การจำแนกอาการไม่พึงประสงค์จากการใช้ยาจากเวชระเบียนอิเล็กทรอนิกส์ โดยใช้การเรียนรู้ของเครื่องและการประมวลผลภาษาธรรมชาติ

Classifying Adverse Drug Reactions from Electronic Medical Records using Machine Learning and Natural Language Processing

ศศิวิมล กอบัว อรรคพล วงศ์กอบลาภ ธรา อั่งสกุล และจิติมนต์ อั่งสกุล*

Sasiwimon Kobua, Akkapon Wongkoblap, Thara Angskun and Jitimon Angskun*

สำนักวิชาศาสตร์และศิลป์ดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี

Institute of Digital Arts and Science, Suranaree University of Technology, Thailand

ABSTRACT

Background and Objectives: Patient drug safety is a priority in the World Health Organization (WHO)'s action plan. In Thailand, psychiatric medications are among the top five drug classes linked to adverse events. However, analyzing adverse drug reactions (ADRs) can be a time-consuming process, and variations may arise due to differences in the expertise of medical personnel. Classifying ADRs in psychiatric patients who use bilingual (Thai-English) electronic medical records is complicated by linguistic structures, transliterations, and non-standard spellings, which can hinder the accuracy of existing classification models. This study aims to develop a model for classifying ADRs associated with psychiatric medications in bilingual text. It will compare machine learning and deep learning techniques to enhance medication safety surveillance within Thailand's healthcare context.

Methodology: The process began with collecting ADE data from psychiatric patients between October 2019 and March 2023. These data were then processed to develop a model for classifying ADRs using natural language processing and five traditional machine learning techniques: Gaussian Naive Bayes, K-Nearest Neighbors, Linear Support Vector Machine, Logistic Regression, and Random Forest. Two deep learning techniques were utilized: Convolutional Neural Networks and Bidirectional Long Short-Term Memory Neural Networks. Finally, the model's performance was evaluated using 70% of the data for training and 30% for testing.

Main Results: The findings indicate that the model developed using the Random Forest technique, combined with feature extraction methods incorporating both 1-gram and 2-gram approaches, achieved the highest level of classification performance. When tested with a dataset of 384 samples, this model achieved an accuracy, precision, recall, and F1-score of 84.6%, 82.6%, 84.6%, and 83.2%, respectively. The developed

ARTICLE INFO

Article history:

Received 24 May 2025

Revised 23 October 2025

Accepted 3 November 2025

Keywords:

Adverse Drug Reaction (อาการไม่พึงประสงค์จากการใช้ยา),

Machine Learning (การเรียนรู้ของเครื่อง),

Deep Neural Network (การเรียนรู้เชิงลึก),

Natural Language Processing (การประมวลผลภาษาธรรมชาติ)

model demonstrated high effectiveness in accurately classifying adverse drug reactions, showcasing its robustness as a machine learning method in the medical domain. Additionally, regarding computational efficiency, the Linear Support Vector Machine and Logistic Regression models exhibited the fastest processing speeds, making them noteworthy options for scenarios requiring rapid decision-making.

Discussions: The findings of this study align with several previous studies, although variations may arise due to the characteristics of the data and the complexity of bilingual (Thai and English) content in electronic medical records. Deep learning methods are resource-intensive and time-consuming; thus, selecting the appropriate technique requires balancing accuracy and speed. Real-time systems may favor Linear Support Vector Machines, even though they are slightly less accurate than Random Forests. Future research should incorporate data balancing techniques, advanced feature extraction, expanding data sources, and experimenting with language models specifically developed for the Thai language and medical contexts.

Conclusions: This study demonstrates the potential of utilizing machine learning and deep learning techniques to detect adverse drug reactions, highlighting important implications for developing an efficient surveillance system. Such a system can enhance the speed and accuracy of identifying events that might otherwise be missed, while reducing the workload for healthcare personnel, allowing them to concentrate more on patient care. The proposed approach aligns with the WHO's guidance and Thailand's digital health policies.

