วารสารการแพทย์และสาธารณสุขเขต 4
Office of Disease Prevention and Control, Region 4 Saraburi
https://he01.tci-thaijo.org/index.php/JMPH4/index

การพยากรณ์การติดเชื้อแบคทีเรียแกรมลบดื้อยาหลายขนานในโรงพยาบาลระดับชุมชน ด้วยอัลกอริทึม eXtreme Gradient Boosting

Prediction of multidrug-resistant gram-negative bacterial infections in community hospitals using the eXtreme Gradient Boosting algorithm

ปรีดา เม่งศิริ[1], รัชฎาภรณ์ อึ้งเจริญ[1,2]*, จินดานุช ผลโยน[3], เสฏฐวิทย์ เกิดผล[4]

Preeda Mengsiri[1], Ratchadaporn Ungcharoen[1,2], Jindanoot Phonyon[3], Sethavidh Gertphol[4]*

[1]หลักสูตรวิทยาศาสตร์ข้อมูลชีวการแพทย์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์, [2]คณะสาธารณสุขศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตเฉลิมพระเกียรติจังหวัดสกลนคร, [3]โรงพยาบาลสมเด็จพระยุพราชธาตุพนม จังหวัดนครพนม,[4]ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

[1]*Biomedical Data Science Program, Faculty of Science, Kasetsart University,* [2]*Faculty of Public Health, Kasetsart University Chalermphrakiat Sakhon Nakhon Province Campus,* [3]*Thatphanom Crown Prince Hospital, Nakhon Phanom Province,* [4]*Department of Computer Science, Faculty of Science, Kasetsart University*

**Corresponding Author: ratchadaporn.un@ku.th*

Received: August 20, 2024 Revised: April 11, 2025 Accepted: May 15, 2025

บทคัดย่อ

การดื้อยาหลายขนานของเชื้อแบคทีเรียแกรมลบ (Multidrug-resistant Gram-negative bacteria, MDR-GNB) ถือเป็นปัญหาทางด้านสาธารณสุขระดับโลก การศึกษานี้มีวัตถุประสงค์เพื่อพยากรณ์การติดเชื้อดื้อยาหลายขนานสำหรับแบคทีเรียแกรมลบ (MDR-GNB) จากสิ่งส่งตรวจโลหิตในโรงพยาบาลสมเด็จพระยุพราชธาตุพนม จังหวัดนครพนม ระหว่างเดือนมกราคม พ.ศ. 2559 ถึง ธันวาคม พ.ศ. 2563 โดยใช้อัลกอริทึม eXtreme Gradient Boosting การศึกษานี้เป็นการศึกษาแบบย้อนหลัง (Retrospective study) โดยมีจำนวนกลุ่มตัวอย่างทั้งหมด 624 ตัวอย่าง พบว่าผู้ป่วยส่วนใหญ่ที่มีการติดเชื้อ MDR-GNB มีอายุมากกว่า 70 ปี คิดเป็นร้อยละ 38 อายุเฉลี่ย 65 ปี (S.D. = 18.42) เชื้อแบคทีเรียที่พบมากที่สุด ได้แก่ Escherichia coli ร้อยละ 63.02 และ Klebsiella pneumoniae ร้อยละ 11.46 การสร้างแบบจำลองการเรียนรู้ของเครื่องด้วยวิธีการ eXtreme Gradient Boosting สามารถจำแนกผู้ป่วยที่ติดเชื้อ MDR-GNB ได้ร้อยละ 71 และมีค่า AUC ร้อยละ 66 และจากการวิเคราะห์ด้วย Logistic regression พบว่าผู้ป่วยที่ติดเชื้อ Escherichia coli มีความเสี่ยงต่อการติดเชื้อ MDR-GNB (OR_{adj} 3.13, 95% CI 2.17–4.53, p-value < 0.001) และมีความเสี่ยงสูงมากเมื่อเป็นเชื้อกลุ่ม Enterobacteriaceae ที่ผลิต Extended spectrum beta-lactamase (OR_{adj} 19.88, 95% CI 5.70–69.27, p-value < 0.001) จากผลการศึกษาพบว่าประสิทธิภาพโดยรวมของแบบจำลอง XGBoost ให้ค่า Sensitivity, Specificity, และ AUC ที่สูงกว่าร้อยละ 60 ซึ่งสะท้อนถึงความสามารถในการจำแนกที่ดีกว่าด้วยแบบจำลอง Logistic Regression จากผลการศึกษาสามารถนำแบบจำลองที่สร้างขึ้นไปใช้เพื่อเฝ้าระวังการติดเชื้อในผู้ป่วยในโรงพยาบาลเพื่อควบคุมการแพร่กระจายของเชื้อแบคทีเรียแกรมลบที่ดื้อยาหลายขนานได้อย่างมีประสิทธิภาพมากขึ้น

คำสำคัญ: การดื้อยาหลายขนาน, การเรียนรู้ของเครื่อง, แบคทีเรียแกรมลบ, XGBoost

Abstract

Multidrug-resistant Gram-negative bacteria (MDR-GNB) constitute a pressing global public health concern. The primary objective of this study was to develop a predictive model for identifying multidrug-resistant Gram-negative bacterial (MDR-GNB) infections using blood samples at Thatphanom Crown Prince Hospital, Nakhon Phanom Province, between January 2016 and December 2020, using the eXtreme Gradient Boosting algorithm. This retrospective study included a total of 624 samples. The findings revealed that the majority of patients with MDR-GNB infections were over 70 years old, accounting for 38%, with an average age of 65 years (S.D. = 18.42). The most frequently identified bacteria were Escherichia coli (63.02%) and Klebsiella pneumoniae (11.46%). The machine learning model developed using the eXtreme Gradient Boosting method demonstrated an ability to accurately classify patients with MDR-GNB infections with 71% accuracy and 66% of AUC. Additionally, logistic regression analysis identified that patients infected with Escherichia coli had a higher risk of MDR-GNB infection (adjusted OR 3.44, 95% CI 2.41 - 4.91, p-value < 0.001), with a significantly higher risk observed in cases involving Enterobacteriaceae producing extended-spectrum beta-lactamases (ESBL) (adjusted OR 19.88, 95% CI 5.70 - 69.27, p-value < 0.001). Based on the study results, the overall performance of the XGBoost model showed Sensitivity, Specificity, and AUC values higher than 60%, reflecting its superior discriminatory ability compared to the logistic regression model. The results suggest that the developed model can be effectively used for surveillance among patients in the hospital setting, thereby improving the control and prevention of the spread of MDR-GNB infections.

Keywords: Multidrug-resistant, Machine learning, Gram-negative bacteria**,** XGBoost

บทนำ

การดื้อยาต้านจุลชีพ (Antimicrobial resistance, AMR) ถือเป็นปัญหาหลักในการรักษาทางการแพทย์[1] โดยองค์การอนามัยโลก (WHO) ได้ระบุว่าทุกประเทศทั่วโลกกำลังเข้าสู่ยุคสมัยที่ยาปฏิชีวนะใช้ไม่ได้ผลต่อเชื้อดื้อยา (post-antibiotic era) นั่นหมายความว่าการติดเชื้อเพียงเล็กน้อยอาจนำไปสู่การเสียชีวิตได้[2] โดยสาเหตุของการเกิดเชื้อดื้อยานั้นมาจากหลายสาเหตุ ได้แก่ การใช้ยาปฏิชีวนะไม่ถูกต้อง หรือใช้ยาอย่างพร่ำเพรื่อ ไม่สมเหตุสมผล[3] รวมถึงการขายยาปฏิชีวนะที่ไม่เหมาะสมที่ไม่ต้องสั่งโดยแพทย์ สุขอนามัยที่ไม่เหมาะสม และการย่อยสลายยาปฏิชีวนะไม่สมบูรณ์ทำให้ส่วนหนึ่งของยาเหล่านั้นกับสารตกค้างถูกขับถ่ายมาพร้อมอุจาระ[4] โดยกลไกการดื้อยาของแบคทีเรียเกิดจากยีนที่ส่งต่อการดื้อยาที่อยู่บนโครโมโซมหรือพลาสมิดของแบคทีเรีย กระบวนการนี้จะเรียกว่า การถ่ายโอนยีนดื้อยา (horizontal gene transfer, HGT) หรือแบคทีเรียมีการปรับตัวต่อยาปฏิชีวนะ เพื่อความอยู่รอด[5] ยีนดื้อยาดังกล่าวจะทำหน้าที่สร้างเอนไซม์ขึ้นมาเพื่อทำลายยา และยีนเหล่านี้สามารถถ่ายทอดในสายพันธุ์เดียวกันหรือถ่ายทอดข้ามสายพันธุ์ได้[6, 7] จากการศึกษาที่ผ่านมารายงานว่าทั่วโลกมีผู้เสียชีวิตจากการติดเชื้อดื้อยาหลายขนาน (Multi-drug resistant, MDR) ประมาณปีละ 700,000 คนต่อปี และคาดการณ์ว่าภายในปี พ.ศ. 2593 จะมีผู้เสียชีวิตเพิ่มขึ้นเป็น 10 ล้านคนต่อปี[8] โดยคิดเป็นความเสียหายทางเศรษฐกิจ 3.8 ของ GPD และผลที่ตามมาคือมีผู้คนถึง 28.3 ล้านคนที่ต้องเผชิญกับความยากจนขั้นรุนแรง[9] สำหรับประเทศไทย มีรายผู้ป่วยติดเชื้อที่เข้ารักษาในโรงพยาบาล 1,023 แห่ง พบว่า 1 ใน 3 ของผู้ป่วยติดเชื้อที่เข้ารับการรักษาในโรงพยาบาลมีการติดเชื้อดื้อยาหลายขนานซึ่งเชื้อก่อโรคส่วนใหญ่เป็นเชื้อแบคทีเรียแกรมลบได้แก่ *Acinetobacter baumannii* ร้อยละ 57.6 และ *Pseudomonas aeruginosa* ประมาณร้อยละ 30[10]

