

## การเปรียบเทียบประสิทธิภาพการจำแนกภาพถ่ายโรคเบาหวานขึ้นจอประสาทตา ด้วยการสกัดคุณสมบัติเชิงพื้นผิวและการเรียนรู้ของเครื่อง

ชัชวาลย์ ศรีมนตรี<sup>1\*</sup> ธนรัฐ โชติพันธ์<sup>2</sup>

Received : June, 1, 2025

Revised : February 6, 2026

Accepted : March 12, 2026

### บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์หลักเพื่อจำแนกภาพถ่ายดวงตาซึ่งมีภาวะโรคเบาหวานขึ้นจอประสาทตาด้วยวิธีการเรียนรู้ของเครื่องร่วมกับการสกัดคุณลักษณะเชิงพื้นผิวและเปรียบเทียบประสิทธิภาพแบบจำลองการเรียนรู้ของเครื่องเพื่อจำแนกภาพถ่ายดวงตาซึ่งมีภาวะโรคเบาหวานขึ้นจอประสาทตา โดยใช้ข้อมูลภาพถ่ายซึ่งเผยแพร่บนเว็บไซต์ [www.kaggle.com](http://www.kaggle.com) ซึ่งรวบรวมโดย Chaudhari (2020) จำนวนทั้งสิ้น 8,834 ภาพ โดยสกัดคุณลักษณะพื้นผิวของภาพถ่ายจอประสาทตาให้อยู่ในรูปแบบตัวเลขเชิงสถิติและลดจำนวนคุณลักษณะซ้ำซ้อนโดยอาศัยค่าสัมประสิทธิ์สหสัมพันธ์อันดับสเปียร์แมน เพื่อปรับปรุงคุณภาพของข้อมูลซึ่งกำหนดค่าค่าสัมประสิทธิ์สหสัมพันธ์ที่ 0.90 และลดจำนวนคุณลักษณะที่ไม่มีประโยชน์ต่อแบบจำลองด้วยอาโนวาเอฟเทสต์ โดยเลือกคุณสมบัติที่มีค่าเอฟสเกอร์สูงสุด 100 ลำดับแรกซึ่งมีความสัมพันธ์กับผลลัพธ์ หลังจากนั้นใช้เทคนิคการหาความสำคัญด้วยแรนดอมฟอเรสต์ เพื่อคัดเลือกคุณลักษณะเชิงสถิติที่มีค่าความสำคัญที่ 0.01 ขึ้นไป ซึ่งถือว่าเป็นคุณลักษณะที่มีนัยสำคัญต่อการสร้างแบบจำลอง จากกระบวนการทั้งหมดทำให้ได้คุณลักษณะทั้งสิ้น 34 คุณลักษณะจาก 316 คุณลักษณะ และนำคุณลักษณะที่ผ่านการคัดเลือกดังกล่าวไปใช้สร้างแบบจำลอง 5 รูปแบบได้แก่ แรนดอมฟอเรสต์ เอ็กตรึมการเดียนบูสต์ ซัพพอร์ตเวกเตอร์แมชชีน นิวรอลเน็ตเวิร์คและเคเนียร์เรสต์เนเบอร์ เพื่อเปรียบเทียบประสิทธิภาพแบบจำลอง ทั้งนี้ผลการวิจัยพบว่าแบบจำลองเอ็กตรึมการเดียนบูสต์มีความแม่นยำสูงสุดที่ 99.34% ค่าประสิทธิภาพโดยรวม 99.35% ค่าความไว 99.98% และค่าจำเพาะ 98.71% แสดงให้เห็นถึงประสิทธิภาพการจำแนกที่สูงและสามารถอธิบายผลได้ง่ายเมื่อเทียบกับวิธีการเรียนรู้เชิงลึกสมัยใหม่ ใช้ชุดข้อมูลที่มีขนาดเล็กกว่าแต่ยังคงประสิทธิภาพการจำแนกที่ใกล้เคียงกัน มีความเหมาะสมในการประยุกต์ใช้ในสภาพแวดล้อมทางคลินิกที่มีทรัพยากรจำกัด

**คำสำคัญ:** การคัดเลือกคุณลักษณะ การประเมินประสิทธิภาพ การเรียนรู้ของเครื่อง  
โรคเบาหวานขึ้นจอประสาทตา

<sup>1</sup> หลักสูตรวิทยาการคอมพิวเตอร์ คณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี  
อีเมล: [chatchawam.s@ubru.ac.th](mailto:chatchawam.s@ubru.ac.th)

<sup>2</sup> หลักสูตรวิทยาการคอมพิวเตอร์ คณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี  
อีเมล: [tanarat.c@ubru.ac.th](mailto:tanarat.c@ubru.ac.th)

\* ผู้นิพนธ์หลัก อีเมล: [chatchawam.s@ubru.ac.th](mailto:chatchawam.s@ubru.ac.th)

## PERFORMANCE COMPARISON OF DIABETIC RETINOPATHY CLASSIFICATION FROM RETINAL IMAGES USING TEXTURE FEATURES AND MACHINE LEARNING

Chatchawarn Srimontree<sup>1\*</sup> Tanarat Chotiphan<sup>2</sup>

### Abstract

This research aims to classify diabetic retinopathy in fundus images using machine learning combined with texture feature extraction and to compare the performance of various machine learning models for diabetic retinopathy classification. The study employed a dataset of 8,834 images published on [www.kaggle.com](http://www.kaggle.com), collected by Chaudhari (2020). According to the study, texture features were extracted from fundus images and converted into numerical statistical features. Redundant features were reduced using Spearman's rank correlation coefficient with a correlation threshold at 0.90 to improve data quality. The number of irrelevant features were further reduced using the ANOVA F-test by selecting the top 100 features with the highest F-scores showing most strongly correlated with the classification outcomes. Subsequently, feature importance analysis using Random Forest were applied to select statistical features with importance values  $\geq 0.01$ , which were considered significant features for model development. This entire process resulted in 34 features selected from the original 316 features. To compare model performance, five machine learning models were trained using these selected features: Random Forest, eXtreme Gradient Boosting, Support Vector Machine, Neural Network, and K-Nearest Neighbors. The results demonstrate that the Extreme Gradient Boosting model achieved the highest performance with 99.34% accuracy, 99.35% overall performance, 99.98% sensitivity, and 98.71% specificity. This approach exhibited superior classification performance with enhanced interpretability compared to modern deep learning methods. Despite utilizing a smaller dataset than typically required by deep learning approaches, it maintained competitive classification performance while being particularly suitable for deployment in resource-constrained clinical environments.

**Keywords:** Feature Selection, Model Performance Evaluation, Machine Learning, Diabetic Retinopathy

---

<sup>1</sup> Computer Science Program, Faculty of Computer Science, Ubon Ratchathani Rajabhat University, e-mail: [chatchawarn.s@ubru.ac.th](mailto:chatchawarn.s@ubru.ac.th)

<sup>2</sup> Computer Science Program, Faculty of Computer Science, Ubon Ratchathani Rajabhat University, e-mail: [tanarat.c@ubru.ac.th](mailto:tanarat.c@ubru.ac.th)

\* Corresponding author, e-mail: [chatchawarn.s@ubru.ac.th](mailto:chatchawarn.s@ubru.ac.th)

## บทนำ

ภาวะโรคเบาหวานขึ้นจอประสาทตา (Diabetic Retinopathy) เป็นภาวะแทรกซ้อนที่สำคัญของโรคเบาหวาน หากไม่ได้รับการวินิจฉัยและรักษาอย่างทันท่วงที อาจนำไปสู่การสูญเสียการมองเห็นและเป็นสาเหตุหนึ่งของอาการตาบอด ภาวะนี้ส่งผลกระทบต่อคุณภาพชีวิตของผู้ป่วยและบุคคลรอบข้าง นอกจากนี้ขั้นตอนการรักษาที่มีความซับซ้อนและปฏิบัติได้ยาก เช่น การหยอดยาและการยิงเลเซอร์ องค์การอนามัยโลกคาดว่าจะมีผู้ป่วยโรคเบาหวานเพิ่มขึ้นถึง 360 ล้านคนในปี พ.ศ. 2573 และผู้ป่วยทุกรายมีความเสี่ยงต่อการเกิดโรคเบาหวานขึ้นจอประสาทตาเกือบทั้งหมด อย่างไรก็ตามกระบวนการคัดกรองผู้ป่วยที่ดีเพียงพอจะสามารถลดอัตราการสูญเสียการมองเห็นได้ถึงร้อยละ 90 (อัจนรา แก้วน้ำเชื้อ, 2564) จากการสำรวจข้อมูลผู้ป่วยเบาหวานโดยโรงพยาบาลศิริราช จำนวน 1,299 ราย มีการตรวจจอประสาทตาด้วยเครื่องตรวจจักษุแบบใช้ลำแสงจำนวน 1,130 ราย คิดเป็นร้อยละ 86.99 ผลการสำรวจพบว่าความชุกของโรคเบาหวานขึ้นจอประสาทตาอยู่ที่ 393 ราย คิดเป็นร้อยละ 34.78 แสดงให้เห็นว่ามีอัตราการเกิดโรคในระดับสูงหลังจากมีภาวะโรคเบาหวาน เมื่อพิจารณาตามช่วงอายุพบว่า ความชุกของผู้ป่วยสูงที่สุดในช่วงอายุ 50–60 ปี โดยมีแนวโน้มเพิ่มขึ้นตามอายุในช่วงอายุก่อน 60 ปี นอกจากนี้ระยะเวลาการเป็นเบาหวานที่ยาวนานขึ้นสัมพันธ์กับความชุกของโรคที่เพิ่มขึ้นอย่างมีนัยสำคัญ ผู้ที่เป็นเบาหวานนานกว่า 20 ปีมีความชุกสูงที่สุดที่ร้อยละ 52.69 อย่างไรก็ตาม ภาวะโรคนี้สามารถตรวจพบได้ตั้งแต่ช่วง 5 ปีแรกของการวินิจฉัยพบโรคเบาหวานในผู้ป่วยทุกประเภท (Boonsaen et al., 2021)

