

Received: 16 เม.ย. 2567

Revised: 30 พ.ค. 2567

Accepted: 4 มิ.ย. 2567

การสร้างตัวแบบการเรียนรู้เชิงลึกเพื่อการวินิจฉัยโรคมาลาเรียอย่างรวดเร็วและแม่นยำ
ในภาพถ่ายฟิล์มเลือด

Development of Deep Learning Models for Rapid and Accurate Malaria Parasite
Detection in Blood Smear Image

เลอศักดิ์ โพธิ์ทอง¹, ภาคินี เหล่ามูล¹, ภัททิยา โมรา¹, และ อนูปงศ์ สุขประเสริฐ^{1*}

¹สาขาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม

Lersak Phothong¹, Pakinee Laomoon¹, Pattiya Mora¹, and Anupong Sukprasert^{1*}

¹Department of Business Computer, Mahasarakham Business School,

Mahasarakham University

* ผู้นิพนธ์ประสานงาน: อนูปงศ์ สุขประเสริฐ อีเมล: anupong.s@acc.msu.ac.th

Abstract

This study aims to develop an efficient model using deep learning techniques for the rapid and accurate diagnosis of malaria in blood smear images. The analysis follows the Cross-Industry Standard Process for Data Mining (CRISP-DM), employing forward selection and backward elimination feature selection techniques combined with four classification methods: k-Nearest Neighbors, Neural Network, Deep Learning, and Random Forest. The performance of these models is compared using metrics such as accuracy, overall performance, speed, and specificity. The results indicate that the deep learning-based feature selection technique achieves an accuracy of 97.75%, an overall performance of 97.73%, a speed of 97.19%, and a specificity of 98.31%. This technique is suitable for developing a malaria patient screening model, reducing delays in patient care in remote and underserved areas. It supports the formulation of treatment policies by multidisciplinary teams and enables rapid and accurate malaria diagnosis using artificial intelligence techniques. This can ensure timely and effective treatment for malaria patients, ultimately reducing malaria-related mortality in the future.

Keywords: Image classification; Deep Learning; Disease Diagnosis; Malaria, Blood Smear Image

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาตัวแบบที่มีประสิทธิภาพโดยใช้เทคนิคการเรียนรู้เชิงลึกสำหรับการวินิจฉัยโรคมาลาเรียในภาพถ่ายฟิล์มเลือดอย่างรวดเร็วและแม่นยำ การวิเคราะห์ตามกระบวนการทำเหมืองข้อมูลมาตรฐานอุตสาหกรรม ใช้เทคนิคการคัดเลือกปัจจัยแบบไปข้างหน้าและการลดตัวแปรร่วมกับเทคนิคการจำแนกประเภท 4 แบบ ได้แก่ เทคนิคเพื่อนบ้านใกล้สุด เทคนิคโครงข่ายประสาทเทียม เทคนิคการเรียนรู้เชิงลึก และเทคนิคป่าสุ่ม โดยเปรียบเทียบประสิทธิภาพของตัวแบบเหล่านี้ด้วยค่าความแม่นยำ ค่าประสิทธิภาพโดยรวม ค่าความเร็ว และค่าจำเพาะ ผลการวิจัยพบว่า เทคนิคการคัดเลือกปัจจัยที่ใช้เทคนิคการเรียนรู้เชิงลึกให้ค่าความแม่นยำ 97.75% ค่าประสิทธิภาพโดยรวม 97.73% ค่าความเร็ว 97.19% และค่าจำเพาะ 98.31% เทคนิคนี้เหมาะสมสำหรับการพัฒนาตัวแบบการคัดกรองผู้ป่วยโรคมาลาเรีย ลดความล่าช้าในการดูแลผู้ป่วยในพื้นที่ห่างไกลและขาดแคลนแพทย์ สนับสนุนการกำหนดนโยบายการรักษาโดยทีมสหวิชาชีพ และสามารถวินิจฉัยโรคมาลาเรียได้อย่างรวดเร็วและแม่นยำด้วยเทคนิคปัญญาประดิษฐ์ ซึ่งช่วยให้การรักษาผู้ป่วยโรคมาลาเรียเป็นไปอย่างทันทั่วถึงและมีประสิทธิภาพ อีกทั้งยังลดอัตราการเสียชีวิตจากโรคมาลาเรียในอนาคต

คำสำคัญ: การจำแนกประเภทข้อมูลภาพ; การเรียนรู้เชิงลึก; การวินิจฉัยโรค; โรคมาลาเรีย; ภาพถ่ายฟิล์มเลือด

บทนำ

โรคมาลาเรีย (Malaria) หรือไข้ป่า เป็นโรคที่เกิดจากการติดเชื้อโปรโตซัว (Protozoa) ในกลุ่มพลาสโมเดียม (Plasmodium species) โดยมียุงก้นปล่องเพศเมียเป็นพาหะนำโรค (มาเรีย นินา จิตะสมบัติ, 2023) เชื้อมาลาเรียเป็นเชื้อโรคที่มีขนาดเล็กมากทำให้ไม่สามารถมองเห็นได้ด้วยตาเปล่า ต้องอาศัยการตรวจหาเชื้อโดยใช้กล้องจุลทรรศน์หรือชุดตรวจหาเชื้อมาลาเรียสำเร็จรูป อีกทั้งเชื้อมาลาเรียบางชนิดเมื่อป่วยแล้วหากไม่ได้รับการรักษาที่ทันทั่วถึงอาจทำให้เสียชีวิตได้ โรคมาลาเรียไม่เพียงเป็นปัญหาสุขภาพเท่านั้น ยังเป็นปัญหาทางสังคมและเศรษฐกิจ โดยเฉพาะในพื้นที่ที่มีการระบาดของสูง โรคมาลาเรียส่งผลกระทบต่อคุณภาพชีวิตและทรัพยากรทางสังคม (โรงพยาบาลศิริราชปิยมหาราชการุณย์, 2022)

องค์การอนามัยโลก (World Health Organization) รายงานว่าในปี พ.ศ. 2565 มีผู้ป่วยโรคมาลาเรียจำนวน 249 ล้านราย และผู้ป่วยที่เสียชีวิตประมาณ 608,000 ราย ซึ่งเมื่อเทียบกับปี พ.ศ. 2564 ที่มีผู้ป่วยโรคมาลาเรียจำนวน 244 ล้านราย และผู้ป่วยที่เสียชีวิตจำนวน 610,000 ราย ถือว่าอัตราการติดเชื้อมีแนวโน้มเพิ่มขึ้น (World Health Organization, 2023) กรมควบคุมโรค กระทรวง

สาธารณสุข ได้รายงานสถานการณ์การแพร่ระบาดของโรคไข้มาลาเรีย ตั้งแต่วันที่ 1 มกราคม ถึง 13 กรกฎาคม 2565 พบผู้ป่วยที่ติดเชื้อจำนวน 4,765 ราย ซึ่งมีจำนวนผู้ติดเชื้อสูงกว่าในปี 2564 ถึง 2.6 เท่า โดยกลุ่มอายุที่พบผู้ป่วยมากที่สุดคือ อายุ 25-44 ปี รองลงมาคือ อายุ 15-24 ปี และอายุ 5-14 ปี โดยส่วนใหญ่เป็นกลุ่มเกษตรกร ร้อยละ 48 รับจ้าง ร้อยละ 25 และเด็ก/นักเรียน ร้อยละ 24 (The Coverage, 2022)

การพัฒนามาตรฐานการรักษาและการควบคุมโรคมาลาเรียในปัจจุบัน โดยการคัดกรองโรคมาลาเรียผ่านการตรวจเลือดภายใต้กล้องจุลทรรศน์โดยบุคลากรที่ผ่านการฝึกอบรมต้องใช้เวลานาน และต้องใช้ความเชี่ยวชาญ นอกจากนี้หากเป็นพื้นที่ที่ขาดความพร้อมในด้านเครื่องมือก็ต้องใช้การตั้งข้อสงสัยในการคัดกรองโรค ซึ่งอาจทำให้เกิดข้อผิดพลาดในการวินิจฉัยผู้ป่วยมาลาเรียได้

นวัตกรรมในเทคนิคการเรียนรู้ของเครื่อง (Machine Learning, ML) ได้มีบทบาทอย่างมากต่อการวิจัยทางการแพทย์ โดยเพิ่มความสามารถในการวินิจฉัยผ่านการวิเคราะห์ชุดข้อมูลขนาดใหญ่ (Rana & Bhushan, 2023) การประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องได้แสดงประสิทธิภาพอย่างมากในการจำแนกภาพเอกซเรย์เพื่อวินิจฉัยโรคปอด (Farhan & Yang, 2023) และการจำแนกเซลล์มะเร็งจากภาพทางพยาธิวิทยา (Ding et al., 2023) ความก้าวหน้าเหล่านี้ช่วยให้ผู้เชี่ยวชาญทางการแพทย์สามารถใช้ประโยชน์จากอัลกอริทึมที่ซับซ้อนในการตรวจจับโรค ระบุรูปแบบ และเพิ่มความสามารถในการวินิจฉัยและการรักษาโดยการปรับกระบวนการอัตโนมัติ เช่น การใช้ Deep Learning ด้วยเทคนิคโครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Networks, CNNs) ช่วยในการจำแนก ตรวจจับ และประเมินภาพเอกซเรย์ทรวงอก ซึ่งในทางการแพทย์ Deep Learning เป็นเครื่องมือที่มีประสิทธิภาพในการวินิจฉัยโรคและตรวจจับภาวะทางการแพทย์ต่างๆ โดยเฉพาะในสถานการณ์ฉุกเฉิน เช่น การระบาดของ COVID-19 การใช้ Deep Learning วิเคราะห์ภาพเอกซเรย์ทรวงอกช่วยให้การวินิจฉัยเป็นไปอย่างรวดเร็วและมีประสิทธิภาพมากขึ้น เพื่อให้การรักษาผู้ป่วยมีโอกาสหายจากโรค (Karim Hammoudi et al., 2021)

