

บทความวิจัย (Research Article)

การเปรียบเทียบประสิทธิภาพการจำแนกภาพถ่ายเนื้องอกในสมองด้วยแบบจำลองการเรียนรู้ของเครื่องแบบรวมกลุ่มร่วมกับการสกัดคุณลักษณะเวฟเลตส์ทรานสฟอร์ม

Comparative Evaluation of Ensemble Models for Brain Tumor Image Classification Using Wavelet Transform-Based Features

ชัชวาลย์ ศรีมนตรี¹ เสาวลักษณ์ ไทยกลาง^{1*} และ ชัยวิชิต แก้วกลม¹

Chatchawarn Srimontree¹, Saowaluk Thaiklang^{1*} and Chaivichit Kaewklom¹

¹ สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี

¹ Department of Computer Science, Faculty of Computer Science, Ubon Ratchathani Rajabhat University

*Corresponding author email: saowaluk.t@ubru.ac.th

วันที่รับบทความ (Received) วันที่ได้รับบทความฉบับแก้ไข (Revised) วันที่ตอบรับบทความ (Accepted)
27 กันยายน 2568 6 พฤศจิกายน 2568 6 พฤศจิกายน 2568

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาและประเมินประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องในการจำแนกภาพถ่ายสมองที่มีภาวะเนื้องอกด้วยภาพถ่ายคลื่นแม่เหล็กไฟฟ้าหรือภาพเอ็มอาร์ไอ โดยกระบวนการเริ่มจากการสกัดคุณลักษณะเชิงพื้นที่และความถี่หลายระดับด้วยวิธีเวฟเลตส์ทรานสฟอร์มแบบไม่ต่อเนื่องประเภท Daubechies 4 ซึ่งประกอบด้วยข้อมูลประมาณค่าของภาพ (LL) จำนวน 1 ซับแบนด์, ข้อมูลขอบแนวนอน (LH) จำนวน 3 ซับแบนด์, ข้อมูลขอบแนวตั้ง (HL) จำนวน 3 ซับแบนด์, ข้อมูลขอบแนวทแยง (HH) จำนวน 3 ซับแบนด์รวมทั้งสิ้น 10 ซับแบนด์ โดยแต่ละซับแบนด์ประกอบด้วยค่าสถิติเชิงปริมาณ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน พลังงาน เอนโทรปี ความเบ้ และความสูงชันของพื้นผิว ซึ่งคุณลักษณะดังกล่าวสามารถแสดงรายละเอียดเชิงโครงสร้างของเนื้อเยื่อสมองได้อย่างชัดเจนจำนวนทั้งสิ้น 60 คุณลักษณะ จากนั้นดำเนินการคัดเลือกคุณลักษณะที่มีความสำคัญด้วยวิธีแรนดอมฟอเรสต์โดยกำหนดเกณฑ์คัดเลือกเฉพาะคุณลักษณะที่มีค่าความสำคัญมากกว่า 0.01 เพื่อเพิ่มความแม่นยำและลดมิติของข้อมูลทำให้ได้คุณลักษณะของเวฟเลตส์ทั้งสิ้น 39 คุณลักษณะจากคุณลักษณะทั้งหมด ซึ่งสามารถลดคุณลักษณะที่ไม่จำเป็นลงถึงร้อยละ 35 ในขั้นตอนการสร้างแบบจำลองการเรียนรู้ของเครื่องได้ประยุกต์วิธีการเรียนรู้ของเครื่องแบบรวมกลุ่มได้แก่ แบบจำลองอาดาดูสต์ (Adaboost หรือ Adaptive Boosting) แบบจำลองเอ็กซ์ตรีมกราดิเอนต์บูสต์ (Extreme Gradient Boost หรือ XGBoost) แบบจำลองแรนดอมฟอเรสต์ (Random Forest) และแบบจำลองแบกกิง (Bagging หรือ Bootstrap Aggregation) เพื่อนำมาเปรียบเทียบประสิทธิภาพของแบบจำลอง ผลการทดลองพบว่าแบบจำลองอาดาดูสต์มีประสิทธิภาพดีที่สุดโดยมีความแม่นยำร้อยละ 93.67 ความไวร้อยละ 94.78 ความจำเพาะร้อยละ 93.00 และประสิทธิภาพโดยรวมร้อยละ 93.81 สะท้อนให้เห็นว่าแนวทางสกัดเวฟเลตส์ทรานสฟอร์มและคัดเลือกคุณลักษณะร่วมกับวิธีการเรียนรู้แบบอาดาดูสต์ มีศักยภาพสูงในการช่วยสนับสนุนการวินิจฉัยทางการแพทย์จากภาพถ่ายสมองได้อย่างมีประสิทธิภาพและแม่นยำสามารถเพิ่มทางเลือกในการวินิจฉัยโรคเนื้องอกในสมอง

คำสำคัญ: การจำแนกภาพถ่ายทางการแพทย์, เนื้องอกในสมอง, เวฟเลตส์ทรานสฟอร์ม, การเรียนรู้ของเครื่องแบบรวมกลุ่ม, การคัดเลือกคุณลักษณะ

Abstract

This study aims to develop and evaluate the performance of machine learning models for classifying magnetic resonance (MRI) brain images with tumor conditions. The process began with the extraction of multi-level texture and frequency features using the Discrete Wavelet Transform (DWT) with Daubechies 4 as the mother wavelet. The decomposition produced a total of ten subbands, including one approximation subband (LL), three horizontal detail subbands (LH), three vertical detail subbands (HL), and three diagonal detail subbands (HH). Each subband was characterized by six statistical parameters: mean, standard deviation, energy, entropy, skewness, and kurtosis, which effectively represented the structural characteristics of brain tissues. In total, 60 features were obtained. Feature selection was then performed using the Random Forest algorithm based on feature importance scores, where only features with importance values greater than 0.01 were retained to enhance accuracy and reduce data dimensionality. This process resulted in 39 selected wavelet features from the all original ones, corresponding to a 35% reduction in unnecessary attributes. For model construction, four ensemble learning techniques were implemented, Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), Random Forest, and Bootstrap Aggregating (Bagging), to compare classification performance. Experimental results revealed that the AdaBoost model achieved the best performance, with an accuracy of 93.67%, sensitivity of 94.78%, specificity of 93.00%, and overall performance of 93.81%. These findings reflect that the integration of wavelet transform-based feature extraction and feature selection with ensemble learning methods, particularly AdaBoost, can effectively and accurately support medical diagnosis from brain MRI images, providing an additional approach for brain tumor detection.

Keywords: Medical Image Classification, Brain Tumor, Wavelet Transform, Ensemble Learning, Feature Selection

บทนำ

เนื้องอกในสมอง คือ การเจริญเติบโตของเซลล์ที่ผิดปกติภายในสมอง ซึ่งอาจมีแหล่งกำเนิดจากเซลล์ประสาทภายในสมอง เยื่อหุ้มสมอง ต่อมใต้สมอง หรือโพรงสมอง เซลล์ภายในระบบประสาทส่วนกลาง หรือเกิดจากการกระจายของเซลล์มะเร็งจากอวัยวะอื่น เช่น ปอดหรือเต้านม เนื้องอกในสมองส่งผลกระทบต่อการทำงานของสมองในตำแหน่งที่เกิดพยาธิสภาพของโรค ซึ่งอาจนำไปสู่ความผิดปกติทางระบบประสาท ความพิการ หรือการเสียชีวิตของผู้ป่วย [1] โรคเนื้องอกในสมองเป็นโรคทางระบบประสาทศาสตร์ที่พบมากที่สุด 1 ใน 5 อันดับแรก โดยมีอัตราการเกิด 18.16 รายต่อประชากร 100,000 รายต่อปี โดยอาการของโรคส่งผลให้ผู้ป่วยมีอาการชัก มีความบกพร่องในการเคลื่อนไหว การรับรู้ในการสัมผัส การมองเห็น การสื่อสาร และความสามารถควบคุมร่างกายและความคิดลดลง ส่งผลกระทบในการใช้ชีวิตประจำวันของผู้ป่วยโดยตรงและต้องพึ่งพาของบุคคลรอบข้าง โดยการรักษาโรคนี้อยู่ในรูปแบบของการฉายรังสี การใช้เคมีบำบัด หรือแม้กระทั่งการผ่าตัด ทั้งนี้หลังกระบวนการรักษาผู้ป่วยร้อยละ 18.5 มีอาการที่แย่ลง [2] โดยการคัดกรองผู้ป่วยนิยมใช้การวิเคราะห์ภาพถ่ายคลื่นแม่เหล็กไฟฟ้า (Magnetic Resonance Imaging หรือ MRI) เพื่อหาภาวะของโรคในระยะเริ่มต้นซึ่งทำให้ผู้ป่วยสามารถเข้ารับการรักษาได้ทันเวลาที่ มีอัตราการรอดชีวิตสูง ลดผลกระทบจากการรักษาได้ดี อย่างไรก็ตามการคัดกรองอาจส่งผลกระทบในด้านลบได้ในเชิงปฏิบัติเช่น ภาวะแทรกซ้อนจากกระบวนการตรวจ ความสิ้นเปลืองงบประมาณและเวลา