จากข้อมูลดังกล่าวพบว่ากระบวนการคัดกรองผู้ป่วยโรคเบาหวานขึ้นจอประสาทตาที่ดีเพียงพอสามารถลดการเกิดโรคและเป็นการป้องกันในเชิงรุก นอกจากกระบวนการวินิจฉัยด้วยแพทย์ผู้ชำนาญการแล้วการจำแนกภาพถ่ายทางการแพทย์ด้วยวิธีการทางคอมพิวเตอร์ยังเป็นทางเลือกที่ได้รับการยอมรับ โดยเฉพาะการใช้เทคนิคการเรียนรู้ของเครื่องและปัญญาประดิษฐ์ในรูปแบบต่างๆ จากการทบทวนงานวิจัยที่เกี่ยวข้องพบว่ามีการนำเทคนิคการเรียนรู้ของเครื่องมาใช้ในการจำแนกภาพถ่ายทางการแพทย์หลากหลายรูปแบบ โดย Kalyani และคณะ (2024) ศึกษาพบว่าคุณสมบัติเชิงพื้นผิวที่หลากหลายสามารถให้ผลการจำแนกรูปภาพที่มีประสิทธิภาพ Kumar (2022) ใช้คุณลักษณะเชิงพื้นผิวร่วมกับแบบจำลองแรนดอมฟอเรสต์ในการจำแนกเนื้องอกสมองได้ความแม่นยำร้อยละ 83.3 คุณานนต์ อ้าทอง และคณะ (2563) ใช้เมตริกซ์ระดับสีเทาร่วมกับซัพพอร์ตเวกเตอร์แมชชีนจำแนกภาพถ่ายทางการแพทย์ได้ความแม่นยำร้อยละ 92.56 (Hassan et al., 2022) ใช้เอ็กตรัมการเคียนบูสต์ร่วมกับคุณลักษณะเชิงลึกจำแนกโรคเบาหวานขึ้นจอประสาทตาได้อัตราความแม่นยำสูง งามลชนก สีพาดิง และคณะ (2566) เปรียบเทียบแบบจำลองการเรียนรู้เชิงลึก นิวรอลเน็ตเวิร์ค และแรนดอมฟอเรสต์ในการจำแนกเนื้องอกสมองจากภาพเอ็มอาร์ไอ พบว่านิวรอลเน็ตเวิร์คให้ความแม่นยำสูงสุดที่ร้อยละ 72.81 อนุพงศ์ สุขประเสริฐ และคณะ (2567) เปรียบเทียบแบบจำลอง 4 รูปแบบในการจำแนกนิวไนด์ด้วยภาพซีทีสแกน พบว่าเคเนียร์สเตเนเบอร์ให้ความแม่นยำสูงสุดที่ร้อยละ 92.64

จากการศึกษาข้อดีข้อเสียของงานวิจัยดังกล่าวส่งผลให้งานวิจัยนี้มุ่งเน้นพัฒนากระบวนการจำแนกภาพถ่ายโรคเบาหวานขึ้นจอประสาทตาโดยอาศัยแบบจำลองการเรียนรู้ของเครื่องร่วมกับการสกัดคุณลักษณะเชิงพื้นผิว กระบวนการเริ่มต้นด้วยการสกัดคุณลักษณะจากภาพถ่ายรวมทั้งสิ้น 316 คุณลักษณะ จากนั้นทำการลดคุณลักษณะซ้ำซ้อนด้วยค่าสหสัมพันธ์อันดับสเปียร์แมนที่ 0.90 และคัดเลือกคุณสมบัติที่มีความสัมพันธ์กับผลลัพธ์ด้วยอาโนวาเอฟเทสต์โดยเลือกคุณสมบัติที่มีค่าเอฟสกอ์สูงสุด 100 ลำดับแรก โดยขั้นตอนต่อมาคือ

การคัดเลือกคุณลักษณะที่มีผลต่อการจำแนกภาพถ่ายด้วยเทคนิคการหาค่าความสำคัญด้วยแรนดอมฟอเรสต์ โดยพิจารณาความสำคัญที่มีค่าไม่น้อยกว่า 0.01 เป็นเกณฑ์ในการคัดเลือก ส่งผลให้ได้ชุดคุณลักษณะที่มีนัยสำคัญจำนวน 34 คุณลักษณะ คุณลักษณะเหล่านี้ถูกนำไปสร้างแบบจำลองเพื่อการจำแนกภาพถ่าย โดยใช้แบบจำลองการเรียนรู้ของเครื่อง 5 แบบจำลอง ได้แก่ แรนดอมฟอเรสต์ ซัพพอร์ตเวกเตอร์แมชชีน นิวรอลเน็ตเวิร์ค เคเนียร์เรสต์เนบอร์ และเอ็กตรึมการเคียนบูสต์ ประสิทธิภาพของแบบจำลองทั้งหมดได้รับการประเมินโดยใช้ตัวชี้วัดมาตรฐาน ได้แก่ ความแม่นยำ ความไว ความจำเพาะ และค่าคะแนนประสิทธิภาพโดยรวม

ผลการศึกษาชี้ให้เห็นว่าคุณลักษณะภาพเชิงพื้นผิวซึ่งผ่านการคัดเลือกคุณสมบัติสามารถใช้จำแนกภาพถ่ายทางการแพทย์ได้อย่างมีประสิทธิภาพ เมื่อเปรียบเทียบกับวิธีการเรียนรู้เชิงลึกที่นิยมในปัจจุบัน แบบจำลองที่มีประสิทธิภาพที่สุดในงานวิจัยนี้ให้ผลลัพธ์ที่ใกล้เคียงกัน แต่สามารถฝึกสอนแบบจำลองได้อย่างมีประสิทธิภาพแม้มีชุดข้อมูลจำกัด มีความสามารถในการอธิบายผลและตีความคุณลักษณะที่มีผลต่อการจำแนก ซึ่งช่วยในการตรวจสอบผลทางคลินิก ลดความต้องการทรัพยากรเนื่องจากความซับซ้อนในการคำนวณที่ต่ำ รวมถึงเพิ่มความเป็นไปได้ในการนำไปประยุกต์ใช้ในสถานพยาบาลที่มีทรัพยากรจำกัด

### วัตถุประสงค์ของการวิจัย

1. จำแนกภาพถ่ายดวงตาซึ่งมีภาวะโรคเบาหวานขึ้นจอประสาทตาด้วยวิธีการเรียนรู้ของเครื่องร่วมกับการสกัดคุณลักษณะเชิงพื้นผิว
2. เปรียบเทียบประสิทธิภาพแบบจำลองการเรียนรู้ของเครื่องเพื่อจำแนกภาพถ่ายดวงตาซึ่งมีภาวะโรคเบาหวานขึ้นจอประสาทตา

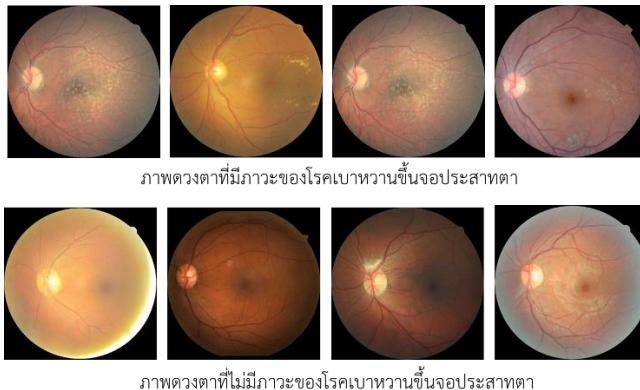
### วิธีดำเนินการวิจัย

คณะผู้วิจัยได้ดำเนินการวิจัยตามมาตรฐานการทำเหมืองข้อมูลด้วยกระบวนการ CRISP-DM ( Cross-Industry Standard Process for Data Mining ) เพื่อจำแนกภาพถ่ายโรคเบาหวานขึ้นจอประสาทตา พร้อมกับเปรียบเทียบประสิทธิภาพของแบบจำลอง เช่นเดียวกับงานวิจัยโดย Garajam ซึ่งใช้แนวทางการจำแนกโรคทางตาโดยใช้กระบวนการเรียนรู้ของเครื่อง (Gajaram, 2022) ซึ่งประกอบด้วยกระบวนการย่อย 6 ขั้นตอน คือการทำความเข้าใจปัญหา การทำความเข้าใจข้อมูล การเตรียมข้อมูล การสร้างแบบจำลอง การประเมินผล การนำไปใช้งาน

**1. การทำความเข้าใจปัญหา** คณะผู้วิจัยมีความเห็นว่าจำนวนผู้ป่วยโรคเบาหวานมีจำนวนเพิ่มมากขึ้น ซึ่งผู้ป่วยทั้งหมดมีความเสี่ยงต่อการเกิดโรคเบาหวานขึ้นจอประสาทตาและอาจนำไปสู่การสูญเสียการมองเห็น หากแต่การคัดกรองผู้ป่วยที่ดีเพียงพอสามารถลดอัตราการสูญเสียการมองเห็นได้ในอัตราที่สูง และโรคเบาหวานขึ้นจอประสาทตาสามารถตรวจพบได้ตั้งแต่ 5 ปีแรกของการวินิจฉัยพบโรคเบาหวาน คณะผู้วิจัยจึงมีความต้องการจำแนกภาพถ่ายโรคเบาหวานขึ้นจอประสาทตาโดยอาศัยคุณลักษณะพื้นผิวจากข้อมูลภาพถ่าย โดยคัดเลือกข้อมูลที่มีนัยสำคัญต่อการจำแนกภาพ เพื่อนำมาผ่านกระบวนการเรียนรู้ของเครื่องจำนวน 5 แบบจำลองพร้อมกับเปรียบเทียบประสิทธิภาพแบบจำลอง เพื่อคัดเลือกแบบจำลองที่ดีที่สุดในการจำแนกภาพถ่ายโรคเบาหวานขึ้นจอประสาทตา

**2. การทำความเข้าใจข้อมูล** ในการศึกษาและวิจัยครั้งนี้ใช้ข้อมูลภาพถ่ายจอประสาทตาซึ่งเผยแพร่บนเว็บไซต์ [www.kaggle.com](http://www.kaggle.com) จำนวน 8,834 ภาพ เป็นชุดข้อมูลรูปภาพจริงซึ่งรวบรวมโดย Chaudhari (Chaudhari, 2020) ซึ่งแบ่งเป็นข้อมูลภาพถ่ายซึ่งมีภาวะโรคเบาหวานขึ้นจอประสาทตา 4,417 ภาพและ

ภาพจอประสาทตาปกติจำนวน 4,417 ภาพ โดยภาพถ่ายแต่ละภาพมีขนาดรูปภาพ 512 x 512 พิกเซลและมีความละเอียดภาพ 96 ดิพีไอ ดังแสดงในภาพที่ 1



ภาพที่ 1 แสดงภาพจอประสาทตาที่มีและไม่มีสภาวะของโรค

### 3. การเตรียมข้อมูล ประกอบด้วยกระบวนการย่อย 3 กระบวนการดังต่อไปนี้

3.1 การรวบรวมคุณลักษณะ งานวิจัยนี้ได้รวบรวมคุณลักษณะของภาพถ่ายที่เป็นมาตรฐานและนิยมใช้ในการจำแนกภาพถ่ายโรคเบาหวานขึ้นจอประสาทตา ซึ่งประกอบด้วยคุณลักษณะพื้นผิวจำนวนทั้งสิ้น 316 คุณลักษณะ โดยใช้คุณลักษณะ 2 ประเภทได้แก่