จากการศึกษาของ Lakshmanaprabu S.K. (2019) ที่พัฒนาโมเดล Optimal Deep Neural Network (ODNN) สำหรับจำแนกมะเร็งปอดจากภาพ CT และการวิจัยของ QinZeng Song (2017) ที่ใช้ Convolutional Neural Network (CNN) ในการจำแนกเนื้องอกในปอดจากภาพเอกซเรย์ ทั้งสองการศึกษานี้ใช้วิธีการเฉพาะเจาะจงต่อชนิดข้อมูลที่แตกต่างกันและมีความแม่นยำในการจำแนกที่แตกต่างกัน นอกจากนี้ การเปรียบเทียบกับการศึกษาของ Hamisu Ismail Ahmad (2019) ที่ใช้เทคนิคการตัดสินใจ (Decision Tree) ในการวินิจฉัยโรคมาลาเรีย โดยใช้ชุดข้อมูลที่มีการสอบถามอาการผู้ป่วย ได้รับผลลัพธ์ที่มีความแม่นยำที่ 88.4% และการศึกษาของ Opeyemi A. Abisoye และ Rasheed G. Jimoh (2018) ที่ใช้เทคนิค Support Vector Machine (SVM) ในการทำนายการเกิดโรคมาลาเรีย โดยใช้ข้อมูลทางสถิติที่เกี่ยวข้องกับอาการของผู้ป่วยและข้อมูลสภาพ

อากาศ ทั้งสองการศึกษานี้ใช้เทคนิคการวิเคราะห์ที่มีความเหมาะสมแต่อาจมีข้อจำกัดในเรื่องของประเภทและความซับซ้อนของข้อมูลที่ประมวลผล

เทคนิคการคัดเลือกปัจจัยแบบไปข้างหน้า (Forward Selection) และเทคนิคการคัดเลือกปัจจัยการลดตัวแปร (Backward Elimination) มีวิธีการทำงานที่แตกต่างกันอย่างชัดเจน แต่ทั้งสองวิธีก็มีจุดมุ่งหมายเดียวกันในการเพิ่มประสิทธิภาพของตัวแบบและลดความซับซ้อนของโมเดล (Pierdzioch, 2023) โดยการคัดเลือกปัจจัยแบบไปข้างหน้าเหมาะสมกับกรณีที่มีคุณสมบัติจำนวนมากและต้องการค้นหาคุณสมบัติที่สำคัญอย่างรวดเร็ว ในขณะที่การคัดเลือกปัจจัยการลดตัวแปรเหมาะสมกับกรณีที่ต้องการกำจัดคุณสมบัติที่ไม่สำคัญและคงไว้เฉพาะคุณสมบัติที่มีความสำคัญ โดยการใช้วิธีการทั้งสองนี้สามารถทำให้ได้ตัวแบบที่มีประสิทธิภาพสูงและง่ายต่อการนำไปใช้งานในการพัฒนาตัวแบบการวิเคราะห์ข้อมูล

การศึกษาของ Julianto และคณะ เปรียบเทียบและปรับปรุงแบบจำลองอัลกอริทึมการวิเคราะห์ข้อมูลโดยใช้ Feature Selection-Wrapper กับข้อมูลเงินดิจิทัล (Binance Coin:BNB) พบว่า หลังจากใช้เทคนิคการคัดเลือกปัจจัยการลดตัวแปร (Backward Elimination) อัลกอริทึมเครือข่ายประสาทเทียม (Artificial Neural Network) สามารถทำงานได้ดีขึ้น โดยมีข้อผิดพลาดรากกำลังสองเฉลี่ย (Root Mean Square Error:RMSE) ลดลง (Julianto et al., 2023)

Giorgos Borboudakis และ Ioannis Tsamardinos ได้เสนอวิธีการฮิวริสติก (Heuristic) เพื่อเพิ่มประสิทธิภาพของอัลกอริทึมการเลือกไปข้างหน้าและถอยหลัง (Forward-Backward Selection Algorithms) โดยเพิ่มความเร็วในการคำนวณ (Computational Speed) ในขณะที่ยังคงความแม่นยำในการทำนาย (Predictive Accuracy)

นอกจากนี้ งานวิจัยของ Damian Kozbur ได้ทดสอบเทคนิคการคัดเลือกปัจจัยแบบไปข้างหน้า (Forward Selection) โดยการใช้การทดสอบ (Testing-Based Forward Model Selection: TBFMS) ในการถดถอยเชิงเส้น (Linear Regression) ซึ่งการเลือกตัวแปร (Covariates) จะขึ้นอยู่กับความสามารถในการทำนาย (Predictive Power) ก่อนที่จะทำการถดถอยขั้นสุดท้าย โดยมีการกำหนดขอบเขตเชิงความน่าจะเป็น (Probabilistic Bounds) สำหรับค่าความผิดพลาดในการทำนาย (Prediction Error) และจำนวนตัวแปรที่เลือก (Number of Selected Covariates)(Kozbur, 2020)

การศึกษาเหล่านี้แสดงให้เห็นถึงความสำคัญและการประยุกต์ใช้ที่หลากหลายของเทคนิคการคัดเลือกปัจจัยแบบไปข้างหน้า (Forward Selection) และเทคนิคการคัดเลือกปัจจัยการลดตัวแปร (Backward Elimination) ในการปรับปรุงการสร้างแบบจำลองเชิงคาดการณ์และกระบวนการคัดเลือกตัวแปรในบริบทการวิจัยที่แตกต่างกัน

การวิเคราะห์ข้อมูลภาพ (Image Analytics) เป็นกระบวนการที่ซับซ้อนซึ่งใช้ในการตรวจสอบและวิเคราะห์ข้อมูลจากรูปภาพ เพื่อแปลงข้อมูลเหล่านี้จากสถานะที่ไม่มีโครงสร้างมาเป็นรูปแบบที่มีโครงสร้างและสามารถนำมาวิเคราะห์ได้ง่ายขึ้น กระบวนการนี้มักจะเริ่มต้นด้วยการแปลงข้อมูลภาพให้กลายเป็นข้อมูลที่สามารถเข้าถึงและจัดการได้ในรูปแบบของตารางข้อมูล โดยข้อมูลแต่ละพิกเซลหรือคุณลักษณะต่างๆ จะถูกแยกออกเป็นแถวและคอลัมน์ที่ชัดเจน เพื่อเตรียมพร้อมสำหรับกระบวนการวิเคราะห์ข้อมูลภาพซึ่งจะใช้เทคนิคทางสถิติและเทคนิคการเรียนรู้ของเครื่องเพื่อจำแนกและวิเคราะห์ข้อมูลเหล่านี้ วิธีการที่นิยมใช้ ได้แก่ การเรียนรู้ภายใต้การดูแล (Supervised Learning) และการเรียนรู้แบบไม่มีผู้ดูแล (Unsupervised Learning) (Hansson & Andersen, 2023) ซึ่งทั้งสองวิธีนี้ช่วยให้ผู้วิจัยสามารถจำแนกและระบุลักษณะเฉพาะของข้อมูลที่สามารถสร้างความแตกต่างระหว่างกลุ่มข้อมูลหรือประเภทข้อมูลที่ต้องการได้ ในส่วนของเทคนิคการจำแนกประเภทข้อมูลภาพ (Image Classification Techniques) วิธีการนี้พิจารณาข้อมูลพิกเซลทั้งหมดภายในพื้นที่ศึกษาและแบ่งข้อมูลเหล่านี้นออกเป็นหมวดหมู่หรือกลุ่มย่อยตามคุณสมบัติทางสถิติที่แสดงความแตกต่าง ช่วยให้สามารถวิเคราะห์ข้อมูลที่มีปริมาณมากอย่างมีประสิทธิภาพและมีการจำแนกที่เฉพาะเจาะจงตามเกณฑ์ที่กำหนดไว้ (อนุพงศ์ สุขประเสริฐ, 2021)

จากประเด็นข้างต้น ผู้วิจัยจึงมีความสนใจศึกษาเรื่องการสร้างตัวแบบสำหรับการคัดกรองผู้ป่วยโรคมะเร็งด้วยภาพถ่ายฟิล์มเลือดจากกล้องจุลทรรศน์ โดยทำการคัดเลือกปัจจัย ด้วยเทคนิคการคัดเลือกปัจจัยแบบไปข้างหน้าและเทคนิคการคัดเลือกปัจจัยการลดตัวแปร ร่วมกับเทคนิคการจำแนกประเภทข้อมูลภาพ 4 เทคนิค ได้แก่ เทคนิคเพื่อนบ้านใกล้สุด (k-Nearest Neighbors: k-NN) เทคนิคโครงข่ายประสาทเทียม (Neural Network: NN) เทคนิคการเรียนรู้เชิงลึก (Deep Learning: DL) และเทคนิคต้นไม้ป่าสุ่ม (Random Forest: RF) (ธนานพ จันภักดี และฐิติรัตน์ ศิริบรรรัตนกุล, 2022) เพื่อสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบที่สร้างขึ้นโดยใช้เทคนิคต่างๆ ด้วยค่าความแม่นยำ (Accuracy) ค่าประสิทธิภาพโดยรวม (F-Measure) ค่าความเร็ว (Sensitivity) และค่าจำเพาะ (Specificity) เพื่อยืนยันว่าตัวแบบที่ใช้เทคนิคการเรียนรู้เชิงลึกมีประสิทธิภาพมากกว่าเทคนิคอื่นๆ สำหรับการคัดกรองผู้ป่วยโรคมะเร็ง การสร้างตัวแบบที่แม่นยำและรวดเร็วยิ่งขึ้นจะช่วยลดการวินิจฉัยที่ผิดพลาด ช่วยให้ผู้ป่วยโรคมะเร็งได้รับการรักษาอย่างทันท่วงทีและมีประสิทธิภาพ และลดจำนวนผู้เสียชีวิตจากโรคมะเร็งในอนาคต