*Corresponding author
E-mail address: jitimon@g.sut.ac.th

บทคัดย่อ

ที่มาและวัตถุประสงค์: ความปลอดภัยจากการใช้ยาผู้ป่วยถือเป็นประเด็นสำคัญในแผนปฏิบัติการขององค์การอนามัยโลก การสำรวจในประเทศไทยพบว่า ยาจิตเวชจัดอยู่ในกลุ่มยา 5 อันดับแรกที่มีการรายงานเหตุการณ์ไม่พึงประสงค์มากที่สุด อย่างไรก็ตาม กระบวนการวิเคราะห์อาการไม่พึงประสงค์จากการใช้ยา ยังคงใช้เวลานานและมีความคลาดเคลื่อนเกิดขึ้น เนื่องจากความเชี่ยวชาญของบุคลากรทางการแพทย์ที่แตกต่างกัน และการวิเคราะห์ข้อความเวชระเบียนอิเล็กทรอนิกส์ที่มีการใช้ภาษาไทยปนภาษาอังกฤษ ซึ่งมีความซับซ้อนด้านโครงสร้างภาษา คำทับศัพท์ และการสะกดคำไม่เป็นมาตรฐาน ทำให้แบบจำลองการเรียนรู้ของเครื่องในปัจจุบันยังไม่สามารถทำงานได้อย่างแม่นยำ งานวิจัยนี้จึงมุ่งพัฒนาแบบจำลองจำแนกอาการไม่พึงประสงค์จากการใช้ยาที่รองรับการประมวลผลข้อความสองภาษา (ไทย-อังกฤษ) พร้อมเปรียบเทียบเทคนิคการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก เพื่อสนับสนุนการเฝ้าระวังความปลอดภัยด้านการใช้ยาที่เหมาะสมกับสภาพแวดล้อมจริงในประเทศไทย

ระเบียบวิธีวิจัย: กระบวนการเริ่มต้นจากการรวบรวมข้อมูลเหตุการณ์ไม่พึงประสงค์ของผู้ป่วยจิตเวชตั้งแต่เดือนตุลาคม พ.ศ. 2562 ถึงเดือนมีนาคม พ.ศ. 2566 นำไปประมวลผลก่อนการพัฒนาแบบจำลองเพื่อจำแนกอาการไม่พึงประสงค์จากการใช้ยาด้วยเทคนิคการประมวลผลภาษาธรรมชาติ ร่วมกับเทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิม 5 เทคนิค ได้แก่ เทคนิคเกาส์เซียนนาอีฟเบย์ เทคนิคเคเนียร์เรสเนเบอร์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น เทคนิคการถดถอยโลจิสติค และเทคนิคแรนดอมฟอเรสต์ และเทคนิคการเรียนรู้เชิงลึก 2 เทคนิค ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชั่น และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง สุดท้ายประเมินประสิทธิภาพของแบบจำลองโดยใช้ข้อมูลชุดฝึกจำนวนร้อยละ 70 และชุดการทดสอบ จำนวนร้อยละ 30

ผลการวิจัย: แบบจำลองที่พัฒนาโดยใช้เทคนิคแรนดอมฟอเรสต์ กับการสกัดคุณลักษณะแบบใช้ 1-gram และ 2-gram ร่วมกัน แสดงให้เห็นถึงประสิทธิภาพในการจำแนกอาการไม่พึงประสงค์จากการใช้ยาสูงที่สุด เมื่อทดสอบกับชุดข้อมูล 384 ตัวอย่าง แบบจำลองดังกล่าวมีความถูกต้องถึงร้อยละ 84.6 ให้ค่าความแม่นยำ ร้อยละ 82.6 ความระลึก ร้อยละ 84.6 และค่าถ่วงดุลร้อยละ 83.2 ในแง่ของความเร็วในการประมวลผล แบบจำลองที่พัฒนาโดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น และเทคนิคการถดถอยโลจิสติค แสดงให้เห็นถึงเวลาในการประมวลผลที่มีประสิทธิภาพสูงสุด