3.1.1 Local Binary Pattern (LBP) เป็นเทคนิคที่ใช้ในการสกัดคุณลักษณะเชิงพื้นผิวจากภาพถ่าย โดยมีหลักการทำงานที่เรียบง่ายแต่มีประสิทธิภาพสูง วิธีการของ LBP คือการเลือกพื้นที่ย่อยของภาพขนาด 3x3 พิกเซล แล้วใช้ค่าความเข้มของพิกเซลตรงกลางเป็นจุดอ้างอิง จากนั้นทำการเปรียบเทียบค่าความเข้มของพิกเซลโดยรอบกับค่าดังกล่าว หากค่าความสว่างของพิกเซลรอบข้างมากกว่าหรือเท่ากับค่าศูนย์กลางจะถูกกำหนดเป็นค่า 1 และหากน้อยกว่าจะถูกกำหนดเป็นค่า 0 ส่งผลให้ได้ลำดับของเลขฐานสองจำนวน 8 บิต ซึ่งสามารถแปลงเป็นตัวเลขฐานสิบ และนำไปสร้างเป็นฮิสโตแกรมเพื่อแทนลักษณะพื้นผิวของบริเวณนั้น เทคนิค LBP นี้มีข้อดีคือคำนวณได้รวดเร็ว และสามารถแสดงความแตกต่างของพื้นผิวในภาพได้อย่างมีประสิทธิภาพแม้ในสภาวะแสงที่แตกต่างกัน (มาลีรัตน์ โสตานิล และอมรศักดิ์ อมรณานันท์, 2555)

3.1.2 GLCM (Gray-Level Co-occurrence Matrix) เป็นเทคนิคการสกัดคุณลักษณะเชิงพื้นผิวซึ่งอาศัยหลักการวิเคราะห์ความสัมพันธ์ด้วยค่าความเข้มของพิกเซลที่อยู่ใกล้กันภายในภาพ โดยกำหนดระยะห่างและทิศทางของพิกเซลทั้งสองตำแหน่งที่ใช้เปรียบเทียบ เช่น แนวนอน แนวตั้ง หรือแนวทแยง เทคนิคนี้จะทำการนับจำนวนครั้งที่ค่าความเข้มระดับต่าง ๆ ของพิกเซลทั้งสองปรากฏร่วมกันในแต่ละตำแหน่งทั่วทั้งภาพ และจัดเก็บข้อมูลเหล่านั้นในรูปแบบของเมทริกซ์ เพื่อใช้แทนลักษณะพื้นผิวของภาพ (อรรถกร มากคงแก้ว, 2557) จากเมทริกซ์ GLCM ที่สกัดได้ สามารถนำมาคำนวณเป็นคุณลักษณะเชิงสถิติต่าง ๆ เช่น วัดความแตกต่างของค่าความเข้ม วัดความคล้ายคลึงของค่าความเข้มที่อยู่ใกล้กัน วัดความสม่ำเสมอของพื้นผิว วัดความสัมพันธ์เชิงเส้นของค่าความเข้มระหว่างพิกเซล GLCM เป็นเทคนิคที่สามารถอธิบายลักษณะพื้นผิวของภาพได้อย่างมีประสิทธิภาพ และมักถูกนำไปประยุกต์ใช้ในงานด้านการวินิจฉัยทางการแพทย์ การแยกประเภทวัตถุ และการประมวลผลภาพดิจิทัลในลักษณะอื่น ๆ

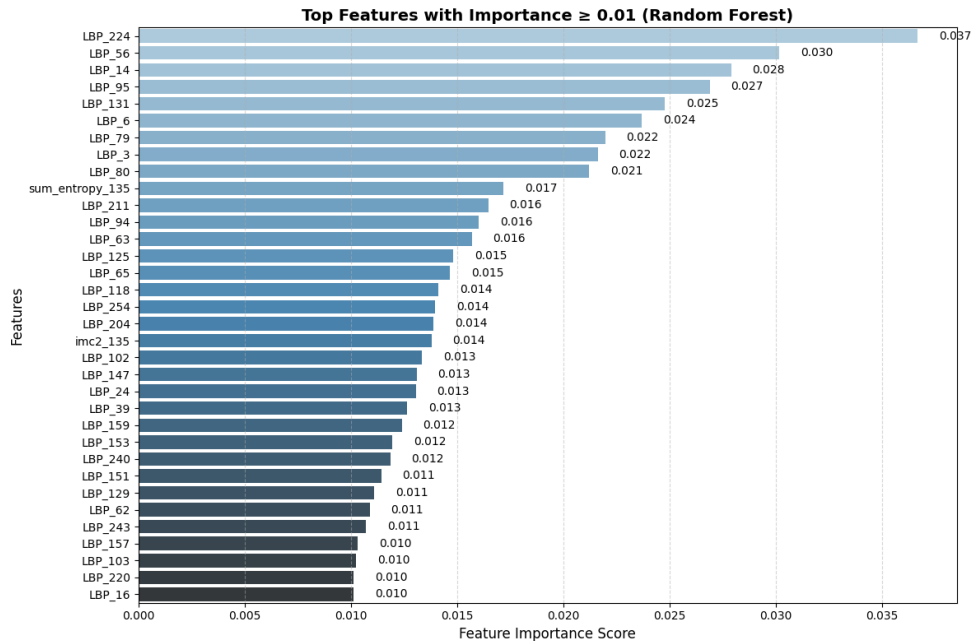
3.2 สกัดคุณลักษณะ กระบวนการสกัดคุณลักษณะจากภาพถ่ายจอประสาทตา กระบวนการใช้งานแบบจำลองและประเมินประสิทธิภาพแบบจำลอง ถูกพัฒนาขึ้นด้วยชุดคำสั่งภาษาไพธอนโดยอาศัยไลบรารีโอเพนซีวีสำหรับภาษาไพธอน เวอร์ชัน 4.11.0.86 (OpenCV Contributors, 2025), ซายคิโทมเมจ เวอร์ชัน 0.25.2 (Scikit-image contributors, 2024), ซายไพ เวอร์ชัน 1.15.3 (SciPy Community, 2024), ซีบอร์น (Waskom, 2021), คีราส เวอร์ชัน 2.15.0 (Chollet et al., 2015) และแพนดาส เวอร์ชัน 2.2.3 (Pandas development team, 2024) ในกระบวนการสกัดคุณลักษณะผู้วิจัยได้เพิ่มกระบวนการปรับคุณภาพของภาพถ่ายด้วยวิธีการปรับสมดุลฮิสโทแกรมแบบจำกัดคอนทราสต์ (Contrast Limited Adaptive Histogram Equalization หรือ CLAHE) เพื่อให้ภาพมีความคมชัดและสามารถสกัดค่าคุณลักษณะเชิงตัวเลขและและสถิติที่มีความแม่นยำ ซึ่งสอดคล้องกับงานวิจัยของกฤษฎิ์ ชินสาร และคณะ (2565) ซึ่งค้นคว้าวิจัยเกี่ยวกับการคัดกรองภาพจอประสาทตาด้วยวิธีการเรียนรู้ของเครื่อง เพื่อหาอาการเส้นเลือดฝอยโป่งพองและค้นหาจุดที่มีเลือดออกบนดวงตาในผู้ป่วยที่มีภาวะโรคเบาหวาน

3.3 คัดเลือกคุณสมบัติที่มีนัยสำคัญ ประกอบด้วยกระบวนการย่อย 3 กระบวนการตามลำดับคือ

3.3.1 การลดคุณลักษณะซ้ำซ้อนด้วยสัมประสิทธิ์สหสัมพันธ์เชิงอันดับของสเปียร์แมน (The Spearman's Rank Correlation Coefficient) เพื่อระบุคุณลักษณะที่มีค่าความสัมพันธ์สูงเกินกว่าค่าที่กำหนดที่ 0.90 ซึ่งบ่งชี้ถึงความซ้ำซ้อนของข้อมูล จากนั้นดำเนินการตัดคุณลักษณะหนึ่งในแต่ละคู่ออกไปเพื่อให้เหลือเฉพาะคุณลักษณะที่ไม่ทับซ้อนกันมากเกินไป กระบวนการนี้ช่วยลดความซับซ้อนของข้อมูลและเพิ่มประสิทธิภาพในการเรียนรู้ของแบบจำลอง สอดคล้องกับงานวิจัยของ Jiang และคณะ (2024) ซึ่งใช้ค่าสัมประสิทธิ์สหสัมพันธ์เชิงอันดับของสเปียร์แมนในการคัดเลือกคุณสมบัติทางคณิตศาสตร์ที่ซับซ้อน ทั้งนี้ผลจากกระบวนการที่ใช้สัมประสิทธิ์สหสัมพันธ์เชิงอันดับของสเปียร์แมนในวิจัยครั้งนี้สามารถลดจำนวนคุณลักษณะลง 50.63% ซึ่งจากเดิมมีคุณลักษณะมากถึง 316 รายการ เหลือเพียง 156 รายการและสามารถคงไว้ซึ่งข้อมูลที่มีความหลากหลายและมีประโยชน์ต่อการจำแนกประเภท

3.3.2 คัดเลือกคุณสมบัติด้วยวิธีการซีเล็คเคเบสต์ (Select K Best) เพื่อลดจำนวนคุณลักษณะที่ไม่มีประโยชน์ต่อแบบจำลองด้วยอาโนวาเอฟเทสต์ โดยเลือกคุณสมบัติที่มีค่าเอฟสกอ์สูงสุด 100 ลำดับแรก ซึ่งมีความสัมพันธ์กับการจำแนกผลลัพธ์ เช่นเดียวกับงานวิจัยของ Ayyanar และคณะ (2022) ซึ่งประยุกต์ใช้วิธีการซีเล็คเคเบสต์ร่วมกับการเรียนรู้ของเครื่องแบบรวมกลุ่มเพื่อวิเคราะห์ข้อมูลทางการแพทย์