โรงพยาบาลชุมชนถือเป็นหน่วยบริการด้านสาธารณสุขที่มีความสำคัญในระบบสุขภาพของประเทศไทย โดยมีการจัดตั้งอยู่ในทุกอำเภอทั่วประเทศ รวมจำนวนทั้งสิ้น 711 แห่ง โรงพยาบาลเหล่านี้มีบทบาทสำคัญในการให้บริการดูแลสุขภาพแก่ประชาชนจำนวนมากในแต่ละวัน

อีกทั้งยังทำหน้าที่รับผู้ป่วยที่ส่งต่อมาจากโรงพยาบาลศูนย์และโรงพยาบาลทั่วไปเพื่อการดูแลต่อเนื่อง[11] ในปัจจุบันโรงพยาบาลที่ทำหน้าที่เป็นศูนย์วิจัยได้รับข้อมูลเกี่ยวกับผลการเพาะเชื้อและผลความไวต่อยาปฏิชีวนะจากการส่งสิ่งส่งตรวจไปยังโรงพยาบาลทั่วไปขนาดใหญ่ในระดับจังหวัด อย่างไรก็ตาม การรายงานข้อมูลดังกล่าวยังคงดำเนินการผ่านระบบออนไลน์ในรูปแบบของภาพถ่ายหรือไฟล์ภาพ ซึ่งส่งผลให้การรวบรวมข้อมูลในระดับภาพรวมของโรงพยาบาลไม่สมบูรณ์เท่าที่ควร นอกจากนี้การติดตามข้อมูลการติดเชื้อดื้อยาหลายขนานมักเป็นเพียงการสุ่มเก็บข้อมูลความชุกของเชื้อ ภายใต้กรอบแผนการควบคุมโรคติดเชื้อ ทำให้ไม่สามารถแสดงแนวโน้มของการติดเชื้อดื้อยาได้อย่างชัดเจน อีกทั้งปริมาณข้อมูลที่รายงานเกี่ยวกับการติดเชื้อดื้อยายังมีข้อจำกัดและไม่สอดคล้องกับการใช้ยาปฏิชีวนะในทางปฏิบัติ โดยเฉพาะอย่างยิ่งในกลุ่มยาที่ต้องควบคุมการใช้ ซึ่งจำเป็นต้องมีการประเมินและติดตามอย่างเคร่งครัด

โรงพยาบาลสมเด็จพระยุพราชธาตุพนม จังหวัดนครพนม ซึ่งเป็นสถานที่ทำการศึกษาวิจัย พบว่าอุบัติการของการติดเชื้อดื้อยาหลายขนานตั้งแต่ปี พ.ศ. 2559 ถึง พ.ศ. 2563 มีจำนวนเพิ่มสูงขึ้นร้อยละ 24.83, 21.64, 22.07, 38.08 และ 30.86 ของสิ่งส่งตรวจที่พบการติดเชื้อแบคทีเรียแกรมลบ[12]

การดื้อยาต้านจุลต้านที่ชีพที่เกิดขึ้นในเชื้อแบคทีเรียเกิดจากการปรับตัวต่อยาปฏิชีวนะที่ได้รับ เพื่อความอยู่รอด การดื้อยาในเชื้อแบคทีเรียสามารถแบ่งออกได้เป็น 2 กลุ่มใหญ่ ๆ คือ การดื้อยาโดยธรรมชาติของแบคทีเรียเอง (intrinsic resistance) แบคทีเรียทุกสายพันธุ์มีความสามารถในการต้านทานต่อยาปฏิชีวนะ[13] และการดื้อยาต้านจุลชีพของแบคทีเรียที่เกิดภายหลัง (acquired resistance) เป็นผลมาจากการปรับตัวและการถ่ายทอดทางพันธุกรรมของแบคทีเรียเพื่อสร้างกลไกการต้านทานต่อยาปฏิชีวนะ ซึ่งมีสาเหตุหลักมาจาก 3 ปัจจัย ได้แก่ 1) การกลายพันธุ์ (mutation) เป็นการเปลี่ยนแปลงลำดับเบสของดีเอ็นเอในเซลล์แบคทีเรีย ทำให้สามารถสร้างกลไกการดื้อยาต่อยาปฏิชีวนะได้ 2) การถ่ายยีนในแนวราบ (Horizontal Gene Transfer) เช่น การถ่ายทอดพลาสมิด การแลกเปลี่ยนดีเอ็นเอระหว่างแบคทีเรียด้วยกระบวนการคอนจูเกชัน เป็นต้น ซึ่งเป็นกระบวนการส่งต่อความสามารถในการดื้อยาของเชื้อแบคทีเรีย การติดเชื้อดื้อยานอกจากจะส่งผลกระทบต่อผู้ป่วยโดยตรงแล้วยังเป็นการเพิ่มค่าใช้จ่ายในการรักษาให้กับผู้ป่วย และโรงพยาบาล[14] จากรายงานของศูนย์เฝ้าระวังการดื้อยาต้านจุลชีพแห่งชาติ พบการติดเชื้อดื้อยาจำนวน 87,751 ครั้ง และในจำนวนนี้มีผู้เสียชีวิต 38,481 คน ระยะเวลาการเข้ารับการรักษาในโรงพยาบาลเพิ่มขึ้นเฉลี่ย 46 วัน ค่าใช้จ่ายเกี่ยวกับยาต้านจุลชีพเพิ่มขึ้นประมาณ 6 พันล้านบาท[15]

ดังนั้นการระบุผู้ป่วยที่มีความเสี่ยงต่อการติดเชื้อดื้อยาหลายขนานก่อนตรวจพบการติดเชื้อถือเป็นสิ่งสำคัญ เนื่องจากช่วยป้องกันการแพร่กระจายเชื้อ บรรเทาภาวะวิกฤต ป้องกันภาวะแทรกซ้อน ฟื้นฟูสภาพ ส่งเสริมสุขภาพ และวางแผนการจำหน่ายอย่างถูกต้องและรวดเร็ว ทำให้ผู้ป่วยมีโอกาสหายขาดจากโรคและลดภาวะแทรกซ้อน การศึกษานี้จึงนำเสนอที่จะพัฒนาแบบจำลองการเรียนรู้ของเครื่อง eXtreme Gradient Boosting (XGBoost) เพื่อจำแนกผู้ป่วยกลุ่มเสี่ยงต่อการติดเชื้อแบคทีเรียแกรมลบที่ดื้อยาหลายขนานในโรงพยาบาลระดับชุมชน เนื่องจาก XGBoost เป็นอัลกอริทึมที่มีประสิทธิภาพสูงและยังจัดการกับข้อมูลที่มีความซับซ้อนได้ดีอีกด้วย ซึ่งเหมาะสมสำหรับการนำมาประยุกต์การจำแนกผู้ป่วยที่ติดเชื้อดื้อยาหลายขนานในโรงพยาบาล[16-19]

วัตถุประสงค์

เพื่อพยากรณ์การติดเชื้อดื้อยาหลายขนานสำหรับแบคทีเรียแกรมลบ (MDR-GNB) จากสิ่งส่งตรวจโลหิตโดยใช้อัลกอริทึม eXtreme Gradient Boosting

วัสดุและวิธีการศึกษา

การศึกษานี้เป็นการเก็บข้อมูลย้อนหลัง (Retrospective study) ในระบบ HOSxP โรงพยาบาลสมเด็จพระยุพราชธาตุพนม จังหวัดนครพนม

ประชากรที่ศึกษา คือ ผู้ป่วยที่เข้ารับการรักษาในแผนกผู้ป่วยในและผู้ป่วยนอกระหว่างเดือนมกราคม พ.ศ. 2559 ถึง ธันวาคม พ.ศ. 2563 มีรายงานการเพาะเชื้อจุลชีพ 4,480 ไอโซเลทในผู้ป่วยของโรงพยาบาลซึ่งเป็น

สถานที่วิจัย เมื่อตัดไอโซเลทที่มีข้อมูลซ้ำและพบเป็นเชื้อเดิมภายใน 1 เดือน เหลือ 1,860 ไอโซเลท เนื่องจากการศึกษานี้วิเคราะห์เฉพาะข้อมูลการเพาะเชื้อจากสิ่งส่งตรวจที่เป็นโลหิตและเชื้อแบคทีเรียแกรมลบ (GNB) โดยมีเกณฑ์การคัดเข้า (Inclusion criteria) ได้แก่ ผู้ป่วยที่มีข้อมูลครบถ้วนทั้งข้อมูลผลการเพาะเชื้อแบคทีเรียและทดสอบความไวต่อยาสำหรับสิ่งส่งตรวจที่เป็นโลหิต ผู้ป่วยที่เข้านอนโรงพยาบาลนานกว่า 24 ชั่วโมง โดยในบันทึกข้อมูลการดื้อยาเป็นผลรายงานครั้งแรกของผู้ป่วย และเกณฑ์การคัดออก (Exclusion criteria) ข้อมูลสิ่งส่งตรวจที่เป็นข้อมูลซ้ำ และข้อมูลผลการเพาะเชื้อที่พบว่าเป็นของผู้ป่วยรายเดิมที่มีผลรายงานเป็นเชื้อเดิมซ้ำภายใน 1 เดือน ข้อมูลครั้งที่ 2 จะถูกคัดออก โดยตัวอย่างสุดท้ายมีจำนวนทั้งสิ้น 624 ตัวอย่าง ทั้งนี้เพราะโลหิตเป็นสิ่งส่งตรวจที่แสดงการติดเชื้อทุกระบบในร่างกายและการติดเชื้อในกระแสโลหิตเป็นสาเหตุการตายลำดับที่ 3 ของโรงพยาบาลซึ่งเป็นสถานที่วิจัย

กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้ถูกคำนวณขนาด โดยกำหนดอัตราส่วนระหว่างกลุ่มศึกษา (case) และกลุ่มควบคุม (control) เท่ากับ 1:2 และใช้ตัวแปรที่ต้องการศึกษา คือ การสัมผัสยาต้านจุลชีพมาก่อนหน้าในช่วง 90 วันที่ผ่านมา โดยมีข้อมูลพื้นฐานดังนี้:

P1: สัดส่วนของกลุ่มศึกษา (case) ที่สัมผัสยาต้านจุลชีพในช่วง 90 วัน มีค่าเท่ากับ 0.62

P2: สัดส่วนของกลุ่มควบคุม (control) ที่สัมผัสยาต้านจุลชีพในช่วง 90 วัน มีค่าเท่ากับ 0.38

N1: จำนวนตัวอย่างขั้นต่ำในกลุ่มศึกษา เท่ากับ 98 คน

N2: จำนวนตัวอย่างขั้นต่ำในกลุ่มควบคุม เท่ากับ 196 คน

จากอัตราส่วนที่กำหนดไว้ (case : control = 1 : 2) ขนาดตัวอย่างขั้นต่ำที่กำหนดจึงประกอบด้วย กลุ่มศึกษาจำนวนอย่างน้อย 98 คน และกลุ่มควบคุมอย่างน้อย 196 คน

ในกระบวนการคัดเลือกประชากรในการศึกษา กลุ่มศึกษา (case) ประกอบด้วยผู้ป่วยที่มีผลเพาะเชื้อจากแบคทีเรียแกรมลบที่ดื้อยาหลายขนาน มีจำนวน 192 คน และกลุ่มควบคุม (control) ประกอบด้วยผู้ป่วยที่มีผลเพาะเชื้อจากแบคทีเรียแกรมลบที่ไม่ดื้อยาหลายขนาน มีจำนวน 432 คน

เมื่อคำนวณอัตราส่วน N1:N2 พบว่าเท่ากับ 1:2 หรือ 192:384 ซึ่งใกล้เคียงกับอัตราส่วนที่กำหนดไว้ ผู้วิจัยจึงใช้ข้อมูลจากผู้ป่วยทั้งหมดในกลุ่มศึกษาและกลุ่มควบคุมตามจำนวนดังกล่าว รวมตัวอย่างทั้งสิ้น 624 คน ในการศึกษา

ข้อมูลผลเพาะเชื้อของทุกสิ่งส่งตรวจและข้อมูลของผู้ป่วยอยู่ในฐานข้อมูล MLAB (version 5.58.22.02.28) ที่มีการตรวจสอบระบบเป็นประจำทุกปี ข้อมูลผลเพาะเชื้อประกอบด้วย วันที่ส่งตรวจ หมายเลขตรวจวิเคราะห์ หมายเลขประจำตัวผู้ป่วย ชื่อ-นามสกุล อายุ เพศ ชื่อโรงพยาบาลที่ส่งตรวจ ประเภทสิ่งส่งตรวจ เชื้อก่อโรคที่ตรวจพบ ผลทดสอบความไวต่อยาปฏิชีวนะที่มีการแปลผลตาม CLSI ที่รายงานผลเป็น susceptible, intermediate และ resistance ผลการตรวจวิเคราะห์ที่ได้รับจะถูกตรวจสอบกับข้อมูลสิ่งส่งตรวจที่ได้ลงทะเบียนไว้ในระบบ Hospital Information System (HIS) ในโปรแกรม BMS-HOSxP XE เวอร์ชั่น 23 พฤษภาคม 2567โดยตรวจสอบในประเด็นวันที่ส่งตรวจ หมายเลขประจำตัวผู้ป่วย ชื่อ-สกุล อายุ เพศ สิ่งส่งตรวจ และชื่อโรงพยาบาล เพื่อให้มั่นใจว่าเป็นข้อมูลของผู้ป่วยจากโรงพยาบาลที่เป็นสถานที่วิจัยจริง ข้อมูลปริมาณการใช้ยาปฏิชีวนะในการวิจัยเป็นยาที่ใช้รักษาโรคติดเชื้อในระหว่างวันที่ 1 มกราคม พ.ศ. 2559 - 31 มกราคม พ.ศ. 2563 ที่ได้จากโปรแกรม BMS-HOSxP XE

การคัดเลือกปัจจัยเสี่ยงสำหรับการสร้างแบบจำลองในงานวิจัยนี้ประกอบด้วย 2 ส่วน ซึ่งแยกกันสำหรับแบบจำลอง Logistic Regression และแบบจำลอง XGBoost ดังนี้

ส่วนแรกแบบจำลองสถิติ Logistic Regression ปัจจัยเสี่ยงที่ใช้สร้างแบบจำลองจะคัดเลือกมาจากปัจจัยเสี่ยงที่มีสหสัมพันธ์กับการติดเชื้อ MDR-GNB ที่ระดับนัยสำคัญน้อยกว่า 0.01 เพื่อคัดเลือกปัจจัยเสี่ยงที่มีผลกระทบต่อการติดเชื้อ MDR-GNB อย่างชัดเจน จากนั้นนำปัจจัยเหล่านี้มาสร้างแบบจำลองต่อไป

แบบจำลอง Logistic Regression เป็นแบบจำลองทางสถิติที่หาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามที่เป็นไปได้ 2 เหตุการณ์ โดยในการศึกษานี้กำหนดให้ 0 เท่ากับผู้ป่วยที่ไม่พบการติดเชื้อแบคทีเรียแกรมลบที่ดื้อยาหลายขนาน และ 1 เท่ากับผู้ป่วยที่ตรวจสอบ

การติดเชื้อแบคทีเรียแกรมลบที่ดื้อยาหลายขนาน โดยสามารถเขียนฟังก์ชันความน่าจะเป็นของการเกิดเหตุการณ์หรือการติดเชื้อดื้อยา P(event) ได้ดังสมการที่ (1)

$$P(\text{event}) = \frac{1}{1+e^{-X}} \quad (1)$$

สำหรับความน่าจะเป็นของการไม่ติดเชื้อดื้อยาหลายขนาน P(no event) จะสามารถเขียนสมการความน่าจะเป็นได้ดังสมการที่ (2)

$$P(\text{no event}) = 1 - \frac{1}{1+e^{-X}} \quad (2)$$

ถ้าชุดข้อมูลมีตัวแปรอิสระ n ตัว เมื่อ X เป็นผลรวมเชิงเส้นของตัวแปรต่าง ๆ ได้ตามสมการที่ (3)

$$X = b_0 + b_1x_1 + b_1x_1 + \ldots + b_nx_n \quad (3)$$

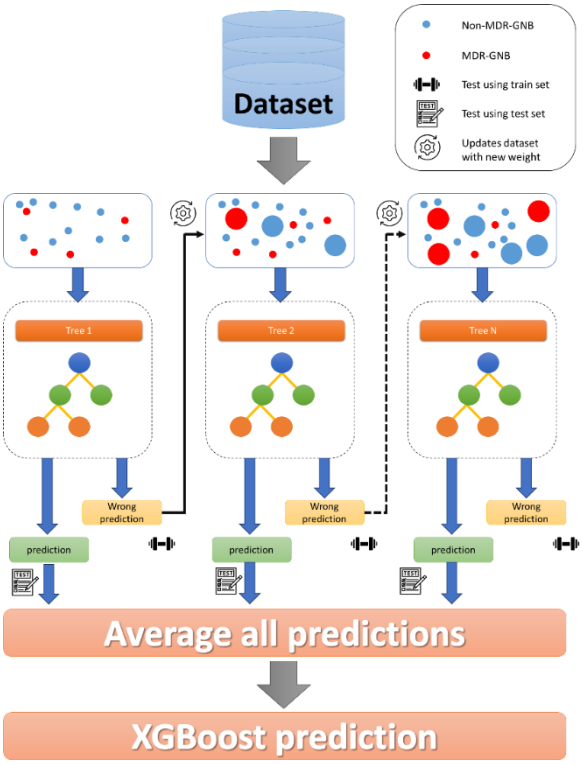
เมื่อ x คือ ตัวแปรอิสระทั้งหมด n ตัวแปร

b คือ สัมประสิทธิ์สำหรับแต่ละตัวแปร

แปลงสมการให้อยู่ในรูปแบบของสมการเชิงเส้น ซึ่งเรียกว่า logit transformation จะได้ log of odds หรือ logit เพื่อคำนวณหาค่าสัมประสิทธิ์ b_i สำหรับแต่ละตัวแปรด้วยสมการที่ (4)

$$\ln\left[\frac{P(\text{event})}{P(\text{no event})}\right] = \log(\text{odds}) = x = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \quad (4)$$

ส่วนที่สองอัลกอริทึม XGBoost กระบวนการสร้างแบบจำลองเป็นไปตามรูปที่ 1 ถูกนำมาใช้ในการจำแนกผู้ป่วยที่มีการติดเชื้อ MDR-GNB โดย XGBoost ไม่ใช้ปัจจัยเสี่ยงที่คัดเลือกจากแบบจำลอง Logistic Regression แต่ประยุกต์ใช้เทคนิคการคัดเลือกปัจจัยเสี่ยง (Feature Selection) ด้วยวิธี Sequential Forward Selection (SFS) เป็นเทคนิคที่ใช้ในการเลือกตัวแปรสำคัญจากชุดข้อมูล โดยเริ่มจากตัวแปรหนึ่งตัวแล้วเพิ่มตัวแปรอื่น ๆ ทีละตัวจนกว่าจะได้ชุดตัวแปรที่ให้ผลลัพธ์ดีที่สุดในการพยากรณ์ ซึ่งมีประสิทธิภาพในการเลือกตัวแปรสำคัญ เพื่อลดความซับซ้อนและเพิ่มความแม่นยำในการจำแนกผู้ป่วย นอกจากนี้ ได้นำอัลกอริทึม Adaptive Synthetic Sampling (ADASYN) มาใช้ในการเพิ่มจำนวนตัวอย่างข้อมูลเพื่อแก้ปัญหาความไม่สมดุลของข้อมูลระหว่างกลุ่มผู้ป่วยที่ไม่มีการติดเชื้อ (Majority Class) และกลุ่มผู้ป่วยที่ติดเชื้อ (Minority Class) โดยอัลกอริทึมจะสังเคราะห์ตัวอย่างข้อมูลใหม่ในกลุ่ม Minority Class ตามความหนาแน่นของข้อมูลในพื้นที่ใกล้เคียง ทำให้ได้ชุดข้อมูลที่สมดุลมากยิ่งขึ้น[20] เพื่อให้ผลการจำแนกด้วย XGBoost ไม่เอนเอียงไปทางกลุ่มข้อมูลที่มีจำนวนมากกว่า ทั้งนี้เพื่อพัฒนาเครื่องมือช่วยตัดสินใจทางคลินิกที่น่าเชื่อถือ



