



วารสารการแพทย์และสาธารณสุขเขต 4

Office of Disease Prevention and Control, Region 4 Saraburi

<https://he01.tci-thaijo.org/index.php/JMPH4/index>

การสร้างตัวแบบต้นไม้ตัดสินใจจากผลการตรวจวิเคราะห์ทางห้องปฏิบัติการ เพื่อการคัดกรองเซลล์มะเร็งจากน้ำในช่องเยื่อหุ้มปอด

Construction of a decision tree model based on laboratory test results for screening lung cancer cells in pleural fluid

จิตติพันธ์ จันทร์แป้น¹, ชลพลสิษฐ์ พันชน², ธันยาสิริ จินคัยก³, นนิงรุทัย นิลศรี^{4*}

Jittipan Chanpaen¹, Chonphasits Panchon², Thanyasiri Jindayok³, Nungruthai Nilsri^{4*}

¹คณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร, ²โรงพยาบาลท่าวังผา จังหวัดน่าน, ³คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร,

⁴คณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร

¹Faculty of Allied Health Sciences, Naresuan University

²Medical Laboratory Department, Thawangpha Hospital

³Faculty of Medicine, Naresuan University

⁴Faculty of Allied Health Sciences, Naresuan University

*Corresponding Author: nungruthaini@nu.ac.th

Received: February 29, 2024 Revised: July 20, 2024 Accepted: July 25, 2024

บทคัดย่อ

การศึกษาพารามิเตอร์จากการตรวจวิเคราะห์ทางห้องปฏิบัติการในการใช้คัดกรองเซลล์มะเร็งจากน้ำในช่องเยื่อหุ้มปอด ยังคงมีข้อจำกัดเรื่องความไวและจำเพาะ รวมถึงค่า cut off ที่เหมาะสมที่สามารถนำมาใช้ในงานประจำวันได้ อีกทั้งการวินิจฉัยด้วยวิธีมาตรฐานทางเซลล์วิทยาใช้ระยะเวลานานในการวิเคราะห์และรายงานผล การวิเคราะห์ข้อมูลโดยการเรียนรู้ของเครื่อง (machine learning) เป็นเครื่องมือที่นิยมนำมาใช้ในยุคของวิทยาศาสตร์ข้อมูล ในการทำนายเพื่อประยุกต์ใช้เป็นเครื่องมือในการคัดกรองโรค ดังนั้นคณะผู้วิจัยจึงสนใจสร้างตัวแบบต้นไม้ตัดสินใจ โดยอาศัยผลการตรวจวิเคราะห์ทางห้องปฏิบัติการ เพื่อใช้คัดกรองมะเร็งจากน้ำในช่องเยื่อหุ้มปอด โดยเก็บข้อมูลย้อนหลัง (retrospective study) จากกลุ่มตัวอย่างที่ตรวจพบและไม่พบเซลล์มะเร็ง จำนวน 357 ราย ทำการสร้างตัวแบบต้นไม้ตัดสินใจ ด้วยโปรแกรม WEKA โดยอาศัยอัลกอริทึม J48 พบว่า พารามิเตอร์ protein, adenosine deaminase (ADA), carcinoembryonic antigen (CEA), ปริมาณ high fluorescence- body fluid cells (HFBF#) และ ร้อยละ high fluorescence- body fluid cells (HFBF%) ทั้งสองกลุ่มตัวอย่างมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p < 0.05$) จากนั้นทำการสร้างตัวแบบต้นไม้ตัดสินใจโดยแบ่งข้อมูล ร้อยละ 90 เป็นข้อมูลชุดฝึกหัด (training dataset) และข้อมูลชุดทดสอบประสิทธิภาพ (testing dataset) ส่วนอีกร้อยละ 10 ใช้เป็นข้อมูลชุดทดสอบแบบอำพราง (blind dataset) ผลการสร้างตัวแบบต้นไม้ตัดสินใจพบว่า ตัวแบบที่มีประสิทธิภาพมากที่สุดประกอบด้วยพารามิเตอร์ CEA, ADA, protein และ HFBF% มีค่าความไวร้อยละ 94.10 ค่าความจำเพาะร้อยละ 72.60 และเมื่อนำตัวแบบดังกล่าวไปทดสอบกับกลุ่มทดสอบแบบอำพราง พบว่าตัวแบบนี้มีค่าความไวร้อยละ 92.31 ค่าความจำเพาะร้อยละ 82.60 ค่าพยากรณ์บวกร้อยละ 75.00 และค่าพยากรณ์ลบร้อยละ 95.00 ซึ่งน่าจะมีประโยชน์ในการนำไปใช้ร่วมคัดกรองเซลล์มะเร็งจากน้ำในช่องเยื่อหุ้มปอดได้

คำสำคัญ: น้ำในช่องเยื่อหุ้มปอด, เซลล์วิทยา, เซลล์มะเร็ง

Abstract

Parameters from laboratory analysis have been studied for screening cancer cells in pleural fluid. However, limitations in sensitivity, specificity, and appropriate cut-off values for routine laboratory use have been identified and the standard cytology methods for diagnosis are long turnaround time. At present, data analysis using machine learning is a popular tool in the era of data science for predicting and applying disease screening tools. Therefore, our research team aimed to create a decision tree model based on laboratory analysis results for screening cancer in pleural fluid. Historical data (retrospective study) from 357 samples, both cancer-positive and cancer-negative were collected, and a decision tree model was developed using the J48 algorithm in the WEKA program. Significantly different parameters ($p < 0.05$) included protein, adenosine deaminase (ADA), carcinoembryonic antigen (CEA), the count of high fluorescence-body fluid cells (HFBF#), and the percentage of high fluorescence-body fluid cells (HFBF%). Subsequently, decision tree modeling divided 90% of the data into a training dataset and a testing dataset, while the remaining 10% served as a blind dataset. The results of the decision tree modeling showed that the most effective model included the parameters CEA, ADA, protein, and HFBF%, achieving a sensitivity of 94.10% and specificity of 72.60%. Testing the model on a blind dataset demonstrated a sensitivity of 92.31% and specificity of 82.60%, with positive and negative predictive values of 75.00% and 95.00%, respectively. This suggests that the model could be beneficial for screening cancer cells in pleural fluid.

Keywords: Pleural fluid, Cytology, Malignant cell

บทนำ

สารน้ำในร่างกาย หรือ body fluid คือของเหลวที่อยู่ในช่องระหว่างเยื่อหุ้มอวัยวะต่างๆ ในร่างกาย สารน้ำเหล่านี้ทำหน้าที่ลำเลียงสารต่างๆ ได้แก่ อาหาร เกลือแร่ ออกซิเจน ไปยังเซลล์ และนำของเสีย ออกสู่ภายนอก ร่างกาย อีกทั้งยังช่วยควบคุมอุณหภูมิ รวมถึงช่วยลดการเสียดสีในขณะที่มีการเคลื่อนไหวของอวัยวะ สารน้ำในร่างกายประกอบไปด้วย น้ำไขสันหลัง (cerebrospinal fluid) น้ำในช่องท้อง (ascites fluid) น้ำจากช่องเยื่อหุ้มปอด (pleural fluid) น้ำจากช่องเยื่อหุ้มหัวใจ (pericardial fluid) และน้ำไขข้อ (synovial fluid) เป็นต้น เมื่อเกิดพยาธิสภาพ เช่น มีการติดเชื้อ การอักเสบ การติดเชื้อไวรัส หรือเกิดภาวะที่มีเซลล์มะเร็ง สารน้ำเหล่านี้จะมีปริมาณเพิ่มสูงขึ้น⁽¹⁾ แพทย์จะทำการเจาะสารน้ำและส่งตรวจทางห้องปฏิบัติการเพื่อค้นหาสาเหตุ วินิจฉัยโรค และบ่งบอกถึงความผิดปกติในร่างกาย โดยทำการตรวจวิเคราะห์ปริมาณเซลล์เม็ดเลือดแดง เซลล์เม็ดเลือดขาว และตรวจแยกชนิดของเซลล์เม็ดเลือดขาว รวมถึงการตรวจหาเซลล์ที่ผิดปกติ อาทิเช่น เซลล์มะเร็ง นอกจากนี้ยังทำการตรวจวิเคราะห์เพื่อหาปริมาณของสารชีวเคมีต่าง ๆ ที่อยู่ในสารน้ำ ได้แก่ protein,

