

Received: 27 เม.ย. 2567

Revised: 9 ก.ค. 2567

Accepted: 11 ก.ค. 2567

การพัฒนาตัวแบบการเรียนรู้ของเครื่องเพื่อการแบ่งชั้นความเสี่ยงและ
การพยากรณ์โรคมะเร็งปอดขั้นสูงในภาพเอกซเรย์คอมพิวเตอร์

Development of Machine Learning Models for Risk Stratification and
Prognostication of Advanced Lung Cancer in Computer Tomography Images

เลอศักดิ์ โพธิ์ทอง¹, อริสรา ผดุงเจริญ¹ และ ศุภกร ศรีสง่า^{1*}

¹สาขาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม

Lersak Phothong¹, Arisara Phadungcharoen¹ and Supakorn Srisnga^{1*}

¹Department of Business Computer, Mahasarakham Business School,
Mahasarakham University

* ผู้นิพนธ์ประสานงาน: ศุภกร ศรีสง่า อีเมล: 010974011@msu.ac.th

Abstract

The accurate identification of lung cancer stages is crucial for effective screening and treatment planning to extend the survival of patients. This study explores the development of machine learning models capable of accurately classifying early-stage and late-stage lung cancer patients using 800 chest computed tomography (CT) images from the dataset on Kaggle.com. The standard CRISM-DM process was used to guide data analysis and model development, aiming to extract useful features from the images that could be potential indicators of cancerous lesions. Subsequently, the performance of three machine learning algorithms, Deep Learning (DL), Artificial Neural Networks (ANN), and k-Nearest Neighbors (k-NN), was compared for the task of lung cancer stage classification. This investigation revealed that the k-NN model outperformed DL and ANN in this specific context. The k-NN model attained an accuracy of 96%, Sensitivity of 96.25%, and Specificity of 95.75%. This research demonstrates the potential of k-NN as a beneficial and efficient tool for classifying stages of lung cancer using CT images. This advancement holds the potential to

significantly accelerate diagnostic processes for medical teams and enhance patient care strategies, thereby improving the efficacy of patient management.

Keywords: Lung Cancer Prognostication; Machine Learning; Risk Stratification; Medical Imaging; Computed Tomography Imaging

บทคัดย่อ

การจำแนกระยะของมะเร็งปอดอย่างแม่นยำมีความสำคัญอย่างยิ่งต่อการตรวจคัดกรองและการวางแผนการรักษาที่มีประสิทธิภาพเพื่อยืดอายุการรอดชีวิตของผู้ป่วย งานวิจัยนี้มุ่งเน้นการพัฒนาตัวแบบการเรียนรู้ของเครื่องที่มีความสามารถในการจำแนกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและระยะลุกลามโดยใช้ภาพเอกซเรย์คอมพิวเตอร์ (CT) จำนวน 800 ภาพ จากชุดข้อมูล Kaggle.com โดยวิเคราะห์ข้อมูลและพัฒนาแบบจำลอง ตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล (CRISMDM) ด้วยการสกัดคุณลักษณะที่เป็นประโยชน์จากภาพ ซึ่งเป็นสัญญาณบ่งชี้ของเนื้อร้ายมะเร็ง จากนั้นทำการเปรียบเทียบประสิทธิภาพการจำแนกระยะของมะเร็งปอดของตัวแบบการเรียนรู้ของเครื่อง 3 เทคนิค ได้แก่ การเรียนรู้เชิงลึก (Deep Learning, DL) , โครงข่ายประสาทเทียม (Artificial Neural Networks, ANN) และเพื่อนบ้านที่ใกล้ที่สุด (k-Nearest Neighbors, k-NN) ผลการวิจัยพบว่า โมเดล k-NN มีประสิทธิภาพเหนือกว่า DL และ ANN ในบริบทเฉพาะนี้ โดยสามารถจำแนกได้อย่างแม่นยำถึง 96% มีความไว 96.25% และความจำเพาะ 95.75% งานวิจัยนี้แสดงให้เห็นถึงศักยภาพของ k-NN ว่ามีแนวโน้มที่เป็นประโยชน์และมีประสิทธิภาพสำหรับการจำแนกระยะของมะเร็งปอดโดยใช้ภาพเอกซเรย์คอมพิวเตอร์ โดยเป็นความก้าวหน้าที่มีศักยภาพในการเพิ่มความรวดเร็วการวินิจฉัยของทีมนแพทย์และการปรับปรุงกลยุทธ์การดูแลผู้ป่วยให้เหมาะสมยิ่งขึ้น

คำสำคัญ: การพยากรณ์โรคมะเร็งปอด; การเรียนรู้ของเครื่อง; การจำแนกความเสี่ยง; ภาพถ่ายทาง การแพทย์; ภาพเอกซเรย์คอมพิวเตอร์

บทนำ

มะเร็งปอดเป็นหนึ่งในปัญหาสุขภาพที่มีความอันตรายและร้ายแรงสูงสุด อีกทั้งเป็นมะเร็งที่มีอัตราการพบได้มากที่สุดตามสถิติมะเร็งของไทย และเป็นสาเหตุของการเสียชีวิตอันดับต้นๆ ของโลก (รัชชัย เหล็กดี, 2023) ด้วยจำนวนผู้ที่เสียชีวิตจากการวินิจฉัยที่ล่าช้าอยู่ที่ประมาณ 1.76 ล้านคนในปี 2561 (Yasemin Gültepe, 2021) ส่งผลกระทบต่อคุณภาพชีวิตทั้งทางด้านร่างกายและจิตใจของ

ผู้ป่วย รวมถึงผู้ที่อยู่รอบข้าง ปัจจัยการเกิดมะเร็งมีสาเหตุมาจากพันธุกรรม สารก่อมะเร็งจากการประกอบอาชีพ และมลพิษในอากาศ (สนหวัช ไชยวงศ์ และคณะ, 2016)

การวินิจฉัยและคัดแยกผู้ป่วยมะเร็งปอดเป็นขั้นตอนที่มีความสำคัญสำหรับการวางแผนการรักษาและดูแลสุขภาพของผู้ป่วย (Wang, et al, 2020) โดยเฉพาะการแยกแยะระหว่างผู้ป่วยมะเร็งปอดระยะเริ่มต้นและระยะลุกลาม ซึ่งหมายถึงการแบ่งชั้นความเสี่ยงของโรคให้ชัดเจน การแบ่งชั้นความเสี่ยงนี้หมายถึงการจำแนกระยะของโรคว่าเป็นระยะเริ่มต้นหรือระยะลุกลาม เพื่อให้สามารถกำหนดยาและเทคนิคการรักษาที่เหมาะสมกับระยะของมะเร็งได้ (รัชนิ นามจินตรา และ สุชาติ รุ่งแจ้ง, 2016)

ความก้าวหน้าในเทคโนโลยีการเรียนรู้ของเครื่อง รวมถึงเทคนิคการเรียนรู้เชิงลึก เช่น เครือข่ายประสาทคอนโวลูชัน (CNNs) (Huang, et al. 2023), เครือข่ายประสาทเทียม (ANNs) (Farhad, et al. 2023), และเทคนิคเพื่อนบ้านที่ใกล้ที่สุด (k-NN) (Malik & Singh, 2024) ได้ปรับปรุงการวินิจฉัยและการรักษาทางการแพทย์อย่างมีนัยสำคัญ การวินิจฉัยโดยการวิเคราะห์ภาพทางการแพทย์ด้วยเทคนิค Machine Learning แตกต่างจากการวินิจฉัยแบบดั้งเดิม เนื่องจากมีความสามารถในการประมวลผลข้อมูลจำนวนมากและสามารถจำแนกและพยากรณ์โรคได้แม่นยำกว่า โดยเฉพาะอย่างยิ่งการใช้เทคนิค Machine Learning กับภาพถ่ายทางการแพทย์ เช่น ภาพเอกซเรย์คอมพิวเตอร์ สามารถตรวจจับและวิเคราะห์ลักษณะเฉพาะของเนื้องอกที่มนุษย์ไม่สามารถมองเห็นได้ด้วยตาเปล่า การวิเคราะห์ภาพเหล่านี้ช่วยในการพยากรณ์โรคมะเร็งปอดขั้นสูงได้อย่างมีประสิทธิภาพและรวดเร็ว