รูปที่ 1 กระบวนการสร้างแบบจำลอง XGBoost

เนื่องจากการศึกษาในครั้งนี้ดำเนินการบนชุดข้อมูลขนาดเล็ก จึงใช้วิธี Stratified K-Fold cross-validator ในการการทดสอบความตรง (Internal validation) ของแบบจำลอง[21] เนื่องจากวิธีการนี้จะทำการรักษาสัดส่วนร้อยละของตัวอย่างในแต่ละคลาสสำหรับแต่ละชุดข้อมูลย่อย ๆ (Folds) จากนั้นทำการเฉลี่ยผลลัพธ์ของประสิทธิภาพของแบบจำลอง

การอธิบายแบบจำลองใช้ SHapley Additive exPlanations (SHAP) เพื่ออธิบายการทำงานของแบบจำลอง XGBoost และระบุความสำคัญของปัจจัยต่างๆ ที่มีผลต่อการพยากรณ์ วิธีการนี้ช่วยเพิ่มความน่าเชื่อถือของแบบจำลองในการคัดแยกผู้ป่วยที่ติดเชื้อ MDR-GNB

การพิทักษ์สิทธิ์กลุ่มตัวอย่าง การศึกษาวิจัยในครั้งนี้ได้ผ่านการรับรองจากคณะกรรมการจริยธรรมวิจัยมนุษย์ มหาวิทยาลัยเกษตรศาสตร์ หมายเลขการรับรอง เลขที่ KUREC-HSR67/012 วันที่อนุมัติ 7 พฤษภาคม พ.ศ. 2567

ผลการศึกษา

ผลการเพาะเชื้อพบ GNB จำนวน 624 ตัวอย่าง พบว่าเป็น GNB ที่ไม่ดื้อยาจำนวน 432 ตัวอย่าง คิดเป็นร้อยละ 69.23 และ GNB ที่ดื้อยาจำนวน 192 ตัวอย่าง คิดเป็นร้อยละ 30.77 โดยกำหนดให้กลุ่มที่ไม่มีการดื้อยาเป็นกลุ่มควบคุม

ผู้ป่วยที่มีผลเพาะเชื้อพบ MDR-GNB ส่วนใหญ่เป็นผู้มีอายุมากกว่า 70 ปี คิดเป็นร้อยละ 38 รองลงมาคืออายุ 50 - 59 ปี คิดเป็นร้อยละ 22.4 ผู้ป่วยส่วนใหญ่ที่มีการติดเชื้อ MDR-GNB เป็นเพศหญิง จำนวน 107 ราย คิดเป็นร้อยละ 55.70 มีค่ามัธยฐานของค่าอายุเท่ากับ 65 ปี (S.D. = 18.42)

การศึกษานี้ได้คัดเลือกชนิดของแบคทีเรียที่เป็นตัวก่อโรค ผลการเพาะเชื้อจากสิ่งส่งตรวจพบ *Escherichia coli (E. coli)*121 ตัวอย่าง คิดเป็นร้อยละ 63.02, *Klebsiella pneumoniae (K. pneumoniae)* 22 ตัวอย่าง คิดเป็นร้อยละ 11.46, *Burkholderia pseudomallei (B. pseudomallei)* 12 ตัวอย่าง คิดเป็นร้อยละ 6.25 และ *Acinetobacter baumannii* 4 ตัวอย่าง คิดเป็นร้อยละ 2.08

ผู้ป่วยส่วนใหญ่ที่ติดเชื้อ MDR-GNB เป็นการติดเชื้อจากชุมชน (Community-acquired) จำนวน 183 คน คิดเป็นร้อยละ 95.31 ในจำนวนนี้ มีผู้ป่วยบางส่วนที่เคยเข้ารักษาตัวในโรงพยาบาลภายใน 90 วันที่ผ่านมา จำนวน 63 คน คิดเป็นร้อยละ 32.81 และผู้ป่วยที่ไม่เคยได้รับยาปฏิชีวนะภายใน 90 วันที่ผ่านมา มีจำนวน 178 คน คิดเป็นร้อยละ 92.70 หอผู้ป่วยที่มีผู้ติดเชื้อมากที่สุด คือ หอผู้ป่วยหญิง จำนวน 84 คน คิดเป็นร้อยละ 43.80

ตารางที่ 1 ข้อมูลทั่วไปของผู้ป่วยที่มีผลการเพาะเชื้อจากสิ่งตรวจที่เป็นโลหิตพบแบคทีเรียแกรมลบ

ลักษณะ	การดื้อยาหลายขนาน		ลักษณะ	การดื้อยาหลายขนาน	
	ไม่พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)	พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)		ไม่พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)	พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)
N	432 (69.20)	192 (30.80)	**เพศ**		
อายุ, ปี			หญิง	218 (50.50)	107 (55.70)
0 - 49	120 (27.80)	40 (20.80)	ชาย	214 (49.50)	85 (44.30)
50 - 59	92 (21.30)	43 (22.40)	**การสูบบุหรี่**		
60 - 69	90 (20.80)	36 (18.80)	ไม่สูบ	319 (73.80)	119 (62.0)
≥70	130 (30.10)	73 (38.00)	สูบ	113 (26.20)	73 (38.0)
อาชีพ			**การดื่มสุรา**		
เกษตรกร	241 (55.80)	105 (54.70)	ไม่ดื่ม	259 (60.00)	102 (53.10)
ว่างงาน	123 (28.50)	54 (28.10)	ดื่ม	173 (40.00)	90 (46.90)
ลูกจ้าง	42 (9.70)	18 (9.40)	**ประวัติเจ็บป่วยในอดีต**		
ค้าขาย	11 (2.50)	4 (2.10)	วัณโรค (TB)	5 (1.16)	7 (3.65)
ข้าราชการ	10 (2.30)	9 (4.70)	**ปัจจัยด้านการใช้ยา**		
นักบวช	5 (1.20)	2 (1.00)	ประวัติการใช้ยาปฏิชีวนะในช่วง 90 วันที่ผ่านมา	16 (22.22)	14(22.95)

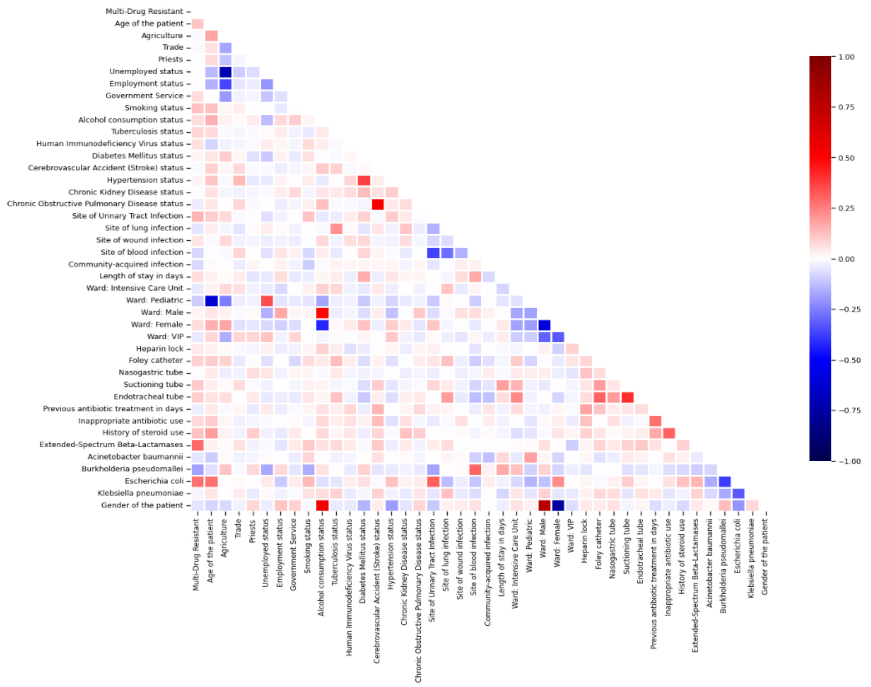
ตารางที่ 1 ข้อมูลทั่วไปของผู้ป่วยที่มีผลการเพาะเชื้อจากสิ่งตรวจที่เป็นโลหิตพบแบคทีเรียแกรมลบ (ต่อ)

ลักษณะ	การดื้อยาหลายขนาน		ลักษณะ	การดื้อยาหลายขนาน	
	ไม่พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)	พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)		ไม่พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)	พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)
โรคประจำตัว			ประวัติการใช้ยาปฏิชีวนะที่ไม่เหมาะสมในช่วง 90 วันที่ผ่านมา	12 (16.67)	11 (18.03)
โรคเบาหวาน (DM)	89 (47.09)	43 (48.86)	ประวัติการใช้สเตียรอยด์ในช่วง 90 วันที่ผ่านมา	44 (61.11)	36 (59.02)
ภาวะความดันโลหิตสูง (HT)	63 (33.33)	31 (35.23)	**ปัจจัยก่อโรค**		
โรคไตวายเรื้อรัง (CKD)	14 (7.41)	6 (6.82)	Escherichia coli	143 (46.28)	121 (64.71)
โรคปอดอุดกั้นเรื้อรัง (COPD)	14 (7.41)	2 (2.27)	Burkholderia pseudomallei	89 (28.80)	12 (6.42)
โรคหลอดเลือดสมอง (CVA)	6 (3.17)	2 (2.27)	Klebsiella pneumoniae	56 (18.12)	22 (11.76)
โรคเอดส์ (HIV)	3 (1.59)	4 (4.55)	Acinetobacter baumannii	18 (5.83)	4 (2.14)
ตำแหน่งที่พบการติดเชื้อ			Extended spectrum beta-lactamase (ESBL)-producing Enterobacteriaceae	3 (0.97)	28 (14.97)
ระบบทางเดินปัสสาวะ	57 (17.50)	48 (31.40)	**แหล่งที่พบการติดเชื้อ**		
ระบบทางเดินหายใจ	70 (21.50)	28 (18.30)	การติดเชื้อในชุมชน	423 (97.92)	183 (95.31)
บาดแผล	12 (3.70)	9 (5.90)	**ประวัติการรักษาในโรงพยาบาลในช่วง 90 วันที่ผ่านมา**	103 (23.84)	63 (32.81)
โลหิต	187 (57.40)	68 (44.40)	**การใช้เครื่องมือหรืออุปกรณ์ทางการแพทย์ในช่วง 90 วันที่ผ่านมา**		
ระยะเวลารักษาตัวในโรงพยาบาล			เฮพารินล็อค (Heparin lock)	92 (49.20)	48 (40.68)
1 - 3 วัน	132 (30.60)	59 (30.70)	สายสวนปัสสาวะ (Urinary catheter หรือ Foley catheter)	60 (32.09)	40 (33.90)