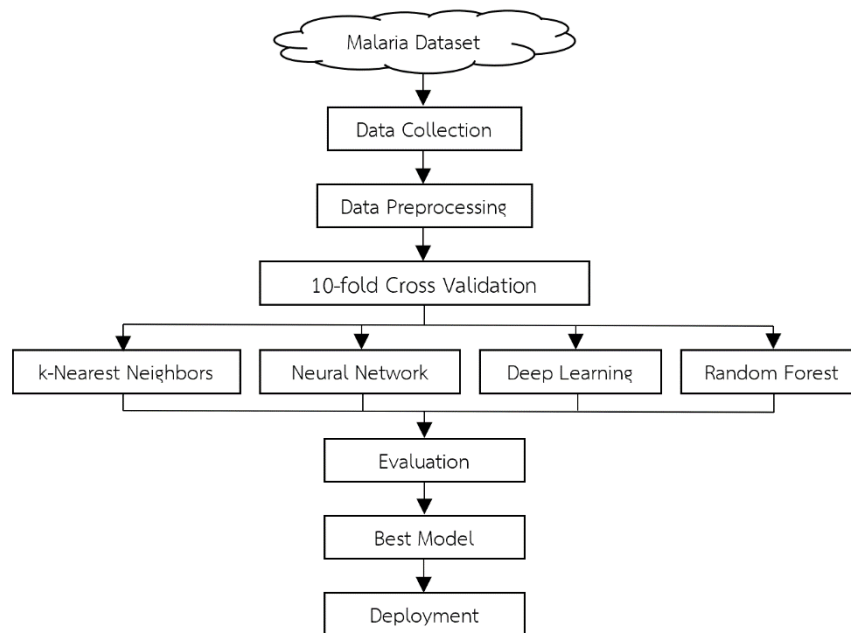
วัตถุประสงค์

1. เพื่อสร้างตัวแบบสำหรับการคัดกรองผู้ป่วยโรคมะเร็งด้วยเทคนิคการเรียนรู้เชิงลึกเพื่อการวินิจฉัยโรคมะเร็งอย่างรวดเร็วและแม่นยำในภาพถ่ายฟิล์มเลือด

2. เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบที่ใช้เทคนิคการเรียนรู้เชิงลึก กับเทคนิคเพื่อนบ้านใกล้สุด เทคนิคโครงข่ายประสาทเทียม และเทคนิคป่าสุ่ม สำหรับการคัดกรองผู้ป่วยโรคมะเร็งด้วยภาพถ่ายฟิล์มเลือด

วิธีดำเนินการวิจัย

งานวิจัยนี้มุ่งเน้นการสร้างตัวแบบการเรียนรู้เชิงลึกเพื่อการวินิจฉัยโรคมะเร็งอย่างรวดเร็วและแม่นยำในภาพถ่ายฟิล์มเลือด โดยกรอบงานวิจัยเริ่มจากการรวบรวมข้อมูลโรคมะเร็งซึ่งประกอบด้วยภาพถ่ายฟิล์มเลือดของผู้ป่วย ขั้นตอนแรกคือการเก็บรวบรวมข้อมูลจากแหล่งข้อมูลที่เกี่ยวข้อง จากนั้นทำการเตรียมข้อมูลโดยการทำความสะอาด ปรับขนาดภาพ และทำการทำให้ข้อมูลอยู่ในรูปแบบมาตรฐาน (normalization) เพื่อให้ข้อมูลพร้อมสำหรับการประมวลผลในขั้นตอนต่อไป การแบ่งข้อมูลออกเป็น 10 ส่วนและทำการทดสอบแบบไขว้ (cross validation) ถูกใช้เพื่อประเมินประสิทธิภาพของตัวแบบอย่างแม่นยำและลดความเอนเอียงในการประเมินผล ตัวแบบที่สร้างและฝึกอบรมประกอบด้วย k-Nearest Neighbors (k-NN), โครงข่ายประสาทเทียม (Neural Network), การเรียนรู้เชิงลึก (Deep Learning) และ ต้นไม้ป่าสุ่ม (Random Forest) หลังจากฝึกอบรมตัวแบบแล้วจะทำการประเมินผลเพื่อหาประสิทธิภาพและความแม่นยำในการวินิจฉัยโรคมะเร็ง จากนั้นเลือกตัวแบบที่มีประสิทธิภาพและความแม่นยำสูงสุดเพื่อนำไปใช้งานจริงในขั้นตอนสุดท้าย ซึ่งกระบวนการทั้งหมดนี้มีเป้าหมายเพื่อพัฒนาตัวแบบการเรียนรู้เชิงลึกที่สามารถวินิจฉัยโรคมะเร็งได้อย่างมีประสิทธิภาพและความแม่นยำสูงสุด เพื่อช่วยให้การวินิจฉัยโรคเป็นไปอย่างรวดเร็วและสามารถช่วยชีวิตผู้ป่วยได้มากขึ้นดังแสดงภาพที่ 1



ภาพที่ 1 แสดงกรอบแนวคิดของการสร้างตัวแบบการเรียนรู้เชิงลึกเพื่อการวินิจฉัยโรคมาลาเรียอย่างรวดเร็วและแม่นยำในภาพถ่ายฟิล์มเลือด

1. การทำความเข้าใจปัญหา (Business Understanding)

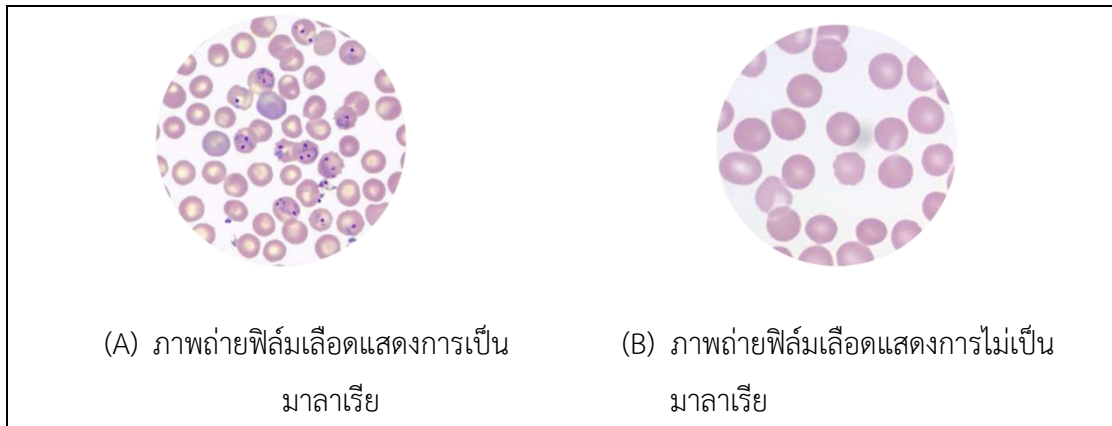
ผู้วิจัยได้ศึกษาข้อมูลเกี่ยวกับโรคมาลาเรียจึงได้มีแนวคิดในการคัดกรองผู้ป่วยโรคมาลาเรียโดยใช้เทคนิคการจำแนกประเภทข้อมูลภาพถ่ายฟิล์มเลือดด้วยกล้องจุลทรรศน์ ซึ่งเป็นอีกวิธีการหนึ่งที่จะสามารถวินิจฉัยโรคได้อย่างรวดเร็ว ในการวินิจฉัยสำหรับการคัดกรองจำแนกผู้ป่วยมาลาเรีย เพื่อสนับสนุนการตัดสินใจของแพทย์อายุรกรรมในการวินิจฉัยโรคได้อย่างรวดเร็วและแม่นยำยิ่งขึ้น ซึ่งแต่เดิมการคัดกรองส่วนใหญ่ใช้เครื่องมือภาพถ่ายฟิล์มเลือดและการคัดกรองโดยการตั้งข้อสงสัยโรค โดยหากเป็นพื้นที่มีเครื่องมือไม่พร้อมก็ต้องใช้ข้อสงสัยในการคัดกรองโรค ซึ่งอาจทำให้เกิดข้อผิดพลาดในการวินิจฉัยผู้ป่วยมาลาเรียได้

2. การทำความเข้าใจข้อมูล (Data Understanding)

งานวิจัยนี้ใช้ชุดข้อมูลภาพถ่ายฟิล์มเลือดของผู้ป่วยโรคมาลาเรียในปี พ.ศ. 2564 จำนวน 26,159 ภาพ ขนาด 32x32 พิกเซล ในรูปแบบไฟล์ .png โดยอ้างอิงจากการศึกษา K. M. Faizullah Fuhad และคณะ จากภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ มหาวิทยาลัย North South, กรุงเทพฯ, บังคลาเทศ (Fuhad et al., 2020)

3. การเตรียมข้อมูล (Data Preparation)

การนำเข้าข้อมูลภาพ โดยนำรูปภาพที่มีนามสกุลไฟล์ .png เข้ามาทำการวิเคราะห์สำหรับสร้างตัวแบบการคัดกรองผู้ป่วยโรคมาลาเรีย ผู้วิจัยได้ทำการแบ่งชุดข้อมูลออกเป็น 2 โพลเดอรีย่อย ประกอบด้วย โพลเดอร A เป็นโพลเดอรที่จัดเก็บภาพถ่ายฟิล์มเลือดด้วยกล้องจุลทรรศน์ที่เป็นมาลาเรีย (Malaria) จำนวน 13,131 ภาพ และโพลเดอร B ที่เป็นโพลเดอรจัดเก็บภาพถ่ายฟิล์มเลือดที่ไม่เป็นมาลาเรีย (Normal) จำนวน 13,028 ภาพ ดังแสดงในภาพที่ 2