3.3.3 การหาค่าความสำคัญด้วยแรนดอมฟอเรสต์ เป็นเทคนิคที่ใช้ลดจำนวนคุณลักษณะที่ไม่จำเป็นออกจากชุดข้อมูลโดยอาศัยหลักการประเมินความสำคัญของแต่ละคุณลักษณะ ซึ่งเป็นเทคนิคการรวมต้นไม้ตัดสินใจหลายต้น สามารถนำมาประยุกต์ใช้จัดอันดับและคัดเลือกเฉพาะคุณลักษณะที่มีบทบาทสำคัญต่อการจำแนกข้อมูล เพื่อนำไปใช้ในการสร้างแบบจำลองที่มีความแม่นยำมากขึ้นและมีความซับซ้อนน้อยลง ช่วยเพิ่มความสามารถในการตีความผลลัพธ์ของแบบจำลองได้อย่างชัดเจนยิ่งขึ้น โดยงานวิจัยนี้ได้ทำการคัดเลือกเฉพาะคุณลักษณะที่มีค่าความสำคัญตั้งแต่ 0.01 ขึ้นไป (Feature Importance  $\geq 0.01$ ) เนื่องจากมีจำนวนคุณลักษณะที่นำมาคัดเลือกจำนวนมาก สอดคล้องกับงานวิจัยของ Wang (Wang, 2023) ซึ่งพบว่า การลดจำนวนคุณสมบัติที่ไม่จำเป็นสามารถเพิ่มประสิทธิภาพให้กับแบบจำลองในการจำแนกประเภทข้อมูลแม้จะเป็นข้อมูลขนาดใหญ่หลังจากเสร็จสิ้นกระบวนการคัดเลือกคุณลักษณะ ทำให้ได้ผลลัพธ์ทั้งสิ้น 34 คุณลักษณะจาก 316 คุณลักษณะ ดังแสดงในภาพที่ 2



**ภาพที่ 2 แสดงคุณลักษณะที่ผ่านการคัดเลือก**

จากภาพที่ 2 แสดงคุณลักษณะ LBP (Local Binary Pattern) ที่ขึ้นต้นด้วย LBP\_ ตามด้วยรหัสตัวเลข เช่น LBP\_224, LBP\_56, LBP\_14 โดยตัวเลขดังกล่าวคือรหัสรูปแบบที่มีค่าตั้งแต่ 0-255 ซึ่งเกิดจากการเปรียบเทียบค่าความเข้มของพิกเซลรอบข้าง 8 พิกเซลกับพิกเซลศูนย์กลาง และแปลงจากรูปแบบไบนารี 8 บิต เป็นเลขฐานสิบ ยกตัวอย่างเช่น LBP\_224 ซึ่งมีรหัสรูปแบบ 11100000 แทนความหมายได้ว่า 3 พิกเซลที่สว่าง ตามด้วย 5 พิกเซลที่มืด แสดงถึงการเปลี่ยนแปลงความเข้มแบบที่มีความแตกต่างกันสูงในพื้นที่ที่ใกล้เคียงกัน อาจบ่งบอกถึงรอยขอบของโรคหรือความผิดปกติในพื้นที่ใกล้เคียงกัน LBP\_254 ซึ่งมีรหัสรูปแบบ 11111110 แทนความหมายได้ว่า เกือบทุกพิกเซลสว่างกว่าพิกเซลตรงกลาง แสดงให้เห็นว่ามีพื้นที่ที่มีจุดมืดขนาดเล็ก ทามกลางพื้นที่สว่างซึ่งอาจเป็นจุดที่มีเลือกออกเป็นต้น โดย LBP\_224 มีค่าความสำคัญสูงสุดที่ 0.037 แสดงให้เห็นว่ารูปแบบพื้นผิวลักษณะนี้เป็นตัวบ่งชี้ที่ดีที่สุดในการแยกภาพที่มีและไม่มีภาวะของโรค ซึ่งคุณลักษณะ LBP สามารถจับลักษณะพื้นผิวที่ซับซ้อนได้ เช่น ความผิดปกติของเส้นเลือด จุดเลือดออก ประกอบกับการที่คุณลักษณะ LBP หลายคุณลักษณะมีค่าความสำคัญสูงบ่งชี้ว่าโรคเบาหวานขึ้นจอประสาทตามีลักษณะพื้นผิวที่แตกต่างจากจอประสาทตาปกติอย่างชัดเจน

#### 4. การสร้างแบบจำลอง

4.1 แบบจำลองแรนดอมฟอเรสต์ (Random Forest) เป็นวิธีการเรียนรู้ของเครื่องแบบมีผู้สอน โดยอาศัยการสร้างต้นไม้ตัดสินใจหลายต้นจากชุดข้อมูลย่อยและคุณลักษณะที่ถูกสุ่มคัดเลือก ผลลัพธ์จากต้นไม้ทั้งหมดจะถูกนำมารวมกันผ่านการโหวตเพื่อใช้จำแนกข้อมูล ทำให้ได้ผลการทำนายที่มีความแม่นยำและมีความแปรปรวนในระดับต่ำ เหมาะสำหรับงานวิจัยที่จัดการข้อมูลซับซ้อน เช่น การวิเคราะห์ข้อมูลด้านการแพทย์หรือสถิติที่ซับซ้อน เช่นเดียวกับงานวิจัยของ Kumar (2022) ซึ่งใช้คุณลักษณะเชิงสถิติด้านพื้นผิวของภาพเอ็กซเรย์ เพื่อคัดแยกโรคเนื้องอกในสมองด้วยเมทริกซ์ระดับสีเทาโดยแบบจำลองแรนดอมฟอเรสต์สามารถให้ความแม่นยำในระดับสูงถึง 83.3% โดยงานวิจัยนี้กำหนดค่าไฮเปอร์พารามิเตอร์โดยใช้จำนวนต้นไม้

ตัดสินใจ 300 ต้น (n\_estimators=300) ความลึกสูงสุดของต้นไม้ที่ 20 ระดับ (max\_depth=20) จำนวนข้อมูลตัวอย่างขั้นต่ำก่อนจะแบ่งโหนด 5 ตัวอย่าง (min\_samples\_split=5) จำนวนข้อมูลตัวอย่างขั้นต่ำในสปีโหนด 2 ตัวอย่าง (min\_samples\_leaf=2) ตั้งค่าจำนวนคุณลักษณะที่ใช้ในการจำแนกแต่ละครั้งจำนวน 6 คุณลักษณะ โดยประมาณค่าจากรากที่สองของคุณลักษณะทั้งหมด (max\_features='sqrt') สุ่มตัวอย่างแบบมีการแทนที่ (bootstrap=True) ปรับน้ำหนักตามสัดส่วนของข้อมูล (class\_weight='balanced') คำนวณคะแนนจากข้อมูลที่ไม่มีถูกใช้ (oob\_score=True) และปรับค่าให้ใช้แทนประมวลผลของหน่วยประมวลผลครบทุกแกน เพื่อเพิ่มความรวดเร็วในการทำงาน (n\_jobs=-1)

4.2 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine หรือ SVM) เป็นวิธีที่ใช้จำแนกข้อมูลออกเป็นกลุ่มโดยการหาเส้นระนาบ ซึ่งสามารถจำแนกข้อมูลสองกลุ่มให้ห่างกันมากที่สุด หากข้อมูลจำแนกได้ยากแบบจำลองจะเปลี่ยนข้อมูลให้อยู่ในรูปแบบใหม่ที่จำแนกได้ง่ายขึ้นโดยใช้ฟังก์ชันพิเศษซึ่งสามารถเปลี่ยนข้อมูลที่จำแนกยากในมิติเดิมไปอยู่ในมิติใหม่ที่สามารถจำแนกง่ายขึ้น เช่นการสร้างระนาบใหม่หรือใช้การจำแนกด้วยเส้นที่มีความโค้งโดยไม่ต้องคำนวณพิกัดใหม่ทั้งหมด ซึ่งในแบบจำลองนี้เรียกว่าเคอร์เนลและมีพารามิเตอร์ซึ่งสามารถกำหนดความสัมพันธ์ระหว่างการจำแนก ซึ่งช่วยปรับสมดุลระหว่างความแม่นยำและความยืดหยุ่น สามารถประยุกต์ใช้กับการจำแนกข้อมูลได้ดี เช่นเดียวกับงานวิจัยของคุณานันต์ อำทองและคณะ (2563) ซึ่งประยุกต์ใช้เมทริกซ์ระดับสี่เหลี่ยมกับซัพพอร์ตเวกเตอร์แมชชีนในการจำแนกภาพถ่ายทางการแพทย์ซึ่งให้ความแม่นยำสูงถึง 92.56% โดยงานวิจัยนี้กำหนดค่าไฮเปอร์พารามิเตอร์โดยใช้เคอร์เนลจำแนกข้อมูลด้วยฟังก์ชันรัศมี (kernel='rbf') ปรับค่าพารามิเตอร์การปรับเทียบ ที่ 10 (C=10.0) กำหนดค่าพารามิเตอร์แกมมาให้ค่าตามอัตโนมัติตามจำนวนคุณลักษณะและความแปรปรวน (gamma='scale') ปรับน้ำหนักตามสัดส่วนของข้อมูล (class\_weight='balanced')

4.3 เอ็กซ์ตรีมกราดิเอนท์บูสต์ (eXtreme Gradient Boost หรือ XGBoost) เป็นแบบจำลองซึ่งพัฒนาขึ้นจากแนวคิดของการเกรเดียนท์บูสต์ จัดอยู่ในกลุ่มของการเรียนรู้แบบรวมกลุ่มโดยอาศัยการฝึกแบบจำลองย่อยซึ่งมีลำดับขั้น โดยแต่ละแบบจำลองย่อยจะพยายามลดข้อผิดพลาดของแบบจำลองย่อยก่อนหน้า ส่งผลให้สามารถปรับปรุงประสิทธิภาพของการทำนายได้อย่างต่อเนื่อง (Chen & Guestrin, 2016) ด้วยเหตุนี้เอ็กซ์ตรีมกราดิเอนท์บูสต์ จึงได้รับความนิยมอย่างแพร่หลายในงานวิจัยด้านการวิเคราะห์ข้อมูลเชิงตัวเลขที่ซับซ้อนรวมถึงการวิเคราะห์ภาพถ่ายทางการแพทย์เช่นเดียวกับงานวิจัยของ Hassan และคณะ (2022) ซึ่งทำการวิจัยเพื่อจำแนกภาพโรคเบาหวานขึ้นจอประสาทตาโดยใช้เอ็กซ์ตรีมกราดิเอนท์บูสต์ร่วมกับคุณลักษณะเชิงลึกซึ่งสกัดจากภาพถ่ายจอประสาทตาและมีผลการจำแนกภาพถ่ายที่ถูกต้องในอัตราที่สูง อย่างไรก็ตามแบบจำลองเอ็กซ์ตรีมกราดิเอนท์บูสต์อาจเกิดปัญหาในประเด็นแบบจำลองมีความแม่นยำสูงกับข้อมูลฝึกแต่มีประสิทธิภาพต่ำเมื่อนำไปใช้กับข้อมูลใหม่หรือข้อมูลทดสอบหากไม่มีการควบคุม ความซับซ้อนของแบบจำลองอย่างเหมาะสม โดยงานวิจัยนี้กำหนดค่าไฮเปอร์พารามิเตอร์โดยใช้จำนวนต้นไม้ตัดสินใจ 300 ต้น (n\_estimators=300) ความลึกสูงสุดของต้นไม้ (max\_depth=6) อัตราการเรียนรู้ที่ 0.1 (learning\_rate=0.1) สัดส่วนของข้อมูลที่สุ่มใช้ในแต่ละรอบที่ 0.8 (subsample=0.8) สัดส่วนคุณลักษณะที่สุ่มใช้ในแต่ละต้นไม้ที่ 0.8 (colsample\_bytree=0.8) สัดส่วนคุณลักษณะที่สุ่มใช้ในต้นไม้ตัดสินใจแต่ละระดับที่ 0.8 (colsample\_bylevel=0.8) น้ำหนักขั้นต่ำที่ต้องมีในโหนดที่ 3 ตัวอย่าง (min\_child\_weight=3) ค่าการลดความสูญเสียขั้นต่ำที่จำเป็นสำหรับการแบ่งข้อมูลที่ 0.1 (gamma=0.1) ตั้งค่าให้น้ำหนักของคุณลักษณะบางตัวเป็น 0 เพื่อคัดเลือเฉพาะคุณลักษณะสำคัญ (reg\_alpha=0.1) ตั้งค่าเพื่อป้องกันไม่ให้คุณลักษณะบางตัวมีน้ำหนักมากเกินไป (reg\_lambda=1.0) ใช้ฟังก์ชันความคลาดเคลื่อนแบบล็อกลอส (eval\_metric='logloss')