glucose, lactate dehydrogenase (LDH), carcinoembryonic antigen (CEA) หรือ adenosine deaminase (ADA) นอกจากนี้ยังทำการส่งตรวจทางเซลล์วิทยาเพื่อวินิจฉัยความผิดปกติของเซลล์ที่พบในสารน้ำโดยพยาธิแพทย์ ซึ่งห้องปฏิบัติการพยาธิวิทยาในประเทศไทยมีจำนวนจำกัด

ในปัจจุบัน มีการพัฒนาวิธีการตรวจวิเคราะห์สารน้ำในร่างกายด้วยเครื่องตรวจวิเคราะห์เซลล์เม็ดเลือดอัตโนมัติ จากพารามิเตอร์ high-fluorescence body fluid cell (HFBF) ซึ่งเป็นการประเมินค่าของเซลล์ที่มีการติดสีฟลูออเรสเซนซ์ของอาร์เอ็นเอ และดีเอ็นเอที่อยู่ในนิวเคลียสของเซลล์ ซึ่งปริมาณของอาร์เอ็นเอ และดีเอ็นเอสูงนี้จะพบมากในเซลล์เม็ดเลือดตัวอ่อน หรือเซลล์มะเร็ง^(2,3) จากการศึกษาวิจัยก่อนหน้านี้พบว่า กลุ่มของสารน้ำจากช่องเยื่อหุ้มปอดที่ตรวจพบเซลล์มะเร็ง (malignant pleural effusions ; MPE) จะมีค่า HFBF สูงกว่ากลุ่มของสารน้ำจากช่องเยื่อหุ้มปอดที่ไม่พบเซลล์มะเร็ง (non-malignant pleural effusions ; Non-MPE) อย่างมีนัยสำคัญทางสถิติ ดังนั้นมีความเป็นไปได้ว่าพารามิเตอร์ HFBF นี้จะสามารถนำมาร่วมใช้ในการคัดกรองมะเร็งในน้ำจะต่าง ๆ ได้⁽⁴⁻⁶⁾ นอกจากนี้งานวิจัย

ของ Chih-Feng Chian และคณะ⁽⁷⁾ ได้ทำการเปรียบเทียบพารามิเตอร์ต่าง ๆ ในสองกลุ่มประชากร คือ กลุ่มของสารน้ำจากช่องเยื่อหุ้มปอดที่ตรวจพบเซลล์มะเร็งและกลุ่มที่ไม่พบเซลล์มะเร็ง พบว่ามีค่า CEA และ LDH ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ และเมื่อคำนวณหาค่า cut off โดยใช้ ROC curve พบว่า CEA มีค่า cut off มากกว่า 4.455 ng/mL มีค่าความไวร้อยละ 71 ความจำเพาะร้อยละ 72.2 ส่วน LDH มีค่า cut off มากกว่า 250.5 U/L มีค่าความไวร้อยละ 60.7 ความจำเพาะร้อยละ 77.8 จากการศึกษาที่กล่าวมาข้างต้นพบว่าพารามิเตอร์ที่ทำการตรวจวิเคราะห์นี้ ยังคงมีข้อจำกัดเรื่องความไวและจำเพาะที่ไม่สูงมาก รวมถึงค่า cut off ที่เหมาะสมที่สามารถนำมาใช้ในงานประจำวัน ซึ่งพบว่ามีประสิทธิภาพแตกต่างกันไปตาม ชนิดสิ่งส่งตรวจ เครื่องตรวจวิเคราะห์ และการวินิจฉัยแยกชนิดของเซลล์มะเร็ง⁽⁸⁻¹¹⁾

Machine Learning หรือการเรียนรู้ของเครื่อง เป็นระบบสมองกลที่สามารถเรียนรู้สิ่งใดสิ่งหนึ่งหรือหลายสิ่งได้จากตัวอย่างที่อยู่ในฐานข้อมูลปัจจุบันหรือตัวอย่างที่เพิ่มเข้ามาในอนาคตด้วยตนเองโดยปราศจากการป้อนคำสั่งจากมนุษย์ โดยคอมพิวเตอร์สามารถเรียนรู้ขีดวิเคราะห์ข้อมูลที่ซับซ้อนมากกว่าการใช้สถิติทดสอบทั่วไป และสามารถผลิตผลลัพธ์ที่แม่นยำออกมาได้ การใช้ตัวแบบต้นไม้ตัดสินใจ J48 (decision tree) ด้วยโปรแกรม WEKA เป็นหนึ่งในการเรียนรู้ของเครื่อง ที่ทำการวิเคราะห์ข้อมูลและแตกกิ่งก้านสาขาจากเหตุไปสู่ผล โดยสามารถให้ผลลัพธ์ที่ดีที่สุดในเวลานั้นและสามารถวิเคราะห์หลายพารามิเตอร์ร่วมกันได้ ซึ่งสามารถนำมาใช้ในการวิเคราะห์ข้อมูลสำหรับคัดกรองหรือวินิจฉัยโรคได้⁽¹²⁾

ดังนั้น หากมีการนำพารามิเตอร์ต่าง ๆ ที่ทำการตรวจวิเคราะห์น้ำในช่องเยื่อหุ้มปอด ในห้องปฏิบัติการมาใช้ร่วมกันหลายพารามิเตอร์ อาจส่งผลให้ค่าความไวและความจำเพาะในการคัดกรองเซลล์มะเร็งจากน้ำในช่องเยื่อหุ้มปอด เพิ่มสูงขึ้น และสามารถคัดกรองผู้ป่วยได้อย่างรวดเร็วยิ่งขึ้น ผู้วิจัยจึงสนใจที่จะศึกษาความสัมพันธ์ระหว่างพารามิเตอร์ของการตรวจน้ำในช่องเยื่อหุ้มปอดกับผลการวินิจฉัยทางเซลล์วิทยาเพื่อเป็นเครื่องมือในการคัดกรองเซลล์มะเร็งด้วยตัวแบบต้นไม้ตัดสินใจ

วัตถุประสงค์

1. ศึกษาความสัมพันธ์ระหว่างพารามิเตอร์ต่าง ๆ ร่วมกับผลการวินิจฉัยทางเซลล์วิทยา เพื่อหาพารามิเตอร์ที่เหมาะสมในการสร้างตัวแบบต้นไม้ตัดสินใจ
2. ออกแบบตัวแบบต้นไม้ตัดสินใจที่เหมาะสมสำหรับการคัดกรองเซลล์มะเร็งจากน้ำในช่องเยื่อหุ้มปอด