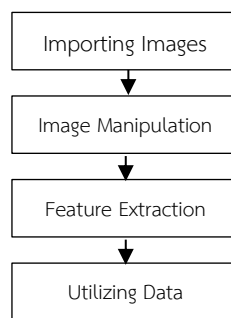
ภาพที่ 2 ตัวอย่างการจำแนกประเภทข้อมูลจากภาพถ่ายฟิล์มเลือดด้วยกล้องจุลทรรศน์

จากนั้นแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลที่มีโครงสร้าง (Structured Data) ดังแสดงในภาพที่ 3 กระบวนการเริ่มจากการนำเข้ารูปภาพ (Importing Images) โดยใช้ operator ชื่อ "Read Images" เพื่อโหลดรูปภาพจากแหล่งข้อมูลທີ່เลือกไว้ จากนั้นเข้าสู่ขั้นตอนการปรับแต่งรูปภาพ (Image Manipulation) ซึ่งใช้ operators เช่น "Crop", "Resize", และ "Color Conversion" เพื่อปรับขนาด ตัด หรือเปลี่ยนแปลงสีของรูปภาพ ทำให้รูปภาพเหมาะสมกับการสกัดคุณลักษณะหรือการวิเคราะห์

ขั้นตอนต่อมาคือการสกัดคุณลักษณะ (Feature Extraction) โดยใช้ operator เช่น "Extract Image Features" เพื่อดึงคุณลักษณะเช่น พื้นผิว ขอบ หรือจุดเด่นอื่นๆ จากรูปภาพ ซึ่งคุณลักษณะเหล่านี้จะถูกแปลงเป็นข้อมูลที่มีโครงสร้าง (Structured Data) ที่สามารถใช้ในการวิเคราะห์ได้ นอกจากนี้ยังมี "Image Processing" ซึ่งเป็นหมวดหมู่ที่รวบรวม operators สำหรับการประมวลผลรูปภาพ เช่น การตรวจจับขอบ การหาศูนย์กลาง หรือการตรวจจับจุดสนใจ ผู้วิจัยได้ใช้การแยกคุณลักษณะโดยธรรมชาติ สำหรับการทำให้เหมือนรูปภาพในระดับสากล (Global Level) เพื่อแยกคุณสมบัติส่วนกลางออกจากรูปภาพ บ่งบอกถึงความแตกต่างของภาพ และการกำหนดคุณสมบัติของภาพโดยรวม การคำนวณค่าต่างๆ จากรูปภาพแบบโหมดขาวดำ (Grayscale) โดยแยกคุณสมบัติของภาพทั้งหมดโดยใช้ค่าคุณสมบัติทั้ง 7 ค่า ได้แก่ ค่าเฉลี่ย (Mean), ค่ามัธยฐาน (Median), ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation), ค่ามาตรฐานสำหรับแกน X (Normalized X), ค่า

มาตรฐานสำหรับแกน Y (Normalized Y), ค่าสีเทาต่ำสุด (Min Gray Value), และค่าสีเทาสูงสุด (Max Gray Value) ดังตารางที่ 1

หลังจากสกัดคุณลักษณะออกมาแล้ว จะนำข้อมูลที่ได้ไปใช้งาน (Utilizing Data) ในการสร้างแบบจำลองการเรียนรู้ของเครื่อง โดยโปรแกรม RapidMiner Studio ที่ติดตั้งส่วนขยาย (IMMI-Image Extension for RapidMiner) การทำเหมืองรูปภาพ (Image Mining Extension) ซึ่งสามารถนำไปใช้ในการจำแนกประเภทรูปภาพ การตรวจจับวัตถุในรูปภาพ หรือการวิเคราะห์อื่นๆ เพื่อให้การวินิจฉัยโรคมะเร็งเป็นไปอย่างรวดเร็วและแม่นยำ



ภาพที่ 3 กระบวนการแปลงภาพให้อยู่ในรูปแบบข้อมูลที่มีโครงสร้าง (Structured Data)

ตารางที่ 1 ตัวอย่างข้อมูลรูปภาพที่ถูกแปลงให้อยู่ในรูปแบบของข้อมูลที่มีโครงสร้าง (Structured Data)

Row No.	Label	Mean	Median	Max Gray Value	Min Gray Value	Normalized X	Normalized Y	Standard Deviation
1	Normal	170.908	169	255	167	0.496	0.496	12.074
2	Normal	167.529	167	255	162	0.497	0.496	8.934
3	Normal	172.885	172	255	169	0.499	0.499	6.663
4	Malaria	153.962	167	255	73	0.499	0.482	32.282
5	Malaria	161.677	163	255	79	0.485	0.468	36.432
6	Malaria	152.372	169	255	79	0.516	0.539	33.813

ขั้นตอนถัดไปผู้วิจัยได้ทำการคัดเลือกปัจจัย (Feature Selection) ด้วยวิธีแรปเปอร์ (Wrapper Method) โดยใช้เทคนิคการคัดเลือกปัจจัยแบบไปข้างหน้า (Forward Selection) และ

การลดตัวแปร (Backward Elimination) เพื่อคัดเลือกปัจจัยที่มีผลต่อการสร้างตัวแปรสำหรับคัดกรองผู้ป่วยโรคมาลาเรีย

เทคนิคการคัดเลือกปัจจัยแบบไปข้างหน้า (Forward Selection) เป็นกระบวนการที่เริ่มต้นจากการไม่มีคุณสมบัติใดๆ ในชุดข้อมูล จากนั้นทำการทดสอบเพิ่มคุณสมบัติทีละตัว โดยวัดประสิทธิภาพของตัวแบบที่เกิดขึ้นแต่ละครั้ง เมื่อพบคุณสมบัติที่ทำให้ตัวแบบมีประสิทธิภาพดีที่สุด คุณสมบัตินั้นจะถูกเพิ่มเข้ามาในชุดคุณสมบัติที่เลือกแล้ว กระบวนการนี้จะดำเนินต่อไปจนกว่าการเพิ่มคุณสมบัติเพิ่มเติมจะไม่ช่วยเพิ่มประสิทธิภาพของตัวแบบอย่างมีนัยสำคัญ ประโยชน์ของวิธีนี้คือช่วยในการค้นหาคุณสมบัติที่มีผลต่อประสิทธิภาพของตัวแบบอย่างแท้จริง ง่ายต่อการทำความเข้าใจ และนำไปใช้งาน และช่วยลดความซับซ้อนของตัวแบบโดยไม่จำเป็นต้องพิจารณาคุณสมบัติที่ไม่สำคัญ (Sputowska, 2023)

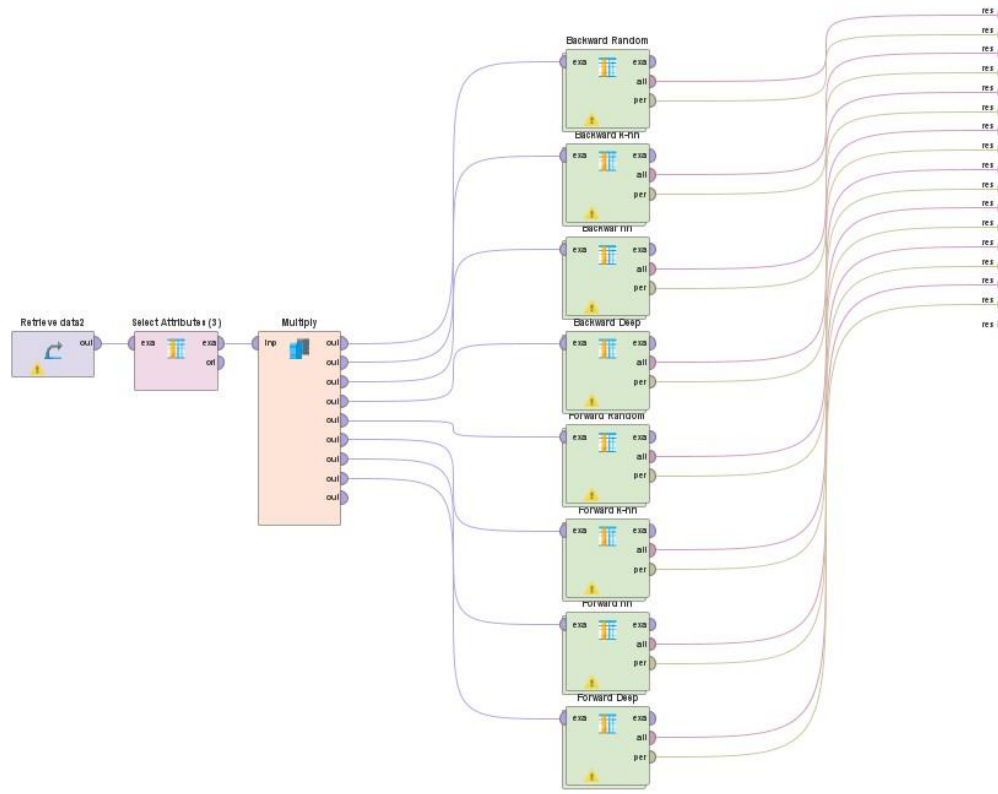
ในทางตรงกันข้าม เทคนิคการคัดเลือกปัจจัยการลดตัวแปร (Backward Elimination) เริ่มต้นจากการใช้คุณสมบัติทั้งหมดที่มีอยู่ในชุดข้อมูล แล้วทำการลบคุณสมบัติทีละตัว โดยลบคุณสมบัติที่มีผลน้อยที่สุดต่อประสิทธิภาพของตัวแบบออกทีละตัว กระบวนการนี้จะดำเนินต่อไปจนกว่าการลบคุณสมบัติเพิ่มเติมจะทำให้ประสิทธิภาพของตัวแบบลดลงอย่างมีนัยสำคัญ วิธีนี้ช่วยในการกำจัดคุณสมบัติที่ไม่สำคัญซึ่งอาจทำให้ตัวแบบมีความซับซ้อนเกินไป เพิ่มความสามารถในการทำนายของตัวแบบโดยคงไว้เฉพาะคุณสมบัติที่สำคัญ และทำให้ตัวแบบง่ายต่อการตีความและการใช้งานในเชิงปฏิบัติ (Gendron et al., 2023)