สร้างต้นไม้ด้วยอัลกอริทึมแบบฮิสโตแกรมแทนการคำนวณจากข้อมูลทั้งหมด (tree\_method='hist') และใช้แกนประมวลผลของหน่วยประมวลผลครบทุกแกนเพื่อเพิ่มความเร็วในการทำงาน (n\_jobs=-1)

4.4 นิวรอลเน็ตเวิร์ค (Neural Networks หรือ NN) หรือเรียกอีกอย่างหนึ่งว่าแบบจำลองเครือข่ายประสาทเทียม เป็นแบบจำลองที่พัฒนาขึ้นโดยเลียนแบบการทำงานของเซลล์ประสาทในร่างกายมนุษย์ โดยมีหลักการคือ เมื่อเซลล์ประสาทได้รับข้อมูลหรือสิ่งเร้าจะประมวลผลภายในเซลล์ และตัดสินใจว่าจะส่งสัญญาณต่อไปยังเซลล์อื่นหรือไม่ แนวคิดนี้ถูกนำมาใช้ในระบบการทางคอมพิวเตอร์เพื่อสร้างระบบที่สามารถเรียนรู้จากข้อมูลเพื่อจำแนกรูปแบบและคาดการณ์สิ่งที่จะเกิดขึ้นในอนาคตแม้ข้อมูลจะมีความซับซ้อนหรือไม่เป็นเชิงเส้น (อนุพงศ์ สุขประเสริฐ, 2564) เช่นเดียวกับงานวิจัยของกมลชนกและคณะ (กมลชนก สีพาดังและคณะ, 2566) ซึ่งทำการวิจัยเกี่ยวกับกระบวนการคัดกรองผู้ป่วยเนื้องอกในสมองด้วยการสกัดคุณลักษณะจากภาพเอ็มอาร์ไอร่วมกับการเรียนรู้ของเครื่องและเปรียบเทียบประสิทธิภาพของแบบจำลองได้แก่ แบบจำลองการเรียนรู้เชิงลึก แบบจำลองนิวรอลเน็ตเวิร์ค และแบบจำลองแรนดอมฟอเรสต์พบว่าแบบจำลองนิวรอลเน็ตเวิร์คมีความแม่นยำที่ร้อยละ 72.81 โดยงานวิจัยนี้กำหนดค่าไฮเปอร์พารามิเตอร์โดยใช้จำนวนนิวรอนแต่ละชั้นซ่อนมีจำนวน 128 64 32 และชั้นเอาต์พุต 1 นิวรอนตามลำดับ ฟังก์ชันกระตุ้นและการปรับเทียบทุกชั้นซ่อนใช้ฟังก์ชันกระตุ้นแบบเรลยู (activation='relu') มีค่าการปรับเทียบแบบแอลทูที่ 0.001 (kernel\_regularizer=l2(0.001)) เพิ่มชั้นแบตช์นอร์มอลไลเซชัน (BatchNormalization) ชั้นที่ 1 และ 2 ใช้ดรอปเอาต์ 0.3 สำหรับชั้นที่ 3 ใช้ดรอปเอาต์ 0.2 (Dropout=0.3, 0.2) ชั้นเอาต์พุตใช้ฟังก์ชันซิกมอยด์ (activation='sigmoid') การปรับค่าน้ำหนักใช้อัลกอริทึมแอดัม (Adam optimizer) โดยมีอัตราการเรียนรู้ 0.001 ค่าเบต้า 1 ที่ 0.9 และเบต้า 2 ที่ 0.999 (learning\_rate=0.001, beta\_1=0.9, beta\_2=0.999) ใช้ฟังก์ชันความสูญเสียแบบไบนารีครอสเอนโทรปี (loss='binary\_crossentropy') การฝึกสอนและควบคุมแบบจำลองกำหนดจำนวนรอบการฝึกสอนสูงสุด 100 รอบด้วยขนาดแบตช์ที่ 64 (epochs=100, batch\_size=64) แบ่งข้อมูลเพื่อใช้ตรวจสอบความถูกต้องของแบบจำลองที่อัตราส่วนร้อยละ 20 (validation\_split=0.2) ใช้เออร์ลี่สโตปปีงที่หยุดการฝึกสอนอัตโนมัติเมื่อความสูญเสียในชุดตรวจสอบไม่ดีขึ้นจำนวน 10 รอบและย้อนกลับไปใช้ค่าของน้ำหนักที่ดีที่สุด (EarlyStopping, monitor='val\_loss', patience=10, restore\_best\_weights=True) ลดอัตราการเรียนรู้ลงครึ่งหนึ่งเมื่อความสูญเสียไม่ดีขึ้นเป็นเวลา 5 รอบโดยมีอัตราต่ำสุดที่ 0.0000001 (ReduceLROnPlateau, factor=0.5, patience=5, min\_lr=1e-7) และปรับน้ำหนักคลาสตามอัตราส่วนของข้อมูลเพื่อจัดการกับข้อมูลที่ไม่สมดุล (class\_weight)

4.5 แบบจำลองเคเนียร์สตันเบอร์ (K-Nearest Neighbors หรือ KNN) เป็นแบบจำลองสำหรับจำแนกข้อมูลโดยอาศัยหลักการเปรียบเทียบข้อมูลใหม่กับข้อมูลในชุดฝึกผ่านการวัดค่าความใกล้เคียงทางคณิตศาสตร์ยกตัวอย่างเช่น ระยะทางยุคลิดเป็นต้น จากนั้นทำการเลือกข้อมูลจำนวน k ตัวอย่างที่อยู่ใกล้ข้อมูลใหม่มากที่สุด และทำการจัดประเภทของข้อมูลใหม่ตามคลาสที่ปรากฏบ่อยที่สุดในกลุ่มข้อมูลเหล่านั้น ส่งผลให้สามารถสะท้อนความคล้ายคลึงผ่านการวัดระยะทางได้อย่างชัดเจน (อรรถสฎาฐ เรื่องสวัสดิ์ และคณะ, 2566) เช่นเดียวกับการค้นคว้าและวิจัยของ อนุพงศ์และคณะ (อนุพงศ์ สุขประเสริฐ และคณะ, 2567) ซึ่งทำการวิจัยเกี่ยวกับการคัดกรองผู้ป่วยโรคหัวใจได้ด้วยการสกัดคุณลักษณะจากภาพซีทีสแกนและกระบวนการเรียนรู้ของเครื่องเพื่อทำการเปรียบเทียบประสิทธิภาพของแบบจำลองจำนวน 4 แบบจำลองได้แก่แบบจำลองเคเนียร์สตันเบอร์ แบบจำลองซัพพอร์ตเวกเตอร์แมชชีน แบบจำลองแรนดอมฟอเรสต์ แบบจำลองนาอิวเบย์ ทั้งนี้จากการวิจัยพบว่าแบบจำลองเคเนียร์สตันเบอร์มีความแม่นยำมากที่สุดที่ร้อยละ 92.64

โดยงานวิจัยนี้กำหนดค่าไฮเปอร์พารามิเตอร์โดยกำหนดจำนวนเพื่อนบ้านให้มีค่าเท่ากับ 7 ( $n\_neighbors=7$ ) กำหนดให้เพื่อนบ้านที่ใกล้กว่ามีน้ำหนักมากกว่า ( $weights='distance'$ ) กำหนดวิธีการวัดระยะทางมินคอฟสกี ( $metric='minkowski'$ ,  $p=2$ ) เลือกอัลกอริทึมที่เหมาะสมแบบอัตโนมัติ ( $algorithm='auto'$ ) และปรับแต่งค่าให้ใช้แกนประมวลผลของหน่วยประมวลผลครบทุกแกนเพื่อเพิ่มความเร็วในการทำงาน ( $n\_jobs=-1$ )

## 5. การประเมินผล

การศึกษาและวิจัยครั้งนี้มีการประเมินประสิทธิภาพแบบจำลองโดยแบ่งข้อมูลออกเป็น 10 ส่วน ในอัตราส่วนที่เท่ากันโดยใช้ 9 ส่วนสำหรับฝึกสอนแบบจำลอง (Training Set) และใช้ 1 ส่วนสำหรับทดสอบแบบจำลอง (Testing Set) โดยทำซ้ำทั้งหมดจำนวน 30 รอบและเฉลี่ยค่าผลลัพธ์ที่ได้ ซึ่งเป็นกระบวนการช่วยลดความแปรปรวนของผลการประเมินแบบจำลอง

ทั้งนี้ผลลัพธ์จากกระบวนการสร้างแบบจำลองสามารถแสดงในรูปแบบตารางเมทริกซ์ความสับสน (Confusion Matrix) ซึ่งกำหนดกลุ่ม Positive คือรูปภาพที่มีภาวะโรคเบาหวานขึ้นจอประสาทตา และกลุ่ม Negative คือ รูปภาพที่ไม่มีภาวะโรคเบาหวานขึ้นจอประสาทตา ทำให้เกิดผลการจำแนก 4 ประเภท ซึ่งนำไปใช้ประเมินประสิทธิภาพของแบบจำลอง

1. True Negative (TN) คือ จำนวนภาพที่ไม่มีภาวะโรคเบาหวานขึ้นจอประสาทตาและแบบจำลองจำแนกได้อย่างถูกต้องว่าภาพถ่ายดังกล่าวไม่มีภาวะโรคเบาหวานขึ้นจอประสาทตา