วัสดุและวิธีการศึกษา

1. เครื่องมือที่ใช้ในการวิจัยและกลุ่มประชากร

การศึกษานี้เป็นการศึกษาเชิงพรรณนา โดยเก็บรวบรวมข้อมูลย้อนหลัง (retrospective study) จากผลการตรวจวิเคราะห์น้ำในช่องเยื่อหุ้มปอด ในงานประจำวัน ตั้งแต่ พ.ศ. 2564 - 2566 โดยดำเนินการที่ งานพยาธิคลินิก กลุ่มงานพยาธิวิทยาคลินิกและเทคนิคการแพทย์ สถาบันโรคทรวงอก กรมการแพทย์ กระทรวงสาธารณสุข จังหวัดนนทบุรี โดยการคำนวณหาขนาดกลุ่มตัวอย่างใช้สูตรของทาโรยามาเน และทำการเก็บข้อมูลจากประชากรทั้งหมด 357 คน โดยการเข้าร่วมการวิจัย มีเกณฑ์ในการคัดเลือกประชากร ดังนี้

เกณฑ์การคัดเลือก (inclusion criteria)

- มีผลการตรวจทางห้องปฏิบัติการ และ ผลการตรวจวินิจฉัยโดยพยาธิแพทย์ ครอบคลุมที่กำหนด
- มีผลการตรวจวินิจฉัยโดยพยาธิแพทย์ชัดเจนได้แก่

กลุ่ม non-malignant pleural effusions (non-MPE) คือ กลุ่มตัวอย่างที่ไม่พบเซลล์มะเร็งจากน้ำในช่องเยื่อหุ้มปอด และผลการตรวจวินิจฉัยโดยพยาธิแพทย์รายงาน negative

กลุ่ม malignant pleural effusions (MPE) คือ กลุ่มตัวอย่างที่พบเซลล์มะเร็งจากน้ำในช่องเยื่อหุ้มปอด และผลการตรวจวินิจฉัยโดยพยาธิแพทย์รายงาน adenocarcinoma, mesothelioma, carcinoma, lymphoma, non-small cell carcinoma, squamous cell carcinoma, small cell carcinoma เป็นต้น

เกณฑ์การคัดออก (exclusion criteria)

- มีผลการตรวจวิเคราะห์ไม่ครอบคลุมที่กำหนด
- มีผลการตรวจวินิจฉัยโดยพยาธิแพทย์ไม่ชัดเจน เช่น สงสัย (suspect)

การดำเนินการวิจัย แบ่งตัวอย่างออกเป็น 2 กลุ่ม ได้แก่ กลุ่ม MPE และกลุ่ม Non-MPE และทำการรวบรวมข้อมูลผลการตรวจวิเคราะห์น้ำตาลในช่องเชื้อหุ้มปอด ทางห้องปฏิบัติการ ได้แก่

ผลการตรวจนับและแยกชนิดของเซลล์โดยใช้เครื่อง Sysmex XN-1000 (Sysmex corporation, Kobe, Japan) โดยมีพารามิเตอร์ดังนี้

- PMN หมายถึง polymorphonuclear cell
- MN หมายถึง mononuclear cell
- Total cell- body fluid (TC-BF) หมายถึง จำนวนเซลล์ที่มีนิวเคลียสทั้งหมดต่อไมโครลิตร
- White blood cell-body fluid (WBC-BF) หมายถึง จำนวนเซลล์เม็ดเลือดขาวต่อไมโครลิตร
- Red blood cell - body fluid (RBC-BF) หมายถึง จำนวนเซลล์เม็ดเลือดแดงต่อไมโครลิตร
- HF-BF# หมายถึง จำนวนเซลล์ที่ติดสีฟลูออเรสเซนส์สูงต่อไมโครลิตร
- HF-BF% หมายถึง จำนวนเซลล์ที่ติดสีฟลูออเรสเซนส์สูงต่อเม็ดเลือดขาว 100 เซลล์

และผลการวิเคราะห์ค่าทางเคมีคลินิกโดยใช้เครื่อง Atellica® CH 930 และ Atellica® IM 1300 Analyzer (Siemens Healthineers, Germany) ได้แก่ พารามิเตอร์ protein, glucose, Lactate dehydrogenase (LDH), adenosine deaminase (ADA) และ carcinoembryonic antigen (CEA) และข้อมูลการรายงานผลเซลล์วิทยาที่วินิจฉัยโดยพยาธิแพทย์จากระบบสารสนเทศโรงพยาบาล (hospital information systems : HIS)

ในการวิเคราะห์ข้อมูล ใช้โปรแกรม waikato environment for knowledge analysis (WEKA) เวอร์ชัน 3.8.6 ซึ่งเป็นโปรแกรมฟรีแวร์ที่นิยมในงานด้านการทำเหมืองข้อมูล

2. การสร้างตัวแบบต้นไม้ตัดสินใจ

2.1 การแบ่งข้อมูล นำข้อมูลทั้งหมดมาแบ่งออกเป็น 3 ส่วน จากประชากรจำนวน 357 คน เพื่อใช้สำหรับการวิเคราะห์ข้อมูล โดยแบ่งข้อมูลออกเป็นสัดส่วนดังนี้

ส่วนที่ 1 : ข้อมูลชุดทดสอบแบบอำพราง (blind dataset) เพื่อนำไปทำนายตัวแบบ สำหรับยืนยันประสิทธิภาพของตัวแบบต้นไม้ตัดสินใจเมื่อนำไปใช้งานจริง มีข้อมูลร้อยละ 10 ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 1 จำนวน 36 คน

นำข้อมูลที่เหลือจำนวน 321 คน ไปทำการทดสอบความแตกต่างของผลการตรวจวิเคราะห์ระหว่างกลุ่มประชากรด้วยวิธีทางสถิติ จากนั้นทำการทดสอบแบบ 10 fold cross validation ซึ่งจะแบ่งข้อมูลออกเป็น 10 ชุด เท่าๆ กัน และแบ่งข้อมูลสำหรับส่วนที่ 2 และ 3 ดังนี้⁽¹³⁾

ส่วนที่ 2 : ข้อมูลชุดฝึกหัด (training dataset) เพื่อนำไปสร้างตัวแบบ มีข้อมูลร้อยละ 70 ของข้อมูลที่เหลือ ซึ่งจะได้ข้อมูลในส่วนที่ 2 จำนวน 225 คน

ส่วนที่ 3 : ข้อมูลชุดทดสอบประสิทธิภาพ (testing dataset) เพื่อนำไปทดสอบความถูกต้องของตัวแบบมีข้อมูลร้อยละ 30 ของข้อมูลข้อมูลที่เหลือ ซึ่งจะได้ข้อมูลในส่วนที่ 3 จำนวน 96 คน

ซึ่งจากการศึกษาก่อนหน้านี้มีการใช้ข้อมูลสำหรับชุดฝึกหัดและชุดทดสอบที่หลากหลายเช่น 70:30, 80:20 และ 90:10 แต่ทั้งนี้ยังไม่มีเกณฑ์กำหนดความเหมาะสมในการใช้ ผู้วิจัยจึงเลือกใช้ สัดส่วน 70:30 เนื่องจากเหมาะกับกลุ่มประชากรที่ไม่มากมีกลุ่มทดสอบที่สูงกว่าสัดส่วนอื่นซึ่งสามารถวัดประสิทธิภาพของแผนผังได้ดีเช่นในงานวิจัยนี้

