] ดังแสดงในภาพที่ 4



ภาพที่ 3 Random Forest Algorithm [22]

6.4.2 เทคนิคเพื่อนบ้านใกล้ที่สุด (K-nearest neighbors: KNN) เป็นอัลกอริทึมที่ใช้ในการจัดกลุ่มข้อมูลที่อยู่ในกลุ่มของการเรียนรู้แบบมีผู้สอน (Supervised learning) [23] เป็นเทคนิคที่ไม่ซับซ้อนและเข้าใจง่ายที่สุดในการจัดกลุ่มข้อมูลโดยหลักการทำงานคือการเปรียบเทียบความคล้ายคลึงของข้อมูลที่สนใจกับข้อมูลอื่น ๆ เพื่อหาข้อมูลที่มีความคล้ายคลึงหรือติดกันมากที่สุด k ตัว จากนั้นทำการตัดสินใจเกี่ยวกับคำตอบของข้อมูลที่สนใจโดยให้เป็นคำตอบเดียวกับข้อมูลที่อยู่ใกล้ที่สุด k ตัวนั้น ๆ ซึ่ง k คือความถี่ของข้อมูลที่ติดกันกับข้อมูลที่สนใจในประการที่มีความคล้ายคลึงหรืออยู่ใกล้กันที่สุด

6.4.3 เทคนิคต้นไม้ตัดสินใจ (Decision tree: C45) การใช้ตัวแบบทางคณิตศาสตร์ที่ใช้โครงสร้างต้นไม้เพื่อหาทางเลือกที่ดีที่สุด โดยการสร้างตัวแบบการพยากรณ์จากข้อมูลที่มีการเรียนรู้แบบมีผู้สอน (supervised learning) ซึ่งสามารถใช้สร้างตัวแบบการจัดกลุ่มได้จากกลุ่มตัวอย่างในชุดข้อมูลฝึกหัด (training data set) ได้โดยอัตโนมัติ และสามารถทำนายกลุ่มของรายการที่ยังไม่เคยถูกจัดกลุ่มได้อีกด้วย [24] ดังแสดงในภาพที่ 5



ภาพที่ 4 Decision tree algorithm [25]

6.5 การประเมินผล (Evaluation) ผู้วิจัยได้ทำการแบ่งข้อมูลออกเป็น 2 ส่วน สำหรับการสร้างตัวแบบและการทดสอบประสิทธิภาพของตัวแบบ ผู้วิจัยเลือกใช้ 10-fold cross validation กล่าวคือ แบ่งข้อมูลออกเป็น 10 ส่วนเท่ากัน ใช้ข้อมูล 9 ส่วนในการสร้างตัวแบบ และอีก 1 ส่วนที่เหลือสำหรับทดสอบ ทำเช่นนี้วนไปจนครบ 10 รอบ เพื่อให้ทุกส่วนของข้อมูลถูกใช้ทดสอบตัวแบบ วิธีนี้เป็นที่นิยมเนื่องจากให้ผลประเมินที่มีความน่าเชื่อถือสูง โดยไม่มีความลำเอียงในการเลือกข้อมูลมาสร้างหรือทดสอบ [26] ด้วยเทคนิค 10-fold cross validation และทำการวัดประสิทธิภาพของตัวแบบโดยใช้ค่าประสิทธิภาพสำหรับการจำแนกประเภทข้อมูล (Classification Performance) 4 ค่า ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าประสิทธิภาพโดยรวม (F-measure) ค่าความไว (Sensitivity) และค่าจำเพาะ (Specificity) โดยมีสูตรการคำนวณ ดังสมการ 1-4

$$\text{ความถูกต้อง (Accuracy)} = \frac{(TP + TN)}{(TP+TN+FP+FN)} \quad (1)$$

$$\text{ความไว (Sensitivity)} = \frac{TP}{(TP + FN)} \quad (2)$$

$$\text{ความจำเพาะ (Specificity)} = \frac{TN}{(FP + TN)} \quad (3)$$

$$\text{ค่าประสิทธิภาพโดยรวม (F-Measure)} = \frac{2*Precision*Recall}{Precision + Recall} \quad (4)$$