2. False Positive (FP) คือ จำนวนภาพที่ไม่มีภาวะโรคเบาหวานขึ้นจอประสาทตา แต่แบบจำลองจำแนกผิดโดยจำแนกว่าภาพถ่ายดังกล่าวมีภาวะโรคเบาหวานขึ้นจอประสาทตา

3. False Negative (FN) คือ จำนวนภาพที่มีภาวะโรคเบาหวานขึ้นจอประสาทตาแต่แบบจำลองจำแนกผิดโดยจำแนกว่าภาพถ่ายดังกล่าวไม่มีภาวะโรคเบาหวานขึ้นจอประสาทตา

4. True Positive (TP) คือ จำนวนภาพที่มีภาวะโรคเบาหวานขึ้นจอประสาทตาและแบบจำลองจำแนกได้อย่างถูกต้อง ว่าภาพถ่ายดังกล่าวมีภาวะโรคเบาหวานขึ้นจอประสาทตา

ค่ามาตรฐานในการวัดประสิทธิภาพของแบบจำลองประกอบด้วย ค่าความแม่นยำ (Accuracy) ค่าประสิทธิภาพโดยรวม (F-Measure) ค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity) ซึ่งมีรายละเอียดดังต่อไปนี้

5.1 ค่าความแม่นยำ (Accuracy) คือ อัตราส่วนของการทำนายที่ถูกต้องทั้งหมดทั้ง Positive และ Negative ต่อจำนวนตัวอย่างทั้งหมดในชุดข้อมูล ค่าความแม่นยำสามารถคำนวณได้ดังสมการต่อไปนี้

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

5.2 ค่าความไว (Sensitivity) คือ อัตราส่วนของตัวอย่าง Positive ที่ถูกทำนายอย่างถูกต้องต่อตัวอย่าง Positive ทั้งหมด ค่าความไวสามารถบ่งบอกความครบถ้วนในกระบวนการจำแนกกลุ่ม Positive

$$Sensitivity = TP / (TP + FN)$$

5.3 ค่าความจำเพาะ (Specificity) คือ อัตราส่วนของตัวอย่าง Negative ที่ถูกทำนายอย่างถูกต้องต่อตัวอย่าง Negative ทั้งหมด ค่าความจำเพาะสามารถบ่งบอกความครบถ้วนในกระบวนการจำแนกกลุ่ม Negative

$$Specificity = TN / (TN + FP)$$

5.4 ค่าประสิทธิภาพโดยรวม (F-Measure) คือ เมตริกซึ่งใช้ประเมินประสิทธิภาพของแบบจำลองเพื่อการจำแนกประเภทโดยคำนวณจากค่าความแม่นยำ (Precision) และความครอบคลุม (Recall) ซึ่งเหมาะสำหรับกระบวนการที่ต้องการวัดความสัมพันธ์ระหว่างความแม่นยำและความครอบคลุม เช่น การประเมินแบบจำลองในงานวิจัยด้านการเรียนรู้ของเครื่อง

$$F1\text{-score} = 2TP / (2TP + FP + FN)$$

โดยที่

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

## 6. การนำไปใช้งาน

ภายหลังจากการดำเนินงานตามมาตรฐานของการทำเหมืองข้อมูลทั้ง 5 ขั้นตอน สามารถนำแบบจำลองที่มีประสิทธิภาพสูงสุดเพื่อประยุกต์ใช้ในกรณีต่างๆ เช่น บันทึกแบบจำลองในรูปแบบแฟ้มข้อมูลเพื่อเก็บไว้ในอนาคต การใช้แบบจำลองกับข้อมูลชุดใหม่ หรือการสร้างระบบสารสนเทศที่สามารถเรียกใช้แบบจำลองผ่านเว็บแอปพลิเคชันหรือเอพีไอเพื่อสร้างระบบวินิจัยที่ทำงานกับข้อมูลใหม่ในทันที และการแปลงแบบจำลองเพื่อใช้ในอุปกรณ์อื่นซึ่งเหมาะกับการค้นคว้าวิจัยในลำดับถัดไป

### ผลการวิจัยและอภิปรายผล

จากผลการทดลองพบว่าแบบจำลองเอ็กตรึมการเคียนบูสต์ ให้ผลการจำแนกที่ดีและสมมูลที่สุด โดยสามารถจำแนกกลุ่มภาพถ่ายซึ่งมีภาวะของโรคเบาหวานขึ้นจอประสาทตาถูกต้อง (TP) สูงถึง 4,416 ภาพ และสามารถจำแนกกลุ่มภาพถ่ายซึ่งไม่มีภาวะของโรคเบาหวานขึ้นจอประสาทตาถูกต้อง (TN) สูงถึง 4,360 ภาพ และมีการจำแนกผิดพลาดในอัตราที่ต่ำ โดยจำแนกกลุ่มภาพถ่ายที่ไม่มีภาวะของโรคผิดพลาด (FP) ต่ำที่สุดเพียง 57 ภาพ และจำแนกกลุ่มภาพถ่ายที่มีภาวะของโรคผิดพลาด (FN) อยู่ในอัตราที่ต่ำเพียง 1 ภาพ สะท้อนให้เห็นถึงประสิทธิภาพในการจำแนกที่มีความเสถียรทั้งในกลุ่มที่มีภาวะของโรคและไม่มีภาวะของโรค เมื่อพิจารณาแบบจำลองนิวรอลเน็ตเวิร์คแสดงให้เห็นว่าแบบจำลองสามารถจำแนกกลุ่มภาพถ่ายซึ่งมีและไม่มีภาวะของโรคได้ถูกต้องในอัตราที่สูงใกล้เคียงกัน โดยสามารถจำแนกกลุ่มภาพถ่ายซึ่งมีภาวะของโรคเบาหวานขึ้นจอประสาทตาถูกต้อง (TP) สูงถึง 4,382 ภาพ และสามารถจำแนกกลุ่มภาพถ่ายซึ่งไม่มีภาวะของโรคเบาหวานขึ้นจอประสาทตาถูกต้อง (TN) สูงถึง 4,286 ภาพ มีการจำแนกกลุ่มที่ไม่มีภาวะโรคผิดพลาด (FP) 131 ภาพ และพบว่ามีจำนวนการจำแนกกลุ่มที่มีภาวะโรคผิดพลาด (FP) ถึง 35 ภาพ ซึ่งอาจเป็นข้อจำกัดในการวินิจฉัยเกี่ยวกับสุขภาพ แบบจำลองเคเนียร์สเนเบอร์มีผลการจำแนกที่แม่นยำและสมมูลแม้ไม่ใช้ข้อมูลในลักษณะเชิงเส้น โดยสามารถจำแนกกลุ่มภาพถ่ายซึ่งมีภาวะของโรคเบาหวานขึ้นจอประสาทตาถูกต้อง (TP) สูงถึง 4,416 ภาพ และสามารถจำแนกกลุ่มภาพถ่ายซึ่งไม่มีภาวะของโรคเบาหวานขึ้นจอประสาทตาถูกต้อง (TN) สูงถึง 4,157 ภาพ มีจำแนกกลุ่มภาพถ่ายที่มีภาวะของโรคผิดพลาด (FN) อยู่ในอัตราที่ต่ำเพียง 1 ภาพ แต่มีการจำแนกกลุ่มที่ไม่มีภาวะโรคผิดพลาด (FP) สูงถึง 260 ภาพ ในขณะที่แบบจำลองแรนดอมฟอเรสต์แม้จะสามารถจำแนกรูปภาพกลุ่มมีภาวะของโรคได้ได้ครบถ้วนและไม่พบข้อผิดพลาดในการจำแนก (FN = 0) แต่มีอัตราการจำแนกกลุ่มที่ไม่มีภาวะโรคผิดพลาด (FP) สูงถึง 2,892 ภาพ ทำให้การจำแนกภาพถ่ายซึ่งไม่มีภาวะของโรคเบาหวานขึ้นจอประสาทตาที่ถูกต้อง (TN) เพียง 1,525 ภาพ จากเหตุปัจจัยดังกล่าวอาจก่อให้เกิดผลข้างเคียงเช่น เกิดการวินิจฉัยที่ผิดพลาดจากผลบวกสูง เกิดภาระงานที่ไม่จำเป็นและสูญเสียทรัพยากรเพิ่มเติมจากการตรวจติดตามผลซ้ำ และแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพในการจำแนกต่ำที่สุด ทั้งนี้การจำแนก

รูปภาพ ที่มีภาวะโรคได้ถูกต้อง (TP) เพียง 2,454 ภาพ และสามารถจำแนกรูปภาพที่ไม่มีโรคถูกต้อง (TN) เพียง 1,994 ภาพ ขณะที่เกิดความผิดพลาดจากการจำแนกในอัตราที่สูงมาก โดยเฉพาะการทำนายกลุ่มที่มีภาวะของโรคผิดพลาด (FN) สูงถึง 1,922 ภาพ และการวินิจฉัยผิดพลาดในกลุ่มที่ไม่มีภาวะของโรค (FP) สูงถึง 2,423 ภาพ ซึ่งแสดงถึงข้อจำกัดที่ชัดเจนของแบบจำลองในการใช้งานจริง โดยเฉพาะในระบบที่ต้องการความแม่นยำสูง ทั้งนี้การเปรียบเทียบผลการจำแนกรูปภาพของแบบจำลองสามารถแสดงในตารางที่ 1

**ตารางที่ 1** แสดงการเปรียบเทียบผลการจำแนกรูปภาพของแบบจำลอง

| แบบจำลอง               | True Negative | False Positive | False Negative | True Positive |
|------------------------|---------------|----------------|----------------|---------------|
| เอ็กตรึมการเคียนบูสต์  | 4360          | 57             | 1              | 4416          |
| นิวโรลเน็ตเวิร์ค       | 4286          | 131            | 35             | 4382          |
| เคเนียร์สเนเบอร์       | 4157          | 260            | 1              | 4416          |
| แรนดอมฟอเรสต์          | 1525          | 2892           | 0              | 4417          |
| ซัพพอร์ตเวกเตอร์แมชชีน | 1994          | 2423           | 1922           | 2495          |