4. การสร้างแบบจำลอง (Modeling)

ในการคัดกรองผู้ป่วยโรคมาลาเรียด้วยภาพถ่ายฟิล์มเลือด ผู้วิจัยได้ใช้เทคนิคการคัดเลือกปัจจัยแบบไปข้างหน้า (Forward Selection) และเทคนิคการคัดเลือกปัจจัยการลดตัวแปร (Backward Elimination) ร่วมกับเทคนิคการจำแนกประเภทข้อมูลภาพ (Image Classification Technique) จำนวน 4 เทคนิค ได้แก่ เทคนิคเพื่อนบ้านใกล้สุด (k-Nearest Neighbors หรือ k-NN), โครงข่ายประสาทเทียม (Neural Network หรือ NN), การเรียนรู้เชิงลึก (Deep Learning หรือ DL) และต้นไม้ป่าสุ่ม (Random Forest หรือ RF) โดยใช้โปรแกรม RapidMiner Studio Version 10 ในการสร้างตัวแบบ

กระบวนการพัฒนาตัวแบบเริ่มต้นด้วยการนำข้อมูลภาพถ่ายฟิล์มเลือดเข้าผ่านโมดูล Retrieve Data เพื่อเตรียมข้อมูลพื้นฐานสำหรับการประมวลผลและวิเคราะห์ จากนั้นใช้โมดูล Select Attribute ในการเลือกคุณสมบัติที่สำคัญต่อการสร้างตัวแบบ ต่อมากระบวนการ Multiply ถูกใช้ในการสร้างชุดข้อมูลเพิ่มเติมเพื่อให้เพียงพอสำหรับการฝึกตัวแบบหลายรูปแบบ

ในการสร้างตัวแบบ k-NN, NN, DL, และ RF ใช้วิธีการเลือกคุณสมบัติแบบ Forward Selection และ Backward Elimination ผลลัพธ์จากการฝึกตัวแบบแต่ละตัวจะถูกนำมาวิเคราะห์และเปรียบเทียบเพื่อหาตัวแบบที่ดีที่สุด โดยภาพที่ 4 แสดงถึงกระบวนการสร้างตัวแบบโดยใช้โปรแกรม RapidMiner Studio Version 10 ซึ่งสรุปขั้นตอนที่กล่าวมาข้างต้น



ภาพที่ 4 กระบวนการสร้างตัวแบบโดยใช้โปรแกรม RapidMiner Studio Version 10

4.1 เทคนิคเพื่อนบ้านใกล้สุด (k-Nearest Neighbors: k-NN)

เทคนิคนี้ใช้วิธีการจัดแบ่งคลาสโดยตัดสินใจว่าคลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้ โดยการตรวจสอบจำนวนบางจำนวน “K” ในขั้นตอนวิธีการหาเพื่อนบ้านใกล้ที่สุดของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะนับจำนวนเงื่อนไขหรือกรณีต่างๆ สำหรับแต่ละคลาสและกำหนดเงื่อนไขใหม่ๆ ให้กับคลาสที่เหมือนกับคลาสที่ใกล้เคียงกันมากที่สุด (อนันต์ชัย ชูติภาสเจริญ และจรัญ แสนราช, 2018) เทคนิค k-NN แสดงดังสมการที่ 1 โดยมีสมการการทำงานดังนี้

$$D_{\text{Euclidian}}(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_{i,k} - x_{j,k})^2} \quad (1)$$

โดยที่ $D_{\text{Euclidian}}(x_i, y_i)$ คือ ระยะห่างระหว่างตัวอย่าง x_i กับตัวอย่าง y_i K คือ คุณสมบัติทั้งหมดของตัวอย่าง (i) เลือกค่าข้อมูลที่มีค่าระยะห่างน้อยที่สุด k ตัว เพื่อนำมาพิจารณาคำตอบ

การคำนวณหาค่า k สามารถทำได้หลายวิธี ขึ้นอยู่กับลักษณะของข้อมูลและปัญหาที่ต้องการแก้ไข หนึ่งในวิธีที่ใช้กันทั่วไปคือวิธีการเชิงวิวัฒนาการ (Evolutionary Algorithms) ซึ่งเป็นกระบวนการคำนวณซ้ำและการเลือกที่เลียนแบบการวิวัฒนาการทางชีววิทยาเพื่อหาค่า k ที่เหมาะสมที่สุด (Yang et al., 2022) กระบวนการนี้เริ่มต้นจากการสร้างกลุ่มประชากรเริ่มต้นของค่าที่เป็นไปได้ จากนั้นใช้การคัดเลือกตามความเหมาะสมเพื่อหาค่าที่มีประสิทธิภาพมากที่สุด ผ่านกระบวนการกลายพันธุ์และการผสมพันธุ์ของค่าที่เลือก จนกว่าจะได้ค่า k ที่เหมาะสมที่สุดสำหรับการทำงานของโมเดล (Wang & Zhang, 2021) ในงานวิจัยนี้ใช้วิธีคำนวณหาค่า k ด้วยวิธีการเชิงวิวัฒนาการ และกำหนดค่า k เท่ากับ 7

4.2 เทคนิคโครงข่ายประสาทเทียม (Neural Network: NN)

เป็นการจำลองการทำงานโครงข่ายประสาทมนุษย์ (Biological Neurons) โดยมีโหนดอินพุตเป็นตัวรับข้อมูลแล้วส่งต่อข้อมูลไปที่นิวรอน (Neuron) โดยผ่านค่าน้ำหนัก (Weight) ซึ่งนิวรอนจะมีการประมวลผลกับฟังก์ชันกระตุ้น (Activate Function) แบบขนานกับอินพุตโหนดตัวอื่นด้วย หลังจากนั้นจะส่งข้อมูลไปยังโหนดถัดไป โครงข่ายประสาทเทียมเรียนรู้ได้จากข้อมูลฝึกสอน (Training Data) โดยผ่านกระบวนการเรียนรู้ (Learning Process) เทคนิค NN แสดงดังสมการที่ (2) โดยมีสมการการทำงาน ดังนี้(ประมัย ชัยวัฒนคุปต์ และทรงยศ กิจธรรมเกษร, 2023)

$$n = \sum_{i=1}^2 X_i W_i + b \quad (2)$$

โดย n คือ ผลรวมที่ได้จากฟังก์ชันผลรวม

X_i คือ ค่าข้อมูลเข้าตัวที่ i

W_i คือ ค่าน้ำหนักของนิวรอนตัวที่ i

Z คือ จำนวนนิวรอนชั้นข้อมูลเข้า

B คือ ค่าความโน้มเอียง

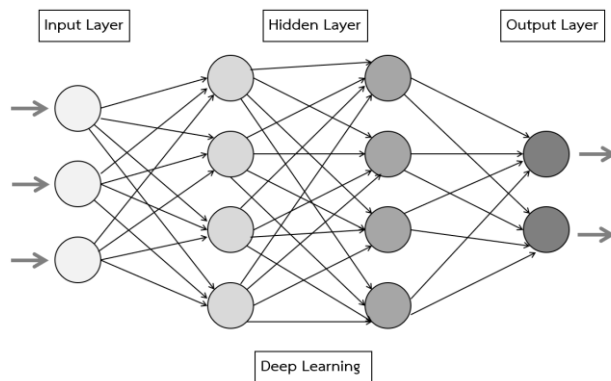
4.3 เทคนิคเรียนรู้เชิงลึก (Deep Learning: DL)

การเรียนรู้เชิงลึก (Deep Learning) เกี่ยวข้องกับการสร้างเครือข่ายประสาทลึก (Deep Neural Network) ที่ประกอบด้วยชั้นอินพุต (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้น

เอาต์พุต (Output Layer) (Ahmed et al., 2023) ดังแสดงในภาพที่ 5 ข้อมูลจะถูกส่งผ่านโหนดในเครือข่าย โดยแต่ละโหนดจะคำนวณผลรวมถ่วงน้ำหนัก (Weighted Sum) และใช้ฟังก์ชันการเปิดใช้งาน (Activation Function) สำหรับการแพร่กระจายไปข้างหน้า (Forward Propagation) ผลลัพธ์จะถูกเปรียบเทียบกับค่าจริง (Actual Value) เพื่อคำนวณการสูญเสีย (Loss) และการแพร่ย้อนกลับ (Backpropagation) จะปรับน้ำหนักโหนดตามไล่ระดับการสูญเสีย (Gradient Descent) (Lippmann, 2023) การฝึกอบรมจะดำเนินการเป็นรอบๆ (Epochs) เพื่อลดการสูญเสีย หลังการฝึกอบรม (Training) โมเดลจะถูกทดสอบด้วยข้อมูลใหม่โดยใช้เมตริก (Metrics) เช่น ความแม่นยำ (Accuracy) ความแม่นยำเชิงลึก (Precision) การเรียกคืน (Recall) และค่า F1-Score (Wang & Guet, 2021)