การอภิปรายผล: งานวิจัยนี้มีข้อค้นพบที่สอดคล้องกับงานวิจัยหลายงาน นอกจากนี้ยังมีความแตกต่างของผลลัพธ์อันเกิดจากลักษณะข้อมูลและความซับซ้อนของภาษาไทยปนอังกฤษในเวชระเบียน การเรียนรู้เชิงลึกพบว่าต้องใช้ทรัพยากรและเวลาสูง การเลือกเทคนิคที่เหมาะสมจึงต้องพิจารณาสมดุลระหว่างความแม่นยำและความเร็ว โดยระบบเรียลไทม์อาจเหมาะกับซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้นมากกว่า แม้แม่นยำน้อยกว่าแรนดอมฟอเรสต์เล็กน้อย การวิจัยในอนาคตเสนอให้เพิ่มเทคนิคสร้างความสมดุลข้อมูล สกัดคุณลักษณะขั้นสูง ขยายแหล่งข้อมูล และทดลองใช้แบบจำลองภาษาที่เหมาะกับภาษาไทยและข้อมูลทางการแพทย์

บทสรุป: การศึกษาวิจัยนี้แสดงให้เห็นถึงศักยภาพของการใช้เทคนิคการเรียนรู้ของเครื่อง และการเรียนรู้เชิงลึกเพื่อจำแนกอาการไม่พึงประสงค์จากการใช้ยา ซึ่งสามารถนำไปต่อยอดเป็นระบบเฝ้าระวังที่มีประสิทธิภาพ เพื่อเพิ่มความรวดเร็ว และความแม่นยำในการตรวจจับอาการที่อาจถูกมองข้ามได้ดียิ่งขึ้น ทั้งยังช่วยลดภาระงานของบุคลากรให้สามารถมุ่งเน้นการดูแลผู้ป่วยได้อย่างเต็มที่ โดยสอดคล้องกับแนวทางขององค์การอนามัยโลกและนโยบายสาธารณสุขดิจิทัลของประเทศไทย

บทนำ

ความปลอดภัยของผู้ป่วยด้านยาเป็นประเด็นสำคัญในแผนปฏิบัติการด้านความปลอดภัยของผู้ป่วยทั่วโลก พ.ศ. 2564-2573 ขององค์การอนามัยโลก ภายใต้ยุทธศาสตร์ที่ 3 ว่าด้วยความปลอดภัยของกระบวนการทางคลินิก ซึ่งมุ่งลดอันตรายจากการใช้ยาและเหตุการณ์ไม่พึงประสงค์ (Adverse Drug Event: AE) ลงร้อยละ 50 70 และ 80 ภายในปี พ.ศ. 2568 2570 และ 2573 ตามลำดับ องค์การอนามัยโลกเสนอแนวทางทั้งในระดับประเทศและหน่วยงาน เช่น การวางกลไก การใช้เทคโนโลยีเพื่อเพิ่มความตระหนักรู้เกี่ยวกับยาและกระบวนการใช้ยา ตลอดจนส่งเสริมการรายงานข้อผิดพลาดและเหตุการณ์ไม่พึงประสงค์ เพื่อพัฒนาระบบการเรียนรู้ที่ช่วยให้บุคลากรทางการแพทย์สามารถติดตาม สืบหาสาเหตุ ประเมิน จัดลำดับความสำคัญ และจัดการปัญหาได้อย่างมีประสิทธิภาพ (World Health Organization, 2021) สถิติการรายงานเหตุการณ์ไม่พึงประสงค์ขององค์การอาหารและยาแห่งสหรัฐอเมริกา (United States Food and Drug Administration : FDA) ระหว่างปี พ.ศ. 2531–2567 แสดงให้เห็นว่าจำนวนกรณีที่ถูกรายงานเพิ่มขึ้นทุกปี โดยแต่ละปีมีสัดส่วนกรณี