6.6 การนำไปใช้งาน (Deployment) จากนั้นผู้วิจัยจะทำการเปรียบเทียบประสิทธิภาพของตัวแบบเพื่อหาตัวแบบที่เหมาะสมที่สุดที่จะนำไปสร้างตัวแบบแล้ว ผู้วิจัยจะนำตัวแบบที่ได้นี้ไปใช้สำหรับการคัดกรองผู้ป่วยกลุ่มที่สุ่มเสี่ยงเป็นโรคไต เพื่อช่วยลดภาระให้กับบุคลากรทางการแพทย์ในการค้นหากลุ่มเสี่ยงและยังสามารถช่วยให้ทีมสหวิชาชีพได้วางแผนการรักษากลุ่มเสี่ยงที่จะเป็นโรคต่อไป

ผลการวิจัย

1. ผลการวิเคราะห์

ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของตัวแบบแต่ละตัวที่ใช้สำหรับการคัดกรองผู้ป่วยกลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไต ด้วยค่าความถูกต้อง ค่าประสิทธิภาพโดยรวม และค่าจำเพาะ โดยทำการทดลองรันทั้งหมด 30 รอบ และสำหรับเทคนิคเพื่อนบ้านใกล้ที่สุด ผู้วิจัยได้ทำการหาค่าที่เหมาะสมที่สุดของพารามิเตอร์ k ด้วยวิธีการ optimization ด้วยเทคนิค Evolutionary Optimization ซึ่งผลการวิเคราะห์จะแสดงไว้ในตารางที่ 2 การหาค่าพารามิเตอร์ k สำหรับเทคนิคเพื่อนบ้านใกล้ที่สุด

ตารางที่ 1 ตารางแสดงค่าที่เหมาะสมที่สุดของพารามิเตอร์ k

จำนวนกลุ่ม K	Accuracy	จำนวนกลุ่ม K	Accuracy
1	0.782	9	0.891
3	0.881	11*	0.892
5	0.891	13	0.891
7	0.891	15	0.891

* คือ ค่าพารามิเตอร์ k ที่มีความเหมาะสมที่สุด

จากตารางที่ 2 ค่าพารามิเตอร์ k ที่มีความเหมาะสมที่สุดK-Nearest Neighbors จำนวนกลุ่มที่ 11 คือค่าที่มีค่าความถูกต้อง สูงสุดที่ 0.892 จากนั้นผู้วิจัยได้นำผลการวิเคราะห์ข้อมูลเพื่อหาประสิทธิภาพของการจำแนกประเภทข้อมูลทั้ง 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม และเทคนิคเพื่อนบ้านใกล้ที่สุด มาทำการเปรียบเทียบประสิทธิภาพโดยมีผลการวิเคราะห์ดังตารางที่ 3

ตารางที่ 2 ประสิทธิภาพการจำแนกประเภทข้อมูล

Classification Techniques	Classification Performance			
	Accuracy	F-measure	Sensitivity	Specificity
Decision Tree	88.93%	15.31%	8.37%	99.77%
Random Forest	89.07%	16.33%	8.95%	99.85%
K-Nearest Neighbors*	89.20%	17.35%	9.51%	99.92%

จากตารางที่ 3 พบว่าเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest-Neighbors) ที่มีค่าพารามิเตอร์ k เท่ากับ 11 ให้ค่าความถูกต้องเท่ากับ 89.20% และค่าประสิทธิภาพโดยรวมเท่ากับ 17.35% ซึ่งบ่งชี้ว่าความสามารถในการทำนายมีความถูกต้องสูงที่สุดเมื่อเปรียบเทียบประสิทธิภาพกับเทคนิคอื่น ๆ ดังนั้นเทคนิคเพื่อนบ้านที่ใกล้ที่สุด จึงเป็นเทคนิคที่มีความเหมาะสมสำหรับการนำไปสร้างตัวแบบสำหรับการคัดกรองผู้ป่วยกลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไต

การวิจัยนี้ได้เสนอความรู้ใหม่ในด้านการคัดกรองกลุ่มเสี่ยงที่อาจเป็นโรคไต โดยเฉพาะอย่างยิ่งในการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในทางการแพทย์ ผลลัพธ์ที่ได้จากการวิจัยนี้คือ การค้นพบว่าเทคนิคเพื่อน