โดยเฉพาะอย่างยิ่งในกรณีของผลตรวจบวกซึ่งก่อให้เกิดความวิตกกังวลที่เพิ่มขึ้นในผู้ที่ได้รับผลตรวจ [3] จากการสังเคราะห์งานวิจัยที่เกี่ยวข้องพบว่าในปัจจุบันการจำแนกภาพเนื้องอกในสมองด้วยเทคนิคการเรียนรู้ของเครื่องได้รับความนิยมน้อยกว่าหลาย โดยเฉพาะแนวทางที่ใช้แบบจำลองแบบรวมกลุ่มร่วมกับการแปลงสัญญาณภาพด้วยเวฟเลตส์ทรานส์ฟอร์ม อย่างไรก็ตามแม้วิธีการดังกล่าวจะให้ความแม่นยำสูงแต่มีข้อจำกัดสำคัญในด้านการใช้ทรัพยากรการคำนวณและประมวลผลที่สูง โดยเฉพาะแบบจำลองแบบรวมกลุ่มที่มีการใช้แบบจำลองพื้นฐานที่ซับซ้อนในการเรียนรู้ข้อมูล ยกตัวอย่างเช่นแบบจำลองเอาดาบัสต์ที่ใช้แบบจำลองแรนดอมฟอเรสต์หรือแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนในการเรียนรู้ข้อมูลส่งผลต่อความเป็นไปได้ในการนำไปใช้งานจริงในสภาพแวดล้อมทางคลินิกที่มีข้อจำกัด เช่นทรัพยากรด้านบุคลากรและด้านเครื่องคอมพิวเตอร์ เป็นต้น จากข้อจำกัดดังกล่าวจึงมีความจำเป็นที่จะต้องพัฒนาแนวทางการจำแนกภาพเนื้องอกในสมองที่มีประสิทธิภาพและความแม่นยำในระดับที่เชื่อถือได้แต่มีความซับซ้อนและใช้ทรัพยากรในการคำนวณและประมวลผลที่น้อยเพื่อให้สามารถประมวลผลได้รวดเร็วและเหมาะสมกับการใช้งานในสถานพยาบาลหรือหน่วยงานที่เกี่ยวข้องที่มีทรัพยากรจำกัด โดยงานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาและประเมินประสิทธิภาพของแบบจำลองการจำแนกภาพเนื้องอกในสมองที่มีสถาปัตยกรรมแบบเรียบง่ายโดยเน้นการสร้างสมดุลระหว่างความแม่นยำในการจำแนก ความเร็วในการประมวลผล และการใช้ทรัพยากรการคำนวณที่เหมาะสม เพื่อเป็นทางเลือกที่มีประสิทธิภาพและใช้งานได้จริงในบริบททางคลินิก [4, 5, 6, 7, 8] โดยงานวิจัยนี้มุ่งเน้นพัฒนาและเปรียบเทียบประสิทธิภาพแบบจำลองแบบรวมกลุ่มในการจำแนกภาพเนื้องอกในสมองโดยอาศัยข้อมูลจากการสกัดคุณลักษณะเวฟเลตส์ทรานส์ฟอร์มซึ่งเป็นที่ยอมรับและนิยมประยุกต์ใช้ในงานวิจัยทั้งนี้ในบริบททางการแพทย์เวฟเลตส์ทรานส์ฟอร์มสามารถสกัดคุณลักษณะที่สะท้อนลักษณะทางพื้นผิวและขอบของรอยโรคซึ่งเป็นข้อมูลสำคัญสำหรับการวินิจฉัยในหลายระดับความละเอียด มีความทนทานต่อสัญญาณรบกวน สามารถแยกสัญญาณออกเป็นส่วนประกอบตามความถี่ต่าง ๆ ซึ่งความหลากหลายนี้เป็นปัจจัยสำคัญที่ทำให้การเรียนรู้แบบรวมกลุ่มทำงานได้ดีเนื่องจากสามารถสร้างเงื่อนไขการแบ่งแยกข้อมูลได้มีอย่างมีประสิทธิภาพ โดยเฉพาะอย่างยิ่งในแบบจำลองซึ่งใช้ต้นไม้ตัดสินใจเป็นแบบจำลองพื้นฐานในการเรียนรู้ข้อมูล โดยข้อมูลคุณลักษณะเวฟเลตส์ทรานส์ฟอร์مدังกล่าวจะถูกนำเข้าสู่วิธีการเรียนรู้ของเครื่องประเภทรวมกลุ่ม ซึ่งมีการปรับปรุงกระบวนการในขั้นตอนการเรียนรู้ของแบบจำลองทำให้แบบจำลองมีความเอนเอียงและความแปรปรวนของผลลัพธ์น้อยและมีความแม่นยำในการจำแนกสูงจำนวน 4 แบบจำลอง ได้แก่ เอาดาบัสต์ เอ็กริมกราเดียนบัสต์ แรนดอมฟอเรสต์ และแบกกิ้ง พร้อมกับวัดและเปรียบเทียบประสิทธิภาพการจำแนกภาพถ่ายโดยอาศัยค่ามาตรฐานตามระเบียบวิธีวิจัยซึ่งประกอบไปด้วย ตารางเมตริกซ์ความสับสน (Confusion Matrix) ค่าความแม่นยำ (Accuracy) ค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) และประสิทธิภาพโดยรวมของระบบ (F-measure) ทั้งนี้เพื่อหาแบบจำลองที่ดีที่สุดในการจำแนกภาพถ่ายที่มีภาวะโรคเนื้องอกและเพื่อสนับสนุนการวินิจฉัยและรักษาโรคเนื้องอกในสมองภายใต้บริบททางคลินิกต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาแบบจำลองในการคัดกรองผู้ป่วยเนื้องอกในสมองด้วยภาพถ่ายทางการแพทย์โดยใช้วิธีการเรียนรู้ของเครื่องแบบรวมกลุ่ม
2. เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองในการคัดกรองผู้ป่วยเนื้องอกในสมองด้วยภาพถ่ายทางการแพทย์โดยใช้วิธีการเรียนรู้ของเครื่องแบบรวมกลุ่ม

วิธีดำเนินการวิจัย

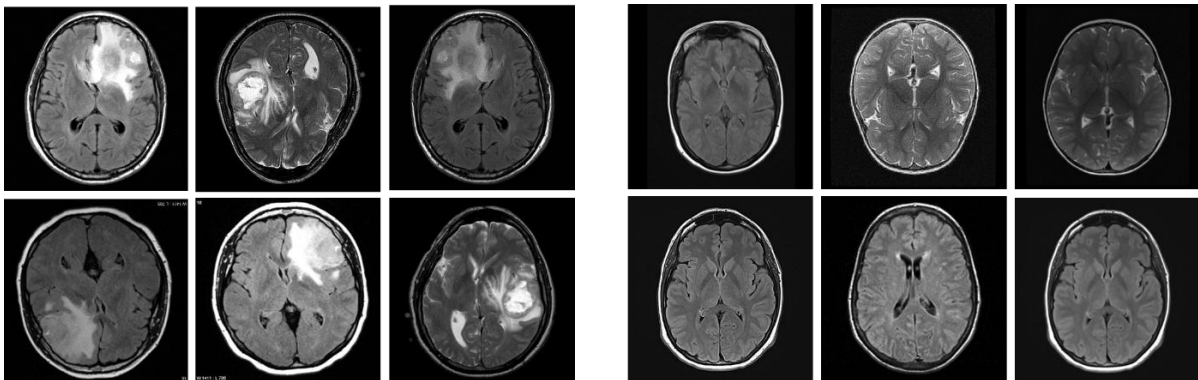
การพัฒนาแบบจำลองดำเนินการตามกระบวนการ CRISP-DM (Cross-Industry Standard Process for Data Mining) ซึ่งประกอบด้วยกระบวนการย่อย 6 ขั้นตอน [9] โดยมีรายละเอียด ดังนี้

การทำความเข้าใจปัญหา

โรคเนื้องอกในสมองเป็นโรคที่สามารถเกิดขึ้นได้ในทุกเพศทุกวัยและส่งผลกระทบต่อการใช้ชีวิตประจำวันของผู้ป่วย และบุคคลรอบข้าง มีการรักษาโรคที่ซับซ้อน และมีผลกระทบจากการรักษาในอัตราที่สูง ก่อให้เกิดความพิการหรือเสียชีวิต ทั้งนี้การมีแนวทางปฏิบัติ เครื่องมือ หรือระบบสารสนเทศสำหรับสนับสนุนการคัดกรองภาวะของโรคสามารถลดความเสี่ยงข้างต้น โดยงานวิจัยนี้มุ่งเปรียบเทียบประสิทธิภาพแบบจำลองเพื่อหาแบบจำลองที่มีประสิทธิภาพสูงที่สุดเพื่อสนับสนุนการคัดกรองภาพถ่ายทางการแพทย์และวินิจฉัยโรคเนื้องอกในสมอง

การทำความเข้าใจข้อมูล

การวิจัยครั้งนี้ใช้ภาพถ่ายสมองในแนวตัดขวางในรูปแบบภาพถ่ายคลื่นแม่เหล็กไฟฟ้า ซึ่งรวบรวมโดย Ahmed Hamada ซึ่งเผยแพร่บนเว็บไซต์ www.kaggle.com จำนวน 314 ภาพ แบ่งเป็นข้อมูลภาพถ่ายสมองซึ่งมีภาวะเนื้องอกในสมอง 157 ภาพ และภาพสมองปกติจำนวน 157 ภาพ โดยภาพถ่ายแต่ละภาพมีขนาดรูปภาพ 512 x 512 พิกเซลและมีความละเอียดภาพ 96 ดิพีไอ [10]



ภาพที่ 1 แสดงตัวอย่างภาพถ่ายสมองโดยภาพซ้ายคือภาพที่มีภาวะโรคเนื้องอกในสมองและภาพขวาคือภาพสมองปกติ

แหล่งที่มา: <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection>

สืบค้นเมื่อวันที่ 9 เมษายน พ.ศ. 2568

การเตรียมข้อมูล

การเตรียมข้อมูลภาพถ่ายก่อนสกัดคุณลักษณะมีความสำคัญเนื่องจากช่วยลดสัญญาณรบกวนและปรับมาตรฐานข้อมูลให้สม่ำเสมอ ทำให้ได้ค่าคุณลักษณะที่มีคุณภาพและสะท้อนลักษณะที่แท้จริงของภาพ ส่งผลให้การจำแนกมีความแม่นยำสูงขึ้น กระบวนการเตรียมข้อมูลของงานวิจัยมีการดำเนินการตามลำดับขั้นตอน ดังนี้

1. การเพิ่มขอบภาพสีดำ กระบวนการเพิ่มขอบภาพด้วยค่า 0 เพื่อสร้างพื้นที่สีดำบริเวณขอบภาพ (Zero Padding) วิธีการนี้เหมาะสมกับลักษณะของภาพถ่ายทางคลื่นแม่เหล็กไฟฟ้า เนื่องจากสามารถปรับขนาดภาพให้มีขนาดมาตรฐานได้โดยไม่กระทบต่อรายละเอียดสำคัญภายในภาพ ช่วยคงสภาพข้อมูลภาพเดิมได้อย่างครบถ้วน [11] ทั้งนี้หลังจากผ่านกระบวนการเพิ่มขอบภาพสีดำแล้วภาพถ่ายทั้งหมดจะถูกปรับให้มีขนาดกว้าง 512 พิกเซล และยาว 512 พิกเซล ซึ่งเป็นขนาดที่เหมาะสม

และเป็นมาตรฐานที่นิยมใช้ในงานวิจัยด้านการวิเคราะห์ภาพถ่ายทางการแพทย์ โดยเฉพาะภาพ เอ็มอาร์ไอหรือภาพซีที เนื่องจากมีความละเอียดเพียงพอในการแสดงโครงสร้างและรายละเอียดเชิงพื้นที่ได้อย่างครบถ้วน

2. แปลงภาพให้อยู่ในรูปแบบสเกลสีเทา ภาพต้นฉบับที่ผ่านกระบวนการเพิ่มขอบภาพสีดำทั้งหมดจะถูกแปลงให้อยู่ในรูปแบบภาพแบบสเกลสีเทา (Grayscale) เพื่อเพิ่มความชัดเจนในการแสดงโครงสร้างภายในเนื้อสมองโดยเฉพาะบริเวณที่มีรอยโรคหรือเนื้องอกซึ่งจะปรากฏให้เห็นได้อย่างชัดเจนมากยิ่งขึ้น ทั้งนี้หลังจากการแปลงภาพเป็นภาพในสเกลสีเทาแล้วจะทำให้ได้ภาพที่มีความแตกต่างเชิงความเข้มสีที่เหมาะสมต่อการแยกขอบด้วยเทคนิคเวฟเลตส์ทรานส์ฟอร์ม ซึ่งการประมวลผลภาพในลักษณะดังกล่าวจะช่วยสนับสนุนให้การสกัดคุณลักษณะในย่านความถี่ต่าง ๆ เป็นไปอย่างมีประสิทธิภาพและแม่นยำมากขึ้น

3. การปรับสมดุลฮิสโตแกรมแบบจำกัดคอนทราสต์ ภาพถ่ายที่ผ่านการแปลงให้อยู่ในรูปเกรย์สเกลจะได้รับการปรับปรุงคุณภาพเพิ่มเติมด้วยเทคนิคการปรับสมดุลฮิสโตแกรมแบบจำกัดคอนทราสต์ (Contrast Limited Adaptive Histogram Equalization: CLAHE) หลังจากประมวลผลด้วยกระบวนการนี้ภาพจะมีคอนทราสต์ที่ดีขึ้นในแต่ละบริเวณ การกระจายความเข้มแสงมีความสมดุล รายละเอียดในบริเวณสว่างและมีมีความชัดเจนขึ้นโดยไม่สูญเสียข้อมูลเหมาะสำหรับงานวิเคราะห์ภาพที่ต้องการความคมชัดสูง สามารถเพิ่มความคมชัดของโครงสร้างภายในภาพ เทคนิคนี้จะช่วยกระจายค่าความเข้มของพิกเซลให้มีการกระจายตัวที่เหมาะสม ลดปัญหาการเกิดบริเวณที่มีคอนทราสต์สูงเกินไปหรือไม่สม่ำเสมอ ซึ่งเป็นประโยชน์โดยตรงต่อกระบวนการสกัดข้อมูลคุณลักษณะด้วยวิธีเวฟเลตส์ทรานส์ฟอร์ม เนื่องจากคุณภาพของภาพที่มีรายละเอียดชัดเจนจะช่วยให้การแยกขอบและการสกัดสัญญาณเชิงพื้นที่ในย่านความถี่ต่าง ๆ มีความถูกต้องมากยิ่งขึ้น ปัจจุบันจึงมีส่วนสนับสนุนให้แบบจำลองการเรียนรู้ของเครื่องสามารถจำแนกข้อมูลได้อย่างแม่นยำและมีประสิทธิภาพสูงขึ้น [12]