การพยากรณ์โรคมะเร็งปอดขั้นสูงหมายถึงการใช้เทคโนโลยี Machine Learning เพื่อทำนายการพัฒนาของโรคและความเสี่ยงในการลุกลามของมะเร็ง ซึ่งสามารถทำได้โดยการวิเคราะห์ภาพเอกซเรย์คอมพิวเตอร์และข้อมูลทางการแพทย์อื่น ๆ นอกจากนี้ การรวมข้อมูลจากภาพ CT ก่อนการผ่าตัดและการประเมินพยาธิสภาพหลังการผ่าตัดผ่านการเรียนรู้เชิงลึกยังช่วยพยากรณ์ความเสี่ยงของการลุกลามของมะเร็งในผู้ป่วยมะเร็งปอดชนิดไม่ใช้เซลล์เล็ก (NSCLC) ระยะแรกที่ยังไม่แพร่กระจายไปยังต่อมน้ำเหลืองได้ดียิ่งขึ้น (Huang et al., 2022)

การวิจัยนี้มีจุดมุ่งหมายเพื่อพัฒนาตัวแบบการเรียนรู้ของเครื่องที่สามารถจำแนกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและระยะลุกลาม เพื่อลดภาระของการแพทย์ ลดความเสี่ยงของการวินิจฉัยที่ไม่ตรงกับระยะของมะเร็ง และเพิ่มโอกาสในการรักษาโรคมะเร็งปอดให้ได้ผลดีที่สุด การพัฒนาและประเมินประสิทธิภาพของตัวแบบการเรียนรู้ของเครื่องในการแบ่งชั้นความเสี่ยงและการพยากรณ์โรคมะเร็งปอดขั้นสูงโดยใช้ภาพเอกซเรย์คอมพิวเตอร์ ผ่านการวิเคราะห์ลักษณะเฉพาะของเนื้องอกและระยะของโรค ความน่าจะเป็นในการใช้เทคนิคการเรียนรู้เชิงลึก โครงข่ายประสาทเทียม และเพื่อนบ้านที่ใกล้ที่สุด นำไปสู่การพัฒนาแบบจำลองที่มีความแม่นยำสูงในการจำแนกและพยากรณ์โรค

ความสำเร็จในการพัฒนาเทคโนโลยีเหล่านี้ไม่เพียงแต่ช่วยเพิ่มความแม่นยำในการวินิจฉัยเท่านั้น แต่ยังช่วยให้แพทย์สามารถวางแผนการรักษาที่เหมาะสมและทันต่วงทีสำหรับผู้ป่วยได้ เนื่องจากการวินิจฉัยที่รวดเร็วและแม่นยำสามารถนำไปสู่การตัดสินใจเลือกวิธีการรักษาที่เหมาะสม ลดเวลาในการรักษา และเพิ่มโอกาสในการรอดชีวิตของผู้ป่วย

วิธีการทดลอง

คณะผู้วิจัยได้นำเทคนิคการทำเหมืองข้อมูลภาพ (Image Mining) โดยการใช้โปรแกรม RapidMiner Studio Version 10 และติดตั้งส่วนเสริม IMMI (Image Mining Extension for RapidMiner) ซึ่งจะมีกระบวนการ การแยกคุณลักษณะ สำหรับการทำเหมืองรูปภาพ (อนุพงศ์ สุขประเสริฐ, 2022) มาใช้เพื่อสร้างตัวแบบจำลองสำหรับการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้น และมะเร็งปอดระยะลุกลามโดยมีการใช้กระบวนการมาตรฐานการทำเหมืองข้อมูล (Cross-industry standard process for data mining : CRISP-DM) ทั้งหมด 6 ขั้นตอน

1.การทำความเข้าใจเกี่ยวกับปัญหา (Business Understanding)

ปัจจุบันโรคมะเร็งปอดเป็นหนึ่งในสาเหตุหลักของการเสียชีวิตทั่วโลก (พญ.ปาริชาติ นิยมทอง, 2018) โดยเฉพาะในกรณีที่พบโรคในระยะขั้นสูง ความล่าช้าหรือความไม่แม่นยำในการวินิจฉัยสามารถทำให้ผู้ป่วยพลาดโอกาสในการรักษาได้อย่างมาก ดังนั้น การพัฒนาตัวแบบการเรียนรู้ของเครื่องที่สามารถแบ่งชั้นความเสี่ยงและพยากรณ์การเกิดโรคได้อย่างแม่นยำ จึงกลายเป็นความต้องการที่เร่งด่วน การศึกษาในปี 2023 โดย Babu และคณะ พบว่าการใช้ภาพเอกซเรย์คอมพิวเตอร์ (CT) สามารถเพิ่มความแม่นยำในการตรวจพบเนื้องอกถึง 90% อย่างไรก็ตาม ความแม่นยำนี้ยังคงพึ่งพาประสบการณ์และความเชี่ยวชาญของแพทย์ในการแปลผลภาพ การนำเทคโนโลยีการเรียนรู้ของเครื่องมาช่วยกระบวนการวินิจฉัยจึงเป็นทางออกที่มีประสิทธิภาพ

การพัฒนาตัวแบบการเรียนรู้ของเครื่องที่สามารถวินิจฉัยโรคมะเร็งปอดได้อย่างรวดเร็วและแม่นยำ ไม่เพียงแต่ช่วยลดภาระของแพทย์ในการแปลผลภาพ แต่ยังช่วยลดเวลาที่ใช้ในการวินิจฉัย และเพิ่มโอกาสในการรักษาที่ประสบความสำเร็จให้กับผู้ป่วย

2.การทำความเข้าใจเกี่ยวกับข้อมูล (Data Understanding)

ข้อมูลที่ผู้วิจัยได้นำมาทำการวิจัยในครั้งนี้ได้นำมาจากเว็บไซต์ Kaggle.com โดยเป็นชุดข้อมูลภาพเอกซเรย์คอมพิวเตอร์ มะเร็งปอดซึ่งถูกจัดแบ่งเป็น 2 กลุ่ม ได้แก่ กลุ่มระยะเริ่มต้น มีจำนวนภาพอยู่ที่ 536 รูป และกลุ่มระยะลุกลาม มีจำนวน 561 รูป ซึ่งเป็นข้อมูลของผู้ป่วยมะเร็งปอด

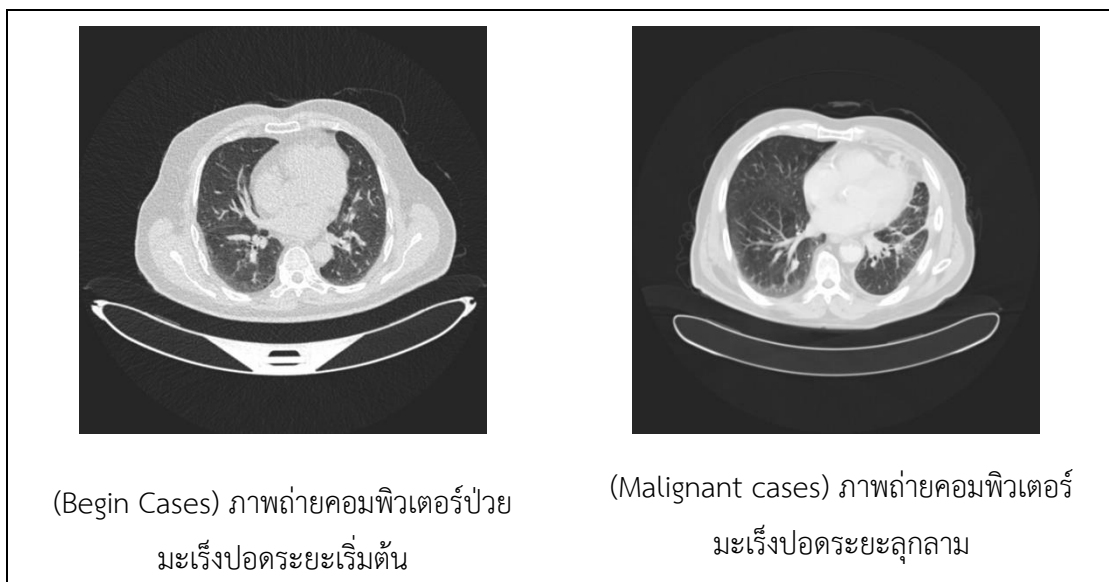
ที่เกิดขึ้นจริง โดยกลุ่มข้อมูลทั้งสองกลุ่มเป็นข้อมูลภาพเอกซเรย์คอมพิวเตอร์ โดยทั้งสองกลุ่มข้อมูลมีทั้งหมด 3 ขนาด 512x512 พิกเซล 623x512 พิกเซล และ 801x512 พิกเซล ทุกภาพที่เป็นภาพที่ถูกจัดเป็นไฟล์ JPEG หรือนามสกุล .jpg เพื่อนำไปใช้ในการวิเคราะห์ข้อมูลเพื่อหาแบบจำลองที่เหมาะสมต่อไป