จากผลการประเมินประสิทธิภาพแบบจำลองพบว่าแบบจำลองเอ็กตรึมการเคียนบูสต์มีประสิทธิภาพการจำแนกที่ดีที่สุดในทุกตัวชี้วัดโดยมีค่าความแม่นยำอยู่ที่ร้อยละ 99.34 และค่าประสิทธิภาพโดยรวมอยู่ที่ 99.35% ซึ่งสะท้อนถึงความสามารถในการจำแนกกลุ่มที่มีภาวะของโรคและกลุ่มที่ไม่มีภาวะของโรคได้อย่างสมดุล นอกจากนี้ค่าความไวที่สูงถึงร้อยละ 99.98 แสดงให้เห็นว่าแบบจำลองสามารถจำแนกรูปภาพที่มีภาวะของโรคได้เกือบทั้งหมดและเกิดกรณีผิดพลาดในการตรวจพบภาวะของโรค (FN) ในอัตราที่น้อยมากเมื่อเทียบกับจำนวนรูปภาพทั้งหมด ขณะที่ค่าความจำเพาะที่สูงถึงร้อยละ 98.71 สามารถบ่งชี้ได้ว่าแบบจำลองสามารถหลีกเลี่ยงการวินิจฉัยเกินจริงในกลุ่มที่ไม่มีภาวะของโรคได้ดีมาก จึงมีความเหมาะสมอย่างยิ่งในการนำไปใช้ในระบบวินิจฉัยทางการแพทย์ที่ต้องการลดความผิดพลาดทั้งกรณีมีและไม่มีภาวะของโรคได้อย่างสมดุล สอดคล้องกับงานวิจัยโดย Hassan และคณะ (2022) ซึ่งประยุกต์ใช้ใช้แบบจำลองเอ็กตรึมการเคียนบูสต์ร่วมกับการสกัดคุณสมบัติเชิงลึกเพื่อจำแนกโรคเบาหวานขึ้นจอประสาทตาซึ่งให้ความแม่นยำสูง แบบจำลองนิวโรลเน็ตเวิร์คให้ผลการประเมินมีประสิทธิภาพสูงรองลงมาจากเอ็กตรึมการเคียนบูสต์โดยมีค่าความแม่นยำอยู่ที่ร้อยละ 98.12 และค่าประสิทธิภาพโดยรวมอยู่ที่ร้อยละ 98.14 จุดเด่นของแบบจำลองนี้คือความสามารถในการเรียนรู้คุณลักษณะจากข้อมูลเชิงซ้อนผ่านกระบวนการเรียนรู้แบบหลายชั้นซ้อน ส่งผลให้สามารถลดอัตราความผิดพลาดทั้งกลุ่มที่มีภาวะของโรคและไม่มีภาวะของโรคได้อย่างมีประสิทธิภาพ โดยมีค่าความไวอยู่ที่ร้อยละ 99.21 และค่าความจำเพาะอยู่ที่ร้อยละ 97.03 แสดงถึงความสามารถในการรักษาความสมดุลและความแม่นยำในการตรวจจับได้ทั้งสองกลุ่ม แม้ว่าจะมีอัตราการจำแนกกลุ่มที่มีภาวะของโรคผิดพลาดแต่นับว่ามีจำนวนน้อยเมื่อเทียบกับแบบจำลองอื่น ประสิทธิภาพแบบจำลองนิวโรลเน็ตเวิร์คในงานวิจัยนี้มีประสิทธิภาพมากกว่างานวิจัยโดย กมลชนก สีพาดัง และคณะ (2566) ซึ่งสกัดค่าทางสถิติของภาพถ่ายทางการแพทย์เช่น ค่าเฉลี่ย ค่ามัธยฐาน ค่าความเข้มของสีเทา แสดงให้เห็นว่าคุณสมบัติเชิงพื้นผิวเป็นทางเลือกที่เหมาะสมมากกว่า แบบจำลองเคเนียร์สเนเบอร์มีลักษณะการจำแนกข้อมูลที่อาศัยลักษณะของข้อมูลรอบข้างส่งผลให้แบบจำลองสามารถให้

ค่าความไวสูงที่สุดในกลุ่มแบบจำลองซึ่งใช้ในการค้นคว้าวิจัยครั้งนี้ โดยมีค่าความไวเท่ากับร้อยละ 99.98 เทียบเท่าเอ็กตริมาร์เดียนบูสต์ซึ่งตรวจพบภาพที่มีภาวะของโรคได้เกือบครบถ้วน แต่อย่างไรก็ตามแบบจำลองมีค่าความจำเพาะอยู่ที่ร้อยละ 94.11 ซึ่งต่ำกว่าแบบจำลองอื่นเล็กน้อย ทำให้แบบจำลองเคเนียร์สตันเบอร์เหมาะสำหรับใช้ในบริบทที่ต้องการลดข้อผิดพลาดในการตรวจพบกลุ่มที่มีภาวะของโรคให้เหลือในอัตราที่น้อยที่สุด ประสิทธิภาพแบบจำลองเคเนียร์สตันเบอร์ในงานวิจัยนี้มีประสิทธิภาพมากกว่างานวิจัยโดย อนุพงศ์ และคณะ (2567) เล็กน้อย ทั้งนี้สาเหตุอาจมาจากประเภทภาพถ่ายทางการแพทย์ที่แตกต่างกัน ในประเด็นแบบจำลองแรนดอมฟอเรสต์แม้แบบจำลองจะให้ค่าความไวที่สูงสุดที่ร้อยละ 100 ซึ่งบ่งชี้ได้ว่าแบบจำลองสามารถตรวจพบผู้ป่วยได้ทั้งหมดแต่แบบจำลองมีค่าความจำเพาะที่ต่ำมากเพียงร้อยละ 34.53 ทำให้เกิดการดำเนินงานกลุ่มรูปภาพที่ไม่มีภาวะของโรคผิดพลาด (FP) ในอัตราที่สูงมาก ส่งผลให้ค่าความแม่นยำโดยรวมลดลงเหลือเพียงร้อยละ 67.26 และค่าประสิทธิภาพโดยรวมอยู่ที่ร้อยละ 75.34 ทั้งนี้ประสิทธิภาพแบบจำลองแรนดอมฟอเรสต์ในงานวิจัยนี้มีประสิทธิภาพน้อยกว่างานวิจัยโดย Kumar (2022) ซึ่งใช้คุณลักษณะเชิงพื้นผิวร่วมกับแบบจำลองแรนดอมฟอเรสต์ในการจำแนกข้อมูลซึ่งผลลัพธ์มีความแม่นยำในระดับสูง แบบจำลองซัพพอร์ทเวกเตอร์แมชชีนเป็นแบบจำลองที่มีประสิทธิภาพต่ำที่สุดในบรรดาแบบจำลองทั้งหมด โดยมีค่าความไวเพียงร้อยละ 56.49 และค่าความจำเพาะเพียงร้อยละ 45.14 สะท้อนถึงความสามารถที่จำกัดในการจำแนกกลุ่มข้อมูลทั้งสองประเภท ส่งผลให้ค่าประสิทธิภาพโดยรวมเพียงร้อยละ 53.45 ซึ่งถือว่าต่ำกว่ามาตรฐานที่ต้องการในบริบทของงานวินิจฉัยทางการแพทย์ และมีประสิทธิภาพน้อยกว่า คุณานันต์ อ่ำทองและคณะ (2563) ซึ่งใช้ GLCM ร่วมกับแบบจำลองซัพพอร์ทเวกเตอร์แมชชีนให้ความแม่นยำสูงที่ร้อยละ 92.56 ซึ่งสูงกว่างานวิจัยนี้ ทั้งนี้อาจมีสาเหตุมาจากการที่งานวิจัยนี้มีข้อมูลมากกว่าและใช้ทั้ง LBP และ GLCM ร่วมกัน ทั้งนี้การเปรียบเทียบประสิทธิภาพของแบบจำลองสามารถแสดงในตารางที่ 2

**ตารางที่ 2** แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลอง

| แบบจำลอง               | ค่าความแม่นยำ | ค่าความไว | ค่าความจำเพาะ | ประสิทธิภาพโดยรวม |
|------------------------|---------------|-----------|---------------|-------------------|
| เอ็กตริมาร์เดียนบูสต์  | 99.34%        | 99.98%    | 98.71%        | 99.35%            |
| นิวรอลเน็ตเวิร์ค       | 98.12%        | 99.21%    | 97.03%        | 98.14%            |
| เคเนียร์สตันเบอร์      | 97.05%        | 99.98%    | 94.11%        | 97.13%            |
| แรนดอมฟอเรสต์          | 67.26%        | 100%      | 34.53%        | 75.34%            |
| ซัพพอร์ทเวกเตอร์แมชชีน | 50.82%        | 56.49%    | 45.14%        | 53.45%            |

## สรุป

งานวิจัยฉบับนี้มุ่งเน้นการพัฒนาและประเมินแบบจำลองการเรียนรู้ของเครื่องสำหรับจำแนกภาพถ่ายจอประสาทตาซึ่งมีภาวะโรคเบาหวานขึ้นจอประสาทตาโดยทำการสกัดคุณลักษณะเชิงพื้นผิวได้แก่ LBP และ GLCM จำนวน 316 คุณลักษณะ โดยนำคุณลักษณะดังกล่าวมาผ่านกระบวนการลดความซ้ำซ้อนของข้อมูลด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์อันดับสเปียร์แมนที่ 0.90 และคัดเลือกคุณสมบัตินี้ที่มีความสัมพันธ์กับผลการจำแนกที่มีค่าเอฟสกอ์สูงสุด 100 คุณลักษณะและทำการคัดเลือกโดยใช้เทคนิคการหาค่าความสำคัญ

ของแต่ละคุณลักษณะที่สกัดได้ด้วยอัลกอริทึมแรนดอมฟอเรสต์ที่มีค่าความสำคัญ 0.01 ขึ้นไป ทำให้ได้คุณลักษณะที่มีความสำคัญสูงและมีนัยสำคัญต่อการจำแนกภาพถ่ายทั้งหมด 34 คุณลักษณะจาก 316 คุณลักษณะ เพื่อใช้สร้างแบบจำลองในการจำแนกภาพถ่าย ทั้งนี้จากการประเมินประสิทธิภาพของแบบจำลอง 5 ประเภท ได้แก่ แรนดอมฟอเรสต์, เอ็กตรึมการเดียนบูสต์, ซัพพอร์ตเวกเตอร์แมชชีน, นิวรอลเน็ตเวิร์ค และ เคเนียร์สเดนเบอร์