4. การสกัดคุณลักษณะเวฟเลตส์ทรานส์ฟอร์ม งานวิจัยนี้ใช้ภาษาไพธอนและไลบรารีที่เกี่ยวข้อง ได้แก่ ไลบรารีโอเพนซีวีสำหรับภาษาไพธอน เวอร์ชัน 4.11.0.86 [13], ซายคิโทมเมจ เวอร์ชัน 0.25.2 [14], ซายไพ เวอร์ชัน 1.15.3 [15], แพนดาส เวอร์ชัน 2.2.3 [16] เป็นเครื่องมือในการสกัดคุณลักษณะจากภาพโดยอาศัยเทคนิคการสกัดเวฟเลตส์แบบไม่ต่อเนื่อง (Discrete Wavelet Transform หรือ DWT) โดยใช้เวฟเลตส์ชนิด Daubechies 4 (db4) สกัดคุณลักษณะ ได้แก่ ข้อมูลประมาณค่าของภาพ (LL3) จำนวน 1 ชั้นแบนด์, ข้อมูลขอบแนวนอน (LH3, LH2, LH1) จำนวน 3 ชั้นแบนด์, ข้อมูลขอบแนวตั้ง (HL3, HL2, HL1) จำนวน 3 ชั้นแบนด์, ข้อมูลขอบแนวทแยง (HH3, HH2, HH1) จำนวน 3 ชั้นแบนด์ ทั้งนี้แต่ละชั้นแบนด์ประกอบด้วยค่าสถิติเชิงปริมาณ ได้แก่ ค่าเฉลี่ย (Mean), ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation), พลังงาน (Energy), เอนโทรปี (Entropy), ความเบ้ (Skewness) และ ความสูงชัน (Kurtosis) ทำให้ได้คุณลักษณะของเวฟเลตส์รวมทั้งสิ้น 60 คุณลักษณะจาก 10 ชั้นแบนด์ ซึ่งคุณสมบัติของเวฟเลตส์เป็นข้อมูลประเภทหนึ่งที่น่าจะใช้ในการจำแนกภาพถ่ายสมองด้วยคลื่นแม่เหล็กไฟฟ้าสอดคล้องกับงานวิจัยของ Mathew และ Anto [17] โดยค่าสถิติแต่ละค่าสามารถประยุกต์ใช้เพื่อบ่งบอกลักษณะทางพยาธิสภาพ ดังรายละเอียดต่อไปนี้

4.1 ค่าเฉลี่ย ค่าเฉลี่ยแสดงความเข้มเฉลี่ยของสัญญาณในภาพเอ็มอาร์ไอหลังการแปลงด้วยเวฟเลตส์ ซึ่งสะท้อนความหนาแน่นของเนื้อเยื่อ เมื่อค่าเฉลี่ยสูงจะบ่งชี้ว่าเนื้อเยื่อที่มีความหนาแน่นสูง มักพบในเนื้องอกซึ่งมีเซลล์หนาแน่นทำให้มีแนวโน้มเป็นเนื้องอก ในขณะที่ค่าเฉลี่ยต่ำจะแสดงเนื้อเยื่อที่มีความหนาแน่นต่ำ เช่น เนื้อเยื่อสมองปกติหรือบริเวณที่เซลล์ตายจึงมีแนวโน้มเป็นเนื้อเยื่อปกติ

4.2 ส่วนเบี่ยงเบนมาตรฐาน ส่วนเบี่ยงเบนมาตรฐานวัดความแปรปรวนของข้อมูลในภาพสะท้อนความเป็นเนื้อเดียวกันของเนื้อเยื่อ ค่าสูงจะบ่งบอกเนื้อเยื่อที่มีความหลากหลายยกตัวอย่างเช่นมีทั้งส่วนแข็งและนิ่มปะปนกัน ซึ่งมีแนวโน้มเป็นเนื้องอกเนื่องจากเนื้องอกมักมีโครงสร้างที่ไม่สม่ำเสมอ ส่วนค่าต่ำจะแสดงเนื้อเยื่อที่มีลักษณะสม่ำเสมอ จึงมีแนวโน้มเป็นเนื้อเยื่อปกติที่มีการจัดเรียงเป็นระเบียบ

4.3 พลังงาน ค่าพลังงานวัดการกระจุกตัวของสัญญาณแสดงความชัดเจนของขอบเขตเนื้องอก ค่าสูงจะบ่งบอกขอบเขตที่ชัดเจน มีการแยกตัวจากเนื้อเยื่อโดยรอบอย่างเด่นชัดทำให้มีแนวโน้มเป็นเนื้องอกที่มีรูปร่างที่แน่นอน ในขณะที่ค่าต่ำจะแสดงขอบเขตที่ไม่เด่นชัดหรือเนื้อเยื่อที่มีลักษณะใกล้เคียงกันกับเนื้อเยื่อโดยรอบจึงมีแนวโน้มเป็นเนื้อเยื่อปกติ

4.4 เอนโทรปี ค่าเอนโทรปีบ่งบอกความไม่เป็นระเบียบของโครงสร้างเนื้อเยื่อ ค่าสูงจะบ่งชี้โครงสร้างที่ซับซ้อนและไม่เป็นระเบียบซึ่งมีแนวโน้มเป็นเนื้องอกโดยเฉพาะเนื้องอกที่รุนแรงที่มีการเจริญเติบโตแบบไม่เป็นระเบียบ ส่วนค่าต่ำจะแสดงโครงสร้างที่ปกติและเป็นระเบียบมีแนวโน้มเป็นเนื้อเยื่อปกติซึ่งมีรูปแบบการจัดเรียงที่สม่ำเสมอเป็นธรรมชาติ

4.5 ความเบ้ ค่าความเบ้แสดงความไม่สมมาตรของการกระจายข้อมูล บ่งบอกการมีจุดผิดปกติในเนื้อเยื่อเมื่อความเข้มมากไม่ว่าจะสูงมากเกินไปหรือต่ำมากเกินไปจะบ่งชี้ว่ามีบริเวณที่ผิดปกติ เช่น มีเลือดออกหรือแคลเซียมสะสมซึ่งมีแนวโน้มเป็นเนื้องอกที่มีส่วนประกอบที่หลากหลาย ในขณะที่ค่าใกล้ศูนย์จะแสดงการกระจายที่สมมาตรทำให้มีแนวโน้มเป็นเนื้อเยื่อปกติที่มีการกระจายแบบปกติ

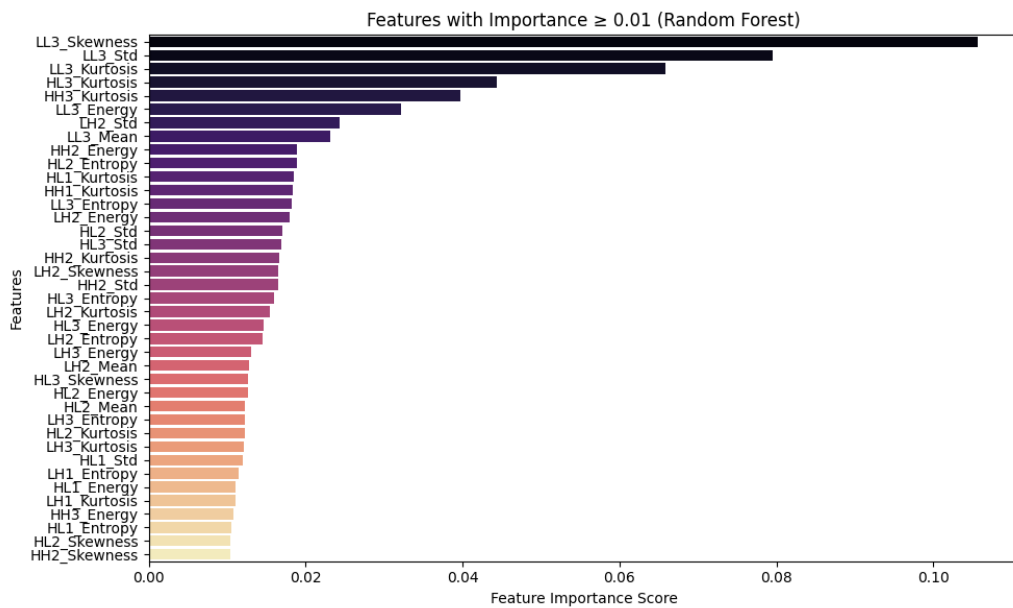
4.6 ความสูงชัน ค่าวัดความสูงของพื้นผิวและแสดงการผิดปกติในเนื้อเยื่อ ค่าสูงเกินไปจะบ่งบอกว่ามีความผิดปกติ เช่น บริเวณที่มีความสูงแตกต่างกันมากจะมีแนวโน้มเป็นเนื้องอกที่มีลักษณะไม่เป็นเนื้อเดียวกัน ส่วนค่าต่ำจะแสดงการกระจายที่สม่ำเสมอโดยไม่มีจุดที่ผิดปกติมีแนวโน้มเป็นเนื้อเยื่อปกติที่มีความสูงชันคงที่

การผสมผสานของค่าสถิติดังกล่าวก่อให้เกิดชุดข้อมูลของคุณลักษณะเชิงปริมาณที่สามารถทำหน้าที่เป็นตัวแทนในการบ่งชี้ลักษณะหรือภาวะเฉพาะของเนื้อเยื่อแต่ละประเภทได้อย่างมีประสิทธิภาพโดยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานให้ข้อมูลพื้นฐานเกี่ยวกับระดับโดยรวมและความแปรปรวนของสัญญาณ ค่าพลังงานและค่าเอนโทรปีสะท้อนความซับซ้อนและความเด่นชัดของโครงสร้าง ในขณะที่ค่าความเบ้และค่าความสูงชันช่วยตรวจจับลักษณะเชิงลึกของการแจกแจงที่ไม่สามารถตรวจพบได้ด้วยการประเมินเชิงสายตาเพียงอย่างเดียว การวิเคราะห์เชิงปริมาณเหล่านี้จึงมีส่วนสำคัญในการเพิ่มศักยภาพของระบบการเรียนรู้ของเครื่องสำหรับการจำแนกและเป็นทางเลือกในการวินิจฉัยความผิดปกติของเนื้อเยื่อจากภาพเอ็มอาร์ไอ

ตารางที่ 1 แสดงตัวอย่างคุณลักษณะของเวฟเลตส์ทรานสฟอร์มซึ่งสกัดจากภาพ

Image	LL3_Mean	LL3_Std	LL3_Energy	LL3_Entropy	LL3_Skewness	LL3_Kurtosis	LH3_Energy	Label
no0.jpg	1.714031	2.320465	40780.05	4.426665	1.098635	-0.12755	702.7213	NoTumor
no1013.jpg	1.705669	2.296569	40099.32	4.517891	1.07321	-0.21863	671.267	NoTumor
y583.jpg	2.093437	2.352014	48580.81	5.301418	0.697567	-0.90742	783.2516	Tumor
y586.jpg	2.461049	2.411077	58163.27	6.652478	0.598346	-0.87857	772.3447	Tumor