3.การเตรียมข้อมูล (Data preparation)

ผู้วิจัยได้ทำการจัดเตรียมข้อมูลสำหรับการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและผู้ป่วยมะเร็งปอดระยะลุกลาม โดยการทำเหมืองข้อมูลภาพด้วยการใช้โปรแกรม RapidMiner Studio เพื่อทำการเปรียบเทียบประสิทธิภาพของแบบจำลองต่างๆ การวิจัยนี้เน้นการเลือกแบบจำลองที่เหมาะสมที่สุดสำหรับการพยากรณ์โรคระยะเริ่มต้นสูง ผู้วิจัยได้ทำการแบ่งชุดข้อมูลออกเป็น 2 โฟลเดอร์ย่อย ประกอบด้วย โฟลเดอร์ “Begin Cases” เป็นโฟลเดอร์ที่จัดเก็บภาพระยะเริ่มต้น จำนวน 400 ภาพ และโฟลเดอร์ “Malignant cases” ที่เป็นโฟลเดอร์จัดเก็บภาพระยะลุกลาม จำนวน 400 ภาพ ดังแสดงในภาพที่ 1 โดยภาพในแต่ละโฟลเดอร์มีขนาดเท่ากันทั้ง 3 ขนาด ได้แก่ 512x512 พิกเซล 623x512 พิกเซล และ 801x512 พิกเซล เนื่องจากข้อมูลในกลุ่มผู้ป่วยมะเร็งปอดระยะเริ่มต้นและระยะลุกลามมีจำนวนไม่เท่ากัน (Unbalanced data) ผู้วิจัยจึงใช้วิธีการสุ่ม (Random sampling) เพื่อให้ได้กลุ่มข้อมูลที่สมดุล (Balanced) โดยสุ่มเลือกภาพจากแต่ละกลุ่มให้มีจำนวนเท่ากันคือกลุ่มละ 400 รูป วิธีการนี้ช่วยลดความอคติที่อาจเกิดขึ้นจากการมีข้อมูลที่ไม่สมดุลและทำให้แบบจำลองมีความแม่นยำในการทำนายสูงขึ้น การเตรียมข้อมูลในลักษณะนี้ช่วยให้การวิจัยสามารถเปรียบเทียบประสิทธิภาพของแบบจำลองได้อย่างมีประสิทธิภาพและเชื่อถือได้ นอกจากนี้ การรักษาขนาดของภาพที่ใช้สำหรับการวิเคราะห์เป็นปัจจัยสำคัญในการรักษาความสม่ำเสมอและความถูกต้องของข้อมูลในการวิจัย

หลังจากการเตรียมรูปภาพเสร็จผู้วิจัยจะทำการแปลงรูปภาพให้อยู่ในรูปแบบของข้อมูลโครงสร้าง (Structured Data) ด้วยวิธีการแยกคุณลักษณะทางธรรมชาติ (Natural Feature Extraction) ในข้อมูลภาพมีบทบาทสำคัญในหลายแอปพลิเคชัน เช่น การจัดประเภทภาพ (Image Classification), การรู้จำวัตถุ (Object Recognition), และการแบ่งส่วนภาพ (Image Segmentation) โดยการใช้โมเดลการเรียนรู้ของเครื่อง เช่น เครือข่ายประสาทเทียมเชิงคอนโวลูชัน (Convolutional Neural Networks - CNN) (Chang, 2023) และอัลกอริธึมการเรียนรู้เชิงลึก (Deep Learning Algorithms) (Hariharan et al., 2022) นักวิจัยสามารถแยกคุณลักษณะที่สำคัญจากภาพได้ เช่น รูปร่าง (Shape), สี (Color), เนื้อสัมผัส (Texture), และลักษณะเชิงพื้นที่ (Spatial Characteristics) คุณลักษณะเหล่านี้มีความสำคัญต่อการตรวจจับทางกายภาพ (Physical

Detection), การจำแนกวัตถุ (Classification of Objectives), และงานด้านล่างอื่น ๆ (Downstream Tasks) นอกจากนี้ เทคนิคต่าง ๆ เช่น การถดถอยเชิงเส้น (Linear Regression), การจัดกลุ่มแบบ K-means (K-means Clustering), และเครือข่ายประสาทเทียม (Neural Networks) ถูกใช้สำหรับการรู้จำรูปแบบในกระบวนการวิเคราะห์ข้อมูลไฮเปอร์สเปกตรัม (Hyperspectral Data Analysis) ซึ่งช่วยในการแยกคุณลักษณะที่มีคุณค่าสำหรับการจัดประเภทและการวิเคราะห์ (Kumar et al., 2023) ยิ่งไปกว่านั้น การใช้กระบวนการทัศน์การเรียนรู้เชิงลึกและการประมวลผลภาพ เช่น OpenCV ช่วยให้สามารถรู้จำกิจกรรมของมนุษย์จากวิดีโอความละเอียดต่ำได้ โดยเน้นถึงความสำคัญของการแยกคุณลักษณะในการปกป้องความเป็นส่วนตัวและการประยุกต์ใช้ในด้านความปลอดภัย (Hassan & Sabha, 2023)



ภาพที่ 1 ตัวอย่างการจำแนกประเภทข้อมูลภาพถ่ายคอมพิวเตอร์
มะเร็งปอดระยะเริ่มต้นและระยะลุกลาม

เพื่อการทำเหมืองข้อมูลภาพในระดับสากล (Global Image Mining) ที่สามารถแยกคุณลักษณะเฉพาะของแต่ละภาพได้ งานวิจัยนี้มุ่งเน้นการคัดแยกสมบัติของภาพเอกซเรย์คอมพิวเตอร์สำหรับมะเร็งปอด โดยการวิเคราะห์คุณลักษณะจากภาพโหมดขาวดำ (Grayscale) ผ่าน 5 คุณลักษณะหลัก ได้แก่ ค่ามัธยฐาน (Median Global Statistics), ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation), ค่าเฉลี่ยทางสถิติระดับสากล (Mean Global Statistics), ค่าสีเทาต่ำสุด (Minimum Gray Value), และค่าสีเทาสูงสุด (Maximum Gray Value) ซึ่งแสดงในตารางที่ 1

ค่าสีเทาสูงสุด (Maximum Gray Value) แสดงถึงความเข้มข้นสูงสุดของพิกเซลในภาพ ซึ่งอาจบ่งชี้ถึงความผิดปกติหรือเนื้อเยื่อที่มีการสะสมของเซลล์มะเร็งสูงสุด การใช้ค่านี้ช่วยให้โมเดลสามารถตรวจจับพื้นที่ที่มีความเข้มข้นผิดปกติสูงได้ ขณะที่ค่าเฉลี่ยทางสถิติระดับสากล (Mean

Global Statistics) แสดงถึงค่าเฉลี่ยของความเข้มข้นของพิกเซลในภาพทั้งภาพ ซึ่งเป็นตัวแทนของภาพรวมของเนื้อเยื่อทั้งหมดในภาพ ค่านี้ช่วยให้โมเดลมองเห็นแนวโน้มทั่วไปของความเข้มข้นในภาพ ซึ่งเป็นข้อมูลพื้นฐานสำคัญในการพยากรณ์