รุนแรงและกรณีเสียชีวิต เกินร้อยละ 50 ของทั้งหมด ทั้งนี้ ในปี พ.ศ. 2565 มีรายงานสูงสุดที่ 2.34 ล้านครั้ง สำหรับประเทศไทยนั้น สำนักงานคณะกรรมการอาหารและยาได้รายงานว่า ในปี พ.ศ. 2564 มีอัตราการรายงานเพิ่มขึ้น 1.5 เท่าจากปีก่อน แม้ว่าจำนวนรายงานลดลงในปี พ.ศ. 2565 แต่ยังคงเพิ่มขึ้นอีกครั้งในปี พ.ศ. 2566 เมื่อพิจารณาจำแนกตามจังหวัดในเขตบริการที่ 9 จังหวัดนครราชสีมามีจำนวนรายงานสูงสุดและเพิ่มขึ้นต่อเนื่องตั้งแต่ปี พ.ศ. 2563

การศึกษาข้อมูลยาของกองนโยบายแห่งชาติด้านยา กระทรวงสาธารณสุข พบว่า ยากลุ่มจิตเวชเป็นกลุ่มยาที่มีความเสี่ยงต่อการเกิดอาการไม่พึงประสงค์สูง สอดคล้องกับสถิติของศูนย์เฝ้าระวังความปลอดภัยด้านผลิตภัณฑ์สุขภาพ สำนักงานคณะกรรมการอาหารและยา ที่ระบุว่ายากระตุ้นระบบประสาทส่วนกลาง (Central Nervous System) อยู่ในอันดับ 4 ของรายงานเหตุการณ์ไม่พึงประสงค์ การใช้ยาจิตเวชมีความสำคัญต่อการดูแลผู้ป่วยโรคจิตเวช เนื่องจากโรคเหล่านี้มักไม่หายขาดและมีแนวโน้มกลับเป็นซ้ำหากการรักษาไม่เพียงพอ ผู้ป่วยจึงจำเป็นต้องปรับยาอย่างต่อเนื่องในระยะยาวหรือทั้งชีวิต แม้จะช่วยควบคุมอาการได้ แต่ยาดังกล่าวก็สามารถก่อให้เกิดผลข้างเคียงที่ไม่พึงประสงค์ต่อร่างกาย (Tiamkaew, 2014) จากการทบทวนวรรณกรรมและงานวิจัยทั้งในและต่างประเทศ ระบุว่า อาการไม่พึงประสงค์จากการใช้ยากลุ่มจิตเวชสามารถแบ่งตามประเภทของยาได้เป็น 4 กลุ่ม ได้แก่ ยารักษาโรคจิต ยารักษาโรคซึมเศร้า ยาทำให้อารมณ์คงที่ และยาคลายกังวล นอกจากนี้ งานวิจัยยังเผยให้เห็นว่าปัจจัยที่ส่งผลต่อการเกิดอาการไม่พึงประสงค์มีความซับซ้อนและหลากหลาย โดยสามารถจำแนกออกเป็น 5 กลุ่ม ได้แก่ ปัจจัยที่เกี่ยวข้องกับผู้ป่วย ปัจจัยที่เกี่ยวข้องกับโรคของผู้ป่วย ปัจจัยด้านสังคม ปัจจัยที่เกี่ยวข้องกับยา และปัจจัยอื่น ๆ ที่เกี่ยวข้องกับสถานพยาบาลและบุคลากรทางการแพทย์ (Alomar, 2014; Guo et al., 2022; Iqbal, 2021; Kobua et al., 2025; Phianthai & Vetchakun-anukun, 2020; Sisay & Wami, 2021; Songsiriphan, 2018) เมื่อเปรียบเทียบกับยากลุ่มอื่น ยากลุ่มจิตเวชมีความซับซ้อนด้านข้อมูลและการรักษามากกว่า เนื่องจากต้องปรับขนาดยาและใช้ยาหลายชนิดร่วมกันอย่างต่อเนื่อง เพิ่มความเสี่ยงต่อปฏิกิริยาระหว่างยาและอาการไม่พึงประสงค์ อีกทั้งผลข้างเคียงของยากลุ่มนี้ยังส่งผลเชิงลึกต่อสมอง จิตใจ และคุณภาพชีวิตของผู้ป่วย ปัจจัยเหล่านี้ทำให้การติดตามและประเมินอาการไม่พึงประสงค์จากการใช้ยาในผู้ป่วยจิตเวชมีความซับซ้อนสูง จึงเป็นกรณีศึกษาที่สำคัญและเหมาะสมต่อการพัฒนางานวิจัยและเทคโนโลยีวิเคราะห์ข้อมูลขั้นสูง