การเรียนรู้เชิงลึกมีความโดดเด่นในหลายงาน เช่น การจดจำภาพ (Image Recognition) การประมวลผลภาษาธรรมชาติ (Natural Language Processing) และการคาดการณ์ผลลัพธ์ (Predictive Modeling) เนื่องจากความสามารถในการเรียนรู้รูปแบบข้อมูลที่ซับซ้อน หนึ่งในโครงข่ายประสาทเทียมที่สำคัญคือเครือข่ายประสาทเทียมแบบคอนโวลูชันนัล (Convolutional Neural Network: CNN) ซึ่งมีบทบาทสำคัญในการประมวลผลภาพและการจำแนกประเภทข้อมูลในหลายด้าน เช่น การจดจำใบหน้า การจำแนกวัตถุ และการวิเคราะห์ภาพจากเครื่องมือการแพทย์ ความสามารถของ CNN ในการเรียนรู้ลักษณะเฉพาะของข้อมูลผ่านเลเยอร์ที่ซ้อนอยู่หลายชั้นช่วยให้สามารถสกัดคุณลักษณะที่สำคัญออกมาได้โดยอัตโนมัติ ทำให้เหมาะสมกับงานที่ต้องการการตีความและการวิเคราะห์ที่ซับซ้อน (Bhatti et al., 2023)

งานวิจัยหลายชิ้นพบว่าจำนวนเลเยอร์ที่ซ้อนอยู่ใน CNN มีผลกระทบโดยตรงต่อประสิทธิภาพของโมเดล โดยเครือข่ายที่มีเลเยอร์ที่ซ้อนอยู่มากขึ้นมักจะสามารถจำแนกประเภทข้อมูลได้ดีกว่าเนื่องจากมีความสามารถในการจับลักษณะเฉพาะที่ซับซ้อนและแตกต่างกันได้มากขึ้น อย่างไรก็ตาม การเพิ่มจำนวนเลเยอร์อาจนำไปสู่ปัญหาการเรียนรู้มากเกินไป (overfitting) ซึ่งส่งผลให้โมเดลทำงานได้ดีในชุดข้อมูลฝึกหัด แต่ทำงานได้ไม่ดีเมื่อพบข้อมูลใหม่ (Deng, 2023) นอกจากนี้ จำนวนรอบที่ข้อมูลทั้งหมดถูกใช้ในการฝึกอบรมโมเดล (Epochs) ก็มีบทบาทสำคัญในการกำหนดประสิทธิภาพของ CNN โดยการเพิ่มจำนวน Epochs ทำให้โมเดลมีโอกาสดัดแปลงน้ำหนักของเครือข่ายได้ดียิ่งขึ้น แต่ก็เสี่ยงที่จะเกิดการปรับแต่งมากเกินไปหากจำนวน Epochs สูงเกินไป ซึ่งอาจทำให้โมเดลไม่สามารถทำนายผลของข้อมูลใหม่ที่ไม่เคยเห็นมาก่อนได้ดี (Marcu & Grava, 2023)



ภาพที่ 5 เทคนิคเรียนรู้เชิงลึก (Deep Learning)

4.4 เทคนิคต้นไม้ป่าสุ่ม (Random Forest: RF)

เป็นเทคนิคที่พัฒนาต่อยอดมาจากเทคนิคต้นไม้ตัดสินใจ โดยจะมีการเพิ่มจำนวนต้นไม้ (Tree) เป็นหลายๆต้น แต่ละต้นจะได้รับคุณลักษณะ (Feature) และข้อมูล (Data) ที่ไม่เหมือนกันทั้งหมด เพื่อให้ได้ต้นไม้ที่มีรูปแบบและอิสระต่อกันมาก ทำให้ประสิทธิภาพของการทำนายสูงขึ้น จากนั้นจะทำการให้คะแนน (Vote) โดยต้นไม้ในป่าหากต้นไม้ใดได้คะแนนมากที่สุดก็นำต้นไม้นั้นมาสร้างเป็นตัวแบบสำหรับการพยากรณ์ต่อไป (ธวัชชัย เหล็กดี รัฐพรพน สันติอนันทยัย และเจษฎา อุดมพิทยาสรรพ์, 2023)

5. การประเมินผล (Evaluation)

ในส่วนการประเมินผล ผู้วิจัยได้แบ่งข้อมูลออกเป็น 2 ส่วน คือ 1) ส่วนข้อมูลฝึกสอนเพื่อสร้างตัวแบบ (training Set) เพื่อทำหน้าที่สร้างตัวแบบ 2) ส่วนข้อมูลทดสอบ (Testing Set) จากนั้นทำการทดสอบประสิทธิภาพของตัวแบบด้วยวิธี Cross Validation โดยแบ่งข้อมูลออกเป็น 10 ส่วน (10-fold cross validation) และทำการทดสอบประสิทธิภาพของการจำแนกประเภทข้อมูลโดยการวัดค่าความแม่นยำ (Accuracy) ค่าประสิทธิภาพโดยรวม (F-Measure) ค่าความเร็ว (Sensitivity) และค่าจำเพาะ (Specificity) ทำการวัดประสิทธิภาพผลลัพธ์แบบจำลองที่เหมาะสมสำหรับการคัดกรองผู้ป่วยโรคมะเร็งจากภาพถ่ายฟิล์มเลือด ซึ่งตัววัดประสิทธิภาพของแบบจำลองประกอบไปด้วย เมทริกซ์ความสับสน (Confusion Matrix) เป็นการเก็บข้อมูลที่เกี่ยวข้องกับการแยกข้อมูลที่เกิดจากการทำนายด้วยระบบแบ่งแยก ดังตารางที่ 2 ซึ่งมี 2 ระดับ

ตารางที่ 2 ตารางความสับสน (Confusion Matrix)

	Predicted Positive (Yes)	Predicted Negatives (No)
Actual Positive (Yes)	True Positive (TP)	False Negative (FN)
Actual Negatives (No)	False Positive (FP)	True Negative (TN)

โดยที่

True Positive (TP) คือ กรณีที่ทำนายว่าจริงและสิ่งที่เกิดขึ้นคือจริง

True Negative (TN) คือ กรณีที่ทำนายว่าไม่จริงและสิ่งที่เกิดขึ้นคือไม่จริง

False Positive (FP) คือ กรณีที่ทำนายว่าจริงแต่สิ่งที่เกิดขึ้นคือไม่จริง

False Negative (FN) คือ กรณีที่ทำนายว่าไม่จริงแต่สิ่งที่เกิดขึ้นคือจริง

1. การวัดค่าความแม่นยำ (Accuracy) เป็นค่าที่ได้จากวิธีการทดสอบเพื่อหาค่าพยากรณ์ความถูกต้องของข้อมูลโดยคิดเป็นร้อยละ (%) คำนวณได้จากสมการที่ (3) ดังนี้

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (3)$$

2. ค่าประสิทธิภาพโดยรวม (F-Measure) คือ ค่าที่เกิดจากการเปรียบเทียบระหว่างค่า Precision และ Recall ของแต่ละคลาสเป้าหมาย คำนวณได้จากสมการที่ (4) และ (5) ดังนี้

F-Measure ของคลาสเป้าหมาย YES

$$\text{F-Measure}_{\text{YES}} = \frac{2 \times \text{Precision}(\text{YES}) \times \text{Recall}(\text{YES})}{\text{Precision}(\text{YES}) + \text{Recall}(\text{YES})} \quad (4)$$

F-Measure ของคลาสเป้าหมาย NO

$$\text{F-Measure}_{\text{NO}} = \frac{2 \times \text{Precision}(\text{NO}) \times \text{Recall}(\text{NO})}{\text{Precision}(\text{NO}) + \text{Recall}(\text{NO})} \quad (5)$$

3. ค่าความเร็ว (Sensitivity) คือ ค่าที่แบบจำลองสามารถพยากรณ์ข้อมูลผู้ป่วยที่เกิดโรคได้อย่างถูกต้องต่อผู้ป่วยที่เกิดโรคขึ้นจริง คำนวณได้จากสมการที่ (6) ดังนี้

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)$$

4. ค่าจำเพาะ (Specificity) คือ ค่าที่แบบจำลองสามารถพยากรณ์ข้อมูลผู้ป่วยที่ไม่เกิดโรคได้อย่างถูกต้องต่อผู้ป่วยที่ไม่เกิดโรคขึ้นจริง คำนวณได้จากสมการที่ (7) ดังนี้

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (7)$$

6. การนำไปใช้ (Deployment)

เมื่อครบทั้ง 5 ขั้นตอนแล้ว จะได้ผลลัพธ์คือเทคนิคที่เหมาะสมสำหรับสร้างตัวแบบการพยากรณ์ในการคัดกรองผู้ป่วยโรคมาลาเรีย เพื่อเป็นประโยชน์ในการคัดกรองผู้ป่วยโรคมาลาเรีย

กำหนดกลยุทธ์ในการลดปริมาณผู้ป่วย สนับสนุนการตัดสินใจสำหรับแพทย์คัดกรองผู้ป่วยโรคมาลาเรีย สนับสนุนนโยบายด้านการวางแผนการรักษาให้เพียงพอต่อจำนวนผู้ป่วยที่จะเกิดขึ้นในอนาคต ทำให้การวินิจฉัยผู้ป่วยโรคมาลาเรียมีความรวดเร็วและแม่นยำมากยิ่งขึ้น สามารถวิเคราะห์การเกิดโรคมาลาเรียและสามารถให้การรักษาเบื้องต้นผู้ป่วยโรคมาลาเรียเพื่อนำไปสู่กระบวนการรักษาต่อไป