บ้านใกล้ที่สุด (K-nearest neighbors: KNN) มีประสิทธิภาพสูงสุดในการจำแนกกลุ่มเสี่ยงของโรคไต เมื่อเปรียบเทียบกับเทคนิคอื่น ๆ เช่น ต้นไม้ป่าสุ่ม (Random Forest) และต้นไม้ตัดสินใจ (Decision Tree) ข้อมูลที่สำคัญที่เกิดขึ้นจากการวิจัยนี้ คือการพบว่า เมื่อใช้ค่า k เท่ากับ 11 สำหรับเทคนิค KNN สามารถบรรลุค่าความถูกต้อง (accuracy) ได้ถึง 89.20% ในการจำแนกผู้ที่มีความเสี่ยงต่อโรคไตจากผู้ที่ไม่มีความเสี่ยง การค้นพบนี้มีความสำคัญในการช่วยแพทย์และบุคลากรทางการแพทย์ในการทำนายกลุ่มเสี่ยงของโรคไตได้ดีขึ้น ซึ่งสามารถนำไปสู่การวางแผนการดูแลรักษาและการป้องกันโรคไตได้อย่างมีประสิทธิภาพมากขึ้น

1. ประสิทธิภาพของเทคนิคที่ใช้ในการคัดกรอง

การศึกษาพบว่าเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors: KNN) มีประสิทธิภาพสูงสุดในการจำแนกกลุ่มผู้ป่วยที่มีความเสี่ยงต่อการเป็นโรคไต โดยให้ค่าความถูกต้องสูงสุดที่ 89.20% เมื่อใช้ค่า k เท่ากับ 11 ในการคัดกรองกลุ่มเสี่ยงที่อาจเป็นโรคไต

2. การเปรียบเทียบกับเทคนิคอื่น

เมื่อเปรียบเทียบกับเทคนิคอื่นที่ใช้ในการวิจัย เช่น ต้นไม้ตัดสินใจ (Decision Tree) และต้นไม้ป่าสุ่ม (Random Forest), KNN แสดงให้เห็นถึงประสิทธิภาพที่สูงกว่าในการจำแนกกลุ่มเสี่ยงของโรคไต

3. การประยุกต์ใช้เพื่อการคัดกรองในทางการแพทย์

ผลลัพธ์จากการวิจัยนี้มีศักยภาพในการช่วยแพทย์และบุคลากรทางการแพทย์ในการคัดกรองกลุ่มเสี่ยงของโรคไตได้อย่างมีประสิทธิภาพ, ซึ่งสามารถนำไปสู่การวางแผนการรักษาและการดูแลผู้ป่วยอย่างเหมาะสม

4. การพัฒนาแนวทางใหม่ในการวิเคราะห์ข้อมูลทางการแพทย์

การศึกษานี้ยังชี้ให้เห็นถึงประโยชน์ของการใช้เทคนิคการทำเหมืองข้อมูลและการวิเคราะห์ข้อมูลขนาดใหญ่ในบริบทของการดูแลสุขภาพ นำมาซึ่งการเข้าใจที่ลึกซึ้งขึ้นเกี่ยวกับการจำแนกประเภทและการคัดกรองโรค

สรุปผล

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อศึกษาการเปรียบเทียบ ประสิทธิภาพเทคนิค การพยากรณ์กลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไตและกลุ่มที่ไม่มีความเสี่ยงต่อการเป็นโรคไต โดยใช้เทคนิคเหมืองข้อมูล Classification เพื่อสร้างตัวแบบพยากรณ์กลุ่มสุ่มเสี่ยงการเป็นโรคไต ซึ่งได้ข้อมูลจาก www.dataworld.com จำนวน 1,500 ระเบียน และ 19 แอตทริบิวต์ และนำข้อมูลมาวิเคราะห์ในแบบ (CRISP-DM) การพยากรณ์เปรียบเทียบประสิทธิภาพ ด้วยเทคนิคการจำแนกประเภทข้อมูลทั้ง 3 ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม และเทคนิคเพื่อนบ้านใกล้ที่สุด จากนั้นนำมาวิเคราะห์เพื่อเปรียบเทียบประสิทธิภาพการจำแนกประเภทข้อมูลทั้ง 4 ค่า ได้แก่ ค่าความถูกต้อง ค่าประสิทธิภาพโดยรวม ค่าความไว และค่าจำเพาะ ผลการวิเคราะห์ข้อมูล พบว่าเทคนิคการพยากรณ์ที่ให้ความถูกต้อง ทำให้สามารถพยากรณ์กลุ่มที่สุ่มเสี่ยงต่อการเป็นโรคไตและกลุ่มที่ไม่มีความเสี่ยงต่อการเป็นโรคไต โดยเทคนิคเพื่อนบ้านใกล้ที่สุดมีค่าความถูกต้องสูงสุด เท่ากับ 89.20% รองลงมาคือเทคนิคต้นไม้ป่าสุ่ม มีค่าเท่ากับ 89.07% และเทคนิคที่ให้ค่าความถูกต้องต่ำสุดสุด คือเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 88.93