5. การคัดเลือกคุณสมบัติที่มีนัยสำคัญด้วยการวัดระดับความสำคัญของแต่ละคุณสมบัติด้วยแรนดอมฟอเรสต์ (Random Forest Importance) โดยคัดเลือกเฉพาะคุณสมบัติที่มีค่าความสำคัญมากกว่า 0.01 ทำให้ได้คุณลักษณะของเวฟเลตส์ทั้งสิ้น 39 คุณลักษณะจาก 60 คุณลักษณะ ซึ่งลดคุณลักษณะที่ไม่จำเป็นลงถึงร้อยละ 35



ภาพที่ 2 กราฟแสดงค่าความสำคัญของเวฟเลตส์จากมากไปหาน้อย

จากการวิเคราะห์คุณสมบัติที่สกัดได้ที่ได้พบว่าคุณสมบัติที่มีความสำคัญสูงสุด ได้แก่ LL3_Skewness (0.1057), LL3_Std (0.0795), และ LL3_Kurtosis (0.0659) ซึ่งสกัดจากซัฟแบนด์ประมาณค่าระดับ 3 รวมถึง HL3_Kurtosis (0.0443) และ HH3_Kurtosis (0.0397) จากซัฟแบนด์รายละเอียดแนวตั้งและแนวทแยง คุณสมบัติเหล่านี้สามารถตรวจจับลักษณะความไม่สมมาตร ความหลากหลายเชิงโครงสร้างและความซับซ้อนของการกระจายข้อมูลที่มากกว่าปกติ ซึ่งสะท้อนถึงลักษณะทางพยาธิวิทยาของเนื้องอกที่แตกต่างจากเนื้อเยื่อสมองปกติ เช่น การเจริญเติบโตของเซลล์ผิดปกติ การสะสมของเหลวหรือแคลเซียมและการเปลี่ยนแปลงของขอบเขตเนื้อเยื่อ จากการวิเคราะห์คุณลักษณะแต่ละซัฟแบนด์พบว่าคุณสมบัติจากซัฟแบนด์ LL3 มีความสำคัญสูงสุดแสดงให้เห็นว่าโครงสร้างหลักและลักษณะโดยรวมของภาพมีอิทธิพลต่อการจำแนกมากกว่ารายละเอียดปลีกย่อย นอกจากนี้ยังพบว่าค่าพลังงานและค่าเอนโทรปีของซัฟแบนด์ต่าง ๆ มีความสำคัญรองลงมา แสดงให้เห็นว่าคุณลักษณะเชิงพื้นที่วิหหลายมิติสามารถช่วยสนับสนุนการจำแนกกลุ่มภาพได้อย่างมีนัยสำคัญ ดังนั้นการคัดเลือกคุณลักษณะจึงควรให้ความสำคัญกับคุณลักษณะเชิงสถิติและพลังงานในย่านความถี่ที่เหมาะสมเพื่อเพิ่มประสิทธิภาพของแบบจำลอง

การสร้างแบบจำลอง

การสร้างแบบจำลอง งานวิจัยนี้ใช้แบบจำลองการเรียนรู้แบบรวมกลุ่มจำนวน 4 แบบจำลองโดยอาศัยซอฟต์แวร์ Altair AI Studio รุ่น 2025.0.1 [18] ในการพัฒนาและประเมินแบบจำลองซึ่งมีรายละเอียด ดังต่อไปนี้

1. อาดาบู้สตี มีกระบวนการทำงานโดยอาศัยการประมวลผลร่วมกันของแบบจำลองที่มีประสิทธิภาพต่ำหลายแบบจำลอง ซึ่งในงานวิจัยนี้ใช้ต้นไม้ตัดสินใจเป็นแบบจำลองการเรียนรู้พื้นฐาน เพื่อร่วมกันสร้างแบบจำลองที่มีประสิทธิภาพสูงกว่าเดิม อาดาบู้สตีใช้หลักการปรับน้ำหนักโดยให้ความสำคัญกับข้อมูลที่ทำนายผิดเพิ่มมากขึ้นเพื่อเพิ่มความแม่นยำในการเรียนรู้ในรอบถัดไป โดยในแต่ละรอบของการเรียนรู้ของแบบจำลองจะมีการกำหนดน้ำหนักใหม่ให้กับชุดข้อมูล

แต่ละชุดและแบบจำลองสามารถเรียนรู้จากน้ำหนักเหล่านั้นเพื่อปรับปรุงสมรรถนะโดยรวม [19] สามารถอธิบายขั้นตอนการทำงานพอสังเขปได้ ดังนี้

- 1) กระบวนการเริ่มต้น กำหนดน้ำหนักเท่ากันให้กับตัวอย่างข้อมูลทุกตัว
- 2) กระบวนการฝึกสอนให้แบบจำลองพื้นฐาน ใช้ข้อมูลที่มีการถ่วงน้ำหนักเพื่อฝึกสอนแบบจำลองพื้นฐานในรูปแบบต้นไม้ตัดสินใจหนึ่งต้น โดยกำหนดความลึกสูงสุดที่ 10 ระดับ ใช้ Gain Ratio เป็นเกณฑ์การแยกโหนด ใช้เทคนิค Pre-pruning โดยกำหนดค่า Minimal Gain ที่ 0.01 ใช้จำนวนตัวอย่างขั้นต่ำในลิฟโหนดที่ 2 ตัวอย่าง ใช้จำนวนตัวอย่างขั้นต่ำสำหรับการแยกโหนดที่ 4 ตัวอย่าง และจำนวนทางเลือกสำหรับ Pre-pruning ที่ 3 ทางเลือก นอกจากนี้ยังใช้เทคนิค Post-pruning ด้วยค่า Confidence ที่ 0.1 เพื่อควบคุมความซับซ้อนและป้องกันการโอเวอร์ฟิตติ้ง
- 3) กระบวนการคำนวณความผิดพลาด หาอัตราความผิดพลาดถ่วงน้ำหนักของตัวจำแนกที่ได้
- 4) กระบวนการกำหนดน้ำหนักตัวจำแนก ตัวจำแนกที่แม่นยำสูงจะได้ค่าน้ำหนักมาก ตัวจำแนกที่แม่นยำต่ำจะได้ค่าน้ำหนักน้อย
- 5) กระบวนการปรับปรุงน้ำหนักตัวอย่าง เพิ่มน้ำหนักให้กับตัวอย่างที่ถูกจำแนกผิด และลดน้ำหนักของตัวอย่างที่ถูกจำแนกถูก เพื่อให้รอบการทำงานถัดไปเกิดการเรียนรู้จากตัวอย่างที่ยากมากขึ้น
- 6) กระบวนการทำซ้ำ วนกลับไปทำขั้นตอนที่สองถึงห้าจนครบจำนวนรอบที่กำหนด
- 7) กระบวนการรวมตัวจำแนก นำผลการทำงานของแบบจำลองพื้นฐานทั้งหมดมารวมกันด้วยการลงคะแนนเสียงถ่วงน้ำหนักโดยตัวจำแนกที่แม่นยำจะมีน้ำหนักเสียงมากกว่าเพื่อสร้างตัวจำแนกที่มีประสิทธิภาพมากยิ่งขึ้น

2. เอ็กริมกราเดียนบูสต์ เป็นแบบจำลองซึ่งอาศัยการรวมแบบจำลองหลายชุดเข้าด้วยกัน โดยรูปแบบการเรียนรู้แบบลำดับซึ่งแต่ละแบบจำลองย่อยมักมีโครงสร้างที่ไม่ซับซ้อน การเรียนรู้ของแบบจำลองจะดำเนินไปโดยนำข้อผิดพลาดที่เกิดจากแบบจำลองก่อนหน้ามาใช้ในการปรับปรุงเพื่อใช้ในแบบจำลองถัดไป โดยกระบวนการนี้จะดำเนินไปอย่างต่อเนื่องในลักษณะไล่ระดับจนกว่าแบบจำลองจะไม่สามารถเพิ่มประสิทธิภาพได้มากกว่าเดิม [20] สามารถอธิบายขั้นตอนการทำงานพอสังเขปได้ ดังนี้

- 1) กระบวนการเริ่มต้น สร้างการทำนายเริ่มต้นด้วยค่าคงที่หรือค่าเฉลี่ยของตัวแปรเป้าหมาย
- 2) หาค่าความแตกต่างระหว่างค่าทำนายกับค่าจริง (Residual) และคำนวณค่า Gradient (อนุพันธ์อันดับหนึ่ง) และค่า Hessian (อนุพันธ์อันดับสอง) ของฟังก์ชันความผิดพลาด
- 3) กระบวนการสร้างต้นไม้ตัดสินใจ สร้างต้นไม้ตัดสินใจใหม่โดยใช้ค่าอนุพันธ์เป็นเป้าหมายในการเรียนรู้ โดยใช้อัลกอริทึม Tree Booster กำหนดวิธีการสร้างต้นไม้เป็นแบบอัตโนมัติและใช้ข้อมูลทั้งหมดในการสร้างต้นไม้ตัดสินใจแต่ละต้น (subsample = 1.0)
- 4) กระบวนการควบคุมความซับซ้อน จำกัดความลึกสูงสุดของต้นไม้ที่ 6 ระดับ (Max Depth = 6) กำหนดน้ำหนักขั้นต่ำของโหนดลูกที่ 1.0 (Min Child Weight = 1.0) กำหนดค่าความสูญเสียขั้นต่ำสำหรับการแยกโหนดที่ 0.0 (Min Split Loss = 0.0) และใช้ค่า Regularization แบบ L2 ที่ 1.0 (lambda = 1.0) และ L1 ที่ 0.0 (alpha = 0.0) เพื่อป้องกันการโอเวอร์ฟิตติ้ง
- 5) กระบวนการปรับน้ำหนักการเรียนรู้ เพิ่มค่าทำนายของต้นไม้ใหม่ด้วยอัตราการเรียนรู้ 0.3 (Learning Rate = 0.3) เพื่อควบคุมการปรับปรุงแบบค่อยเป็นค่อยไป
- 6) กระบวนการปรับปรุงการทำนาย เพิ่มค่าทำนายจากต้นไม้ใหม่ให้กับค่าทำนายสะสมจากรอบก่อนหน้า
- 7) กระบวนการทำซ้ำ วนกลับไปทำขั้นตอนที่สองถึงหกจนครบ 25 รอบ (rounds = 25) โดยไม่ใช้เทคนิค Early Stopping

8) กระบวนการทำนายสุดท้าย รวมการทำนายจากต้นไม้ทั้งหมดเข้าด้วยกันเพื่อได้ผลลัพธ์สุดท้าย

3. แรนดอมฟอร์เรสต์ เป็นเทคนิคการสร้างแบบจำลองโดยใช้ต้นไม้ตัดสินใจจำนวนหลายต้น ซึ่งแต่ละต้นจะถูกสร้างขึ้นจากการสุ่มเลือกข้อมูลและตัวแปรบางส่วน เมื่อต้นไม้ตัดสินใจแต่ละต้นทำการจำแนกข้อมูลเสร็จจะนำผลลัพธ์ทั้งหมดมารวมกันโดยใช้วิธีการนับเสียงข้างมากเพื่อเลือกคำตอบ แบบจำลองแรนดอมฟอร์เรสต์สามารถให้ผลการจำแนกที่แม่นยำและลดปัญหาในกรณีที่แบบจำลองมีความสามารถในการจำแนกข้อมูลฝึกได้แม่นยำสูงเกินไปได้ดีแต่มีประสิทธิภาพต่ำเมื่อประมวลผลกับข้อมูลใหม่หรือปัญหาโอเวอร์ฟิตต์ตั้งเนื่องจากใช้การรวมผลจากหลายแบบจำลองเข้าด้วยกัน [21] สามารถอธิบายขั้นตอนการทำงานพอสังเขปได้ ดังนี้