ตารางที่ 1 ข้อมูลทั่วไปของผู้ป่วยที่มีผลการเพาะเชื้อจากสิ่งตรวจที่เป็นโลหิตพบแบคทีเรียแกรมลบ (ต่อ)

ลักษณะ	การดื้อยาหลายขนาน		ลักษณะ	การดื้อยาหลายขนาน	
	ไม่พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)	พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)		ไม่พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)	พบการติดเชื้อดื้อยาหลายขนาน (ร้อยละ)
4 - 7 วัน	141 (32.60)	50 (26.00)	ท่อสายยางผ่านระหว่างรูจมูกลงไปยังกระเพาะอาหาร (Nasogastric tube)	14 (7.49)	6 (5.08)
8 - 14 วัน	91 (21.10)	44 (22.90)	สายดูดเสมหะ (Suction tube)	3 (1.60)	7 (5.93)
>14 วัน	68 (15.70)	39 (20.30)	ท่อช่วยหายใจ (Endotracheal tube)	18 (9.62)	17 (14.41)
หอผู้ป่วย					
หอผู้ป่วยหญิง	157 (36.30)	84 (43.80)			
หอผู้ป่วยชาย	154 (35.60)	73 (38.00)			
หอผู้ป่วยพิเศษ	67 (15.50)	24 (12.50)			
หอผู้ป่วยเด็ก	30 (6.90)	4 (2.10)			
หอผู้ป่วยหนัก	24 (5.60)	7 (3.60)			

จากตารางที่ 1 สามารถหาสหสัมพันธ์แบบเพียร์สันเพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรในชุดข้อมูลได้ จากนั้นสร้างแผนภูมิความร้อน (heatmap) ตามรูปที่ 2 หากค่าความสัมพันธ์ (correlation values) มีค่าเข้าใกล้ 1 (สีแดง) หมายความว่าตัวแปรทั้งสองมีค่าสัมพันธ์ไปในทิศทางที่เพิ่มขึ้น แต่ถ้าหากเข้าใกล้ -1 (สีน้ำเงิน) หมายความว่าตัวแปรทั้งมีความสัมพันธ์กันในเชิงลดลง ซึ่งโดยภาพรวมในชุดข้อมูลมีความสัมพันธ์กันค่อนข้างน้อย



รูปที่ 2 แผนภูมิความร้อนแสดงสหสัมพันธ์ระหว่างตัวแปรในชุดข้อมูล

ปัจจัยเสี่ยงสำหรับการติดเชื้อ MDR-GNB การวิเคราะห์ด้วย Univariate logistic regression พบปัจจัยที่มีความสัมพันธ์กับการติดเชื้อ MDR-GNB ดังนี้

ผู้ป่วยที่มีประวัติการสูบบุหรี่ (OR = 1.73, 95% CI: 1.20 - 2.48, p-value < 0.01) มีโอกาสเสี่ยงที่จะติดเชื้อดื้อยาเพิ่มขึ้น ผู้ที่มีประวัติการเป็นวัณโรค (OR = 3.23, 95% CI: 1.01 - 10.31, p-value < 0.05) ก็มีความเสี่ยงเพิ่มขึ้นเช่นกัน นอกจากนี้ การติดเชื้อที่ระบบทางเดินปัสสาวะ (OR = 2.19, 95% CI: 1.42 - 3.36, p-value < 0.001) และการใช้ยาสเตียรอยด์ในช่วง 90 วันที่ผ่านมา (OR = 2.03, 95% CI: 1.26 - 3.28, p-value < 0.01) เป็นปัจจัยเสี่ยงที่สำคัญ

การเข้ารักษาในโรงพยาบาลในระยะเวลา 90 วันที่ผ่านมา (OR = 1.55, 95% CI: 1.07 - 2.26, p-value < 0.05) ก็เป็นอีกปัจจัยหนึ่งที่มีความสัมพันธ์กับการติดเชื้อ MDR-GNB นอกจากนี้ ผู้ป่วยที่มีประวัติการใช้ Urinary catheter (OR = 1.63, 95% CI: 1.04 - 2.53, p-value < 0.05) การใช้ Endotracheal tube (OR = 2.23, 95% CI: 1.12 - 4.43, p-value < 0.05) และ Suction tube (OR = 5.41, 95% CI: 1.38 - 21.15, p-value < 0.05) ยังเป็นกลุ่มที่มีความเสี่ยงสูงต่อการติดเชื้อ

สำหรับปัจจัยเสี่ยงที่เกี่ยวข้องกับตัวก่อโรค พบว่า *E. coli* (OR = 3.44, 95% CI: 2.41 - 4.91, p-value < 0.001) และ *Enterobacteriaceae* ที่ผลิต ESBL (OR = 24.41, 95% CI: 7.32 - 81.39, p-value < 0.001) มีความสัมพันธ์กับการติดเชื้อ MDR-GNB อย่างมีนัยสำคัญ

การวิเคราะห์ Multivariate logistic regression ด้วยเทคนิค forward stepwise สามารถคัดเลือกปัจจัยที่มีความเกี่ยวข้องกับการติดเชื้อดื้อยาหลายขนานได้ 2 ปัจจัยที่มีความเกี่ยวข้องอย่างมีนัยสำคัญ นั่นคือ การติดเชื้อ *Enterobacteriaceae* ที่ผลิต ESBL (OR = 19.88, 95% CI: 5.70 - 69.27, p-value < 0.001) และการติดเชื้อ *E. coli* (OR = 3.13, 95% CI: 2.17 - 4.53, p-value < 0.001)

การวิเคราะห์ด้วย Logistic regression พบว่าปัจจัยเสี่ยงทั้งหมดเป็นปัจจัยที่ก่อโรค ซึ่งอาจทำให้การคัดแยกผู้ป่วยเกิดความล่าช้า และส่งผลให้ไม่สามารถควบคุมการแพร่ระบาดของเชื้อ MDR-GNB ได้อย่างมีประสิทธิภาพ นอกจากนี้ยังมีค่าใช้จ่ายสูงในการส่งตรวจสิ่งส่งตรวจ

การใช้อัลกอริทึม XGBoost ในการจำแนกผู้ป่วยที่มีความเสี่ยงต่อการติดเชื้อ MDR-GNB ได้นำอัลกอริทึม Sequential Forward Selection (SFS) มาช่วยคัดเลือกปัจจัยเสี่ยงที่ส่งผลต่อการพยากรณ์

นอกจากนี้ ได้มีการตรวจสอบและแก้ไขปัญหาความไม่สมดุลของข้อมูล (Class Imbalance) ระหว่างกลุ่มผู้ป่วยที่มีการติดเชื้อ MDR-GNB และกลุ่มผู้ป่วยที่ไม่มีการติดเชื้อ ซึ่งเป็นปัญหาที่พบได้บ่อยในการวิเคราะห์ข้อมูลทางการแพทย์ โดยพบว่าอัตราส่วนความไม่สมดุลของข้อมูล (Imbalance Ratio, IR) ในชุดข้อมูลที่ใช้ในการศึกษานี้มีค่าเท่ากับ 2.30 เท่า หมายความว่า จำนวนผู้ป่วยในกลุ่มที่ไม่มีการติดเชื้อมีมากกว่ากลุ่มที่มีการติดเชื้อ MDR-GNB ประมาณ 2.30 เท่า

เพื่อแก้ไขปัญหานี้ ได้มีการใช้อัลกอริทึม ADASYN ในการปรับปรุงความไม่สมดุลของข้อมูล โดยสามารถลด Imbalance Ratio ลงเหลือเท่ากับ 1 ทำให้จำนวนผู้ป่วยในทั้งสองกลุ่มมีความสมดุลกัน (ตารางที่ 2)

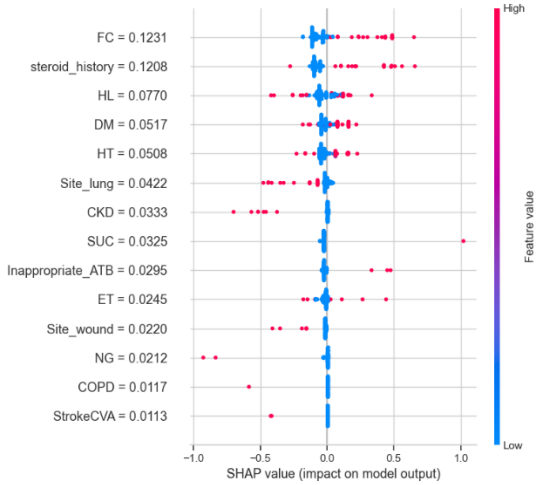
ตารางที่ 2 เปรียบเทียบการกระจายข้อมูลก่อนและหลังการใช้งาน ADASYN