2.2 การสร้างตัวแบบ ทำการสังเคราะห์ข้อมูลในส่วนที่ 2 เพิ่มเพื่อปรับความสมดุลของข้อมูล (synthetic minority oversampling technique; SMOT) ด้วยโปรแกรม WEKA คำสั่ง filters > supervised > instance > SMOTE โปรแกรมจะทำการสังเคราะห์ข้อมูลระหว่างกลุ่ม MPE และ Non-MPE ให้มีปริมาณใกล้เคียงกันเพื่อเพิ่มประสิทธิภาพในการวิเคราะห์ข้อมูล กระบวนการสร้างตัวแบบต้นไม้ตัดสินใจ วนทำเป็นจำนวน 10 รอบ โดยเปลี่ยนกลุ่มทดสอบไปจนครบ ข้อมูลการวัดประสิทธิภาพของการจำแนกประเภทข้อมูล ได้แก่ ร้อยละของความไว (sensitivity) ร้อยละค่าความจำเพาะ (specificity) ร้อยละค่าพยากรณ์ผลบวก (positive predictive value) ร้อยละ

ค่าพยากรณ์ผลลบ (negative predictive value) ค่าถ่วงดุล (F-Measure) ค่า Area under the curve (AUC) และ ร้อยละค่าความถูกต้อง (Accuracy)

ทำการคัดเลือกตัวแบบต้นไม้มัดสันใจที่เหมาะสมที่สุด โดยพิจารณาจากค่าร้อยละความไวเป็นหลัก และตัวแบบจะถูกทดสอบกับกลุ่มทดสอบแบบอำพรางเพื่อทดสอบประสิทธิภาพของตัวแบบต้นไม้มัดสันใจเมื่อนำไปใช้งานจริง^(14, 15) แผนผังการสร้างตัวแบบต้นไม้มัดสันใจ แสดงดังรูปที่ 1

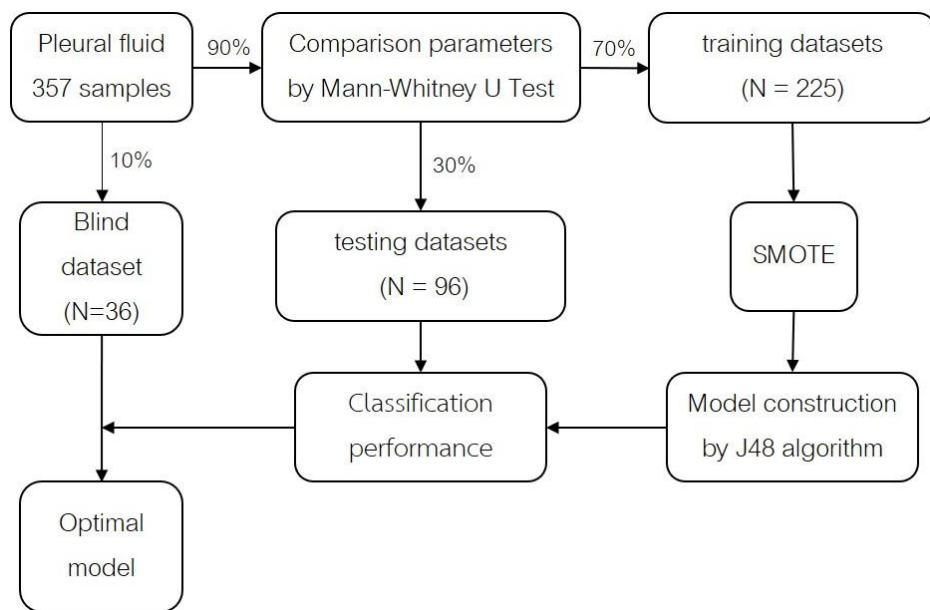
3. สถิติที่ใช้ในการวิเคราะห์

ทดสอบการกระจายตัวของประชากรด้วยสถิติ Kolmogorov-Smirnov test เนื่องจากประชากรที่ต้องการศึกษามีจำนวนมากกว่า 50 ราย ซึ่งพบว่ามี การกระจายตัว

แบบไม่ปกติ จึงแสดงผลด้วยค่า median และ interquartile range (IQR) จากนั้นทำการทดสอบความแตกต่างของข้อมูลระหว่างกลุ่มประชากรด้วยสถิติ Mann-Whitney U Test และ chi-square test (สำหรับข้อมูลเพศ) โดยใช้โปรแกรม IBM SPSS Statistics v.29

4. การพิจารณาด้านจริยธรรมการวิจัยในมนุษย์

การศึกษานี้ได้ผ่านการรับรองด้านจริยธรรมการวิจัยในมนุษย์จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยนครสวรรค์ เลขที่โครงการ IRB No. P1-0112/2565 วันที่รับรอง 25 ตุลาคม 2565 และคณะกรรมการจริยธรรมการวิจัยในมนุษย์ สถาบันโรคทรวงอก เลขที่โครงการ REC 002/2566 วันที่รับรอง 5 กันยายน 2565



รูปที่ 1 แผนผังการสร้างตัวแบบต้นไม้มัดสันใจ

ผลการศึกษา

จากการรวบรวมข้อมูลการตรวจวิเคราะห์จากผู้ป่วยของสถาบันโรคทรวงอกระหว่าง ปี พ.ศ.2564 - 2566 เมื่อทดสอบการกระจายตัวของประชากรด้วยสถิติ Kolmogorov-Smirnov test พบว่ากลุ่มตัวอย่างมีการกระจายตัวแบบไม่ปกติ จึงใช้สถิติ Mann-Whitney U Test ในการทดสอบความแตกต่างของผลการตรวจวิเคราะห์ระหว่างกลุ่มประชากร การทดสอบพบว่า พารามิเตอร์ protein, ADA, CEA, HFBF#, HFBF% ในกลุ่ม MPE

แตกต่างกับกลุ่ม Non-MPE อย่างมีนัยสำคัญทางสถิติที่ $p < 0.05$ ดังตารางที่ 1

โปรแกรมจะทำการสร้างตัวแบบต้นไม้มัดสันใจ และเลือกจุดตัดที่เหมาะสมที่สุดให้โดยอัตโนมัติ และโปรแกรมจะทำการสร้างตัวแบบทั้งหมด 10 ตัวแบบจากชุดฝึกหัด นำตัวแบบที่ได้ไปทดสอบกับชุดทดสอบประสิทธิภาพ พบว่าตัวแบบที่ 3 และ 9 เป็นตัวแบบที่มีค่าความไวมากกว่าร้อยละ 90.00 ซึ่งมีความเหมาะสมต่อ

การนำไปใช้คัดกรองหาเซลล์มะเร็งจากน้ำในช่องเยื่อหุ้มปอด โดยตัวแบบที่ 3 ชุดฝึกหัดมีค่าความไวร้อยละ 98.70 เมื่อนำไปทดสอบกับชุดทดสอบประสิทธิภาพ พบว่าค้นไม่ตัดสินใจตัวแบบที่ 3 มีค่าความไวร้อยละ 94.10 และในตัวแบบที่ 9 ชุดฝึกหัดมีค่าความไวร้อยละ 100.00 เมื่อนำไปทดสอบด้วยชุดทดสอบประสิทธิภาพ พบว่า ค้นไม่ตัดสินใจตัวแบบที่ 9 มีค่าความไวร้อยละ 91.70