ค่ามัธยฐานทางสถิติระดับสากล (Median Global Statistics) แสดงถึงค่ากลางของความเข้มข้นของพิกเซล ซึ่งลดผลกระทบจากค่าผิดปกติ (outliers) ในภาพ ค่านี้ช่วยให้โมเดลพิจารณาความหนาแน่นของเนื้อเยื่อมะเร็งได้แม่นยำขึ้นเมื่อมีค่าผิดปกติ ในขณะที่ค่าสีเทาต่ำสุด (Minimum Gray Value) แสดงถึงความเข้มข้นต่ำสุดของพิกเซลในภาพ บ่งชี้ถึงบริเวณที่มีการทำลายของเนื้อเยื่อหรือพื้นที่ที่มีกิจกรรมผิดปกติบ่อย ค่านี้ช่วยให้โมเดลสามารถตรวจจับพื้นที่ที่มีความผิดปกติหรือเนื้อเยื่อที่มีการทำลายน้อย

ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) แสดงถึงการกระจายตัวของค่าความเข้มข้นในภาพ บ่งบอกถึงความแปรปรวนของเนื้อเยื่อภายในภาพ ค่านี้ช่วยให้โมเดลพิจารณาความผันผวนของความเข้มข้นในภาพ ซึ่งอาจบ่งบอกถึงความไม่สม่ำเสมอของเนื้อเยื่อมะเร็ง การรวมกันของทั้ง 5 คุณลักษณะนี้มีความสำคัญในการสร้างโมเดลที่มีความแม่นยำและสามารถแยกแยะความแตกต่างของเนื้อเยื่อมะเร็งปอดในภาพเอกซเรย์คอมพิวเตอร์ได้อย่างมีประสิทธิภาพ

ตารางที่ 1 หลังจากเปลี่ยนรูปภาพให้เป็นข้อมูลโครงสร้าง (Structured Data)

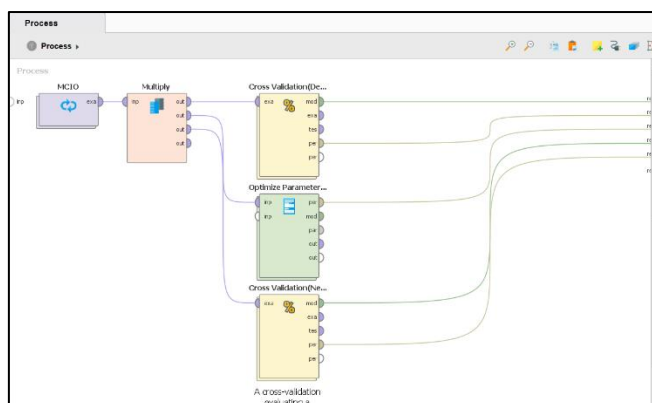
Row NO.	Label	Max Gray Value	Mean Global Statistics	Median Global Statistics	Minimum Gray Value	Standard Deviation
1	normal	255	99.647	50	37	69.840
2	normal	255	87.922	48	37	67.742
3	normal	255	89.270	57	37	71.859
4	malignant	255	146.957	196	0	83.973
5	malignant	255	143.511	194	0	83.738
6	malignant	255	135.405	95	0	79.057

4. การสร้างแบบจำลอง (Modeling)

ในการเตรียมข้อมูลสำหรับจำแนกภาพเอกซเรย์คอมพิวเตอร์ เราได้เลือกใช้เทคนิคการเรียนรู้เชิงลึก (DL), โครงข่ายประสาทเทียม (ANN) และ k-ใกล้เคียงที่สุด (k-NN) ข้อมูลชุดนี้ได้มาจากเว็บไซต์ kaggle.com ซึ่งเป็นรูปภาพเอกซเรย์คอมพิวเตอร์ทั้งหมด 800 รูป แบ่งออกเป็นสองกลุ่ม

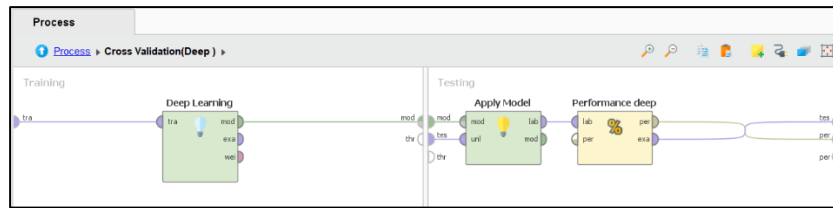
กลุ่มละ 400 รูป โดยรูปภาพทั้งหมดจัดเก็บในรูปแบบไฟล์ JPEG (.jpg) โดยการจัดเก็บรูปภาพนี้ได้ถูกแบ่งเป็นสองโฟลเดอร์ โฟลเดอร์แรกชื่อว่า "Begin cases" ใช้สำหรับจัดเก็บภาพเอกซเรย์คอมพิวเตอร์ของผู้ป่วยมะเร็งปอดระยะเริ่มต้น จำนวน 400 รูป และโฟลเดอร์ที่สองชื่อว่า "Malignant cases" ใช้สำหรับจัดเก็บรูปภาพมะเร็งปอดระยะลุกลาม จำนวน 400 รูป ข้อมูลเหล่านี้จะถูกนำมาใช้ในการสร้างแบบจำลองเพื่อการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและมะเร็งปอดระยะลุกลาม โดยใช้เทคนิคทั้งสามที่ได้กล่าวมา ซึ่งจะช่วยให้เพิ่มความแม่นยำและประสิทธิภาพในการวินิจฉัยโรคนี้

ภาพที่ 2 แสดงการใช้โปรแกรม RapidMiner Studio ในการเปรียบเทียบประสิทธิภาพของแบบจำลองต่าง ๆ สำหรับการทำนายข้อมูล โดยเริ่มจากการนำเข้าชุดข้อมูลภาพเอกซเรย์คอมพิวเตอร์ จากนั้นทำซ้ำข้อมูลเพื่อสร้างแบบจำลองหลาย ๆ แบบพร้อมกัน ต่อมาใช้เทคนิคการตรวจสอบความถูกต้องข้ามกลุ่ม (Cross Validation) เพื่อทดสอบและปรับค่าพารามิเตอร์ของแบบจำลองเพื่อหาโครงสร้างที่เหมาะสมที่สุด ผลลัพธ์ที่ได้จะนำมาเปรียบเทียบประสิทธิภาพของแบบจำลองต่าง ๆ



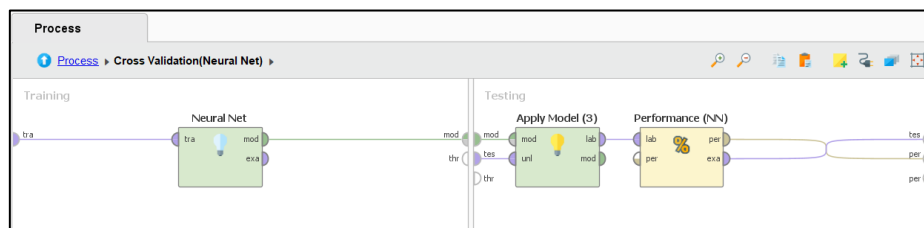
ภาพที่ 2 แบบจำลองต่างๆที่ใช้สำหรับการเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยโปรแกรม RapidMiner studio

เทคนิค DL เป็นเทคนิคหนึ่งของการประมวลผลข้อมูลที่ใช้โครงข่ายประสาทเทียม (Neural Networks) ที่มีโครงสร้างซับซ้อนเพื่อให้ระบบคอมพิวเตอร์สามารถเรียนรู้และทำนายผลลัพธ์ได้โดยไม่ต้องระบุชั้นคำตอบหรือลักษณะของข้อมูลล่วงหน้า ความลึก (Deep) ใน DL หมายถึงจำนวนชั้น (Layer) ในโครงข่ายประสาทเทียมที่มีมากและซับซ้อน โครงข่ายประสาทเทียมจะต้องประมวลผลข้อมูลผ่านหลายชั้นเพื่อให้ระบบสามารถค้นพบลักษณะเฉพาะของกลุ่มข้อมูลแต่ละกลุ่มได้ ในการวิจัยนี้ได้นำเทคนิค DL มาใช้ในการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและมะเร็งปอดระยะลุกลาม โดยกำหนดค่า Hidden layer sizes เท่ากับ 320 และกำหนด epochs เท่ากับ 20,000.0 ซึ่งสอดคล้องกับการวิจัยของ Kumjit และคณะ (2022) ที่ได้นำเทคนิค DL ไปใช้ในการพยากรณ์การเกิดโรคหลอดเลือดในสมอง การสร้างแบบจำลอง DL ด้วยโปรแกรม RapidMiner Studio แสดงดังภาพที่ 3