ขั้นตอน	Non-MDR-GNB	MDR-GNB	IR
Original	345	154	2.30
ADASYN	345	343	1.01

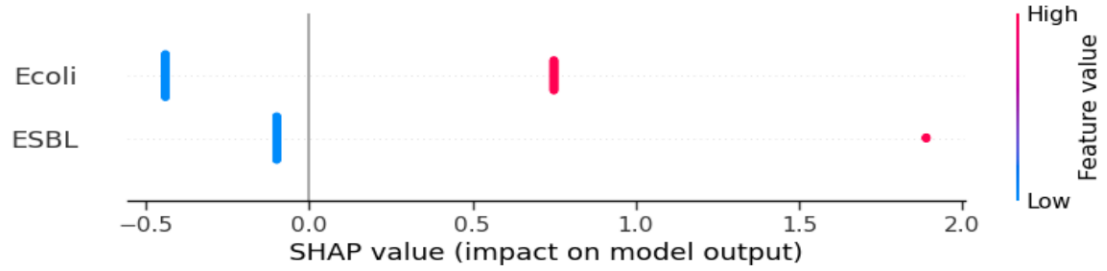
หลังจากการปรับความสมดุลของชุดข้อมูลด้วย ADASYN แล้ว ได้มีการนำชุดข้อมูลที่ปรับปรุงแล้วไปใช้ในการสร้างแบบจำลอง XGBoost จากผลการจำแนกผู้ป่วยที่ติดเชื้อ MDR-GNB แสดงให้เห็นว่าแบบจำลอง XGBoost ที่ผ่านการใช้อัลกอริทึม SFS และ ADASYN สามารถปรับปรุงประสิทธิภาพในการคัดกรองผู้ป่วยที่มีติดเชื้อ

MDR-GNB ได้อย่างมีประสิทธิภาพ (ตารางที่ 3) และเมื่อเปรียบเทียบประสิทธิภาพกับแบบจำลองทางสถิติสามารถแสดงได้ดังตารางที่ 4

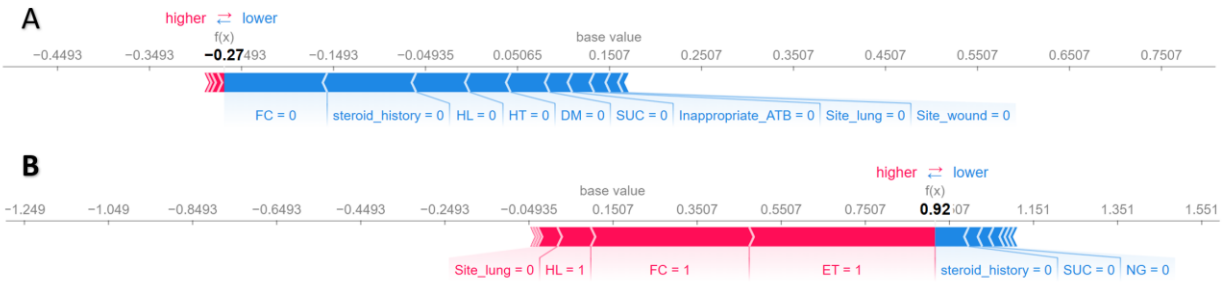
นอกจากนี้ ได้มีการวิเคราะห์ความสำคัญของปัจจัยต่าง ๆ ที่ส่งผลต่อการพยากรณ์การติดเชื้อ MDR-GNB โดยใช้ค่า SHAP ซึ่งพบว่า ปัจจัยเสี่ยงที่สำคัญ 5 อันดับแรกเป็นปัจจัยเสี่ยงต่อการติดเชื้อ MDR-GNB ได้แก่ การใส่สายสวนปัสสาวะ มีประวัติการใช้ยาสเตียรอยด์ การใส่เฮพาริน ล็อค มีประวัติการเป็นโรคเบาหวาน และมีประวัติการเป็นโรคความดันโลหิตสูง (รูปที่ 3) และขยายความผลลัพธ์การพยากรณ์ สำหรับแต่ละตัวอย่าง (รูปที่ 4)



รูปที่ 3 ค่า SHAP สำหรับความเสี่ยงการติดเชื้อ MDR-GNB ของแบบจำลอง XGBoost



รูปที่ 4 ค่า SHAP สำหรับความเสี่ยงการติดเชื้อ MDR-GNB ของแบบจำลอง Logistic regression



รูปที่ 5 กราฟ SHAP เพื่อขยายความผลการพยากรณ์ของตัวอย่างที่ผู้ไม่พบติดเชื้อ MDR-GNB (A) และผู้ป่วยที่พบการติดเชื้อ MDR-GNB (B)

ตารางที่ 3 ร้อยละของประสิทธิภาพของแบบจำลอง XGBoost

ขั้นตอน	Sen (ร้อยละ)	Spec (ร้อยละ)	Acc (ร้อยละ)	AUC (ร้อยละ)
Original	38	70	60	54
SFS + ADASYN	71	61	64	66

หมายเหตุ Sen, Sensitivity; Spec, Specificity; Acc, Accuracy

ตารางที่ 4 เปรียบเทียบร้อยละของประสิทธิภาพ XGBoost และ Multivariate logistic regression (MLR)

อัลกอริทึม	Sen (ร้อยละ)	Spec (ร้อยละ)	Acc (ร้อยละ)	AUC (ร้อยละ)
MLR	16	100	74	66
XGBoost	71	61	64	66

หมายเหตุ Sen, Sensitivity; Spec, Specificity; Acc, Accuracy

จากผลลัพธ์ที่แสดงใน

ตารางที่ แสดงให้เห็นว่าการใช้อัลกอริทึม XGBoost สำหรับการพยากรณ์ผู้ป่วยที่คาดว่ามีการติดเชื้อ MDR-GNB โดยใช้ปัจจัยพื้นฐานของผู้ป่วย ได้แก่ เพศ โรคประจำตัว ประวัติการใช้ยา ประวัติการติดเชื้อที่บริเวณต่างๆ ประวัติหัตถการทางการแพทย์สามารถให้ผลลัพธ์การคัดกรองผู้ป่วยเทียบเท่ากับการพยากรณ์ด้วย MLR ที่ใช้ปัจจัยเสี่ยงที่พบจากการวิเคราะห์ด้วย Multivariate logistic regression ซึ่งเป็นปัจจัยเสี่ยงที่เป็นปัจจัยตัวก่อโรคทั้งหมด นั่นคือ การติด Enterobacteriaceae ที่ผลิต ESBL และการติดเชื้อ *E. coli* แต่การใช้การเรียนรู้ของเครื่องสามารถให้ประสิทธิภาพทั้งค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) และค่า AUC สูงกว่าร้อยละ 60 นอกจากนี้แบบจำลอง XGBoost แสดงให้เห็นถึงประสิทธิภาพที่สมดุล โดยค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity) มีค่าใกล้เคียงกันอย่างมาก ทั้งนี้ XGBoost เป็นอัลกอริทึมที่พัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) โดยใช้กระบวนการเรียนรู้ข้อผิดพลาดจากต้นไม้ตัดสินใจก่อนหน้าและปรับปรุงแนวทางการตัดสินใจอย่างต่อเนื่อง ส่งผลให้แบบจำลองมีความทนทานต่อออคติในชุดข้อมูล (Data Bias) ได้ดีกว่าแบบจำลอง Logistic Regression ซึ่งมักมีข้อจำกัดใน%กรณีที่ข้อมูลมีความซับซ้อนหรือมีการกระจายตัวของข้อมูลที่ไม่สมดุล

อภิปรายผลการวิจัย

การศึกษาครั้งนี้ศึกษาเป็นการศึกษาข้อมูลย้อนหลังระหว่างวันที่ 1 มกราคม พ.ศ. 2559 ถึง 31 ธันวาคม พ.ศ. 2563 ในโรงพยาบาลศูนย์ระดับ A ขนาด 120 เตียง ในโรงพยาบาลสมเด็จพระยุพราชธาตุพนมที่เป็นสถานที่ทำการศึกษาครั้งนี้ พบอัตราการดื้อยาต้านจุลชีพหลายขนานร้อยละ 30.77 ผู้ป่วยส่วนใหญ่มีการติดเชื้อ MDR-GNB มีอายุมากกว่า 70 ปี คิดเป็นร้อยละ 38 โดยค่าเฉลี่ยอายุของกลุ่มผู้ป่วยมีการติดเชื้อ MDR-GNB เท่ากับ 61 ปี ผลการศึกษาครั้งนี้สอดคล้องกับการศึกษา[22] ในโรงพยาบาลสุราษฎร์ธานี พบว่าผู้ป่วยที่มีอายุมากกว่า 60 ปีขึ้นไป มีแนวโน้มที่จะเกิดอุบัติการณ์การติดเชื้อดื้อยามากกว่าผู้ป่วยที่มีอายุน้อยกว่า 60 ปี คิดเป็นร้อยละ 42.55 และการศึกษาของในโรงพยาบาลชัยภูมิ[14] พบว่าผู้ป่วยที่ติดเชื้อ MDR-GNB เป็นผู้มีอายุมากกว่า 60 ปี คิดเป็นร้อยละ 66.7 ซึ่งต่างจากการศึกษาในโรงพยาบาลสุราษฎร์ธานีที่ผู้ป่วยที่ติดเชื้อ MDR-GNB ส่วนใหญ่อายุน้อยกว่า 60 ปี

สำหรับระดับความชุกของเชื้อแบคทีเรียแกรมลบในโรงพยาบาลชุมชนที่เป็นสถานที่วิจัย พบว่าเชื้อ *E. coli* ที่ดื้อยาร้อยละ 64.71 รองลงมาคือ *K. pneumoniae* ร้อยละ 11.76 ซึ่งสอดคล้องกับผลการศึกษาของ Ponyon et al. เชื้อที่พบมากที่สุดคือ *E. coli* ร้อยละ 34.85 ของไอโซเลทที่พบ GNB) รองลงมาเป็น *K. pneumoniae* ร้อยละ 14.19 และ *B. pseudomallei* ร้อยละ 3.19 ของไอโซเลทที่พบ GNB ตามลำดับ