ข้อเสนอแนะ

1. งานวิจัยในอนาคตอาจพิจารณานำผลการวิเคราะห์ไปพัฒนาเป็นรายงานแบบ Business Intelligence (BI) โดยใช้เครื่องมือเช่น Power BI เพื่อนำเสนอข้อมูลในรูปแบบที่เข้าใจง่ายและสามารถนำไปใช้ประโยชน์ได้จริงสำหรับกลุ่มผู้ใช้งานที่เฉพาะเจาะจง เช่น ผู้บริหารโรงพยาบาลหรือหน่วยงานด้านการดูแลสุขภาพ อย่างไรก็ตาม ควรดำเนินการภายใต้กรอบการรักษาความปลอดภัยและความเป็นส่วนตัวของข้อมูลผู้ป่วยอย่างเคร่งครัด [27]

2. เพื่อให้ได้ผลการวิจัยที่ครอบคลุมและน่าเชื่อถือมากยิ่งขึ้น งานวิจัยในอนาคตอาจพิจารณาเพิ่มขนาดของกลุ่มตัวอย่างและใช้ข้อมูลที่ทันสมัยมากขึ้น รวมถึงเปรียบเทียบประสิทธิภาพของอัลกอริทึมการทำเหมืองข้อมูลอื่น ๆ เช่น Support Vector Machines (SVM) หรือ Neural Networks เพื่อหาแนวทางที่เหมาะสมที่สุดในการคัดกรองและพยากรณ์โรคไต

3. การนำตัวแบบที่พัฒนาขึ้นไปใช้งานจริงในการคัดกรองและพยากรณ์โรคไต ควรมีการติดตามและประเมินผลอย่างต่อเนื่อง เพื่อปรับปรุงประสิทธิภาพของตัวแบบและให้สอดคล้องกับบริบทและความต้องการของแต่ละพื้นที่หรือหน่วยงาน รวมถึงมีการฝึกอบรมบุคลากรที่เกี่ยวข้องให้สามารถใช้งานและแปลผลตัวแบบได้อย่างถูกต้องและมีประสิทธิภาพ [28]

4. จากผลการวิจัยที่พบว่า เทคนิค K-Nearest Neighbor (KNN) ให้ประสิทธิภาพสูงสุดในการพยากรณ์ผู้ป่วยกลุ่มเสี่ยงโรคไต ผู้วิจัยเสนอแนะให้มีการนำตัวแบบที่ได้ไปพัฒนาเป็นระบบสนับสนุนการตัดสินใจทางคลินิก (clinical decision support system) ในรูปแบบของเว็บแอปพลิเคชันหรือโมบายแอปพลิเคชัน เพื่อช่วยแพทย์และบุคลากรทางการแพทย์ในการคัดกรองและติดตามกลุ่มเสี่ยงได้อย่างมีประสิทธิภาพ ซึ่งจะช่วยให้สามารถวางแผนการดูแลและให้คำแนะนำแก่ผู้ป่วยได้อย่างเหมาะสมและทันท่วงที อันจะช่วยชะลอความเสื่อมของไตและลดอัตราการเกิดภาวะไตวายเรื้อรังระยะสุดท้าย [29]

5. ควรมีการศึกษาเพิ่มเติมเพื่อวิเคราะห์และระบุปัจจัยเสี่ยงที่สำคัญของการเกิดโรคไต โดยอาศัยตัวแบบที่พัฒนาขึ้นจากการวิจัยนี้ ซึ่งอาจใช้เทคนิคการวิเคราะห์ความสำคัญของคุณลักษณะ (feature importance) เช่น ค่า Gini importance จาก Random Forest เป็นต้น ผลลัพธ์ที่ได้จะช่วยให้เกิดความเข้าใจที่ลึกซึ้งยิ่งขึ้นเกี่ยวกับสาเหตุของโรคไต ซึ่งจะเป็นประโยชน์ต่อการวางแผนป้องกันและควบคุมโรคในระดับประชากร รวมถึงการให้ความรู้แก่ประชาชนเพื่อปรับเปลี่ยนพฤติกรรมสุขภาพที่เหมาะสมอีกด้วย