จากการศึกษาเอกสาร (2017-2024) และสัมภาษณ์เภสัชกร (Amornrat Chuanchaisit, personal communication, 2022) พบว่ากระบวนการประเมินอาการไม่พึงประสงค์จากการใช้ยาในสถานพยาบาลมีขั้นตอนคล้ายคลึงกันและใช้เวลาเฉลี่ย 30 นาทีต่อรายงาน โดยเริ่มจากการแจ้งเหตุจากแพทย์ หรือพยาบาล จากนั้นเภสัชกรระบุอาการที่สงสัยว่าเกิดจากยาตัวใด ซึ่งต้องอาศัยความรู้เกี่ยวกับโปรไฟล์ยาแต่ละตัว อีกทั้งยังต้องพิจารณาปัจจัยอื่น ๆ อาทิ ประวัติผู้ป่วย โรคทางกาย และการใช้ยารักษาโรคดังกล่าวจากข้อมูลรายงานบันทึกอาการไม่พึงประสงค์จากการใช้ยาร่วมกับข้อมูลเวชระเบียนอิเล็กทรอนิกส์ (Electronic Medical Records) จากระบบฮอสเอกซ์พี (HOSxP) ซึ่งเป็นระบบสารสนเทศสำหรับสถานพยาบาล ต่อมาจึงประเมินความสัมพันธ์ระหว่างยาและอาการโดยใช้เกณฑ์ที่ต่างกันตามแต่ละสถานพยาบาล ได้แก่ Naranjo's algorithm (Naranjo et al., 1981), WHO-UMC Causality Assessment ขององค์การอนามัยโลก และ Thai

algorithm (Suwankesawong et al., 2016) ก่อนสรุประดับความรุนแรงเพื่อเสนอแนวทางรักษาและตรวจสอบความป้องกันได้ตาม Schumock and Thornton (1992) ต่อแพทย์ผู้รักษา

อย่างไรก็ตาม กระบวนการประเมินอาการไม่พึงประสงค์จากการใช้ยาในปัจจุบันยังเผชิญข้อจำกัดสำคัญ ได้แก่ (1) การรายงานที่อาศัยบุคลากร ทำให้ข้อมูลบางส่วนอาจตกหล่นและมีความล่าช้า (2) การพิจารณาความสัมพันธ์ระหว่างยาและอาการต้องอาศัยความรู้และประสบการณ์เฉพาะบุคคล ทำให้ผลลัพธ์มีความแปรปรวน และ (3) การใช้ข้อมูลจากเวชระเบียนอิเล็กทรอนิกส์ในการวิเคราะห์ยังไม่ได้รับการใช้ประโยชน์เต็มศักยภาพ ช่องว่างเหล่านี้สะท้อนถึงความจำเป็นในการพัฒนาแบบจำลองในเพื่อจำแนกอาการไม่พึงประสงค์จากการใช้ยาแบบอัตโนมัติ โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) และการประมวลผลภาษาธรรมชาติ (NLP) เพื่อเพิ่มความรวดเร็ว ความแม่นยำ และลดความแปรปรวนในการวิเคราะห์ข้อมูลดังกล่าว