ทำการคัดเลือกตัวแบบที่เหมาะสมโดยมีเกณฑ์พิจารณาจากค่าความไวที่สูงที่สุด และค่าความไวระหว่างชุดฝึกหัดและชุดทดสอบประสิทธิภาพในตัวแบบนั้นต้องแตกต่างกันไม่เกินร้อยละ 5 ซึ่งบ่งบอกถึงการรักษาสภาพของตัวแบบนั้นเหมาะสมสำหรับการคัดกรอง จากเกณฑ์ที่กำหนดข้างต้นพบว่าค้นไม่ตัดสินใจตัวแบบที่ 3 เป็นตัวแบบที่มีประสิทธิภาพสูงที่สุด ดังแสดงในตารางที่ 2

ตารางที่ 1 ข้อมูลทั่วไปและผลการตรวจน้ำในช่องเยื่อหุ้มปอดของผู้ป่วย เปรียบเทียบระหว่างกลุ่มที่พบเซลล์มะเร็ง (n=112) กับกลุ่มที่ไม่พบเซลล์มะเร็ง (n=209)

ข้อมูล	ค่ามัธยฐาน (ค่าพิสัยระหว่างควอไทล์)		p-value
	MPE	Non-MPE	
เพศ			0.016
จำนวน (ร้อยละ)			
ชาย	54.00(16.82)	130.00(40.50)	
หญิง	58.00(18.07)	79.00(24.61)	
อายุ(ปี)	65.00(21.00)	61.00(29.00)	0.009
ผลตรวจน้ำในช่องเยื่อหุ้มปอด			
Protein (g/dL)	4.85(1.28)	4.50(1.70)	0.022
Glucose (mg/dL)	108.50(72.75)	98.00(50)	0.077
LDH (U/L)	401.50(399.25)	371.00(635.50)	0.302
ADA (U/L)	32.35(11.70)	37.30(43.15)	<.001
CEA (ng/mL)	218.05(568.13)	1.20(2.00)	<.001
WBC (10 ³ /μL)	1.07(1.60)	0.95(2.80)	0.695
RBC (10 ⁶ /μL)	0.01(0.06)	0.01(0.04)	0.053
MN# (10 ³ /μL)	0.81(1.13)	0.67(1.51)	0.277
PMN# (10 ³ /μL)	0.15(0.35)	0.10(0.68)	0.180
MN%	82.80(27.80)	83.50(41.75)	0.934
PMN%	17.20(27.80)	16.50(41.75)	0.934
TCBF (10 ³ /μL)	1.27(1.64)	1.03(2.92)	0.517
HFBF# (10 ³ /μL)	0.05(0.12)	0.01(0.05)	<.001
HFBF%	4.75(9.40)	1.40(3.50)	<.001

ตารางที่ 2 ผลการทดสอบระหว่างชุดฝึกหัดและชุดทดสอบประสิทธิภาพ ด้วยโปรแกรม WEKA

ตัว แบบ	กลุ่ม	ความไว (ร้อยละ)	ความ แม่นยำ (ร้อยละ)	พยากรณ์ ผลบวก (ร้อยละ)	พยากรณ์ ผลลบ (ร้อยละ)	Area under the curve	ค่าถ่วงดุล F-measure	ความถูกต้อง (ร้อยละ)
1	Train	95.90	83.60	84.80	95.50	0.946	0.896	90.30
	Test	89.70	63.20	62.50	90.00	0.821	0.740	78.80
2	Train	96.80	90.50	91.40	96.40	0.943	0.937	93.90
	Test	85.70	72.10	63.80	89.80	0.739	0.755	80.30
3	Train	98.70	85.00	87.50	98.40	0.848	0.920	92.80
	Test	94.10	72.60	65.30	95.70	0.871	0.806	85.00
4	Train	95.10	88.90	89.00	95.10	0.928	0.919	92.10
	Test	85.00	89.30	85.00	89.30	0.836	0.875	87.50
5	Train	89.20	91.80	92.20	88.70	0.912	0.905	90.50
	Test	78.80	76.20	63.40	87.30	0.759	0.775	79.10
6	Train	100.00	78.00	82.00	100.00	0.930	0.889	91.00
	Test	86.50	81.40	74.40	90.60	0.871	0.835	84.30
7	Train	94.00	86.50	89.30	92.40	0.921	0.906	90.70
	Test	75.00	86.80	70.00	89.40	0.895	0.835	83.70
8	Train	98.80	87.90	90.70	98.40	0.955	0.938	94.20
	Test	78.60	77.90	59.50	89.80	0.793	0.789	81.00
9	Train	100.00	82.60	85.40	100.00	0.950	0.913	92.60
	Test	91.70	78.30	71.70	94.00	0.857	0.836	85.70
10	Train	98.30	86.80	90.70	97.50	0.939	0.933	93.60
	Test	78.30	78.10	52.90	91.90	0.811	0.793	82.60

เมื่อพิจารณาต้นไม้ตัดสินใจตัวแบบที่ 3 ซึ่งเป็นตัวแบบที่มีประสิทธิภาพสูงที่สุด ซึ่งประกอบไปด้วยพารามิเตอร์ CEA, ADA, Protein และ HFBF% ดังแสดงในรูปที่ 2 โดยสามารถสร้างเป็นกฎการตัดสินใจได้ 9 กฎ ดังนี้

กฎที่ 1 ถ้าผู้ป่วยมีปริมาณ CEA น้อยกว่าหรือเท่ากับ 1.80 ng/mL และ HFBF% น้อยกว่าหรือเท่ากับ 1.20 ng/ml สามารถพยากรณ์ได้ว่ามีโอกาสให้ผลการตรวจพบเซลล์มะเร็งเป็นลบ จากการทดลองให้ผลลบจริง จำนวน 62 ราย

กฎที่ 2 ถ้าผู้ป่วยมีปริมาณ CEA น้อยกว่าหรือเท่ากับ 1.80 ng/mL และ มี HFBF% มากกว่าร้อยละ 1.61 สามารถพยากรณ์ได้ว่ามีโอกาสให้ผลการตรวจพบเซลล์มะเร็ง

เป็นลบ จากการทดลองให้ผลลบจริง จำนวน 39 ราย และผลลบปลอม จำนวน 2 ราย

กฎที่ 3 ถ้าผู้ป่วยมีปริมาณ CEA น้อยกว่าหรือเท่ากับ 1.80 ng/mL มี HFBF% มากกว่าร้อยละ 1.20 ถึงร้อยละ 1.61 และค่า ADA มากกว่า 30.80 U/L สามารถพยากรณ์ได้ว่ามีโอกาสให้ผลการตรวจพบเซลล์มะเร็งเป็นลบ จากการทดลองให้ผลลบจริง จำนวน 5 ราย

กฎที่ 4 ถ้าผู้ป่วยมีปริมาณ CEA น้อยกว่าหรือเท่ากับ 1.80 ng/mL มี HFBF% มากกว่าร้อยละ 1.20 ถึงร้อยละ 1.61 และค่า ADA น้อยกว่าหรือเท่ากับ 30.80 U/L สามารถพยากรณ์ได้ว่ามีโอกาสให้ผลการตรวจพบเซลล์มะเร็งเป็นบวก จากการทดลองให้ผลบวกจริง จำนวน 4 ราย