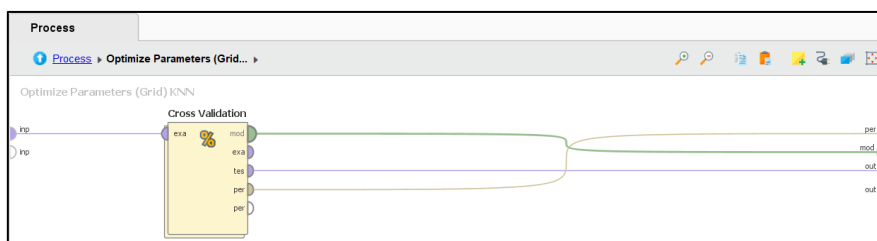
ภาพที่ 3 แบบจำลองการเรียนรู้เชิงลึก (DL)

โครงข่ายประสาทเทียม (Artificial Neural Networks: ANNs) เป็นแบบจำลองทางคณิตศาสตร์ที่ได้แรงบันดาลใจจากโครงสร้างประสาทของสมองมนุษย์ ประกอบด้วยโหนดที่เชื่อมต่อกันเพื่อจำลองเซลล์ประสาทและประมวลผลข้อมูล ANNs เช่นเดียวกับสมองมนุษย์ เรียนรู้ผ่านตัวอย่างและสามารถจัดการกับงานที่ซับซ้อน ทำให้เหมาะสมสำหรับการใช้งานต่างๆ รวมถึงการทำนายการเกิดมะเร็งปอด (Dimitrov, 2023) การศึกษาโดย Eduarda Vieira และคณะ (2021) อาจใช้เทคนิคโครงข่ายประสาทเทียมในการทำนายการเกิดมะเร็งปอด แสดงให้เห็นถึงความหลากหลายและประสิทธิภาพของ ANNs ในการวิจัยทางการแพทย์และการสร้างแบบจำลองเชิงพยากรณ์ การสร้างแบบจำลองโครงข่ายประสาทเทียมด้วยโปรแกรม RapidMiner Studio แสดงในภาพที่ 4



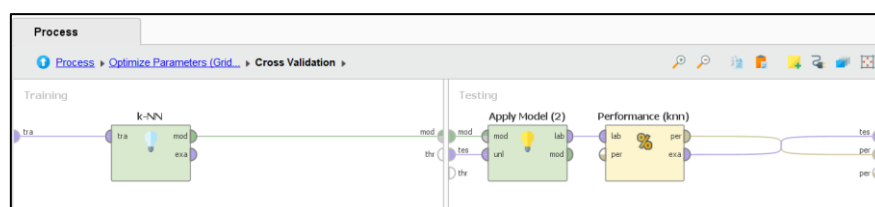
ภาพที่ 4 แบบจำลอง โครงข่ายประสาทเทียม (NN)

เทคนิค k-NN (k-Nearest Neighbors) เป็นแบบจำลองหรืออัลกอริทึมที่ใช้ในการจำแนกข้อมูลหรือทำนายค่า โดยพิจารณาข้อมูลที่ใกล้เคียงกับข้อมูลที่ต้องการจำแนกหรือทำนายผล โดยใช้หลักการความคล้ายคลึงกันระหว่างข้อมูล เช่นเดียวกับงานวิจัยของ Maleki, Zeinali, และ Niaki (2021) ที่ใช้เทคนิคเพื่อนบ้านที่ใกล้ที่สุดในการสร้างแบบจำลองสำหรับการทำนายมะเร็งปอดระยะเริ่มต้น การทำงานของเทคนิค k-NN คือ เมื่อมีข้อมูลหรือชุดข้อมูลที่ต้องการจำแนกหรือต้องการทำนายค่า ระบบจะหาข้อมูลที่ใกล้เคียงกันโดยคำนวณระยะห่าง (distance) ระหว่างข้อมูลหรือชุดข้อมูลแต่ละชุด แล้วเลือกข้อมูลหรือชุดข้อมูลที่ใกล้เคียงที่สุดจำนวน K ชุด หลังจากนั้นจะใช้การโหวตของค่าประเภทข้อมูลของ K ข้อมูลนั้น ๆ เพื่อกำหนดประเภทหรือค่าที่ต้องการจำแนกหรือทำนาย (Progga et al., 2023)



ภาพที่ 5 กระบวนการปรับพารามิเตอร์ของแบบจำลอง k-Nearest Neighbors (k-NN)

เทคนิคเพื่อนบ้านที่ใกล้ที่สุดเหมาะสำหรับการพยากรณ์โรค โดยเฉพาะในกรณีของข้อมูลเกี่ยวกับผู้ป่วยที่มีโรคและไม่มีโรค ในการวิจัยครั้งนี้ ผู้วิจัยได้นำแบบจำลองเพื่อนบ้านที่ใกล้ที่สุดมาใช้ในการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและระยะลุกลาม เนื่องจากเทคนิคนี้สามารถช่วยในการวินิจฉัยและคัดแยกผู้ป่วยได้อย่างแม่นยำ ภาพที่ 5 แสดงกระบวนการปรับพารามิเตอร์ของแบบจำลอง k-Nearest Neighbors (k-NN) โดยใช้เทคนิค Cross Validation ในโปรแกรม RapidMiner Studio ข้อมูลถูกนำเข้า (input) และแยกเป็นชุดฝึกอบรมและชุดทดสอบเพื่อประเมินประสิทธิภาพของแบบจำลอง กระบวนการนี้ใช้การปรับพารามิเตอร์ (optimize parameters) โดยการทดลองหลายครั้งเพื่อหาค่าที่เหมาะสมที่สุด Cross Validation ช่วยให้การประเมินผลมีความเที่ยงตรงและแม่นยำมากขึ้น เนื่องจากการใช้ข้อมูลทั้งหมดในการฝึกอบรมและทดสอบแบบจำลอง



ภาพที่ 6 กระบวนการสร้างแบบจำลอง k-Nearest Neighbors (k-NN)

การปรับพารามิเตอร์ในขั้นตอนนี้มีความสำคัญอย่างยิ่งในการเพิ่มประสิทธิภาพและความแม่นยำของแบบจำลอง k-NN ในการคาดการณ์หรือแบ่งชั้นข้อมูลที่ซับซ้อน เช่น การวินิจฉัยโรคจากภาพเอกซเรย์คอมพิวเตอร์ ดังนั้น การใช้เทคนิค Cross Validation และการปรับพารามิเตอร์จึงเป็นขั้นตอนสำคัญที่ช่วยเพิ่มความน่าเชื่อถือและประสิทธิภาพของผลลัพธ์ในการวิจัยนี้ ภาพที่ 6 แสดงกระบวนการสร้างแบบจำลอง k-Nearest Neighbors (k-NN) โดยใช้โปรแกรม RapidMiner Studio ขั้นตอนประกอบด้วย การฝึกอบรม (Training) และการทดสอบ (Testing) ในส่วนของการฝึกอบรม ข้อมูลถูกนำเข้า (tra) เพื่อสร้างแบบจำลอง k-NN (mod) และประเมินผล (eva) จากนั้นในส่วนของการทดสอบ แบบจำลองที่สร้างขึ้นถูกนำไปใช้ (Apply Model) กับข้อมูลทดสอบ (tes) เพื่อประเมินประสิทธิภาพของแบบจำลอง (Performance) โดยมีการวัดผลลัพธ์และเปรียบเทียบกับข้อมูลจริง

กระบวนการนี้ช่วยให้ได้แบบจำลองที่มีความแม่นยำและประสิทธิภาพสูงในการพยากรณ์หรือแบ่งชั้นข้อมูลตามที่ต้องการ

5. การประเมินผล

การประเมินผลของงานวิจัยนี้ กลุ่มผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลอง โดยที่ทุกแบบจำลองจะมีพารามิเตอร์การตรวจสอบแบบข้ามกลุ่ม (Cross Validation) เพื่อที่จะนำค่าความแม่นยำ (Accuracy) ค่าประสิทธิภาพรวม (F-Measure) ค่าความไว (Sensitivity) และความจำเพาะ (Specificity) ไปใช้สำหรับการเปรียบเทียบประสิทธิภาพของแบบจำลอง เพื่อที่จะหาแบบจำลองที่เหมาะสมที่สุดสำหรับการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและผู้ป่วยมะเร็งปอดระยะลุกลาม โดยที่การเปรียบเทียบประสิทธิภาพของแบบจำลองจะใช้หลักการ เมทริกซ์ความสับสน (Confusion Matrix) คือหลักการที่จะทำการคัดแยกข้อมูลที่เกิดขึ้นจากการพยากรณ์หรือการทำนายด้วยระบบการคัดแยก 2 ระดับ ดังตารางที่ 2