อย่างไรก็ตาม ก่อนการนำตัวแบบไปใช้งานจริง ควรมีการศึกษาเพิ่มเติมเกี่ยวกับความเป็นไปได้ในการพัฒนาและการบูรณาการเข้ากับระบบสารสนเทศของหน่วยงาน รวมถึงการฝึกอบรมผู้ใช้งานให้สามารถใช้ระบบและตีความผลลัพธ์ได้อย่างถูกต้อง นอกจากนี้ ควรมีการติดตามและประเมินประสิทธิผลของการใช้งานระบบในสถานการณ์จริงอย่างสม่ำเสมอ เพื่อให้สามารถปรับปรุงและพัฒนาให้ตอบสนองต่อบริบทและความต้องการของผู้ใช้งานได้ดียิ่งขึ้น [30]

เอกสารอ้างอิง

[1] *Chronic Kidney Disease Initiative.* (n.d.). *The page you were looking for has moved.*
https://www.cdc.gov/kidneydisease/basics.html

[2] Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abedi, A., Abedi, P., Abegaz, K. H., Abolhassani, H., Abosetugn, A. E., Aboyans, V., Abrams, E. M., Abreu, L. G., Abrigo, M. R. M., ... Murray, C. J. L. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet,* 396(10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

[3] โรงพยาบาลพระรามเก้า. (2563, 11 มีนาคม). *โรคไตเรื้อรัง*. Https://Www.Praram9.Com/. https://www.praram9.com/โรคไตเรื้อรัง-2

[4] Suriyong, P., Ruengorn, C., Shayakul, C., Anantachoti, P., & Kanjanarat, P. (2022). Prevalence of chronic kidney disease stages 3–5 in low- and middle-income countries in Asia: A systematic review and meta-analysis. *PLOS ONE,* 17(2), e0264393. https://doi.org/10.1371/JOURNAL.PONE.0264393

[5] Jin, D., Sergeeva, E., Weng, W. H., Chauhan, G., & Szolovits, P. (2021). Explainable Deep Learning in Healthcare: A Methodological Survey from an Attribution View. *WIREs Mechanisms of Disease,* 14(3). https://doi.org/10.1002/wsbm.1548

[6] Kaur, P., Sharma, M., & Mittal, M. (2018). Big Data and Machine Learning Based Secure Healthcare Framework. *Procedia Computer Science,* 132, 1049–1059. https://doi.org/10.1016/J.PROCS.2018.05.020

[7] Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. *International Journal of Bio-Science and Bio-Technology,* 5(5), 241–266. https://doi.org/10.14257/ijbsbt.2013.5.5.25

[8] อนุพงศ์ สุขประเสริฐ. (2563). *คู่มือการทำเหมืองข้อมูลด้วยโปรแกรม RapidMiner Studio* (3rd ed.). สาขาวิชาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม.

[9] Han, J., Kamber, M., & Pei, J. (2012). 8 - Classification: Basic Concepts. In J. Han, M. Kamber, & J. Pei (Eds.), Data Mining (Third Edition) (pp. 327–391). *Morgan Kaufmann*. https://doi.org/https://doi.org/10.1016/B978-0-12-381479-1.00008-3

[10] M, H., & M.N, S. (2015). A Review On Evaluation Metrics For Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process,* 5(2), 01–11. https://doi.org/10.5121/IJDKP.2015.5201

[11] Sanmarchi, F., Fanconi, C., Golinelli, D., Gori, D., Hernandez-Boussard, T., & Capodici, A. (2023). Predict, diagnose, and treat chronic kidney disease with machine learning: a systematic literature review. *Journal of Nephrology,* 36(4), 1101–1117. https://doi.org/10.1007/S40620-023-01573-4

[12] Polat, H., Danaei Mehr, H., & Cetin, A. (2017). Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods. *Journal of Medical Systems,* 41(4). https://doi.org/10.1007/S10916-017-0703-X

[13] Salekin, A., & Stankovic, J. (2016). Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes. Proceedings - 2016 IEEE International Conference on Healthcare Informatics, ICHI 2016, 262–270. https://doi.org/10.1109/ICHI.2016.36