งานวิจัยที่เกี่ยวข้อง

การศึกษางานวิจัยที่ผ่านมา พบว่าได้มีการนำเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) มาใช้วิเคราะห์ข้อมูลอาการไม่พึงประสงค์จากยาหลากหลายประเภท เช่น มะเร็ง หัวใจ เบาหวาน คอเลสเตอรอล และไทรอยด์ ด้วยเทคนิคประมวลผลภาษาธรรมชาติหลายภาษา ร่วมกับวิธีเรียนรู้ของเครื่องทั้งแบบดั้งเดิม และการเรียนรู้เชิงลึก โดยมุ่งเน้นจำแนกการเกิดเหตุการณ์ไม่พึงประสงค์บนข้อมูลเวชระเบียนอิเล็กทรอนิกส์ โดยข้อมูลมีทั้งรูปแบบมีโครงสร้าง (เช่น ข้อมูลส่วนบุคลของผู้ป่วย ข้อมูลประวัติการใช้ยา ข้อมูลผลตรวจทางห้องปฏิบัติการ เป็นต้น) และไม่มีโครงสร้าง (เช่น บันทึกทางคลินิก บันทึกการรักษาที่เป็นคำบรรยาย บันทึกการจำหน่ายผู้ป่วย เป็นต้น) ดังตารางที่ 1 (Table 1)

งานวิจัยที่ใช้เทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิมหรือการเรียนรู้เชิงลึก ที่วิเคราะห์ข้อมูลอยู่ในรูปแบบมีโครงสร้าง ได้แก่ งานวิจัยที่ศึกษาการตรวจหาภาวะกล้ามเนื้อสลายตัวจากยาสแตตินโดยใช้เทคนิคเหมืองข้อมูล ได้นำข้อมูลจากระบบเวชระเบียนอิเล็กทรอนิกส์ มาพัฒนาแบบจำลองด้วยเทคนิคการถดถอยโลจิสติก เครือข่ายเบย์เซียน ป่าสุ่ม และเอ็กซ์ตรีมเกรเดียนต์บูสต์ติ้ง จากคุณลักษณะ 122 รายการ คัดเลือกเหลือรายการสำคัญ 18 รายการ ผลการศึกษาพบว่าแม้เทคนิคป่าสุ่มและเอ็กซ์ตรีมเกรเดียนต์บูสต์ติ้งให้ความไวสูงสุด แต่การถดถอยโลจิสติกมีประสิทธิภาพโดยรวมดีที่สุดเนื่องจากมีความจำเพาะสูง (Kunakorntham et al., 2022) งานวิจัยที่เปรียบเทียบ 18 เทคนิคการเรียนรู้ของเครื่องร่วมกับวิธีการเติมข้อมูลที่หายไป การจัดการข้อมูลที่ไม่สมดุล และการคัดเลือกตัวแปร การวิจัยใช้ค่า SHAP วิเคราะห์ปัจจัยที่มีผลต่อการเกิดอาการไม่พึงประสงค์จากการใช้ยา จากนั้นทดสอบกับชุดข้อมูล 60 แบบ ผลการวิจัยพบว่า เทคนิคการเรียนรู้แบบรวมกลุ่มซึ่งรวมแบบจำลอง 5 เทคนิคที่มีค่าพื้นที่ใต้กราฟ ROC (Area Under ROC Curve : AUC) สูงสุดให้ผลลัพธ์ดีที่สุด โดยมีค่า AUC สูงถึงร้อยละ 0.9141 ค่าความถูกต้องร้อยละ 0.8947 ค่าความแม่นยำร้อยละ 0.75 และค่าความระลึกร้อยละ 0.6667 (Wu et al., 2022) นอกจากนี้ยังมีงานวิจัยที่วิเคราะห์ข้อมูลที่ไม่มีโครงสร้างเป็นข้อความภาษาธรรมชาติที่พัฒนาวิธีผสมผสานระหว่างเทคนิคฐานกฎ และการเรียนรู้ของ