ปัจจัยที่ใช้สำหรับการสร้างแบบจำลอง XGBoost ผ่านอัลกอริทึม SFS สามารถแบ่งออกได้เป็น 4 กลุ่ม ได้แก่ 1) หัตการทางการแพทย์ ประกอบด้วย การใส่สายสวนปัสสาวะ (FC) การใส่เฮพารินล็อค (HL) การใส่ท่อสายยางผ่านระหว่างรูจมูกลงไปยังกระเพาะอาหาร (NG) สายดูดเสมหะ (SUC) และการใส่ท่อช่วยหายใจ (ET) 2) ปัจจัยด้านการใช้ยา ประกอบด้วย ประวัติการใช้ยาสเตียรอยด์ และประวัติการใช้ยาปฏิชีวนะที่ไม่เหมาะสม 3) ปัจจัยด้านโรคประจำตัว ประกอบด้วย โรคเบาหวาน (DM) โรคความดันโลหิตสูง (HT) โรคไตเรื้อรัง (CKD) โรคปอดอุดกั้นเรื้อรัง (COPD) และโรคหลอดเลือดสมอง (CVA/Stroke)

4) ปัจจัยเกี่ยวกับบริเวณที่พบการติดเชื้อ ประกอบด้วย การติดเชื้อระบบทางเดินหายใจและแผลติดเชื้อ

สำหรับผลการพยากรณ์ผู้ป่วยที่คาดการณ์ว่าจะมีการติดเชื้อ MDR-GNB สำหรับงานวิจัยนี้เสนอที่จะใช้การเรียนรู้ของเครื่องแทนแบบจำลองทางสถิติ ด้วยอัลกอริทึม XGBoost โดยแบบจำลองมีประสิทธิภาพที่ดีด้วยค่า AUC เท่ากับ 0.66 ความแม่นยำของแบบจำลองเท่ากับ 0.64 สำหรับการศึกษาก่อนหน้านี้สำหรับการใช้การเรียนรู้ของเครื่องการพยากรณ์หรือคัดกรองผู้ป่วยที่คาดการณ์ว่ามีการติดเชื้อ MDR-GNB ให้ประสิทธิภาพที่แตกต่างกันเล็กน้อย ค่า AUC อยู่ระหว่าง 0.63 ถึง 0.76[23-27]

การขยายความแบบจำลองโดยใช้อัลกอริทึม SHAP โดยใช้ค่า SHAP และกราฟ SHAP เพื่อพิสูจน์ว่า XGBoost สามารถใช้ปัจจัยในการสร้างแบบจำลองพยากรณ์การติดเชื้อ MDR-GNB ได้โดยมีการแสดงความสำคัญของปัจจัยต่าง ๆ และอธิบายภาพรวมของความสำคัญของปัจจัยทำให้แพทย์เข้าใจได้ถึงตัวบ่งชี้ที่สำคัญที่เป็นพื้นฐานในการพยากรณ์การติดเชื้อดื้อยาหลายขนานที่เกิดขึ้น

ผลการศึกษานี้เน้นย้ำถึงความจำเป็นในการควบคุมปัจจัยเสี่ยงที่เกี่ยวข้องกับการติดเชื้อ MDR-GNB โดยเฉพาะในผู้ป่วยสูงอายุและผู้ป่วยที่มีโรคประจำตัว การใช้ยาปฏิชีวนะอย่างเหมาะสมและการลดการใช้หัตถการที่ไม่จำเป็นเป็นประเด็นสำคัญที่จะช่วยลดอัตราการติดเชื้อดื้อยา นอกจากนี้ การใช้แบบจำลอง XGBoost ในการคัดกรองผู้ป่วยที่มีความเสี่ยงต่อการติดเชื้อ MDR-GNB ยังช่วยเพิ่มประสิทธิภาพของการป้องกันและควบคุมการติดเชื้อในโรงพยาบาลและชุมชน

จุดเด่นสำคัญของงานวิจัยนี้คือการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องร่วมกับกระบวนการคัดเลือกฟีเจอร์หรือปัจจัยเสี่ยง และการสังเคราะห์ข้อมูล (Data Augmentation) อย่างมีประสิทธิภาพ ผลการทดลองชี้ให้เห็นว่า แบบจำลองที่พัฒนาขึ้นสามารถให้ผลลัพธ์ที่มีประสิทธิภาพสูงสำหรับการวิเคราะห์ข้อมูลที่มีขนาดเล็ก อีกทั้งยังสามารถอธิบายกระบวนการตัดสินใจในการพยากรณ์ได้อย่างชัดเจน ซึ่งถือเป็นข้อได้เปรียบเมื่อเปรียบเทียบกับงานวิจัยก่อนหน้า ที่แม้จะมีการรายงานประสิทธิภาพของแบบจำลอง แต่ยังขาดการบูรณาการอัลกอริทึมเพื่ออธิบายกระบวนการตัดสินใจของแบบจำลองที่สร้างขึ้นอย่างชัดเจนและเป็นระบบ

ข้อจำกัดสำหรับการศึกษานี้คือเป็นการศึกษาการติดเชื้อดื้อยาหลายขนานที่ดำเนินการในศูนย์การแพทย์เดียว (Single-center) ในโรงพยาบาลระดับชุมชนและเป็นการศึกษาแบบย้อนหลัง ซึ่งอาจทำให้ข้อมูลบางอย่างไม่สมบูรณ์ นอกจากนี้ ขนาดข้อมูลผู้ป่วยค่อนข้างเล็ก เมื่อแบ่งข้อมูลออกเป็นชุดข้อมูลฝึกและชุดข้อมูลทดสอบส่งผลให้มีขนาดที่เล็กลงอีก การประเมินแบบจำลองใช้ค่า AUC-ROC เพื่อวัดความแม่นยำของแบบจำลอง

สรุป

การศึกษานี้เป็นการศึกษาแบบย้อนหลัง ระยะเวลา 5 ปี เพื่อศึกษาปัจจัยเสี่ยงและพยากรณ์ผู้ป่วยที่คาดการณ์ว่ามีการติดเชื้อ MDR-GNB เพื่อป้องกันและควบคุมการแพร่ระบาดของเชื้อดื้อยาภายในโรงพยาบาลได้อย่างมีประสิทธิภาพ ซึ่งเป็นเครื่องมือที่ช่วยบุคลากรทางการแพทย์ในการตัดสินใจคัดแยกผู้ป่วย

เครื่องมือสำหรับการคัดกรองผู้ป่วยที่ติดเชื้อ MDR-GNB มีความสามารถในการจำแนกได้ในระดับปานกลาง ผลการศึกษาปัจจัยเสี่ยงพบสองปัจจัย ซึ่งเป็นปัจจัยตัวก่อโรคทั้งหมด สำหรับปัจจัยที่ใช้สร้างแบบจำลอง XGBoost ใช้ 14 ตัวแปร สามารถแบ่งออกได้เป็น 4 กลุ่ม ได้แก่ หัตถการทางการแพทย์ ปัจจัยด้านการใช้ยา ปัจจัยด้านโรคประจำตัว และปัจจัยเกี่ยวกับบริเวณที่พบการติดเชื้อ

แบบจำลอง XGBoost แสดงให้เห็นถึงประสิทธิภาพในการจำแนกผู้ป่วยที่ติดเชื้อ MDR-GNB ได้ดีกว่า เนื่องจากใช้ปัจจัยเสี่ยงที่ไม่ใช่ตัวก่อโรค ซึ่งแตกต่างจากแบบจำลอง Logistic Regression ที่ใช้ปัจจัยเสี่ยงที่เกี่ยวข้องกับสิ่งก่อโรค ซึ่งได้จากการส่งสิ่งตรวจ ในด้านประสิทธิภาพโดยรวม แบบจำลอง XGBoost ให้ค่า Sensitivity, Specificity, และ AUC ที่สูงกว่า 0.60 ซึ่งสะท้อนถึงความสามารถในการจำแนกที่ดีกว่า

เพื่ออธิบายแบบจำลอง จึงใช้ SHAP ซึ่งเป็นอัลกอริทึมที่ช่วยให้เข้าใจผลการพยากรณ์สำหรับผู้ป่วยแต่ละคน โดยอธิบายถึงปัจจัยที่เพิ่มความน่าเชื่อถือสำหรับผลการพยากรณ์ สามารถใช้เป็นแนวทางในการตัดสินใจทางการแพทย์ได้อย่างมีประสิทธิภาพ

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

จากปัจจัยที่มีความสัมพันธ์กับการดื้อยาหลายขนานที่พบในการศึกษาครั้งนี้ คณะกรรมการป้องกันโรคติดเชื้อในโรงพยาบาล ควรนำแบบจำลองการเรียนรู้ของเครื่องไปประยุกต์ใช้ในการจำแนกผู้ป่วยที่ติดเชื้อ MDR-GNB เพื่อคัดกรองและแยก ลดการแพร่กระจายเชื้อดื้อยาหลายขนานไปสู่ผู้ป่วยรายอื่น เจ้าหน้าที่ และสิ่งแวดล้อมในโรงพยาบาล

การขยายความแบบจำลองโดยใช้อัลกอริทึม SHAP โดยใช้ค่า SHAP และกราฟ SHAP เพื่อพิสูจน์ว่า XGBoost สามารถใช้ปัจจัยในการสร้างแบบจำลองพยากรณ์การติดเชื้อ MDR-GNB ได้ โดยมีการแสดงความสำคัญของปัจจัยต่าง ๆ และอธิบายภาพรวมของความสำคัญของปัจจัยทำให้แพทย์เข้าใจได้ถึงตัวบ่งชี้ที่สำคัญที่เป็นพื้นฐานในการพยากรณ์การติดเชื้อดื้อยาหลายขนานที่เกิดขึ้น