จากผลการทดลองพบว่าแบบจำลองเอ็กตรึมการเดียนบูสต์ให้ผลการจำแนกที่ดีและสมมูลที่สุด โดยมีค่าความแม่นยำอยู่ที่ร้อยละ 99.34 เทียบกับนิวรอลเน็ตเวิร์คที่ได้ร้อยละ 98.12 แสดงให้เห็นความแตกต่างร้อยละ 1.22 ค่าประสิทธิภาพโดยรวมอยู่ที่ร้อยละ 99.35 ค่าความไวสูงถึงร้อยละ 99.98 และค่าความจำเพาะสูงถึงร้อยละ 98.71 ผลลัพธ์ดังกล่าวสะท้อนถึงความสามารถในการจำแนกกลุ่มที่มีภาวะของโรคและกลุ่มที่ไม่มีภาวะของโรคได้อย่างสมมูล โดยแบบจำลองสามารถจำแนกภาพถ่ายที่มีภาวะของโรคถูกต้อง (TP) สูงถึง 4,416 ภาพ และจำแนกภาพถ่ายที่ไม่มีภาวะของโรคถูกต้อง (TN) สูงถึง 4,360 ภาพ พร้อมทั้งมีการจำแนกผิดพลาดในอัตราที่ต่ำมาก โดยจำแนกผิดพลาดในกลุ่มที่ไม่มีภาวะของโรค (FP) เพียง 57 ภาพ และจำแนกผิดพลาดในกลุ่มที่มีภาวะของโรค (FN) เพียง 1 ภาพเท่านั้น ความสามารถนี้ทำให้แบบจำลองเอ็กตรึมการเดียนบูสต์เหมาะสมอย่างยิ่งในการนำไปใช้ในระบวินิจัยทางการแพทย์ที่ต้องการลดความผิดพลาดทั้งกรณีมีและไม่มีภาวะของโรคได้อย่างสมมูล

## ข้อเสนอแนะ

### ข้อเสนอแนะเกี่ยวกับงานวิจัย

ควรทดลองปรับค่าไฮเปอร์พารามิเตอร์เพื่อช่วยเพิ่มความแม่นยำในการจำแนกข้อมูลเช่นทดลองหาค่าที่เหมาะสมของพารามิเตอร์ในแบบจำลองซัพพอร์ตเวกเตอร์แมชชีน เป็นต้น

### ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. การนำแบบจำลองไปใช้งานแบบเรียลไทม์โดยการออกแบบระบบที่ระบบสามารถรับข้อมูลใหม่และส่งผลการทำนายได้ทันทีโดยไม่ต้องรอการประมวลผลข้อมูลทั้งกลุ่ม ซึ่งเหมาะสำหรับสถานการณ์ที่ต้องการผลลัพธ์รวดเร็วเช่น การสร้างเอพพิโอวิเคราะห์ภาพถ่ายทางการแพทย์เพื่อสนับสนุนการวินิจฉัยและการคัดกรองความเสี่ยงในระดับปฐมภูมิ การออกแบบระบบลักษณะนี้ มีประเด็นที่น่าสนใจในด้านเวลาในการตอบสนอง ความเสถียรของระบบและการจัดการกับข้อมูลเพื่อให้ประสิทธิภาพการประมวลผลที่ดี

2. การแปลงแบบจำลองเพื่อใช้ในอุปกรณ์อื่น ซึ่งในบางกรณีอาจมีความจำเป็นต้องนำแบบจำลองไปประยุกต์ใช้บนอุปกรณ์ที่มีข้อจำกัดด้านทรัพยากร เช่น อุปกรณ์พกพา รัสเบอร์รี่พาย หรืออุปกรณ์ประมวลผลขนาดเล็กซึ่งใช้ในระบบสุขภาพเคลื่อนที่สำหรับพื้นที่ห่างไกลร่วมกับการแปลงแบบจำลองให้อยู่ในรูปแบบที่เหมาะสม

## กิตติกรรมประกาศ

ขอขอบพระคุณคณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี ซึ่งสนับสนุนอุปกรณ์สถานที่ในการทำวิจัยและให้คำแนะนำที่เป็นประโยชน์ในการวิจัย ขอขอบคุณ Rutuja Chaudhari ซึ่งเอื้อเฟื้อข้อมูลในการทำวิจัย

## เอกสารอ้างอิง

- กมลชนก สีพาดัง, นนทวัฒน์ เปรณาม, วีระศักดิ์ สว่างโลก, และอนุพงศ์ สุขประเสริฐ(2566). การจำแนกภาพ MRI สำหรับการคัดกรองผู้ป่วยเนื้องอกในสมอง. **วารสารวิทยาการสารสนเทศและเทคโนโลยีประยุกต์**, 5(2), 100–115.
- กฤษณะ ชินสาร, สุวรรณ รัชมิขวิญ, เบญจภรณ์ จันทรวงกุล, ภูสิต กุลเกษม, อัมมพันธ์ รอดทุกข์, จรรยา อ้นปัส ฐนินท์ อินทรมณี, และ Khoeun, R. (2559). **โครงการการคัดเลือกด้วยวิธีการเรียนรู้รวมแบบปรับตัวได้สำหรับการตรวจพบอาการเส้นเลือดฝอยโป่งพองและมีจุดเลือดออกในตาสำหรับผู้ป่วยเบาหวานในระยะเริ่มต้น**. รายงานการวิจัย, คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา.
- คุณานนต์ อ่ำทอง, คชินทร์ บุญทรัพย์, กษิตินาถ บุญพันธ์, และพิศณุ คูมีชัย. (2563). **การพัฒนาแบบคัดแยกของเนื้องอกในสมองโดยใช้เมตริกซ์ของระดับสีเทาที่เกิดขึ้นร่วมกันและการแปลงเวฟเลต**. การประชุมวิชาการระดับชาติ ประจำปี 2563 เรื่อง การจัดการในยุคเทคโนโลยีนำการเปลี่ยนแปลง (30 สิงหาคม 2561). มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา.
- มาลีรัตน์ โสตานิล, และอมรศักดิ์ อมรณานันท์. (2555). การรู้จำใบหน้าแบบหลายมุมมองโดยใช้เทคนิคผสมผสานการแบ่งภาพและการจับคู่ภาพมุมมองจริง. **วารสารเทคโนโลยีสารสนเทศ**, 8(2), 33–38.
- อนุพงศ์ สุขประเสริฐ. (2564). **คู่มือการทำเหมืองข้อมูลด้วยโปรแกรม RapidMiner Studio**. มหาสารคาม: คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม.
- อนุพงศ์ สุขประเสริฐ, น้ำฝน ดังโพนทอง, และธีรยุทธ จันทรา. (2567). **การสร้างตัวแบบสำหรับการคัดกรองผู้ป่วยโรคหัวใจโดยใช้เทคนิคการจำแนกประเภทข้อมูลภาพ**. **วารสารนวัตกรรมวิทยาศาสตร์เพื่อการพัฒนาอย่างยั่งยืน**, 5(1), 12–23.
- อรรถกร มากคงแก้ว. (2555). **การประยุกต์ใช้ลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็มในการนับนิวเคลียสเซลล์มะเร็งเต้านม** (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). มหาวิทยาลัยสงขลานครินทร์.
- อรรถภาวูธ เรืองสวัสดิ์, อนุพงศ์ สุขประเสริฐ, ทิพา สันธฤฎ, และศิริลักษณ์ ไกยวินิจ. (2566). **การเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์โรคมะเร็งปอด**. **วารสารวิทยาศาสตร์ เทคโนโลยี และนวัตกรรม มหาวิทยาลัยกาฬสินธุ์**, 2(2), 39–52.
- อัจฉรา แก้วน้ำเชื้อ. (2564). **การพยาบาลผู้ป่วยโรคเบาหวานขึ้นจอประสาทตา**. **วารสารสุขภาพและสิ่งแวดล้อมศึกษา**, 6(4), 177–180.
- Ayyanar, M., Jeganathan, S., Parthasarathy, S., Jayaraman, V., & Lakshminarayanan, A. R. (2022). **Predicting cardiac diseases using SelectKBest method equipped Light Gradient Boosting Machine**. In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 117–122). IEEE.
- Boonsaen, T., Choksakunwong, S., & Lertwattanak, R. (2021). Prevalence of and factors associated with diabetic retinopathy in patients with diabetes mellitus at Siriraj Hospital – Thailand’s largest national tertiary referral center. **Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy**, 14, 4945–4957. <https://doi.org/10.2147/DMSO.S346719>
- Chaudhari, R. (2020). **Diabetic retinopathy dataset** [Data set]. Kaggle. Retrieved from <https://www.kaggle.com/datasets/rutujachaudhari/diabetic-retinopathy>

- Chen, T., & Guestrin, C. (2016). **XGBoost: A scalable tree boosting system**. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp.785–794). Association for Computing Machinery.
- Chollet, F., et al. (2015). **Keras: Deep learning for humans** (Version 2.15.0) [Computer software]. <https://keras.io>
- Gajaram, A. M. (2022). **An approach to classify ocular diseases using machine learning and deep learning** (Master's thesis). National College of Ireland.
- Hassan, D., Gill, H. M., Happe, M., Bhatwadekar, A. D., Hajrasouliha, A. R., & Janga, S. C. (2022). Combining transfer learning with retinal lesion features for accurate detection of diabetic retinopathy. **Frontiers in Medicine**, 9, 1050436. <https://doi.org/10.3389/fmed.2022.1050436>
- Jiang, J., Zhang, X., & Yuan, Z. (2024). Feature selection for classification with Spearman's rank correlation coefficient-based self-information in divergence-based fuzzy rough sets. **Expert Systems with Applications**, 249(Part B), 123633. <https://doi.org/10.1016/j.eswa.2024.123633>
- Kalyani, D., Reddy, J., Ramesh, C., Saraf, S., Kaushal, A., Ben Slimane, J., & Maqbool, A. (2025). Hybrid statistical-textural and intensity feature-based approach for accurate discrimination of retinal diseases. **Journal of Information Systems Engineering & Management**, 10, 245–263. <https://doi.org/10.52783/jisem.v10i3s.385>
- Kumar, A. (2022). Learning texture features from GLCM for classification of brain tumor MRI images using random forest classifier. **WSEAS Transactions on Signal Processing**, 18, 60–63. <https://doi.org/10.37394/232014.2022.18.8>
- OpenCV Contributors. (2025). **opencv-python** (Version 4.11.0.86) [Computer software]. Retrieved from <https://pypi.org/project/opencv-python/>
- Pandas Development Team. (2024). **Pandas: Python data analysis library** (Version 2.2.3) [Computer software]. Retrieved from <https://pandas.pydata.org/>
- Scikit-Image Contributors. (2025). **Scikit-image** (Version 0.25.2) [Computer software]. Retrieved from <https://pypi.org/project/scikit-image/>
- SciPy Community. (2025). **SciPy** (Version 1.15.3) [Computer software]. Retrieved from <https://pypi.org/project/scipy>
- Wang, H. (2023). Research on the application of random forest-based feature selection algorithm in data mining experiments. **International Journal of Advanced Computer Science and Applications**, 14(10), 466–473. <https://doi.org/10.14569/IJACSA.2023.0141054>
- Waskom, M. L. (2021). seaborn: Statistical data visualization. **Journal of Open Source Software**, 6(60), Article 3021. <https://doi.org/10.21105/joss.0302>