ผลการวิจัย

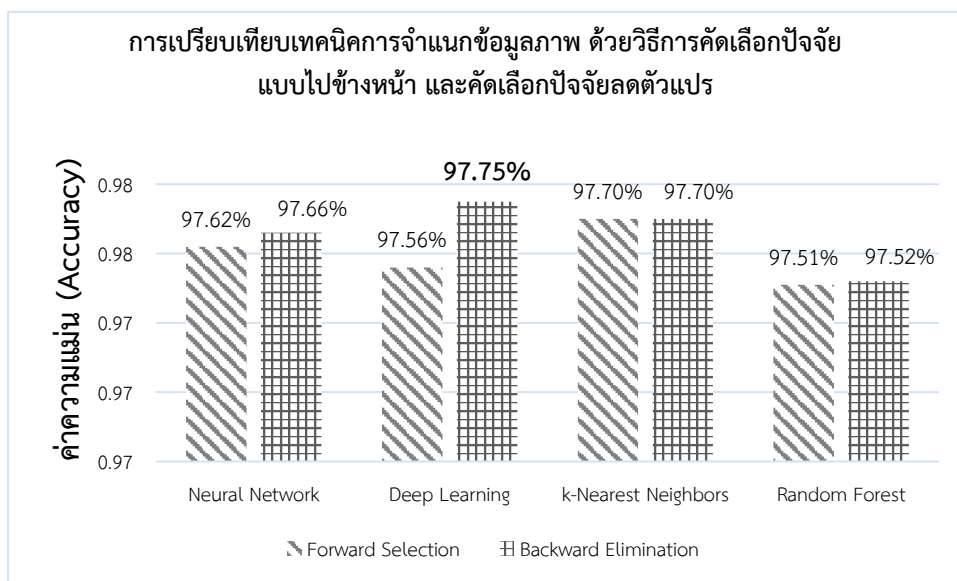
ผลการวิเคราะห์การเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับการคัดกรองผู้ป่วยโรคมาลาเรียด้วยภาพถ่ายฟิล์มเลือดที่ผ่านกระบวนการแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลที่มีโครงสร้างและนำปัจจัยที่ผ่านการคัดแยกคุณสมบัติเด่นของรูปภาพออกมาคัดเลือกปัจจัยด้วยเทคนิคการคัดเลือกปัจจัยแบบไปข้างหน้า (Forward selection) และการคัดเลือกปัจจัยลดตัวแปร (Backward Elimination) จากนั้นสร้างตัวแบบการพยากรณ์ด้วยเทคนิคการจำแนกประเภทข้อมูลภาพทั้ง 4 เทคนิค ประกอบด้วย เทคนิคเพื่อนบ้านใกล้สุด (k-NN) เทคนิคโครงข่ายประสาทเทียม (NN) เทคนิคเรียนรู้เชิงลึก (DL) และเทคนิคต้นไม้ป่าสุ่ม (RF) นำมาประเมินประสิทธิภาพการจำแนกประเภทข้อมูลภาพด้วยการแบ่งข้อมูลออกเป็น 2 ชุด (Training set and testing set) ด้วยวิธีการ 10-Fold Cross Validation และวัดค่าประสิทธิภาพการจำแนกประเภทข้อมูลด้วยค่าความแม่นยำ (Accuracy) ค่าประสิทธิภาพโดยรวม (F-Measure) ค่าความเร็ว (Sensitivity) และค่าจำเพาะ (Specificity) ดังแสดงผลลัพธ์ของค่าประสิทธิภาพการจำแนกข้อมูลภาพของแต่ละเทคนิค ในตารางที่ 3

ตารางที่ 3 ตารางเปรียบเทียบค่าประสิทธิภาพทั้ง 5 เทคนิค

Feature Selection Techniques	Image Classification Techniques	Classification Performance			
		Accuracy	F-Measure	Sensitivity	Specificity
Forward Selection	NN	97.62%	97.60%	96.88%	98.36%
	DL	97.56%	97.54%	97.23%	97.89%
	k-NN	97.70%	97.67%	97.03%	98.36%
	RF	97.51%	97.48%	96.68%	98.32%
Backward Elimination	NN	97.66%	97.64%	97.11%	98.20%
	DL*	97.75%	97.73%	97.19%	98.31%
	k-NN	97.70%	97.67%	97.03%	98.36%
	RF	97.52%	97.49%	96.70%	98.34%

*คือ เทคนิคการคัดเลือกปัจจัยและเทคนิคการจำแนกประเภทข้อมูลภาพที่มีความเหมาะสมที่สุดในการนำไปใช้สร้างตัวแบบสำหรับคัดกรองผู้ป่วยโรคมะเร็ง

จากตารางที่ 3 แสดงผลการวิเคราะห์หาค่าประสิทธิภาพการจำแนกข้อมูลโดยใช้การคัดเลือกปัจจัย 2 เทคนิค และการจำแนกประเภทข้อมูลภาพทั้ง 4 เทคนิค พบว่าเทคนิคเรียนรู้เชิงลึก (DL) ที่ใช้การคัดเลือกปัจจัยลดตัวแปร (Backward Elimination) แสดงความแม่นยำสูงสุด (Accuracy) 97.75% บ่งบอกว่าตัวแบบการเรียนรู้เชิงลึกนี้มีความสามารถในการจำแนกโดยรวมทั้งตัวอย่างที่เป็นและไม่เป็นโรคมะเร็งได้ดีที่สุด และมีค่าประสิทธิภาพโดยรวม (F-Measure) ที่ 97.73% แสดงถึงความสามารถของตัวแบบนี้ในการรักษาสมดุลระหว่างความแม่นยำของการจำแนกผู้ป่วยจริง (Precision) และความสามารถในการค้นพบผู้ป่วย (Recall) สูงเมื่อเปรียบเทียบกับเทคนิคอื่นๆ งานวิจัยนี้มีจุดมุ่งหมายในการพัฒนาเทคนิคที่ทำให้การตรวจจับโรคมะเร็งจากภาพถ่ายฟิล์มเลือดนั้นเร็วขึ้นและความแม่นยำสูงขึ้น โดยมีการคัดเลือกปัจจัยที่ต่างกันสองวิธีเพื่อดูผลกระทบต่อประสิทธิภาพ จากผลการวิจัยแสดงให้เห็นว่าเทคนิคเรียนรู้เชิงลึก (DL) ใช้การคัดเลือกปัจจัยลดตัวแปร (Backward Elimination) มีประสิทธิภาพที่ดีที่สุดในเกือบทุกด้าน ดังในภาพที่ 6 แสดงการเปรียบเทียบเทคนิคการจำแนกข้อมูลภาพ ด้วยวิธีการคัดเลือกปัจจัยแต่ละแบบ



ภาพที่ 6 การเปรียบเทียบค่าทดสอบประสิทธิภาพของการจำแนกข้อมูลภาพ

อภิปรายผล

งานวิจัยนี้ศึกษาเทคนิคการประมวลผลภาพและเทคนิคการจำแนกภาพเพื่อคัดกรองโรคมาลาเรียจากภาพถ่ายฟิล์มเลือด ผลลัพธ์ชี้ให้เห็นว่าเทคนิคการเรียนรู้เชิงลึก (Deep Learning: DL) ที่ใช้การคัดเลือกปัจจัยลดตัวแปร (Backward Elimination) มีประสิทธิภาพสูงสุด โดยมีค่าความแม่นยำ (Accuracy) สูงที่สุด และมีค่าประสิทธิภาพโดยรวม (F-Measure) ที่เหนือกว่าเทคนิคอื่นๆ การศึกษานี้สอดคล้องกับการศึกษาเทคนิคการเรียนรู้เชิงลึกที่ผ่านมา ซึ่งมีประสิทธิภาพในการวินิจฉัยโรคมาลาเรียดีกว่าเทคนิคอื่นๆ การเปรียบเทียบนี้เสริมความเข้าใจว่าเทคนิคการเรียนรู้เชิงลึกเมื่อประยุกต์ใช้กับข้อมูลภาพถ่ายจากกล้องจุลทรรศน์สามารถเพิ่มความแม่นยำในการวินิจฉัยได้มากกว่าเทคนิคการวิเคราะห์ข้อมูลดั้งเดิม ซึ่งช่วยยืนยันถึงความสำคัญของการใช้ข้อมูลภาพคุณภาพสูงและเทคนิคการเรียนรู้เชิงลึกในการพัฒนาเครื่องมือวินิจฉัยในสถานที่ที่มีทรัพยากรจำกัด และยังมีศักยภาพในการส่งเสริมการวินิจฉัยโรคอื่นๆ ผ่านภาพถ่ายทางการแพทย์

ในการวิจัยนี้ การใช้เทคนิคการคัดเลือกปัจจัยลดตัวแปรร่วมกับเทคนิคการเรียนรู้เชิงลึกสามารถสร้างโมเดลที่มีประสิทธิภาพสูงสุดในการคัดกรองผู้ป่วยโรคมาลาเรีย ผลการวิจัยนี้ชี้ให้เห็นถึงความสำคัญของการพัฒนาเทคนิคที่เหมาะสมกับข้อมูลที่มีโครงสร้างซับซ้อนเพื่อเพิ่มความแม่นยำในการวินิจฉัยโรค อย่างไรก็ตาม การใช้งานข้อมูลภาพถ่ายจากต่างประเทศอาจไม่สอดคล้องกับบริบทในประเทศไทย การวิจัยในอนาคตควรพิจารณาเปรียบเทียบกับชุดข้อมูลที่มีลักษณะเฉพาะตามภูมิภาคเพื่อให้เข้ากับบริบทของพื้นที่และผู้คนได้มากขึ้น นอกจากนี้ ขนาดของภาพถ่ายฟิล์มเลือดที่แตกต่าง