เครื่องเพื่อจับความสัมพันธ์ระหว่างยากับอาการไม่พึงประสงค์จากบันทึกการจำหน่ายผู้ป่วย (Discharge Summaries) และฐานข้อมูล MIMIC III วิธีการเริ่มจากใช้ฐานกฎคัดกรองประโยคที่มีคำเกี่ยวกับยา อาการ และสัญญาณบ่งชี้ความสัมพันธ์จากพจนานุกรม จากนั้นทำการกำกับข้อมูล (Data Annotation) โดยผู้เชี่ยวชาญ พัฒนาแบบจำลองหลายรูปแบบ โดยใช้ป่าสุ่ม (Random Forests) เอ็กซ์ตรีมเกรเดียนต์บูสต์ติ้ง (Extreme Gradient Boosting : XGBoost) นาอีฟเบย์ (Naïve Bayes) การลดเกรเดียนต์เชิงสุ่ม (Stochastic Gradient Descent) ร่วมกับเทคนิคความถี่ของคำ-ส่วนกลับความถี่ของเอกสาร (TF-IDF) ใช้เอ็กซ์ตรีมเกรเดียนต์บูสต์ติ้ง (XGBoost) ร่วมกับการแปลงคำเป็นเวกเตอร์ (Word2Vec) และหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory) ร่วมกับการฝังคำ (Word embeddings) พบว่าเมื่อทดสอบกับชุดข้อมูลที่จำกัดด้วยฐานกฎ แบบจำลองให้ค่าความระลึกสูงร้อยละ 0.76-0.80 แต่เมื่อทดสอบกับชุดข้อมูลที่ไม่จำกัดด้วยฐานกฎ ค่าความระลึกลดลง 12-14% เหลือ 0.62-0.68 (Tan et al., 2022)

งานวิจัยที่ใช้การเรียนรู้เชิงลึกและแบบจำลองทางภาษา เพื่อวิเคราะห์ข้อมูลภาษาธรรมชาติซึ่งอยู่ในรูปแบบที่ไม่มีโครงสร้าง ได้แก่ งานวิจัยที่ศึกษาการจัดการข้อมูลไม่สมดุลในงานประมวลผลภาษาธรรมชาติด้วยการสร้างประโยคสังเคราะห์ (Data augmentation) โดยแทนที่คำบางคำด้วยคำที่คล้ายกัน งานวิจัยนี้สร้างตัวแทนเชิงกระจายของข้อมูลประวัติผู้ป่วยด้วยแบบจำลองข้ามแกรม (Skip-gram) และนำมาสร้างข้อมูลสังเคราะห์โดยแทนที่ข้อมูลประวัติด้วยข้อมูลที่คล้ายกัน ผลการประเมินพบว่า แบบจำลองการเรียนรู้เชิงลึกด้วยเทคนิคเซลฟ์-แอทเทนชัน (Self-attention) ร่วมกับวิธีการสุ่มลด (Under sampling) มีประสิทธิภาพดีกว่า นอกจากนี้ การประยุกต์ใช้วิธีการเพิ่มข้อมูลที่นำเสนอร่วมกับการสุ่มลดสามารถปรับปรุงค่าความถูกต้องได้ถึง 80% และปรับปรุงค่าความแม่นยำและคะแนน F1 ได้ถึง 40% (Ishikawa et al., 2022) และงานวิจัยที่พัฒนาอัลกอริธึมการประมวลผลภาษาธรรมชาติด้วยการเรียนรู้เชิงลึกเพื่อจำแนกบันทึกสรุปการจำหน่ายผู้ป่วยที่มีการกล่าวถึงอาการไม่พึงประสงค์จากการใช้ยา โดยเริ่มจากการเตรียมข้อมูลโดยระบุตำแหน่งของชื่อยาและปฏิกิริยาไม่พึงประสงค์ ใช้เครื่องมือ Prodigy สำหรับชุดแรก และ Med7 เพื่อแนะนำป้ายกำกับแก่ผู้เชี่ยวชาญทางการแพทย์ จากนั้นนำข้อมูลที่ระบุตำแหน่งแล้วมาฝึกสอนแบบจำลองทางภาษาในการจดจำเอนทิตี้ใหม่ในแต่ละรอบเพื่อเพิ่มประสิทธิภาพ จนกระทั่งข้อมูลทั้งหมดได้รับการระบุตำแหน่งครบถ้วน ต่อมาได้ประมวลผลข้อมูลด้วย Word Piece Tokenizer และพัฒนาแบบจำลอง 3 แบบเปรียบเทียบกัน ได้แก่ เทคนิค DeBERTa, MeDeBERTa (การฝึกฝนล่วงหน้าด้วยข้อมูลบันทึกทางการแพทย์ของสถาบัน) และ RoBERTa (การฝึกฝนล่วงหน้าด้วยข้อมูลบันทึก MIMIC-III) ผลการวิจัยพบว่าเทคนิค MeDeBERTa มีประสิทธิภาพดีที่สุดในการตรวจจับอาการไม่พึงประสงค์จากการใช้ยา (McMaster et al., 2023)