[14] Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic Kidney Disease. *The Lancet*, 389(10075), 1238–1252. https://doi.org/10.1016/S0140-6736(16)32064-5

[15] Luyckx, V. A., Tonelli, M., & Stanifer, J. W. (2018). The global burden of kidney disease and the sustainable development goals. *Bulletin of the World Health Organization,* 96(6), 414-422C. https://doi.org/10.2471/BLT.17.206441

[16] Kovesdy, C. P., Furth, S. L., & Zoccali, C. (2017). Obesity and kidney disease: hidden consequences of the epidemic. *Journal of Nephrology,* 30(1). https://doi.org/10.1007/S40620-017-0377-Y

[17] Xia, J., Wang, L., Ma, Z., Zhong, L., Wang, Y., Gao, Y., He, L., & Su, X. (2017). Cigarette smoking and chronic kidney disease in the general population: a systematic review and meta-analysis of prospective cohort studies. *Nephrology Dialysis Transplantation*, 32(3), 475–487. https://doi.org/10.1093/NDT/GFW452

[18] data.world (n.d.). The Data Catalog Platform. https://data.world/home

[19] Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science,* 181, 526–534. https://doi.org/10.1016/J.PROCS.2021.01.199

[20] ณัฐพล แสนคำ, ธนากร ปุรารัมย์, และทิพวัลย์ แสนคำ. (2560). ระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูล. *วารสารวิชาการ โรงเรียนนายร้อยพระจุลจอมเกล้า*, 15(1), 161–170. https://ph01.tci-thaijo.org/index.php/crma-journal/article/view/243110

[21] IBM. (n.d.). *What is Random Forest?* https://www.ibm.com/topics/random-forest

[22] Spotfire. (n.d.). Demystifying the Random Forest Algorithm for Accurate Predictions. https://www.spotfire.com/glossary/what-is-a-random-forest

[23] GlurGeek.Com. (2562, 14 กรกฎาคม). KNN หรือ K-Nearest Neighbors คืออะไร. https://www.glurgeek.com/Education/Knn/

[24] สุรวัชร ศรีเปารยะ, และสายชล สินสมบูรณ์ทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 839–853. https://li01.tci-thaijo.org/index.php/tstj/article/view/85101

[25] Hafeez, M. A., Rashid, M., Tariq, H., Abideen, Z. U., Alotaibi, S. S., & Sinky, M. H. (2021). Performance Improvement of Decision Tree: A Robust Classifier Using Tabu Search Algorithm. Applied Sciences 2021, Vol. 11, Page 6728, 11(15), 6728. https://doi.org/10.3390/APP11156728

[26] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). Statistical Learning. In G. James, D. Witten, T. Hastie, & R. Tibshirani (Eds.), An Introduction to Statistical Learning: with Applications in R (pp. 15–57). Springer US. https://doi.org/10.1007/978-1-0716-1418-1_2

[27] Strachna, O., & Asan, O. (2021). Systems Thinking Approach to an Artificial Intelligence Reality within Healthcare: From Hype to Value. ISSE 2021 - 7th IEEE International Symposium on Systems Engineering, Proceedings. https://doi.org/10.1109/ISSE51541.2021.9582546

[28] Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. Npj Digital Medicine 2020 3:1, 3(1), 1–10. https://doi.org/10.1038/s41746-020-0221-y

[29] Thilly, N., Chanliau, J., Frimat, L., Combe, C., Merville, P., Chauveau, P., Bataille, P., Azar, R., Laplaud, D., Noël, C., & Kessler, M. (2017). Cost-effectiveness of home telemonitoring in chronic kidney disease patients at different stages by a pragmatic randomized controlled trial (eNephro): rationale and study design. BMC Nephrology, 18(1). https://doi.org/10.1186/S12882-017-0529-2

[30] Connell, A., Montgomery, H., Martin, P., Nightingale, C., Sadeghi-Alavijeh, O., King, D., Karthikesalingam, A., Hughes, C., Back, T., Ayoub, K., Suleyman, M., Jones, G., Cross, J., Stanley, S., Emerson, M., Merrick, C., Rees, G., Laing, C., & Raine, R. (2019). Evaluation of a digitally-enabled care pathway for acute kidney injury

management in hospital emergency admissions. Npj Digital Medicine 2019 2: 1, 2(1) , 1–9. https://doi.org/10.1038/s41746-019-0100-6