ตารางที่ 2 เมทริกซ์ความสับสน (Confusion Matrix)

	Predicted Positive (yes)	Predicted Negatives (no)
Actual Positive	True Positive TP	False Negatives FN
Actual Negatives (no)	False Positive FP	True Negatives TN

เมทริกซ์ความสับสนจะเป็นการใช้เทคนิควิเคราะห์ ทำนาย พยากรณ์ โดยการเปรียบเทียบค่าความเป็นจริงกับค่าที่ไม่เกิดขึ้นจริง

โดย True Positive (TP) คือชั้นของเป้าหมายที่เกิดขึ้นจริง และการพยากรณ์แล้วเกิดขึ้นจริง

False Negatives (FN) คือชั้นของเป้าหมายที่เกิดขึ้นจริง และการพยากรณ์แล้วไม่เกิดขึ้น

True Negatives (TN) คือชั้นของเป้าหมายที่ไม่เกิดขึ้น และการพยากรณ์แล้วไม่เกิดขึ้น

False Positive (FP) คือชั้นของเป้าหมายที่ไม่เกิดขึ้น และการพยากรณ์แล้วเกิดขึ้นจริง

ค่าความแม่นยำ (Accuracy) เป็นค่าที่ได้มาจากการทดสอบหาค่าที่ถูกต้องของการพยากรณ์ของข้อมูลโดยคำนวณเป็นร้อยละ (%) ดังสมการต่อไปนี้

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

สมการที่ (1) แสดงค่าความแม่นยำ (Accuracy) จะเป็นค่าสำหรับการวัดพื้นฐานของการวัดประสิทธิภาพของแบบจำลองต่างๆ โดยวัดที่ความถูกต้องของการจำแนกประเภทของชุดข้อมูลต่อข้อมูลทั้งหมด

ค่าประสิทธิภาพรวม (F-Measure) เป็นค่าที่ได้จากการนำค่า recall และ Precision มาทำคำนวณแล้วจะได้ค่า ค่าประสิทธิภาพรวม (F-Measure) ของแต่ละชั้นเป้าหมาย ดังสมการต่อไปนี้

$$F \text{ measure} = \text{ชั้นเป้าหมาย } yes = \frac{(2 * precision(yes) + recall(yes))}{(precision(yes) + recall(yes))} \quad (2)$$

$$F \text{ measure} = \text{ชั้นเป้าหมาย } no = \frac{(2 * precision(no) + recall(no))}{(precision(no) + recall(no))} \quad (3)$$

สมการที่ (2) และ (3) แสดง ค่าประสิทธิภาพรวม (F-Measure) จะเป็นการวัดค่าความแม่นยำจากการจำแนกประเภทของข้อมูล (classification) โดยทั่วไปจะเกี่ยวกับการประมวลผลทางสถิติ โดยที่จำคำนวณจาก Precision และ Recall โดยใช้ทั้งสองสูตรด้านบนซึ่งสามารถแสดงผลลัพธ์ของประสิทธิภาพได้อย่างครอบคลุม

สมการที่ (4) แสดงค่าความไว (Sensitivity) เป็นค่าที่ได้มาจากความสามารถของแบบจำลองที่ทำการพยากรณ์หรือทำนายผลข้อมูลของการคัดแยกผู้ป่วยมะเร็งระยะลุกลามได้อย่างถูกต้องว่ารูปภาพของผู้ป่วยแต่ละคนอยู่ในกลุ่มผู้ป่วยมะเร็งระยะลุกลาม

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (4)$$

สมการที่ (5) แสดงค่าความจำเพาะ (Specificity) เป็นค่าที่ได้มาจากความสามารถของแบบจำลองที่ทำการพยากรณ์หรือทำนายผลข้อมูลของการคัดแยกผู้ป่วยมะเร็งระยะเริ่มต้นได้อย่างถูกต้อง

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (5)$$

โดยการนำ ค่าความไว และ ค่าความจำเพาะ มาใช้ในงานวิจัยในครั้งนี้ด้วยเพราะว่า ค่าความไวกับ ค่าความจำเพาะ สามารถนำมาวินิจฉัยเกี่ยวกับโรคทางการแพทย์ได้ เช่นเดียวกับ วนิดา ผาตาล.(2021) ที่ใช้ค่าความไว และ ค่าความจำเพาะ สำหรับการตรวจทางสัลยพยาธิวิทยาในมะเร็งต่อมลูกหมากในโรงพยาบาลอุดรธานี

6. การปรับใช้ (Deployment)

เมื่อผู้วิจัยได้ดำเนินการครบ 5 กระบวนการแล้ว ผู้วิจัยจึงได้ผลลัพธ์ที่เหมาะสมที่สุดสำหรับการสร้างแบบจำลองการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและผู้ป่วยมะเร็งปอดระยะลุกลาม เพื่อที่จะสามารถช่วยลดขั้นตอนทางการแพทย์ทำให้สามารถแบ่งเบาภาระของทางการแพทย์และทรัพยากรค่าใช้จ่ายในการรักษาลดลงได้ ลดความเสี่ยงให้น้อยลง เพิ่มความแม่นยำให้ตรงจุดการรักษามากยิ่งขึ้น

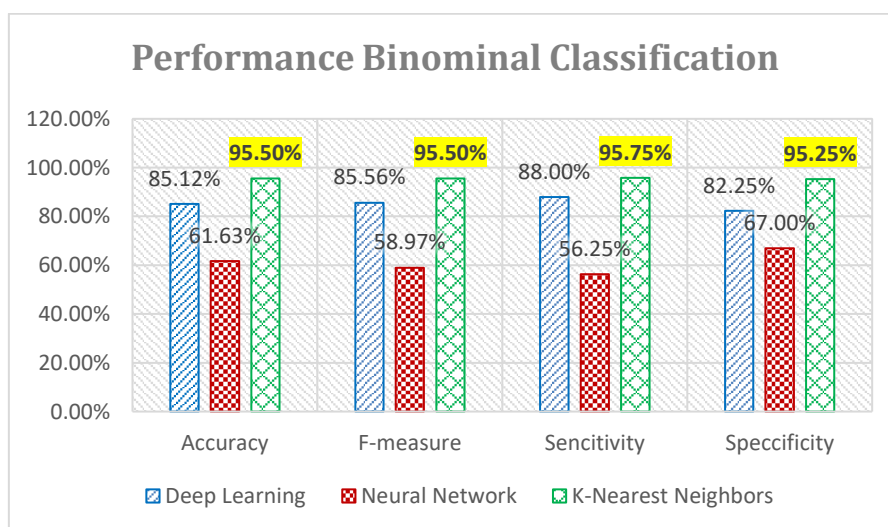
ผลการทดลอง

หลังจากที่ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพด้วยการทำเหมืองข้อมูลภาพโดยการนำภาพเอกซเรย์คอมพิวเตอร์ ของผู้ป่วยมะเร็งปอดระยะเริ่มต้นและผู้ป่วยมะเร็งปอดระยะลุกลาม มาจากเว็บไซต์ www.kaggle.com ผู้วิจัยจึงนำชุดข้อมูลภาพมาทำการเปรียบเทียบประสิทธิภาพอย่างละ 400 ภาพ รวมเป็น 800 ภาพ โดยเป็นภาพมะเร็งปอดระยะเริ่มต้น 400 รูป และภาพมะเร็งปอดระยะลุกลาม 400 มาใช้สำหรับการทำเหมืองข้อมูลภาพโดยใช้กระบวนการมาตรฐานการทำเหมืองข้อมูล (CRISP-DM) สำหรับการสร้างแบบจำลองเพื่อการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและผู้ป่วยมะเร็งปอดระยะลุกลาม ด้วยการนำพารามิเตอร์การตรวจสอบแบบข้ามกลุ่ม cross validation มาใช้ในการวิจัยครั้งนี้จะสามารถแบ่งข้อมูลออกเป็น 10 ส่วนได้ (10-fold cross validation) มาใช้กับเทคนิคการทำเหมืองข้อมูลภาพทั้ง 3 เทคนิคต่อไปนี้ เทคนิค DL ANN และ k-NN จากนั้นผู้วิจัยจะทำการคำนวณเพื่อหาค่ามัธยฐาน ค่าเบี่ยงเบนมาตรฐาน และค่าเฉลี่ยของแบบจำลองทั้ง 3 เพื่อทำการเปรียบเทียบประสิทธิภาพของการจำแนกข้อมูลภาพเอกซเรย์คอมพิวเตอร์ เพื่อทำการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและผู้ป่วยมะเร็งปอดระยะลุกลามด้วยการนำค่าเฉลี่ยอย่าง ค่าความแม่นยำ (Accuracy) ค่าประสิทธิภาพรวม (F-measure) ค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity) ด้วยการใช้โปรแกรมสำหรับการทำเหมืองข้อมูล RapidMiner Studio version 10 มาใช้สำหรับการวิเคราะห์ข้อมูลรูปภาพและทำการเปรียบเทียบประสิทธิภาพของแบบจำลอง ทั้ง 3 เทคนิค