1) กระบวนการสุ่มตัวอย่างข้อมูล ดำเนินการสุ่มเลือกตัวอย่างข้อมูลจากชุดข้อมูลต้นฉบับแบบมีการแทนที่ (Bootstrap Sampling) โดยต้นไม้ตัดสินใจแต่ละต้นจะได้รับชุดข้อมูลที่แตกต่างกัน ในงานวิจัยนี้กำหนดให้ไม้ต้นไม้ตัดสินใจ 100 ต้น ความลึกสูงสุด 10 ระดับ และใช้ Gain Ratio เป็นเกณฑ์การแยกโหนด

2) กระบวนการสุ่มเลือกคุณลักษณะ ในแต่ละโหนดของต้นไม้จะสุ่มเลือกคุณลักษณะเพียงส่วนหนึ่งเพื่อพิจารณาในการแยกโหนดทั้งนี้เพื่อสร้างความหลากหลายระหว่างต้นไม้ตัดสินใจแต่ละต้น

3) กระบวนการสร้างต้นไม้ตัดสินใจ ดำเนินการสร้างต้นไม้ตัดสินใจจากชุดข้อมูลที่สุ่มได้ในขั้นตอนที่หนึ่ง โดยในแต่ละโหนดจะเลือกคุณลักษณะที่ให้ค่า Gain Ratio สูงสุดจากชุดคุณลักษณะที่สุ่มมาในขั้นตอนที่สองกระบวนการนี้ทำซ้ำจนกว่าจะหยุดตามเงื่อนไข ได้แก่ ความลึกสูงสุด ความบริสุทธิ์ของโหนด หรือจำนวนตัวอย่างขั้นต่ำในโหนด

4) กระบวนการควบคุมความซับซ้อน จำกัดความลึกสูงสุดของต้นไม้ตัดสินใจไว้ที่ 10 ระดับเพื่อควบคุมความซับซ้อนของแบบจำลองและป้องกันปัญหาการโอเวอร์ฟิตต์ การจำกัดความลึกนี้เป็นกลไกการควบคุมความซับซ้อนแบบ Pre-pruning ซึ่งแตกต่างจาก Post-pruning ที่ต้องปรับแต่งต้นไม้ตัดสินใจหลังจากสร้างเสร็จแล้ว วิธีการนี้ช่วยให้ต้นไม้ตัดสินใจแต่ละต้นมีขนาดที่เหมาะสมและยังคงรักษาความหลากหลายในการเรียนรู้รูปแบบของข้อมูล

5) กระบวนการทำซ้ำ วนซ้ำขั้นตอนที่หนึ่งถึงสี่จนได้ต้นไม้ตัดสินใจครบ 100 ต้น โดยต้นไม้ตัดสินใจแต่ละต้นสร้างอย่างเป็นอิสระ

6) กระบวนการรวมการทำนาย สำหรับการจำแนกใช้วิธีการลงคะแนนเสียงข้างมาก (Majority Voting) จากผลของต้นไม้ทั้งหมด โดยคลาสที่ได้รับคะแนนเสียงมากที่สุดจะเป็นผลการทำนายสุดท้าย

7) กระบวนการประเมินความสำคัญของคุณลักษณะ คำนวณค่าความสำคัญของแต่ละคุณลักษณะโดยพิจารณาจากค่าเฉลี่ยการลดลงของค่าความผิดพลาดหรือความไม่บริสุทธิ์ของโหนดเมื่อใช้คุณลักษณะนั้นในการแยกโหนดทั่วทั้งต้นไม้ทั้งหมดในป่าสุ่ม คุณลักษณะที่มีค่าความสำคัญสูงแสดงว่ามีส่วนสำคัญในการสร้างแบบจำลองการทำนาย

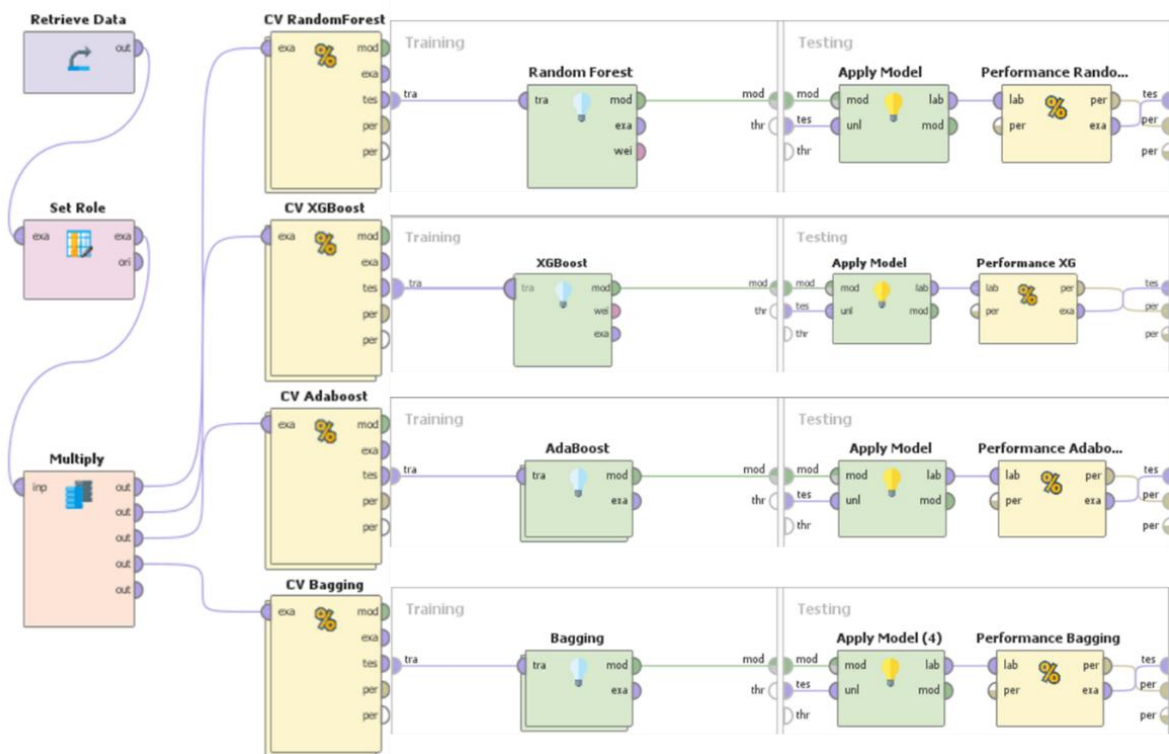
4. แบกกิ้ง เป็นแบบจำลองซึ่งอาศัยการสุ่มข้อมูลจากชุดฝึกสอนแบบจำลองด้วยการสุ่มแบบมีการแทนที่เพื่อสร้างชุดข้อมูลย่อยหลายชุด จากนั้นนำข้อมูลย่อยแต่ละชุดไปสร้างแบบจำลองย่อยด้วยแบบจำลองพื้นฐานที่เหมือนกัน ซึ่งในงานวิจัยนี้ใช้ต้นไม้ตัดสินใจเป็นแบบจำลองพื้นฐานพื้นฐาน และกระบวนการสุดท้ายแบบจำลองจะทำการรวมผลลัพธ์จากแบบจำลองทั้งหมดเพื่อเพิ่มความเสถียรและความแม่นยำในการพยากรณ์ [22] ทั้งนี้สามารถอธิบายขั้นตอนการทำงานพอสังเขปได้ ดังนี้

1) กระบวนการสุ่มตัวอย่างข้อมูล สุ่มเลือกตัวอย่างข้อมูลจากชุดข้อมูลต้นฉบับแบบมีการแทนที่โดยแต่ละชุดข้อมูลจะมีขนาดเท่ากับชุดข้อมูลต้นฉบับ แต่อาจมีตัวอย่างบางตัวซ้ำกันและบางตัวไม่ถูกเลือก

2) กระบวนการฝึกสอนแบบจำลองพื้นฐาน ใช้ชุดข้อมูลที่สุ่มได้แต่ละชุดเพื่อฝึกสอนแบบจำลองต้นไม้ตัดสินใจหนึ่งต้นอย่างเป็นอิสระ โดยกำหนดความลึกสูงสุดที่ 10 ระดับ ใช้ Gain Ratio เป็นเกณฑ์การแยกโหนด ใช้เทคนิค

Pre-pruning โดยกำหนดค่า Minimal Gain ที่ 0.01 จำนวนตัวอย่างขั้นต่ำในลิฟโหนดที่ 2 ตัวอย่าง จำนวนตัวอย่างขั้นต่ำสำหรับการแยกโหนดที่ 4 ตัวอย่าง และจำนวนทางเลือกสำหรับ Pre-pruning ที่ 3 ทางเลือก นอกจากนี้ยังใช้เทคนิค Post-pruning ด้วยค่า Confidence ที่ 0.1 เพื่อควบคุมความซับซ้อนและป้องกันการโอเวอร์ฟิตติ้ง

- 3) กระบวนการสร้างแบบจำลองพื้นฐานหลายตัว ทำซ้ำขั้นตอนที่หนึ่งและสองจนได้แบบจำลองครบตามจำนวนที่กำหนด โดยแต่ละแบบจำลองจะเรียนรู้จากข้อมูลที่มีความแตกต่างกันเล็กน้อย
- 4) กระบวนการรวมการจำแนก ใช้การลงคะแนนเสียงข้างมากจากแบบจำลองทั้งหมดในการจำแนก
- 5) กระบวนการประเมินความไม่แน่นอน สามารถวิเคราะห์ความแตกต่างของการทำนายจากแต่ละแบบจำลองเพื่อความมั่นใจในผลลัพธ์



ภาพที่ 3 แสดงการสร้างแบบจำลองและประเมินประสิทธิภาพแบบจำลอง

การประเมินผล

การศึกษาและวิจัยในครั้งนี้มีกระบวนการประเมินประสิทธิภาพของแบบจำลองโดยใช้เทคนิคการแบ่งชุดข้อมูลแบบไขว้ (Cross-Validation) โดยแบ่งข้อมูลทั้งหมดออกเป็น 10 ส่วนที่มีขนาดเท่ากัน โดยในการเรียนรู้แต่ละรอบแบบจำลองจะสุ่มเลือก 1 ส่วนเป็นชุดทดสอบ (Testing Set) และใช้ส่วนที่เหลืออีก 9 ส่วนเป็นชุดฝึกสอนแบบจำลอง กระบวนการนี้ถูกทำซ้ำทั้งหมด 30 รอบและนำผลลัพธ์ที่ได้จากแต่ละรอบมาเฉลี่ยเพื่อประเมินประสิทธิภาพของแบบจำลอง โดยวิธีการนี้เป็นมาตรฐานที่ได้รับการยอมรับอย่างกว้างขวางและเหมาะสมที่สุดในการสร้างสมดุลระหว่างความเอนเอียงและความแปรปรวนของการประมาณค่าสอดคล้องกับงานวิจัยโดย Allgaier & Pryss [23] และ Kohavi [24] ซึ่งช่วยลดปัญหาอัตราความแปรปรวนที่มีสูงจากการใช้จำนวนรอบน้อยเกินไปและหลีกเลี่ยงความเอนเอียงในอัตราที่สูงจากการใช้จำนวนรอบที่มากเกินไป พร้อมทั้งสร้างสมดุลระหว่างการคำนวณและความแม่นยำของการประเมิน นอกจากนี้การใช้ข้อมูลฝึกสอนถึงร้อยละ 90 ในแต่ละรอบ ทำให้ข้อมูลทั้งหมดถูกใช้ทั้งในกระบวนการฝึกสอนและกระบวนการทดสอบอย่างครบถ้วน

ส่งผลให้การประเมินสะท้อนความสามารถแบบจำลองได้อย่างแท้จริงสอดคล้องกับ Rodriguez et al. [25] และ Varoquaux et al. [26] และตัวชี้วัดในการประเมินประสิทธิภาพของแบบจำลองในครั้งนี้ประกอบด้วย ค่าความแม่นยำ ค่าความไว ค่าความจำเพาะ และค่าประสิทธิภาพโดยรวมของระบบซึ่งคำนวณร่วมกับตารางเมตริกซ์ความสับสนซึ่งมีรายละเอียดดังต่อไปนี้