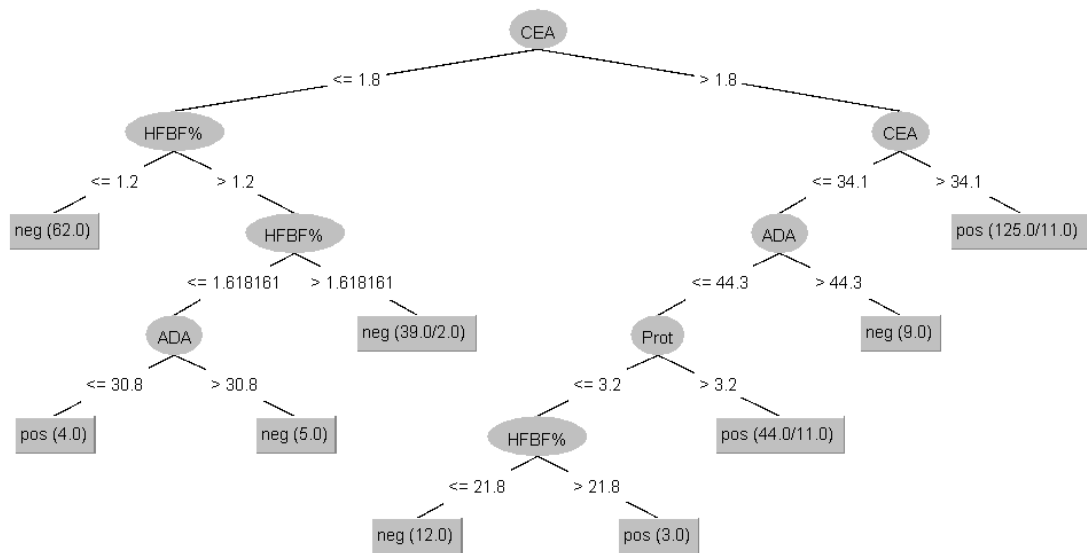
กฎที่ 5 ถ้าผู้ป่วยมีปริมาณ CEA มากกว่า 34.10 ng/mL สามารถพยากรณ์ได้ว่ามีโอกาสให้ผลการตรวจพบเซลล์มะเร็งเป็นบวก จากการทดลองให้ผลบวกจริง จำนวน 125 ราย และผลบวกปลอม จำนวน 11 ราย

กฎที่ 6 ถ้าผู้ป่วยมีปริมาณ CEA มากกว่า 1.80 ng/mL ถึง 34.10 ng/mL และ ADA มากกว่า 44.30 U/L สามารถพยากรณ์ได้ว่ามีโอกาสให้ผลการตรวจพบเซลล์มะเร็งเป็นลบ จากการทดลองให้ผลลบจริง จำนวน 9 ราย

กฎที่ 7 ถ้าผู้ป่วยมีปริมาณ CEA มากกว่า 1.80 ng/mL ถึง 34.10 ng/mL มี ADA น้อยกว่าเท่ากับ 44.30 U/L และ protein มากกว่า 3.20 g/dL สามารถพยากรณ์ได้ว่ามีโอกาสให้ผลการตรวจพบเซลล์มะเร็งเป็นบวก จากการทดลองให้ผลบวกจริง จำนวน 44 ราย และผลบวกปลอม จำนวน 11 ราย

กฎที่ 8 ถ้าผู้ป่วยมีปริมาณ CEA มากกว่า 1.80 ng/mL ถึง 34.10 ng/mL มี ADA น้อยกว่าเท่ากับ 44.30 U/L มี protein น้อยกว่าเท่ากับ 3.20 g/dL และมี HFBF% มากกว่าร้อยละ 21.80 สามารถพยากรณ์ได้ว่ามีโอกาสให้ผลการตรวจพบเซลล์มะเร็งเป็นบวก จากการทดลองให้ผลบวกจริง จำนวน 3 ราย

กฎที่ 9 ถ้าผู้ป่วยมีปริมาณ CEA มากกว่า 1.80 ng/mL ถึง 34.10 ng/mL มี ADA น้อยกว่าเท่ากับ 44.30 U/L มี protein น้อยกว่าเท่ากับ 3.20 g/dL และมี HFBF% น้อยกว่าเท่ากับร้อยละ 21.80 สามารถพยากรณ์ได้ว่ามีโอกาสให้ผลการตรวจพบเซลล์มะเร็งเป็นลบ จากการทดลองให้ผลลบจริง จำนวน 12 ราย



รูปที่ 2 ต้นไม้ตัดสินใจตัวแบบที่ 3

เมื่อนำต้นไม้ตัดสินใจตัวแบบที่ 3 ไปทดสอบกับกลุ่มทดสอบแบบอำพราง พบว่าให้ค่าความไวร้อยละ 92.31 ค่าความจำเพาะร้อยละ 82.60 ค่าพยากรณ์ผลบวก

ร้อยละ 75.00 และค่าพยากรณ์ผลลบร้อยละ 95.00 ซึ่งเป็นการยืนยันถึงประสิทธิภาพของตัวแบบอย่างแท้จริง ดังแสดงในตารางที่ 3

ตารางที่ 3 ผลการทดสอบค่าความไว ความจำเพาะ ค่าพยากรณ์ผลบวก และค่าพยากรณ์ผลลบของต้นไม้ตัดสินใจตัวแบบที่ 3 กับกลุ่มทดสอบอำพราง

		Disease		
		MPE	Non-MPE	
Test	Positive	12	4	PPV = 75.00%
	Negative	1	19	NPV = 95.00%
		Sensitivity = 92.31%		Specificity = 82.60%

อภิปรายผลการวิจัย

Decision tree หรือ ตัวแบบต้นไม้ตัดสินใจ เป็นสถิติทดสอบชนิดหนึ่งที่จำลองวิธีการตัดสินใจโดยมีเงื่อนไขต่างๆ ประกอบการตัดสินใจในแต่ละครั้งโดยพิจารณาแบ่งแยกข้อมูลโดยพิจารณาจากข้อมูลทั้งหมดที่มี เพื่อให้สามารถเรียนรู้การตัดสินใจที่ซับซ้อนคล้ายการตัดสินใจของมนุษย์ ซึ่งจากการทบทวนวรรณกรรมพบว่าการนำต้นไม้ตัดสินใจมาใช้ในทางด้านสาธารณสุข เช่น งานวิจัยของคุณ Massimo Giotta และคณะ⁽¹⁶⁾ ที่ศึกษาและประยุกต์ใช้แบบจำลองการตัดสินใจเพื่อพยากรณ์ผลลัพธ์สุดท้ายของผู้ป่วย COVID-19 ที่เข้ารับการรักษาในโรงพยาบาล แต่ทั้งนี้ยังไม่พบการนำต้นไม้ตัดสินใจมาใช้ในการช่วยคัดกรองมะเร็งจากน้ำในช่องเยื่อหุ้มปอด เหมือนกับงานวิจัยในครั้งนี้ ตัวแบบต้นไม้ตัดสินใจที่ถูกสร้างขึ้นด้วยโปรแกรม WEKA โดยใช้อัลกอริทึม J48 ในงานวิจัยนี้พบว่า ตัวแบบต้นไม้ตัดสินใจตัวแบบที่ 3 ซึ่งประกอบไปด้วยพารามิเตอร์ CEA, ADA, protein และ HFBF# มีค่า sensitivity สูงถึงร้อยละ 94.10 ค่า specificity ร้อยละ 72.60 ค่า positive predictive value ร้อยละ 65.30 และ negative predictive value ร้อยละ 95.70 และเมื่อทดสอบประสิทธิภาพเมื่อนำไปใช้งานจริงกับกลุ่มทดสอบอำพรางพบว่า ตัวแบบต้นไม้ตัดสินใจของตัวแบบที่ 3 สามารถรักษาประสิทธิภาพของตัวแบบในการคัดกรองไว้ได้ น่าจะมีประโยชน์ในการนำไปใช้ร่วมในการคัดกรองมะเร็งจากน้ำในช่องเยื่อหุ้มปอดได้ ซึ่งสอดคล้องกับการศึกษาก่อนหน้าของคุณ Chih-Feng Chian และคณะ⁽⁷⁾ ที่ทำการศึกษาร่วมเปรียบเทียบพารามิเตอร์ระหว่างกลุ่ม Non-MPE กับ MPE พบว่าค่า CEA cut off ที่ 4.455 ng/ml ให้ค่า sensitivity ที่ร้อยละ 71.00 specificity ที่ร้อยละ 72.20 และค่า LDH cut off ที่ 250.50 ให้ค่า sensitivity อยู่ที่ร้อยละ 60.70 specificity อยู่ที่ร้อยละ 77.80 แต่เมื่อใช้ LDH ร่วมกับ CEA พบว่าให้ค่า sensitivity สูงขึ้นอยู่ที่ร้อยละ 90.30 และ specificity อยู่ที่ร้อยละ 66.70 และในงานวิจัยของคุณ Mei Feng และคณะ⁽¹⁷⁾ พบว่าค่า CEA, CYFRA 21-1, CA 19-9 จากน้ำในช่องเยื่อหุ้มปอด จะเพิ่มสูงขึ้นอย่างมีนัยสำคัญทางสถิติในกลุ่มของคนที่เป็นมะเร็ง โดยใช้ค่า CEA cut off ที่ 4.55 ng/ml เพียงพารามิเตอร์เดียว พบว่าให้ค่า sensitivity เท่ากับ