ข้อเสนอแนะในการวิจัยครั้งต่อไป

การศึกษาเพิ่มเติมเกี่ยวกับการติดเชื้อดื้อยาหลายขนานมีความสำคัญอย่างยิ่ง โดยควรมุ่งเน้นการวิเคราะห์ปัจจัยที่เกี่ยวข้องกับการเกิดเชื้อดื้อยา โดยจำแนกตามประเภทของเชื้อก่อโรคและหอผู้ป่วยที่พบการติดเชื้อดื้อยาบ่อยครั้ง นอกจากนี้ ควรศึกษาผลกระทบของการติดเชื้อดื้อยาที่มีต่อผู้ป่วยและโรงพยาบาล เพื่อให้เข้าใจถึงความรุนแรงของปัญหาอย่างครอบคลุม รวมถึงใช้ข้อมูลจากหลายศูนย์การแพทย์เพื่อให้แบบจำลองมีมาตรฐาน (Normalization) มากยิ่งขึ้น

การวิจัยเชิงลึกเกี่ยวกับการใช้ยาปฏิชีวนะอย่างสมเหตุผลเป็นอีกประเด็นสำคัญที่ควรได้รับการศึกษา โดยควรมุ่งเน้นการวิเคราะห์ความสัมพันธ์ระหว่างโรคปัจจัยที่เกี่ยวข้องกับการใช้ยาปฏิชีวนะ ตลอดจนชนิดและปริมาณของยาปฏิชีวนะที่ใช้ กับการเกิดเชื้อแบคทีเรียดื้อยา ทั้งนี้เพื่อนำไปสู่แนวทางการใช้ยาปฏิชีวนะที่เหมาะสมและลดปัญหาการดื้อยา

การจัดการข้อมูลอย่างเป็นระบบเป็นสิ่งจำเป็นในการติดตามและแก้ไขปัญหาการดื้อยา โรงพยาบาลควรพัฒนาฐานข้อมูลเฉพาะสำหรับการเก็บรวบรวมข้อมูลเกี่ยวกับการดื้อยาของเชื้อแบคทีเรีย ซึ่งจะช่วยให้สามารถวิเคราะห์แนวโน้มการดื้อยาได้อย่างทันท่วงที อันจะนำไปสู่การกำหนดนโยบายและมาตรการป้องกันที่มีประสิทธิภาพ

กิตติกรรมประกาศ

คณะผู้วิจัยขอแสดงความขอบคุณอย่างสูงต่อโรงพยาบาลสมเด็จพระยุพราชธาตุพนม ที่อนุญาตให้ใช้เป็นสถานที่ดำเนินการวิจัยครั้งนี้ ขอขอบคุณงานจุลชีววิทยา กลุ่มงานชันสูตร โรงพยาบาลนครพนม ที่กรุณาให้ความอนุเคราะห์ข้อมูลผลการเพาะเชื้อ รวมถึงกลุ่มงานเทคโนโลยีสารสนเทศและกลุ่มงานเวชระเบียน ที่ช่วยรวบรวมข้อมูลประวัติผู้ป่วย ตลอดจนคณะกรรมการควบคุมโรคติดเชื้อในโรงพยาบาล และกลุ่มงานเภสัชกรรมและคุ้มครองผู้บริโภคโรงพยาบาลสมเด็จพระยุพราช-ธาตุพนม ที่ให้ข้อเสนอแนะและคำแนะนำอันเป็นประโยชน์ต่องานวิจัย ท้ายที่สุดนี้ คณะผู้วิจัยขอขอบพระคุณผู้มีพระคุณทุกท่านที่ให้คำปรึกษาและความช่วยเหลือจนทำให้การวิจัยครั้งนี้สำเร็จลุล่วงไปด้วยดี

เอกสารอ้างอิง

1. Petersen E, Shan Lee S, Blumberg L, Levison ME. Antimicrobial resistance - A global problem in need of global solutions. IJID Reg. 2023;9:102-3.
2. Sumpradit N, Wongkongkathep S, Poonpolsup S, Janejai N, Paveenkittiporn W, Boonyarit P, et al. New chapter in tackling antimicrobial resistance in Thailand. BMJ. 2017;358:j3415. (in Thai)
3. Khoka A. Antibiotics and antibiotic resistance. Journal of Medicine and Health Sciences. 2020;27(2):125-39. (in Thai)
4. Samreen, Ahmad I, Malak HA, Abulreesh HH. Environmental antimicrobial resistance and its drivers: a potential threat to public health. Journal of Global Antimicrobial Resistance. 2021;27:101-11.

5. Yosboonruang A, Jitsatthra C, Tarawet L. Incidence of Resistant Bacteria Isolated from Medical Devices in the Intensive Care Unit of a Hospital. Journal of Public Health. 2017;47:222-34. (in Thai)
6. Boonrod P, Pornsiriwatanakul A. Prevalence of Carbapenem Resistant Enterobacteriaceae and Carbapenemase Genes in Lerdsin Hospital. Journal of The Department of Medical Services. 2023;48(2):13-20. (in Thai)
7. Molton JS, Tambyah PA, Ang BSP, Ling ML, Fisher DA. The Global Spread of Healthcare-Associated Multidrug-Resistant Bacteria: A Perspective From Asia. Clinical Infectious Diseases. 2013;56(9):1310-8.
8. Naghavi M, Vollset SE, Ikuta KS, Swetschinski LR, Gray AP, Wool EE, et al. Global burden of bacterial antimicrobial resistance 1990– 2021: a systematic analysis with forecasts to 2050. The Lancet. 2024;404(10459):1199-226.
9. Cooperation OfE, Development. Stemming the superbug tide: just a few dollars more. OECD Paris; 2018.
10. Phumart P, Phodha T, Thamlikitkul V, Riewpaiboon A, Prakongsai P, Limwattananon S. Health and Economic Impacts of Antimicrobial Resistant Infections in Thailand : A Preliminary Study. Journal of Health Systems Research. 2555;6(3):352–60. (in Thai)
11. Baramee J, Unahalekhaka A, Klunklin P. Implementation and Barriers in Prevention of Multidrug-Resistant Organisms Transmission Among Community Hospitals. Nursing Journal CMU. 2021;48(2):95-106. (in Thai)
12. Ponyon J, Ungcharoen R. Incidence of Multidrug-resistant Gram-Negative Bacteria in Blood Specimens in a Community Hospital in Northeast. Pharmacy Practice. 2566;15(1):96-109. (in Thai)
13. Olivares J, Bernardini A, Garcia-Leon G, Corona F, M BS, Martinez JL. The intrinsic resistome of bacterial pathogens. Front Microbiol. 2013;4:103.
14. Acheewakullamas M, Sawasrak K, Wanthana maneekul W, Chaiyabutra S, Thanwiset T. Risk Factors for Multi-Drug Resistant Acinetobacter baumannii Nosocomial Infection, Chaiyaphum Hospital. Chaiyaphum medical journal. 2016;38(1):5-15. (in Thai)
15. Sripromma N. Nursing care for COVID-19 with ventilator-associatedpneumonia and multidrug resistance. The Academic and Nursing Journal of Boromarajonani College of Nursing,Chakriraj. 2023;3(2):E000518. (in Thai)
16. Fatima S, Hussain A, Amir SB, Ahmed SH, Aslam SMH. XGBoost and Random Forest Algorithms: An in Depth Analysis. Pakistan Journal of Scientific Research. 2023;3(1):26-31.
17. Duman E. Implementation of XGBoost Method for Healthcare Fraud Detection. Scientific Journal of Mehmet Akif Ersoy University. 2022;5(2):69-75.
18. Garcia-Vidal C, Puerta-Alcalde P, Cardozo C, Orellana MA, Besanson G, Lagunas J, et al. Machine Learning to Assess the Risk of Multidrug-Resistant Gram-Negative Bacilli Infections in Febrile Neutropenic Hematological Patients. Infect Dis Ther. 2021;10(2):971-83.
19. Moran E, Robinson E, Green C, Keeling M, Collyer B. Towards personalized guidelines: using machine-learning algorithms to guide antimicrobial selection. Journal of Antimicrobial Chemotherapy. 2020;75(9):2677-80.
20. Liu J, Gao Y, Hu F. A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. Computers & Security. 2021;106:102289.
21. Humbert-Vidan L, Hansen CR, Patel V, Johansen J, King AP, Guerrero Urbano T. External validation of a multimodality deep-learning normal tissue complication probability model for mandibular osteoradionecrosis trained on 3D radiation distribution maps and clinical variables. Physics and Imaging in Radiation Oncology. 2024;32:100668.

22. Boonlum S. Risk Factors of the Multidrug-resistant in Surgical II Intensive Care Unit, Suratthani Hospital. Journal of Health Sciences and Pedagogy. 2021;1(1):28-39. (in Thai)
23. Lee ALH, To CCK, Lee ALS, Chan RCK, Wong JSH, Wong CW, et al. Deep learning model for prediction of extended-spectrum beta-lactamase (ESBL) production in community-onset Enterobacteriaceae bacteraemia from a high ESBL prevalence multi-centre cohort. Eur J Clin Microbiol Infect Dis. 2021;40(5):1049-61.
24. Sick-Samuels AC, Goodman KE, Rapsinski G, Colantouni E, Milstone AM, Nowalk AJ, et al. A Decision Tree Using Patient Characteristics to Predict Resistance to Commonly Used Broad-Spectrum Antibiotics in Children With Gram-Negative Bloodstream Infections. Journal of the Pediatric Infectious Diseases Society. 2019;9(2):142-9.
25. Rich SN, Jun I, Bian J, Boucher C, Cherabuddi K, Morris JG, Jr., et al. Development of a Prediction Model for Antibiotic-Resistant Urinary Tract Infections Using Integrated Electronic Health Records from Multiple Clinics in North-Central Florida. Infect Dis Ther. 2022;11(5):1869-82.
26. Liang Q, Ding S, Chen J, Chen X, Xu Y, Xu Z, et al. Prediction of carbapenem-resistant gram-negative bacterial bloodstream infection in intensive care unit based on machine learning. BMC Medical Informatics and Decision Making. 2024;24(1):123.
27. Ponyon J, Kerdsin A, Preeprem T, Ungcharoen R. Risk Factors of Infections Due to Multidrug-Resistant Gram-Negative Bacteria in a Community Hospital in Rural Thailand. Trop Med Infect Dis. 2022;7(11). (in Thai)