1. ตารางเมตริกซ์ความสับสน (Confusion Matrix) คือ ผลของการจำแนกภาพถ่ายซึ่งประกอบด้วยการจำแนกคลาสจำนวน 2 คลาสได้แก่คลาส Positive คือกลุ่มภาพถ่ายซึ่งมีภาวะโรคเนื้องอกในสมองและคลาส Negative คือกลุ่มภาพถ่ายซึ่งไม่มีภาวะโรคเนื้องอกในสมองทั้งนี้ผลการจำแนกประกอบด้วย 4 รูปแบบ ได้แก่

- 1.1 True Negative (TN) คือ แบบจำลองจำแนกว่าไม่มีเนื้องอกในสมองและไม่มีเนื้องอกในสมองจริง
- 1.2 False Positive (FP) คือ แบบจำลองจำแนกว่ามีเนื้องอกในสมองแต่ไม่มีเนื้องอกในสมองจริง
- 1.3 False Negative (FN) คือ แบบจำลองจำแนกว่าไม่มีเนื้องอกในสมองแต่มีเนื้องอกในสมองจริง
- 1.4 True Positive (TP) คือ แบบจำลองจำแนกว่ามีเนื้องอกในสมองและมีเนื้องอกในสมองจริง

2. ค่าความแม่นยำ (Accuracy) คือ สัดส่วนของจำนวนครั้งที่แบบจำลองจำแนกผลทั้งกลุ่มมีเนื้องอกและไม่มีเนื้องอกในสมองได้ถูกต้องเปรียบเทียบกับจำนวนข้อมูลทั้งหมดที่ใช้ในการประเมินแบบจำลอง

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

3. ค่าความไว (Sensitivity) คือ สัดส่วนที่แบบจำลองสามารถพยากรณ์ผลได้ถูกต้องในกรณีที่มีภาวะโรคเนื้องอกในสมองเมื่อเทียบกับจำนวนที่มีภาวะโรคเนื้องอกในสมองทั้งหมด

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

4. ค่าความจำเพาะ (Specificity) คือ สัดส่วนที่แบบจำลองสามารถจำแนกกรณีที่ไม่มีเนื้องอกในสมองได้อย่างถูกต้องเมื่อเทียบกับจำนวนไม่มีเนื้องอกในสมองทั้งหมด

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

5. ค่าประสิทธิภาพโดยรวมของระบบ (F-measure) คือค่าที่ใช้ประเมินประสิทธิภาพของแบบจำลอง โดยวัดความสมดุลระหว่างความแม่นยำในการจำแนกภาพถ่ายได้ถูกต้อง (Precision) และ ความครบถ้วนในการตรวจจับภาพ (Recall) ในแต่ละคลาส

$$\text{Precision (Positive)} = TP / (TP + FP) \quad (4)$$

$$\text{Recall (Positive)} = TP / (TP + FN) \quad (5)$$

$$\text{Precision (Negative)} = TN / (TN + FP) \quad (6)$$

$$\text{Recall (Negative)} = TN / (TN + FN) \quad (7)$$

$$\text{F1-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

การนำไปใช้งาน

หลังจากได้แบบจำลองที่ดีที่สุดแล้วแบบจำลองดังกล่าวสามารถนำไปประยุกต์ใช้กับข้อมูลใหม่ได้โดยตรง เพื่อทำการจำแนกภาพถ่ายทางการแพทย์ตามวัตถุประสงค์ที่กำหนดไว้ โดยแบบจำลองจะถูกจัดเก็บในรูปแบบดิจิทัลหรือไฟล์แบบจำลองเพื่อให้สามารถเรียกใช้งานซ้ำในสภาพแวดล้อมจริงได้อย่างมีประสิทธิภาพ เช่น การประมวลผลภาพถ่ายทางการแพทย์ การคัดกรองความผิดปกติ หรือการตัดสินใจเชิงระบบ ทั้งนี้การนำไปใช้งานจริงควรดำเนินการควบคู่กับกระบวนการตรวจสอบคุณภาพและประเมินผลการทำงานอย่างต่อเนื่องเพื่อรักษาความถูกต้องและความเชื่อถือได้

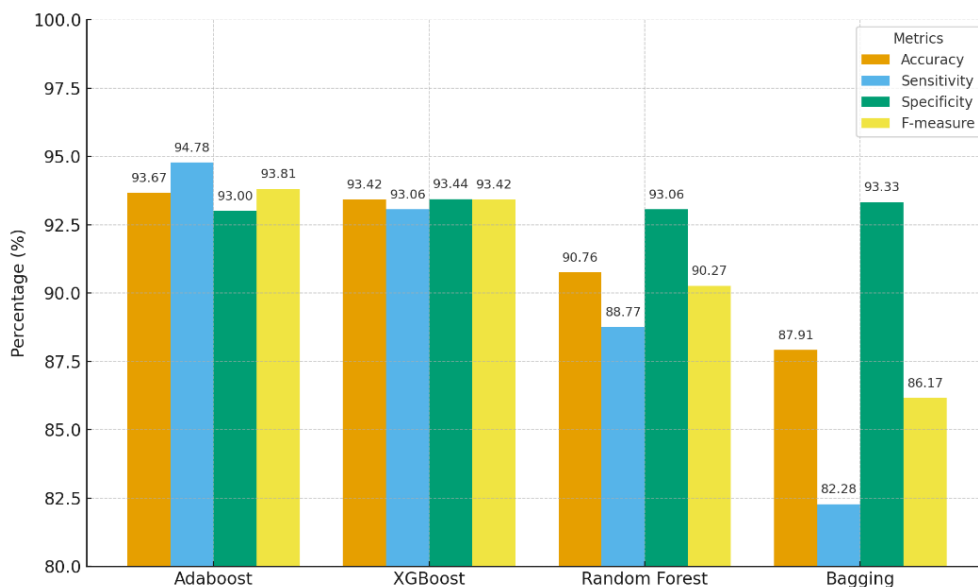
ผลการวิจัย

จากผลการจำแนกภาพถ่ายดังแสดงในตารางที่ 2 พบว่าแบบจำลองอาดาบุสต์มีความถูกต้องในการจำแนกภาพถ่ายกลุ่มที่มีภาวะเนื้องอกในสมองถูกต้อง (TP) ในอัตราที่สูงที่สุดที่ 148 ภาพ และสามารถจำแนกภาพถ่ายกลุ่มที่ไม่มีภาวะเนื้องอกในสมอง (TN) ได้ถูกต้อง 146 ภาพ ขณะเดียวกันมีการจำแนกภาพถ่ายกลุ่มที่มีภาวะเนื้องอกผิดพลาดการจำแนก (FN) เพียง 9 ราย ซึ่งต่ำที่สุดเมื่อเปรียบเทียบกับแบบจำลองซึ่งใช้ในงานวิจัยทั้งหมดและมีจำนวนการจำแนกภาพถ่ายกลุ่มที่ไม่มีภาวะเนื้องอกผิดพลาด (FP) เพียง 11 ภาพ แสดงให้เห็นว่าแบบจำลองอาดาบุสต์มีสมรรถนะที่ดีระหว่างความสามารถในการจำแนกภาพที่มีและไม่มีภาวะเนื้องอก แบบจำลองเอ็กตรีมกราเดียนบุสต์มีความสามารถในการจำแนก ภาพที่ไม่มีภาวะเนื้องอกได้ถูกต้องมากที่สุด (TN) ที่ 147 ภาพ และมีจำนวนการจำแนกภาพที่ไม่มีภาวะเนื้องอกผิดพลาด (FP) เพียง 10 ภาพ ซึ่งอยู่ในอัตราที่ต่ำที่สุดเมื่อเปรียบเทียบกับแบบจำลองอื่น ซึ่งบ่งชี้ถึงความแม่นยำในการจำแนกภาพถ่ายกลุ่มที่ไม่มีภาวะเนื้องอกในสมอง อย่างไรก็ตามอัตราการจำแนกภาพที่มีภาวะเนื้องอกผิดพลาด (FN) ที่ 11 ภาพ ซึ่งมีมากกว่าแบบจำลองอาดาบุสต์เล็กน้อยแสดงให้เห็นว่าแม้แบบจำลองเอ็กตรีมกราเดียนบุสต์สามารถลดความผิดพลาดในการจำแนกกลุ่มภาพที่ไม่มีภาวะเนื้องอกได้ดีแต่ยังมีแนวโน้มที่จะพลาดการตรวจจับภาพที่มีภาวะเนื้องอกในบางกรณี แบบจำลองแรนดอมฟอรัเรสต์แสดงผลการจำแนกภาพถ่ายกลุ่มที่ไม่มีภาวะเนื้องอกในสมองถูกต้อง (TN) ได้ใกล้เคียงกับแบบจำลองอื่นที่ 146 ภาพ และอัตราการจำแนกภาพที่ไม่มีภาวะเนื้องอกผิดพลาด (FP) ในระดับเดียวกับแบบจำลองอาดาบุสต์ที่ 11 ภาพ อย่างไรก็ตามอัตราการจำแนกภาพที่มีภาวะเนื้องอกผิดพลาด (FN) มีสูงถึง 18 ภาพ ซึ่งสูงกว่าแบบจำลองอาดาบุสต์อย่างชัดเจน บ่งบอกถึงข้อจำกัดในการจำแนกภาพที่มีภาวะเนื้องอกส่งผลให้ประสิทธิภาพโดยรวมในการวินิจฉัยโรคลดลง โดยเฉพาะในกรณีที่ต้องการความแม่นยำในการคัดกรองภาพถ่ายสมองที่มีภาวะของโรค แบบจำลองแบกกิ้งมีจำนวนการจำแนกภาพที่มีภาวะเนื้องอกในอัตราที่น้อยที่สุด (TP) ที่ 130 ภาพ และมีอัตราการจำแนกภาพที่มีภาวะเนื้องอกผิดพลาดสูงที่สุด (FN) ที่ 27 ภาพ ซึ่งสะท้อนถึงข้อจำกัดในการจำแนกภาพที่มีภาวะเนื้องอกอย่างชัดเจน ทั้งนี้อาจกล่าวได้ว่าแม้จะมีจำนวนการจำแนกภาพถ่ายกลุ่มที่ไม่มีภาวะเนื้องอกในสมองถูกต้อง (TN) เทียบเท่ากับแบบจำลองอาดาบุสต์และแบบจำลองแรนดอมฟอรัเรสต์ที่ 146 ภาพ แต่จำนวนภาพซึ่งมีภาวะเนื้องอกที่แบบจำลองแบกกิ้งจำแนกผิดพลาดอยู่ในอัตราที่สูงจึงอาจไม่เหมาะสมสำหรับงานวินิจฉัยทางการแพทย์ที่ต้องการลดการพลาดให้อยู่ในอัตราที่ต่ำที่สุด

ตารางที่ 2 แสดงผลการจำแนกของแบบจำลอง

แบบจำลอง	TN	FP	FN	TP
อาดาบุสต์	146	11	9	148
เอ็กตรีมกราเดียนบุสต์	147	10	11	146
แรนดอมฟอรัเรสต์	146	11	18	139
แบกกิ้ง	146	11	27	130