ร้อยละ 83.95 ส่วนค่า specificity เท่ากับร้อยละ 96.88 แต่หากมีการใช้พารามิเตอร์หลายชนิดร่วมพิจารณาคัดกรองมะเร็ง พบว่าสามารถเพิ่มค่า sensitivity ของการตรวจวิเคราะห์ได้ โดยใช้ CEA ร่วมกับ CYFRA 21-1 พบว่าให้ค่า sensitivity ร้อยละ 87.65 specificity ร้อยละ 93.75 และหากใช้ CEA ร่วมกับ CA 19-9 พบว่าให้ค่า sensitivity ร้อยละ 88.89 specificity ร้อยละ 93.75 และถ้าใช้ทั้ง CEA, CA19-9 และ CYFRA 21-1 ร่วมกัน พบว่าให้ค่า sensitivity เพิ่มสูงถึงร้อยละ 95.06 specificity ร้อยละ 87.50 โดยทั้งสองงานวิจัยดังกล่าวใช้เพียง ROC curve มาใช้ในการหาค่า cut off จะเห็นได้ชัดเจนว่าการนำพารามิเตอร์หลายชนิดมาช่วยในการคัดกรองหรือวินิจฉัยโรคส่งผลให้ค่า sensitivity สูงขึ้นแต่จะมีค่า specificity ที่ลดลง ซึ่งแตกต่างกับการวิจัยในครั้งนี้ที่พบว่าการใช้ตัวแบบต้นไม้ตัดสินใจที่มีการนำพารามิเตอร์มาร่วมกันพิจารณาส่งผลให้ประสิทธิภาพ sensitivity และ specificity ที่สูงกว่าการใช้ CEA เพียงอย่างเดียว ซึ่งเกิดจากการใช้การเรียนรู้โดยคอมพิวเตอร์ด้วยโปรแกรม WEKA ที่สามารถคิดวิเคราะห์โดยใช้พารามิเตอร์มาร่วมพิจารณามากขึ้น ซึ่งจะช่วยเพิ่มประสิทธิภาพในการคัดกรองเพิ่มขึ้นเช่นกัน แสดงให้เห็นว่าการนำ Machine Learning มาใช้ในการช่วยวิเคราะห์ข้อมูลทำให้สามารถดึงประสิทธิภาพของพารามิเตอร์ต่างๆออกมาได้ดีมากขึ้น ซึ่งหากมีการนำตัวแบบต้นไม้ตัดสินใจดังกล่าวไปใช้คัดกรองมะเร็งจากน้ำในช่องเยื่อหุ้มปอด จะเป็นประโยชน์ต่อผู้ป่วยและทีมรักษาต่อไป

นอกจากนี้งานวิจัยนี้ทำให้ทราบถึงความแตกต่างของพารามิเตอร์ HFBF% และ HFBF# ระหว่างกลุ่ม MPE และกลุ่ม Non-MPE ซึ่งอาจจะสามารถนำค่าดังกล่าวไปประยุกต์ใช้สำหรับการคัดกรองเซลล์มะเร็งหรือเซลล์ที่ผิดปกติในสิ่งส่งตรวจชนิดสารน้ำจากร่างกายในงานตรวจประจำวันได้ ซึ่งสอดคล้องกับวิจัยก่อนหน้า^(1, 3, 6, 18) เนื่องจากเครื่องตรวจวิเคราะห์อัตโนมัติจะใช้หลักการย้อม DNA และ RNA ภายในเซลล์ด้วยสีฟลูออเรสเซนซ์ ซึ่งเซลล์ที่มีการติดสีฟลูออเรสเซนซ์ในปริมาณมาก จะหมายถึงเซลล์มีอัตราการแบ่งตัวเพิ่มจำนวนสูง อาทิเช่น เซลล์ตัวอ่อน (blast cell)

และเซลล์มะเร็ง (malignant cells)^(2,3) และจากตัวแบบต้นไม้ตัดสินใจพบว่าโดยปกติผลการตรวจวิเคราะห์ที่มีค่าสูงมักจะถูกนำไปใช้พยากรณ์ว่าเป็นโรค ส่วนผลการตรวจวิเคราะห์ที่มีค่าต่ำมักถูกนำไปใช้พยากรณ์ว่าไม่เป็นโรค แต่จากตัวแบบต้นไม้ตัดสินใจตัวแบบที่ 3 ที่ทางผู้วิจัยได้นำเสนอพบว่า ADA ที่มีค่าสูง ถูกพยากรณ์ว่าไม่เป็นโรค ส่วนค่า ADA ที่มีค่านั้นถูกพยากรณ์ว่าเป็นโรค ซึ่งเกิดจากความจำเพาะของพารามิเตอร์นั้นๆ เช่น พารามิเตอร์ CEA เป็นสารบ่งชี้มะเร็งที่มีความจำเพาะกับเซลล์มะเร็งสูง จึงมักนำมาเป็นพารามิเตอร์แรกในการคัดกรองเหมือนตัวแบบต้นไม้ตัดสินใจตัวแบบที่ 3 ส่วน ADA เป็นพารามิเตอร์ที่ใช้ในการคัดกรองวัณโรคเชื้อหุ้มปอด จึงจำเพาะกับวัณโรคมากกว่า ซึ่งสอดคล้องกับงานวิจัยของคุณปญญาภา⁽¹⁹⁾ ที่พบว่าสามารถใช้การตรวจระดับของ ADA ในน้ำเชื้อหุ้มปอด ที่จุดตัด 38.60 U/L ในการแยกกลุ่มที่เป็นวัณโรคเชื้อหุ้มปอดและไม่เป็นวัณโรคเชื้อหุ้มปอดได้ในตัวแบบต้นไม้ตัดสินใจตัวแบบที่ 3 ค่า ADA มากกว่า 30.80 และ 44.30 U/L จึงพยากรณ์ว่าไม่เป็นโรค ซึ่งหากมีการนำพารามิเตอร์ในกลุ่มโรคหรือพยาธิสภาพที่เกี่ยวข้องกับน้ำในเชื้อหุ้มปอดมาร่วมในตัวแบบต้นไม้ตัดสินใจอาจสามารถเพิ่มความไวและความจำเพาะได้มากขึ้น