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพของแบบจำลอง ด้วยภาพเอกซเรย์คอมพิวเตอร์ เพื่อคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและมะเร็งปอดระยะลุกลาม

Classification Techniques	Performance Binominal Classification			
	Accuracy	F-measure	Sensitivity	Specificity
<i>k</i> -NN*	95.50%	95.50%	95.75%	95.25%
DL	85.12%	85.56%	88.00%	82.25%
ANN	61.63%	58.97%	56.25%	67.00%

จากผลการทดลองที่แสดงในตารางที่ 3 พบว่าเทคนิค *k*-NN ให้ผลลัพธ์ที่โดดเด่นมากกว่าเทคนิคอื่นๆ โดยได้ค่าความแม่นยำ (Accuracy) อยู่ที่ 95.50% ค่าประสิทธิภาพรวม (F-measure) อยู่ที่ 95.50% ค่าความไว (Sensitivity) อยู่ที่ 95.75% และค่าความจำเพาะ (Specificity) อยู่ที่ 95.25% ทำให้เป็นแบบจำลองที่เหมาะสมสำหรับการคัดแยกผู้ป่วยมะเร็งปอดระยะเริ่มต้นและผู้ป่วยมะเร็งปอดระยะลุกลาม สำหรับแบบจำลองจากเทคนิค DL ได้ค่าความแม่นยำอยู่ที่ 85.12% ค่าประสิทธิภาพรวมอยู่ที่ 85.56% ค่าความไวอยู่ที่ 88.00% และค่าความจำเพาะอยู่ที่ 82.25% ซึ่งมีผลลัพธ์ที่ดีเช่นกันแต่ยังต่ำกว่าเทคนิค *K*-NN และแบบจำลองที่ใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network) มีผลลัพธ์ที่ต่ำที่สุด โดยได้ค่าความแม่นยำอยู่ที่ 61.63% ค่าประสิทธิภาพรวมอยู่ที่ 58.97% ค่าความไวอยู่ที่ 56.25% และค่าความจำเพาะอยู่ที่ 67.00% ซึ่งแสดงให้เห็นถึงความต้องการการปรับปรุงเพิ่มเติมเพื่อให้แบบจำลองมีประสิทธิภาพสูงขึ้นดังแสดงในภาพที่ 7



ภาพที่ 7 การเปรียบเทียบประสิทธิภาพแบบจำลองในรูปแบบกราฟข้อมูล

วิจารณ์และสรุปผลการทดลอง

การวิจัยนี้มุ่งเน้นไปที่การพัฒนาตัวแบบการเรียนรู้ของเครื่องเพื่อการวินิจฉัยมะเร็งปอด โดยใช้เทคนิค k-NN ซึ่งแสดงผลลัพธ์ที่แม่นยำ การศึกษาได้เปรียบเทียบผลลัพธ์กับงานวิจัยก่อนหน้านี้ โดยเฉพาะงานของ P. Bhuvaneswari และคณะในปี 2015 ที่ใช้เทคนิค Genetic k-NN ในการตรวจหามะเร็งปอด โดยมีความแม่นยำ 90% นอกจากนี้ การศึกษาของ Kumar และ Priyanka (2023) ที่มุ่งเน้นการปรับปรุงความแม่นยำของระบบการระบุโรคมะเร็งปอดโดยใช้เทคนิค k-NN และเปรียบเทียบกับอัลกอริทึม Logistic Regression พบว่า k-NN มีความแม่นยำ 89.56% ซึ่งเหนือกว่า Logistic Regression ที่มีความแม่นยำ 80.11% งานวิจัยนี้แสดงให้เห็นว่าเทคนิค k-NN มีประสิทธิภาพสูงสุดในการตรวจหามะเร็งปอด และช่วยลดขั้นตอนทางการแพทย์ รวมถึงมอบการรักษาที่เหมาะสมสำหรับผู้ป่วยมะเร็งปอดได้ดียิ่งขึ้น

การวิจัยนี้จึงสามารถเสนอแนวทางในการพัฒนาเครื่องมือทางการแพทย์เพื่อการวินิจฉัยมะเร็งปอดได้อย่างมีประสิทธิภาพและแม่นยำสูง ซึ่งจะเป็นประโยชน์อย่างยิ่งต่อการรักษาผู้ป่วยมะเร็งปอดในระยะเริ่มต้นและระยะที่ลุกลาม การทดสอบและประเมินผลโดยใช้ชุดข้อมูลภาพเอกซเรย์คอมพิวเตอร์จำนวนมากเป็นหลักฐานที่เชื่อถือได้และให้ข้อมูลที่ครอบคลุมเกี่ยวกับประสิทธิภาพของเทคนิคต่างๆ

จากผลการวิจัย พบว่าเทคนิค k-NN มีค่าประสิทธิภาพรวม (F-measure) ความไว (Sensitivity) และความจำเพาะ (Specificity) สูงถึงมากกว่า 95% ซึ่งเหนือกว่าผลการวิจัยอื่นๆ ที่เปรียบเทียบมา ทำให้เทคนิคนี้เป็นทางเลือกที่ดีในการประเมินและตรวจสอบผู้ป่วยมะเร็งปอดด้วยความแม่นยำสูง

เพื่อเพิ่มความน่าเชื่อถือและความสามารถในการประยุกต์ใช้ในทางปฏิบัติ ผู้วิจัยควรพิจารณาทำการทดลองเพิ่มเติมเกี่ยวกับประสิทธิภาพของเทคนิคนี้ในสภาพแวดล้อมทางคลินิกจริง และทดสอบกับข้อมูลจากฐานข้อมูลผู้ป่วยมะเร็งปอดที่หลากหลาย การศึกษาเพิ่มเติมเหล่านี้จะช่วยให้มั่นใจได้ว่าเทคนิค k-NN สามารถช่วยในการตัดสินใจทางการแพทย์ได้อย่างไร และมีความเกี่ยวข้องต่อการรักษาและดูแลผู้ป่วยมะเร็งปอดอย่างไรให้เหมาะสมและมีประสิทธิภาพมากที่สุดในแต่ละระยะของโรค

ทั้งนี้ การวิจัยที่เน้นไปที่การปรับปรุงเทคนิคและเพิ่มความแม่นยำในการวินิจฉัยเป็นสิ่งสำคัญที่จะช่วยให้การรักษามะเร็งปอดเป็นไปอย่างมีประสิทธิภาพ และช่วยให้ผู้ป่วยมีโอกาสหายจากโรคมามากขึ้น การเพิ่มประสิทธิภาพของเทคนิค k-NN ในการคัดแยกประเภทและระยะของมะเร็งปอด ทำให้มีโอกาสในการตรวจจับโรคได้เร็วขึ้น ส่งผลให้การเริ่มการรักษาเกิดขึ้นในช่วงที่โรคมะเร็งยังไม่ได้ลุกลามอย่างหนัก ซึ่งเป็นประโยชน์ต่อการควบคุมและรักษาโรคให้มีประสิทธิภาพสูงสุด