จากการประเมินและเปรียบเทียบประสิทธิภาพแบบจำลองพบว่าแบบจำลองอาดาบัสต์มีความแม่นยำโดยรวมสูงสุดที่ร้อยละ 93.67 สามารถจำแนกภาพถ่ายกลุ่มที่มีภาวะเนื้องอกในสมองได้ดีที่สุดโดยมีค่าความไวที่ร้อยละ 94.78 และมีค่าความจำเพาะที่ร้อยละ 93.00 บ่งบอกได้ว่าจะสามารถจำแนกภาพถ่ายกลุ่มที่ไม่มีภาวะเนื้องอกในสมองได้ดีเช่นกันส่งผลให้มีค่าประสิทธิภาพโดยรวมสูงสุดที่ร้อยละ 93.81 สะท้อนถึงความสมดุลระหว่างความแม่นยำและความครอบคลุมในอัตราที่สูง แบบจำลองเอ็กซ์ตรีมกราดิเียนบู้สต์ความแม่นยำและความไวโดยรวมที่ร้อยละ 93.42 และร้อยละ 93.06 ซึ่งใกล้เคียงกับอาดาบัสต์เป็นอย่างมาก และมีค่าความจำเพาะสูงสุดที่ร้อยละ 93.44 บ่งบอกถึงความสามารถในการจำแนกภาพถ่ายกลุ่มที่ไม่มีภาวะเนื้องอกในสมองได้ดีและมีความสมดุลระหว่างความแม่นยำและความครอบคลุมในอัตราที่สูง โดยมีค่าประสิทธิภาพโดยรวมที่ร้อยละ 93.42 แบบจำลองแรนดอมฟอเรสต์มีค่าความจำเพาะสูงที่ร้อยละ 93.06 บ่งบอกถึงความสามารถในการจำแนกภาพถ่ายกลุ่มที่ไม่มีภาวะเนื้องอกในสมองได้ดี แต่มีความแม่นยำที่ร้อยละ 90.76 มีค่าความไวที่ร้อยละ 88.77 และมีค่าประสิทธิภาพโดยรวมที่ร้อยละ 90.27 ซึ่งต่ำกว่าเอ็กซ์ตรีมกราดิเียนบู้สต์และอาดาบัสต์อย่างมีนัยสำคัญ ทั้งนี้เมื่อพิจารณาแบบจำลองแบกกิ้งพบว่าแม้จะมีค่าความจำเพาะในอัตราที่สูงที่ร้อยละ 93.33 ซึ่งบ่งบอกว่าสามารถจำแนกภาพถ่ายกลุ่มที่ไม่มีภาวะเนื้องอกในสมองได้ดี แต่มีความแม่นยำและค่าความไวโดยรวมต่ำที่สุดที่ร้อยละ 87.91 และร้อยละ 82.28 ซึ่งบ่งชี้ได้ว่าจะเกิดข้อผิดพลาดในการจำแนกภาพถ่ายกลุ่มที่มีภาวะเนื้องอกในสมองจำนวนมาก ส่งผลให้ประสิทธิภาพโดยรวมของระบบต่ำที่สุดที่ร้อยละ 86.17 สะท้อนถึงความไม่สมดุลในการจำแนกภาพถ่ายของแบบจำลองทั้งนี้ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองสามารถนำเสนอด้วยภาพที่ 4



ภาพที่ 4 กราฟแสดงการเปรียบเทียบประสิทธิภาพของแบบจำลอง

อภิปรายผล

งานวิจัยนี้มุ่งพัฒนาและเปรียบเทียบประสิทธิภาพการจำแนกภาพเนื้องอกในสมองในรูปแบบภาพถ่ายคลื่นแม่เหล็กไฟฟ้า โดยมีการปรับปรุงขนาดภาพให้มีขนาดกว้าง 512 พิกเซลและยาว 512 พิกเซล ปรับภาพให้อยู่ในรูปแบบในรูปแบบเกรย์สเกล และปรับปรุงความคมชัดของภาพด้วยเทคนิคการปรับสมดุลฮิสโตแกรมแบบจำกัดคอนทราสต์ ก่อนกระบวนการสกัดคุณลักษณะเวฟเลตส์ซึ่งสามารถบ่งบอกลักษณะพื้นผิวของภาพ โดยงานวิจัยนี้ประยุกต์ใช้เวฟเลตส์ชนิด Daubechies 4 ซึ่งประกอบด้วยข้อมูลพื้นฐานของภาพ (LL) จำนวน 1 ซับแบนด์, ข้อมูลขอบแนวนอน (LH) จำนวน 3

ซับบ้านด์, ข้อมูลขอบแนวตั้ง (HL) จำนวน 3 ซับบ้านด์, ข้อมูลขอบแนวทแยง (HH) จำนวน 3 ซับบ้านด์รวมทั้งสิ้น 10 ซับบ้านด์ โดยแต่ละซับบ้านด์ประกอบด้วยค่าสถิติเชิงปริมาณ ได้แก่ ค่าเฉลี่ย, ส่วนเบี่ยงเบนมาตรฐาน, พลังงาน, เอนโทรปี, ความเบ้ และความสูงชันทำให้ได้คุณลักษณะของเวฟเล็ตส์รวมทั้งสิ้น 60 คุณลักษณะ และคัดเลือกเฉพาะคุณสมบัติที่มีนัยสำคัญด้วยการวัดระดับความสำคัญด้วยแรนดอมฟอเรสต์ที่มีค่าความสำคัญมากกว่า 0.01 ทำให้เหลือคุณลักษณะของเวฟเล็ตส์ทั้งสิ้น 39 คุณลักษณะ และนำข้อมูลคุณลักษณะดังกล่าวเข้าสู่กระบวนการเรียนรู้ของเครื่องด้วยแบบจำลอง 4 แบบจำลองได้แก่ อาดาบุสต์ เอ็กริมกราเดียนบุสต์ แรนดอมฟอเรสต์ และแบกกิ้ง ทั้งนี้ผลการพัฒนาและเปรียบเทียบแบบจำลองพบว่าแบบจำลองอาดาบุสต์แสดงให้เห็นถึงประสิทธิภาพโดยรวมที่ดีที่สุดในทุกตัวชี้วัด เหมาะกับงานที่ต้องการความแม่นยำสูงทั้งด้านการตรวจพบภาวะโรคและการลดข้อผิดพลาดในการจำแนกข้อมูล มีความเหมาะสมในการประยุกต์ใช้ในการจำแนกภาพถ่ายโรคเนื้องอกในสมอง แบบจำลองเอ็กริมกราเดียนบุสต์เป็นแบบจำลองที่มีประสิทธิภาพทั้งด้านความสามารถในการจำแนกภาพที่มีและไม่มีภาวะโรคเนื้องอกในสมอง สามารถลดอัตราการจำแนกที่ผิดพลาดได้ดีหากแต่มีประสิทธิภาพน้อยกว่าแบบจำลองอาดาบุสต์เล็กน้อย แบบจำลองแรนดอมฟอเรสต์และแบบจำลองแบกกิ้งแม้มีระดับค่าความสามารถในการจำแนกกลุ่มที่ไม่มีภาวะของโรคที่สูง แต่มีค่าความสามารถในการจำแนกกลุ่มที่มีภาวะของโรคที่ต่ำกว่าแบบจำลองอื่นอาจไม่เหมาะในการถ่ายภาพโรคเนื้องอกในสมองเนื่องจากมีอัตราการจำแนกกลุ่มภาพที่มีภาวะโรคผิดพลาดในอัตราที่สูง

ทั้งนี้เมื่อเปรียบเทียบผลการวิจัยกับงานวิจัยที่มีระเบียบวิธีที่ใกล้เคียงกันพบว่างานวิจัยนี้มีประสิทธิภาพน้อยกว่างานวิจัยโดย Gopi Krishna et al. [4] ซึ่งใช้คุณสมบัติของเวฟเล็ตส์ที่คล้ายคลึงกันและมีความแม่นยำร้อยละ 98.7 โดยงานวิจัยนี้มีค่าความแม่นยำน้อยกว่าคิดเป็นร้อยละ 5.09 และมีความสอดคล้องกับ Nayak et al. [5] ซึ่งสกัดคุณลักษณะเวฟเล็ตส์จากภาพเอ็มอาร์ไอโดยใช้ข้อมูลรูปภาพ 3 ชุด จากนั้นใช้การลดมิติข้อมูลด้วยการวิเคราะห์องค์ประกอบหลักภายใต้กรอบความน่าจะเป็น (Probabilistic Principal Component Analysis หรือ PPCA) ซึ่งสามารถลดคุณลักษณะเหลือ 13 คุณลักษณะจาก 1,024 คุณลักษณะ และใช้แบบจำลองอาดาบุสต์ซึ่งมีแรนดอมฟอเรสต์เป็นอัลกอริทึมการเรียนรู้พื้นฐานผลการจำแนกภาพสมองที่ปกติและผิดปกติให้ความแม่นยำเฉลี่ยร้อยละ 99.84 ซึ่งมีความแม่นยำมากกว่างานวิจัยนี้ร้อยละ 6.18 โดยความแตกต่างนี้อาจเกิดจากชุดข้อมูลหรือการตั้งค่าแบบจำลอง ทั้งนี้เมื่อเปรียบเทียบความแม่นยำกับงานวิจัยที่ใช้แบบจำลองอาดาบุสต์ซึ่งมีอัลกอริทึมการเรียนรู้พื้นฐานที่แตกต่างกันโดยอ้างอิงงานวิจัย Nayak et al. [6] ซึ่งใช้ซัพพอร์ตเวกเตอร์แมชชีนเป็นอัลกอริทึมการเรียนรู้พื้นฐานให้ความแม่นยำเฉลี่ยร้อยละ 99.45 ซึ่งมีความแตกต่างของผลลัพธ์เพียงเล็กน้อย เมื่อเปรียบเทียบความแม่นยำกับงานวิจัยที่เกี่ยวข้องซึ่งใช้แบบจำลองเพียงแบบเดียวพบว่าแบบจำลองอาดาบุสต์ซึ่งให้ผลดีที่สุดในงานวิจัยนี้มีประสิทธิภาพต่ำกว่างานวิจัยของ Yadav & Yadav [7] ซึ่งใช้กระบวนการปรับปรุงประสิทธิภาพของเครือข่ายประสาทเทียมร่วมกับการจำแนกภาพถ่ายเนื้องอกในสมองมีความแม่นยำร้อยละ 98.78 ซึ่งแม่นยำกว่างานวิจัยนี้ที่ร้อยละ 5.17 อย่างไรก็ตามงานวิจัยนี้ใช้การเรียนรู้ของเครื่องแบบรวมกลุ่มซึ่งให้ความเสถียรและลดความเสี่ยงของปัญหาซึ่งแบบจำลองเรียนรู้ข้อมูลชุดฝึกได้ดีแต่ให้ความถูกต้องในจำแนกข้อมูลใหม่ในอัตราที่ต่ำได้ดีกว่าการใช้แบบจำลองเพียงแบบเดียว ทั้งนี้ในประเด็นการใช้งานทรัพยากรและความซับซ้อนของกระบวนการโดยอ้างอิงงานวิจัยของ Patil & Kirange [8] ซึ่งใช้วิธีการเรียนรู้แบบรวมกลุ่มในการจำแนกเนื้องอกสมองโดยนำแบบจำลองการเรียนรู้เชิงลึกหลายแบบจำลองมาประมวลผลรวมกันเพื่อเพิ่มประสิทธิภาพในการเรียนรู้และการทำนาย ส่งผลให้มีความแม่นยำร้อยละ 97.77 ซึ่งแม่นยำกว่างานวิจัยนี้ที่ร้อยละ 4.19 ซึ่งกระบวนการทำงานลักษณะนี้ต้องการทรัพยากรการประมวลผลสูง ทั้งนี้ประสิทธิภาพของแบบจำลองอาดาบุสต์ในงานวิจัยนี้มีประสิทธิภาพต่ำกว่างานวิจัยที่เปรียบเทียบซึ่งอาจเกิดจากปัจจัยหลายประการ ได้แก่ ความแตกต่างของชุดข้อมูลที่ใช้ในการทดลอง ประเภทและจำนวนคุณลักษณะที่สกัดได้ การตั้งค่าไฮเปอร์พารามิเตอร์ของแบบจำลอง และการเลือกใช้อัลกอริทึมการเรียนรู้พื้นฐานที่แตกต่างกัน โดยงานวิจัยที่นำมาอ้างอิงซึ่งให้ผลลัพธ์ดีกว่ามีการใช้เทคนิคที่ซับซ้อน หรือใช้การเรียนรู้เชิงลึกหลายแบบจำลองร่วมกัน