ข้อเสนอแนะ

จากการศึกษาในครั้งนี้ ห้องปฏิบัติการทางการแพทย์สามารถนำตัวแบบต้นไม้ตัดสินใจที่ได้ ไปใช้ประโยชน์ในการคัดกรองหาเซลล์มะเร็งจากน้ำในช่องเชื้อหุ้มปอดได้ แต่ตัวแบบต้นไม้ตัดสินใจในงานวิจัยนี้ยังให้ค่าความจำเพาะไม่สูงมาก ซึ่งเกิดจากกลุ่มประชากรที่นำมาทดสอบ เป็นเพียงการศึกษาจากข้อมูลในโรงพยาบาลเพียงแห่งเดียว และมีข้อจำกัดจากกลุ่มประชากรที่ใช้ในการศึกษาครั้งนี้คือไม่ได้ทำการแยกโรคที่อาจจะเกิดร่วมกับการพบเซลล์มะเร็งจากน้ำในช่องเชื้อหุ้มปอด เช่น วัณโรค การติดเชื้อแบคทีเรีย การอักเสบ ซึ่งอาจให้ผลในบางพารามิเตอร์ไม่ชัดเจนเท่าที่ควร จึงควรมีการศึกษาจำแนกผู้ป่วยและควรศึกษาในกลุ่มประชากรที่หลากหลาย เพื่อเพิ่มความจำเพาะของการทดสอบ ทั้งนี้เนื่องจากความรู้ที่ได้จากการสร้างตัวแบบต้นไม้ตัดสินใจนี้ จะสามารถใช้เป็นต้นแบบ

เพื่อศึกษาความสัมพันธ์ระหว่างผลตรวจทางห้องปฏิบัติการกับการวินิจฉัยโรคอื่น ๆ ต่อไปในอนาคตได้

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณผู้อำนวยการ คณะกรรมการจริยธรรมการวิจัยและเจ้าหน้าที่ห้องปฏิบัติการสถาบันโรคทรวงอก ที่ให้ความอนุเคราะห์ในการเก็บรวบรวมข้อมูลในการวิจัยครั้งนี้ ขอขอบคุณ ผศ.ดร.กิตติศักดิ์ แถวนาชม ที่ให้ข้อเสนอแนะเรื่องการสร้างตัวแบบต้นไม้ตัดสินใจและการใช้โปรแกรม WEKA ขอขอบคุณหน่วยวิจัยด้านภูมิคุ้มกัน วิทยาระดับเซลล์และโมเลกุล คณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร และขอบคุณทุนอุดหนุนการวิจัยจากคณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร

เอกสารอ้างอิง

1. Thaikerd P, Bhummichitra K, Sanchaisuriya K, Kitcharoen K, Srivorakun H, Khemtonglang N. Seros Fluid Parameters Generated on the Sysmex XN-1000 in Malignant and Non-malignant Samples. J Med Tech Assoc Thailand. 2023;51(1):8422-33 (in thai)
2. Labaere D, Boeckx N, Geerts I, Moens M, Van den Driessche M. Detection of malignant cells in serous body fluids by counting high-fluorescent cells on the Sysmex XN-2000 hematology analyzer. Int J Lab Hematol. 2015;37(5):715-22.
3. Wong-Arteta J, Gil-Rodriguez E, Cabezon-Vicente R, Bereciartua-Urbieto E, Bujanda L. High fluorescence cell count in pleural fluids for malignant effusion screening. Clin Chim Acta. 2019;499:115-7.
4. Buoro S, Mecca T, Azzara G, Seghezzi M, Candiago E, Gianatti A, et al. Mindray BC-6800 body fluid mode, performance of nucleated cells, and differential count in ascitic and pleural fluids. Int J Lab Hematol. 2016;38(1):90-101.
5. Rastogi L, Dass J, Arya V, Kotwal J. Evaluation of high-fluorescence body fluid (HF-BF) parameter as a screening tool of malignancy in body fluids. Indian J Pathol Microbiol. 2019;62(4):572-7.

6. Xu W, Yu Q, Xie L, Chen B, Zhang L. Evaluation of Sysmex XN-1000 hematology analyzer for cell count and screening of malignant cells of serous cavity effusion. *Medicine (Baltimore)*. 2017;96(27):e7433.
7. Chian CF, Wu FP, Tsai CL, Peng CK, Shen CH, Perng WC, et al. Echogenic swirling pattern, carcinoembryonic antigen, and lactate dehydrogenase in the diagnosis of malignant pleural effusion. *Sci Rep*. 2022;12(1):4077.
8. Ai T, Tabe Y, Takemura H, Kimura K, Takahashi T, Yang H, et al. Novel flowcytometry-based approach of malignant cell detection in body fluids using an automated hematology analyzer. *PLoS One*. 2018;13(2):e0190886.
9. Zhang H, Li C, Hu F, Zhang X, Shen Y, Chen Y, et al. Auxiliary diagnostic value of tumor biomarkers in pleural fluid for lung cancer-associated malignant pleural effusion. *Respir Res*. 2020;21(1):284.
10. Li H, Huang L, Tang H, Zhong N, He J. Pleural fluid carcinoembryonic antigen as a biomarker for the discrimination of tumor-related pleural effusion. *Clin Respir J*. 2017;11(6):881-6.
11. Radjenovic-Petkovic T, Pejic T, Nastasijević-Borovac D, Rancic M, Radojkovic D, Radojkovic M, et al. Diagnostic value of CEA in pleural fluid for differential diagnosis of benign and malign pleural effusion. *Med Arh*. 2009;63:141-2.
12. Palwisut P. Improving decision tree technique in imbalanced data sets using SMOTE for internet addiction disorder data. *Information Technology Journal*. 2016;12(1):54-63(in thai).
13. Pongsanguan W, Thinsungnoen T, Thinsungnoen M. Development of model for diabetes mellitus using decision tree technique. *Journal of Science and Technology, Rajabhat Maha Sarakham University*. 2018;1(1):1-8.
14. Makond B, Pornsawad P, Thawnashom K. Decision Tree Modeling for Osteoporosis Screening in Postmenopausal Thai Women. *Informatics*. 2022;9(4):83.
15. Banluesapy S, Jirapanthong W. Towards machine learning algorithm for screening prediction of COVID-19 patients. *JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY*. 2022;12(1):47-60.
16. Giotta M, Trerotoli P, Palmieri VO, Passerini F, Portincasa P, Dargenio I, et al. Application of a decision tree model to predict the outcome of non-intensive inpatients hospitalized for COVID-19. *Int J Environ Res Public Health*. 2022;19(20).
17. Feng M, Zhu J, Liang L, Zeng N, Wu Y, Wan C, et al. Diagnostic value of tumor markers for lung adenocarcinoma-associated malignant pleural effusion: a validation study and meta-analysis. *Int J Clin Oncol*. 2017;22(2):283-90.
18. Cho YU, Chi HS, Park SH, Jang S, Kim YJ, Park CJ. Body fluid cellular analysis using the Sysmex XN-2000 automatic hematology analyzer: focusing on malignant samples. *Int J Lab Hematol*. 2015;37(3):346-56.
19. Tanwarawutthikul P. Optimum and effectiveness of adenosine deaminase level for pleural tuberculosis screening. *Journal of Medical and Public Health Region 4*. 2023;13(3):45-58.