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณสำหรับชุดข้อมูลภาพเอกซเรย์คอมพิวเตอร์จากเว็บไซต์ www.kaggle.com ซึ่งใช้เป็นชุดข้อมูลในการทดลองสำหรับงานวิจัยครั้งนี้ รวมถึงขอบคุณคณะกรรมการ บัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม ที่ให้การสนับสนุนการวิจัยนี้จนเป็นผลสำเร็จ

เอกสารอ้างอิง

- ธวัชชัย เหล็กดี รัฐพรธรณ สันติอนันต์ และ เจษฎา อุดมพิทยาสรรพ์. (2024). การเปรียบเทียบประสิทธิภาพของแบบจำลองการทำนายความเสี่ยงโรคมะเร็งปอดด้วยเทคนิคเหมืองข้อมูล. *วิทยาศาสตร์และเทคโนโลยีสู่ชุมชน*, 2(1), 22-35. <https://li02.tci-thaijo.org/index.php/STC/article/view/705>
- ภัทรนัย ไชยพรม สมชาย จาดศรี ดุริยา พองมูล และ สันหวีช ไชยวงศ์. (2016). ปัจจัยเกี่ยวข้อง มะเร็งปอด: การศึกษาเชิงวิเคราะห์ย้อนหลัง. *Health Science, Science and Technology Reviews*, 9(3), 47-49. <https://li01.tci-thaijo.org/index.php/journalup/article/view/73694>
- วนิดา ผาตาล. (2021). ค่าความไวและค่าความจำเพาะของ Serum prostate Specific antigen เมื่อเทียบกับการตรวจทางศัลยพยาธิวิทยาในมะเร็งต่อมลูกหมากในโรงพยาบาลอุดรธานี. *วารสารโรงพยาบาลมหาสารคาม*, 18(3), 113-123.
- สุชาวดี รุ่งแจ้ง, และ รัชนี นามจันทรา. (2016). การจัดการอาการในผู้ป่วยมะเร็งปอดระยะลุกลาม. *Thai Journal of Cardio-Thoracic Nursing*, 27(2), 43-57. <https://he02.tci-thaijo.org/index.php/journalthaicvtnurse/article/view/79374>
- อนุพงศ์ สุขประเสริฐ . (2022). *คู่มือการทำเหมืองข้อมูลด้วยโปรแกรม RapidMiner Studio* (พิมพ์ครั้งที่ 5). คณะการบัญชีและการจัดการ: มหาวิทยาลัยมหาสารคาม
- Ali, M. H., Jaber, M. M., Abd, S. K., Alkhayyat, A., & Jasim, A. D. (2022). Artificial Neural Network-Based Medical Diagnostics and Therapeutics. *International Journal of Pattern Recognition and Artificial Intelligence*, 36(16), 2240007. <https://doi.org/10.1142/S0218001422400079>
- Babu, N. S. S., Ashwini, N. G., & Ramashri, N. D. T. (2023). Automatic Benign and Malignant Classification of Pulmonary Nodules in Thoracic Computed Tomography Images based on CNN Algorithm. *International Journal of Scientific Research in Science and Technology*, 644-649.

<https://doi.org/10.32628/ijrst52310340>

- Bhuvaneswari, P., & Therese, A. B. (2015). Detection of cancer in lung with k-NN classification using genetic algorithm. **Procedia Materials Science**, 10, 433-440. <https://www.sciencedirect.com/science/article/pii/S2211812815003156>
- Chang, X. (2023). Research on image features extraction based on machine learning algorithms. In **Eighth International Conference on Electronic Technology and Information Science (ICETIS 2023)** (Vol. 12715, pp. 399-404). SPIE. <https://doi.org/10.1117/12.2682449>
- Dimitrov, P. (2023). Neural network models. **arXiv preprint** arXiv:2301.02987. <https://doi.org/10.48550/arXiv.2301.02987>
- Farhad, M., Masud, M. M., Beg, A., Ahmad, A., & Ahmed, L. (2023). A review of medical diagnostic video analysis using deep learning techniques. **Applied Sciences**, 13(11), 6582. <https://doi.org/10.3390/app13116582>
- Gültepe, Y. (2021). Performance of Lung Cancer Prediction Methods Using Different Classification Algorithms. **Computers, Materials & Continua**, 67(2). https://www.researchgate.net/publication/349530874_Performance_of_Lung_Cancer_Prediction_Methods_Using_Different_Classification_Algorithms
- Hariharan, J., Ampatzidis, Y., Abdulridha, J., & Batuman, O. (2022). Useful Feature Extraction and Machine Learning Techniques for Identifying Unique Pattern Signatures Present in Hyperspectral Image Data. In **Hyperspectral Imaging-A Perspective on Recent Advances and Applications**. **IntechOpen**. <https://doi.org/10.5772/intechopen.107436>
- Hassan, A., & Sabha, M. (2023). Feature extraction for image analysis and detection using machine learning techniques. **International Journal of Advanced Networking and Applications**, 14(4), 5499-5508. <https://doi.org/10.35444/ijana.2023.14401>
- Huang, P., Illei, P. B., Franklin, W., Wu, P. H., Forde, P. M., Ashrafinia, S., ... & Gabrielson, E. (2022). Lung Cancer Recurrence Risk Prediction through Integrated Deep Learning Evaluation. **Cancers**, 14(17), 4150. <https://doi.org/10.3390/cancers14174150>
- Huang, T., Ma, L., Zhang, B., & Liao, H. (2023). Advances in deep learning: From

- diagnosis to treatment. **BioScience Trends**, 17(3), 190-192.
<https://doi.org/10.5582/bst.2023.01148>
- Kumar, K. V., Ayyappa, Y., Kumar, B. V., Jonnadula, H., Tondepu, Y., & Vidya, N. S. (2023). A scheme of long short-term memory (LSTM)-CNN features extraction and spatial relationship to anonymize image and video for Privacy-Preserving Human Activity Recognition of multiple sequences of data. In **2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)** (pp. 1-5). IEEE. <https://doi.org/10.1109/AISP57993.2023.10134958>
- Kumar, Y. K., & Priyanka, R. (2023, January). Lung Cancer Identification System to Improve the Accuracy Using Novel K Nearest Neighbour in Comparison with Logistic Regression Algorithm. In **2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)** (pp. 1-5). IEEE.
<https://doi.org/10.1109/ICECONF57129.2023.10084340>
- Kumjit, K., Jaikoomkao, D., Phumirang, W., Sattanako, A., & Sukprasert, A. (2022, September 22). ประสิทธิภาพของเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเกิดโรคหลอดเลือดในสมอง. **ph01.tci-thaijo.org**. <https://doi.org/10.14456/jait.2022.7>
- Maleki, N., Zeinali, Y., & Niaki, S. T. A. (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. **Expert Systems with Applications**, 164, 113981.
<https://www.sciencedirect.com/science/article/abs/pii/S0957417420307594>
- Malik, P., & Singh, S. (2024). Deep Learning Approaches and Biomarkers in Medical Diagnosis. **Recent Patents on Engineering**, 18(3), 13-26.
10.2174/1872212117666230130100048
- Niyomthong, P. (2018). โรค มะเร็ง ปอด รู้ เท่าทัน ป้องกัน ได้. (PMJCS) **Phrae Medical Journal and Clinical Sciences**, 26(1), 61-80.
<https://thaidj.org/index.php/jpph/article/download/7062/6547>
- Progga, P. H., Rahman, M. J., Biswas, S., Ahmed, M. S., & Farid, D. M. (2023, April). K-Nearest Neighbour Classifier for Big Data Mining based on Informative Instances. In **2023 IEEE 8th International Conference for Convergence in Technology (I2CT)** (pp. 1-7). IEEE.

<https://doi.org/10.1109/I2CT57861.2023.10126147>

The IQ-OTH/NCCD lung cancer dataset. (2020, May 24). **Kaggle**.

<https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset/data>

Vieira, E., Ferreira, D., Neto, C., Abelha, A., & Machado, J. (2021, March). Data mining approach to classify cases of lung cancer. **In World Conference on Information Systems and Technologies** (pp. 511-521). Cham: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-72657-7_49

Wang, C., Long, Y., Li, W., Dai, W., Xie, S., Liu, Y., ... & Duan, Y. (2020). Exploratory study on classification of lung cancer subtypes through a combined K-nearest neighbor classifier in breathomics. **Scientific reports**, 10(1), 5880. <https://www.nature.com/articles/s41598-020-62803-4>