ซึ่งต้องการทรัพยากรการประมวลผลสูงกว่า ขณะทำงานวิจัยนี้เน้นการพัฒนาแบบจำลองที่มีความซับซ้อนต่ำและเหมาะสมสำหรับการนำไปประยุกต์ใช้ในสภาพแวดล้อมที่มีข้อจำกัดด้านทรัพยากร แม้จะมีความแม่นยำต่ำกว่าแต่ยังคงให้ความสมดุลงที่ระหว่างประสิทธิภาพและความเป็นไปได้ในการนำไปใช้งานจริงในขณะที่งานวิจัยนี้เสนอแนวทางที่มีความซับซ้อนน้อยกว่าเหมาะในการนำไปประยุกต์ใช้ในสภาพแวดล้อมที่มีข้อจำกัดด้านทรัพยากร

สรุปผลการวิจัย

ผลการวิจัยนี้แสดงให้เห็นว่าแบบจำลองเอาดาบุดีมีประสิทธิภาพในการจำแนกเนื้ออกในสมองด้วยภาพถ่ายทางการแพทย์ที่เหนือกว่าแบบจำลองอื่นในทุกตัวชี้วัดอย่างมีนัยสำคัญ ผลการคัดเลือกคุณสมบัติด้วยค่าความสำคัญของแรนดอมฟอเรสต์ที่มากกว่า 0.01 ขึ้นไปแสดงให้เห็นว่าคุณสมบัติทางสถิติจากซับบนดัดประมาณค่าระดับ 3 โดยเฉพาะค่าความเบ้ ค่าส่วนเบี่ยงเบนมาตรฐาน และ ค่าความสูงชันมีบทบาทสำคัญในการจำแนกเนื้ออกในสมอง และแบบจำลองที่พัฒนาขึ้นสามารถสร้างสมดุลงระหว่างประสิทธิภาพและความเป็นไปได้ในการนำไปใช้จริง แม้จะมีความแม่นยำต่ำกว่างานวิจัยที่ใช้เทคนิคซับซ้อนแต่งานวิจัยนี้ให้ผลลัพธ์ที่ยอมรับได้ด้วยความแม่นยำที่ร้อยละ 93.67 ด้วยความซับซ้อนที่น้อยกว่าแสดงให้เห็นว่าการสกัดคุณลักษณะเวฟเลตส์จากภาพถ่ายทางการแพทย์สามารถประยุกต์ใช้ร่วมกับการเรียนรู้ของเครื่องแบบรวมกลุ่มและสามารถให้ความแม่นยำในอัตราที่สูงเหมาะสำหรับการประยุกต์ใช้ในสภาพแวดล้อมจริงโดยเฉพาะอย่างยิ่งการประยุกต์ใช้ในระบบสุขภาพที่มีข้อจำกัดด้านเทคโนโลยี ในประเด็นการศึกษาและวิจัยในอนาคตคณะผู้วิจัยมีแนวทางในการวิจัยเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองพื้นฐาน (Base Learner) ที่ใช้ร่วมกับแบบจำลองเอาดาบุดี โดยพิจารณาทั้งแบบจำลองที่มีความซับซ้อนต่ำ อาทิเช่นแบบจำลองต้นไม้ตัดสินใจที่มีความลึกแตกต่างกันแบบจำลองโลจิสติกส์เกรสชัน ตลอดจนแบบจำลองที่มีความซับซ้อนสูงเช่น แบบจำลองซัพพอร์ตเวกเตอร์แมชชีนแบบจำลองนาอีฟเบย์ รวมถึงแบบจำลองอื่นๆที่มีความเหมาะสมต่อลักษณะข้อมูล โดยการศึกษาวิจัยจะครอบคลุมถึงกระบวนการปรับแต่งค่าไฮเปอร์พารามิเตอร์ของแต่ละแบบจำลองเพื่อให้ประสิทธิภาพสูงสุดเมื่อประมวลผลร่วมกับแบบจำลองเอาดาบุดี โดยคำนึงถึงปัจจัยสำคัญ ได้แก่ ความรวดเร็วในการประมวลผล ความสามารถในการเพิ่มประสิทธิภาพตามจำนวนรอบการทำงานและความสามารถในการลดความแปรปรวนของกระบวนการเรียนรู้ ทั้งนี้เพื่อยกระดับประสิทธิภาพในการจำแนกภาพถ่ายทางการแพทย์ให้มีความแม่นยำและความน่าเชื่อถือมากยิ่งขึ้น

กิตติกรรมประกาศ

ขอขอบพระคุณคณะวิทยาการคอมพิวเตอร์มหาวิทยาลัยราชภัฏอุบลราชธานีในการอำนวยความสะดวกเอื้อเฟื้อสถานที่และให้คำแนะนำที่เป็นประโยชน์ และอุปกรณ์คอมพิวเตอร์สมรรถนะสูงในการทำงานวิจัย ขอขอบคุณ Ahmed Hamada ที่เอื้อเฟื้อภาพถ่ายทางการแพทย์และชุดข้อมูลในการทำวิจัยครั้งนี้

เอกสารอ้างอิง

- 1 พิมพ์ภาภรณ์ ทรัพย์มีชัย, สุนทรา เลี้ยงเขวงวงศ์ และ ภัทรวินัย รักษ์กุล. ปัจจัยทำนายสภาวะส่วนบุคคลความทุกข์ทรมานจากอาการและพฤติกรรมจัดการอาการต่อคุณภาพชีวิตของผู้ป่วยเนื้ออกสมอง. วารสารพยาบาลศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. 2561; 33(2): 1-12. สืบค้น จาก <https://he01.tci-thaijo.org/index.php/CUNS/article/view/254383>

- 2 Mukdaprawat P, Danaidutsadeekul S, Chanruangvanich W, Itthimathin P. Relationships between brain tumor grading, severity of neurological deficit, nutritional status and functional status in brain tumor patients during hospital stay. *Nurs Sci J Thai* [Internet]. 2012 Sep 28 [cited 2025 Sep.27] ; 30(3): 46-54. Available from: <https://he02.tci-thaijo.org/index.php/ns/article/view/10468>
- 3 Neugut AI, Sackstein P, Hillyer GC, Jacobson JS, Bruce J, Lassman AB, et al. Magnetic resonance imaging-based screening for asymptomatic brain tumors: a review. *Oncologist*. 2019; 24(3): 375-84.
- 4 Gopi Krishna T, Sunitha KVN, Mishra S. Detection and classification of brain tumor from MRI medical image using wavelet transform and PSO based LLRBFNN algorithm. *Int J Comput Sci Eng*. 2018; 6(1): 18-23. doi: 10.26438/ijcse/v6i1.1823. Available from: https://www.ijcseonline.org/full_paper_view.php?paper_id=1627
- 5 Nayak DR, Dash R, Majhi B. Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. *Neurocomputing*. 2016;177:188-97.
- 6 Nayak DR, Dash R, Majhi B. Stationary wavelet transform and AdaBoost with SVM based pathological brain detection in MRI scanning. *CNS Neurol Disord Drug Targets*. 2017; 16(2): 137-49.
- 7 Yadav AK, Yadav V. Brain tumor recognition from multimodal magnetic resonance images using wavelet texture features and optimized artificial neural network. *Multimed Tools Appl*. 2024; 83: 72975-96.
- 8 Patil S, Kirange D. Ensemble of deep learning models for brain tumor detection. *Procedia Comput Sci*. 2023; 218: 2468-79.
- 9 อนุพงศ์ สุขประเสริฐ. คู่มือการทำเหมืองข้อมูลด้วยโปรแกรม RapidMiner Studio. พิมพ์ครั้งที่ 3. มหาสารคาม: ภาควิชาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม; 2563. น. 16-41.
- 10 Hamada A. Br35H: brain tumor detection 2020 [Internet]. Kaggle; 2020 [cited 2025 Apr 9]. Available from: <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection>
- 11 นราทิพย์ ณ ระนอง. การจำแนกภาพขอบแบบเซตเปิดด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชัน [วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ]. กรุงเทพมหานคร: มหาวิทยาลัยศรีนครินทรวิโรฒ; 2562.
- 12 Naknaem K, Kaewlek T. A comparative study of pre-processing methods to improve glioma segmentation performance in brain MRI using deep learning. *J Assoc Med Sci*. 2024; 57(2): 132-40.
- 13 OpenCV. opencv-python [Internet]. PyPI; [cited 2025 Feb 19]. Available from: <https://pypi.org/project/opencv-python/>
- 14 Scikit-Image. Scikit-image: image processing in python [Internet]. PyPI; [cited 2025 Feb 6]. Available from: <https://pypi.org/project/scikit-image>
- 15 SciPy. SciPy: scientific computing tools for python [Internet]. PyPI; [cited 2025 Feb 14]. Available from: <https://pypi.org/project/scipy>
- 16 Pandas development team. Pandas: python data analysis library [Internet]. [cited 2025 Jul 9]. Available from: <https://pandas.pydata.org>

- 17 Mathew AR, Anto PB. Tumor detection and classification of MRI brain image using wavelet transform and SVM. In: 2017 International Conference on Signal Processing and Communication (ICSPC); 2017 Jul 28-29; Coimbatore, India. Piscataway (NJ): IEEE; 2017. p. 75-8.
- 18 Altair. Altair AI Studio [Internet]. [cited 2025 Mar 7]. Available from: <https://altair.com/altair-ai-studio>
- 19 เดช ธรรมศิริ และ พยุง มีสัจ. การเรียนรู้แบบรวมกลุ่มด้วยโครงข่ายประสาทเทียมเอาดาบูทสำหรับการจำแนกข้อมูล. วารสารเทคโนโลยีสารสนเทศมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. 2554; 7(2) :7-12. สืบค้น จาก https://ph01.tci-thaijo.org/index.php/IT_Journal/article/view/54513
- 20 พชร ดอกชะเอม. แบบจำลองทำนายโรคหลอดเลือดสมองด้วยเทคนิคการเรียนรู้ของเครื่อง [วิทยานิพนธ์ วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล]. กรุงเทพมหานคร: มหาวิทยาลัยศรีนครินทรวิโรฒ; 2566.
- 21 กฤตกนก ศรีพิมพ์สอ และ กิตติพล วิแสงวิแสง. การพยากรณ์โรคเบาหวานด้วยเทคนิคเหมืองข้อมูล. วารสารการจัดการเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม. 2023; 10(1): 51-63. สืบค้น จาก <https://ph02.tci-thaijo.org/index.php/itm-journal/article/view/248575>
- 22 ปพิชญา กลางนอก และ จารีย์ ทองคำ. การประยุกต์ใช้เทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพของแบบจำลองตามกฎในเหมืองข้อมูล. วารสารเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏอุบลราชธานี. 2562; 9(1): 97-108. สืบค้น จาก <https://ph01.tci-thaijo.org/index.php/jitubru/article/view/183422>
- 23 Allgaier J, Pryss R. Cross-validation visualized: a narrative guide to advanced methods. Mach Learn Knowl Extr. 2024; 6(2): 1378-88. doi: 10.3390/make6020065. Available from: <https://www.mdpi.com/2504-4990/6/2/65>
- 24 Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence; 1995. p.1137-45.
- 25 Rodriguez JD, Perez A, Lozano JA. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell. 2010; 32(3): 569-75. doi: 10.1109/tpami.2009.187. Available from: https://www.researchgate.net/publication/224085226_Sensitivity_Analysis_of_k-Fold_Cross_Validation_in_Prediction_Error_Estimation
- 26 Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. NeuroImage. 2017; 145: 166-79.