กันอาจส่งผลต่อความแม่นยำของผลการทดลอง จึงจำเป็นต้องปรับเทคนิคการประมวลผลภาพให้สอดคล้องกับชนิดของข้อมูลที่ใช้ในการวิเคราะห์เพื่อเพิ่มความแม่นยำและความเชื่อถือได้ของโมเดล อีกทั้งควรพิจารณาศึกษาเทคนิคการเรียนรู้แบบมีผู้สอนและไม่มีผู้สอนอื่นๆให้หลากหลายมากขึ้น

การวิจัยเพิ่มเติมในด้านการประยุกต์ใช้เทคนิคเหล่านี้อาจช่วยพัฒนาความสามารถในการวินิจฉัยโรคมะเร็งและโรคอื่นๆ ผ่านภาพถ่ายทางการแพทย์ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น การสร้างฐานข้อมูลที่ครอบคลุมและการปรับปรุงเทคนิคการประมวลผลภาพให้ทันสมัยสามารถเพิ่มศักยภาพในการนำไปใช้จริงในภาคสนามได้อย่างมีประสิทธิภาพ

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จได้ด้วยความเอื้อเฟื้อข้อมูลที่เป็นประโยชน์ จากทีมวิจัย K.M.Faizullah Fuhadet ประเทศบังคลาเทศ พร้อมด้วยการมอบโอกาสและให้การสนับสนุนในการทำงานวิจัยครั้งนี้ จากคณะกรรมการบัณฑิตและการจัดการ มหาวิทยาลัยมหาสารคาม จนทำให้สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยขอขอบพระคุณและหวังเป็นอย่างยิ่งว่างานวิจัยนี้จะก่อให้เกิดประโยชน์แก่ผู้ที่สนใจทุกท่าน

เอกสารอ้างอิง

- โรงพยาบาลศิริราชปิยมหาราชารุณย์. (2022) ไข้มาลาเรีย โรคติดต่อจากยุงก้นปล่อง. Retrieved January 8, 2024, from <https://www.siphphospital.com/th/news/article/share/573>
- ธนานพ จันภักดี และฐิติรัตน์ ศิริบรรณกุล. (2018). การเรียนรู้เชิงลึกสำหรับตรวจจับความผิดปกติจากภาพในวิดีโอการเต้นสำหรับเด็ก. วารสารวิทยาศาสตร์ประยุกต์, 21(2), 1-16.
- ธวัชชัย เหล็กดี รัฐพรหม สันติอนันต์ และเจษฎา อุดมพิทยาสรรพ์. (2023). การเปรียบเทียบประสิทธิภาพของแบบจำลองการทำนายความเสี่ยงโรคมะเร็งปอดด้วยเทคนิคเหมืองข้อมูล. Science and Technology to Community, 2(1), 22-35.
- ประมัย ชัยวัฒนคุปต์ และทรงยศ กิจธรรมเกษร. (2023). การวิเคราะห์ประสิทธิภาพการควบคุมสัญญาณไฟจราจรบริเวณทางแยกกันสามทิศทางโดยใช้วิธีโครงข่ายประสาทเทียม. การประชุมวิชาการวิศวกรรมโยธาแห่งชาติ ครั้งที่ 28, (47), 1-8.
- มาเรีย นิน่า จิตะสมบัติ. (2023) มาลาเรีย (Malaria). Retrieved January 8, 2024, from <https://www.medparkhospital.com/disease-and-treatment/malaria>
- อนันต์ชัย ชูติภาสเจริญ และจรัญ แสนราช. (2018). การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้าม

ประเทศของบุคคลทั่วไป. วารสารวิจัย มหาวิทยาลัยขอนแก่น (ฉบับบัณฑิตศึกษา) สาขา
มนุษยศาสตร์และสังคมศาสตร์, 6(3), 105-113.

อนุพงศ์ สุขประเสริฐ. (2021). คู่มือการทำเหมืองข้อมูล โปรแกรม RAPIDMINER STUDIO พิมพ์
ครั้งที่ 4. มหาสารคาม : สาขาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัย

มหาสารคามAn, Q., Rahman, S., Zhou, J., & Kang, J. J. (2023). A comprehensive
review on machine learning in healthcare industry: classification, restrictions,
opportunities and challenges. **Sensors**, 23(9), 4178.

<https://doi.org/10.3390/s23094178>

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiaq, T., Rafa, N., ... &
Gandomi, A. H. (2023). Deep learning modelling techniques: current progress,
applications, advantages, and challenges. **Artificial Intelligence Review**,
56(11), 13521-13617. <https://doi.org/10.1007/s10462-023-10466-8>

Bhatti, U. A., Tang, H., Wu, G., Marjan, S., & Hussain, A. (2023, February 28). Deep
Learning with Graph Convolutional Networks: An Overview and Latest
Applications in Computational Intelligence. **International Journal of
Intelligent Systems**, 2023, 1–28. <https://doi.org/10.1155/2023/8342104>

Borboudakis, G., & Tsamardinos, I. (2019). Forward-backward selection with early
dropping. **Journal of Machine Learning Research**, 20(8), 1-39.

Deng, Z. (2023, June 14). Investigation of influence of additional convolutional and
max-pooling layers in general adversarial network. **Applied and
Computational Engineering**, 5(1), 390–397. <https://doi.org/10.54254/2755-2721/5/20230604>

Ding, Y., Feng, H., Yang, Y., Holmes, J., Liu, Z., Liu, D., ... & Liu, W. (2023). Deep-
learning based fast and accurate 3D CT deformable image registration in lung
cancer. **Medical physics**, 50(11), 6864-6880. <https://doi.org/10.1002/mp.16548>

Farhan, A. M. Q., & Yang, S. (2023). Automatic lung disease classification from the
chest X-ray images using hybrid deep learning algorithm. **Multimedia Tools
and Applications**, 82(25), 38561-38587. <https://doi.org/10.1007/s11042-023-15047-z>

Fuhad, K. F., Tuba, J. F., Sarker, M. R. A., Momen, S., Mohammed, N., & Rahman, T.
(2020). Deep learning based automatic malaria parasite detection from blood

- smear and its smartphone based application. **Diagnostics**, 10(5), 329.
<https://doi.org/10.3390/diagnostics10050329>
- Hamisu, I. A. (2019). **Malaria Prediction using Bayesian and other Machine Learning Techniques** (Doctoral dissertation).
<http://repository.aust.edu.ng/xmlui/bitstream/handle/123456789/4965/Hamisu%20Ismail%20Ahmad.pdf?sequence=1&isAllowed=y>
- Hansson, A., & Andersen, M. (2023). **Optimization for Learning and Control**. John Wiley & Sons. (pp. 243-296)
- Julianto, I. T., Kurniadi, D., Fauziah, F. A., & Rohmanto, R. (2023). Improvement of Data Mining Models using Forward Selection and Backward Elimination with Cryptocurrency Datasets. **Journal of Applied Intelligent System**, 8(1), 100-109. <https://www.doi.org/10.33633/jais.v8i1.7568>
- Kozbur, D. (2020). Analysis of testing-based forward model selection. **Econometrica**, 88(5), 2147-2173. <https://www.doi.org/10.3982/ECTA16273>
- Lakshmanaprabu S.K. et al. (2019). Optimal deep learning model for classification of lung cancer on CT images. **Future generation computer systems**, 374-382.
- Lippmann, R. (2023, February 17). Understanding and Applying Deep Learning. **Neural Computation**. https://doi.org/10.1162/neco_a_01518
- Marcu, D. C., & Grava, C. (2023, June 9). The Importance of Data Quality in Training a Deep Convolutional Neural Network. **2023 17th International Conference on Engineering of Modern Electric Systems (EMES)**.
<https://doi.org/10.1109/emes58375.2023.10171785>
- Opeyemi A. Abisoye and Rasheed G. Jimoh (2018). comparative study on the prediction of symptomatic and climatic based malaria parasite counts using machine learning models.I.J. **Modern Education and Computer Science**, 18-25.
- Pierdzioch, C. (2023). On the efficiency of growth forecasts for Germany: an application of forward and backward predictor variable selection. **Applied Economics Letters**, 1-4. <https://www.doi.org/10.1080/13504851.2023.2207811>
- QinZeng Song et al. (2017). Using deep learning for classification of lung nodules on computed tomography images. **Journal of healthcare engineering**, 1-7.

- Rana, M., & Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection. **Multimedia Tools and Applications**, 82(17), 26731-26769. <https://doi.org/10.1007/s11042-022-14305-w>
- The Coverage. (2022) **สร.พยากรณ์โรคและภัยร้ายลับตาห้ คัดพบผู้ป่วย โรคมาลาเรีย จากยุ่ง ก้น ปล่องเพิ่มขึ้น**. Retrieved January 8, 2024, from <https://www.thecoverage.info/news/content/3799>.
- Wang, H., & Zhang, J. (2021). An $O(n \log n)$ -Time Algorithm for the k-Center Problem in Trees. **SIAM Journal on Computing**, 50(2), 602-635. <https://www.doi.org/10.1137/18M1196522>
- Wang, Z., & Guet, C. (2021). Deep learning in physics: a study of dielectric quasi-cubic particles in a uniform electric field. **IEEE Transactions on Emerging Topics in Computational Intelligence**, 6(3), 429-438. <https://doi.org/10.1109/TETCI.2021.3086237>
- World Health Organization. (2023) **Malaria**. Retrieved January 8, 2024, from <https://www.who.int/news-room/fact-sheets/detail/malaria>.
- Yang, M., Pang, J., Zhang, Z., Fu, J., Fan, H., Zhang, Y., ... & Han, J. (2022). K value: An indicator that can characterize the cold and hot properties of traditional Chinese medicines. **Frontiers in Pharmacology**, 13, 877102. <https://www.doi.org/10.3389/fphar.2022.877102>