งานวิจัยที่ใช้เทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิม ร่วมกับการเรียนรู้เชิงลึก และแบบจำลองทางภาษา เพื่อวิเคราะห์ข้อมูลภาษาธรรมชาติซึ่งอยู่ในรูปแบบที่ไม่มีโครงสร้าง มีงานวิจัยที่ได้พัฒนาระบบจำแนกคำศัพท์อาการจากรายงานอาการไม่พึงประสงค์จากยาแบบสองภาษา (ไทย-อังกฤษ) โดยเปรียบเทียบเทคนิคการประมวลผลภาษาธรรมชาติหลายวิธี ได้แก่ เทคนิคหาความถี่คำ-ความถี่เอกสารผกผัน (TF-IDF) กับเทคนิคนาอีฟเบย์-ซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น (NB-SVM) ใช้วิธีตัดแบ่งคำด้วย Sentence Piece ร่วมกับแบบจำลองทางภาษาต่าง ๆ ได้แก่ ULMFit, mBERT, XLM-RoBERTa, WangchanBERTa และ

Byte-Level BPE กับ AllergyRoBERTa เพื่อจับคู่กับคำศัพท์อาการมาตรฐาน 36 ข้อ และบันทึกลงระบบบันทึกข้อมูลสุขภาพอิเล็กทรอนิกส์ ผลพบว่าแบบจำลองแบบจำลองแบบรวมกลุ่ม (Ensemble) มีประสิทธิภาพดีที่สุด โดยมีการจับคู่ถูกต้อง 95.33% คะแนน F1 98.88% และความแม่นยำเฉลี่ย 97.07% (Chaichulee et al., 2022) และงานวิจัยของ Zitu และคณะ (2023) ประเมินประสิทธิภาพของวิธีการเรียนรู้ของเครื่องในการตรวจหาเหตุการณ์ไม่พึงประสงค์จากการใช้ยาในเวชระเบียนอิเล็กทรอนิกส์ โดยรวบรวมข้อมูลจากโรงพยาบาลมะเร็งเจมส์ (James Cancer Hospital) และชุดข้อมูล n2c2 ของมหาวิทยาลัยฮาร์วาร์ดมาสร้างคลังข้อมูลความสัมพันธ์ยาและเหตุการณ์ไม่พึงประสงค์ โดยมีผู้เชี่ยวชาญ 3 คนระบุเอนทิตี้ ได้แก่ ชื่อยา เหตุการณ์ไม่พึงประสงค์ วันที่เกี่ยวข้อง คำแสดงความสัมพันธ์เชิงสาเหตุ และระดับความรุนแรง จากนั้นเปรียบเทียบประสิทธิภาพเทคนิคต่าง ๆ (SVM, CNN, BiLSTM, BERT, ClinicalBERT) ในการจำแนกคู่ยา-ผลข้างเคียงที่มีความสัมพันธ์กันหรือไม่ ผลพบว่า การใช้ ClinicalBERT มีประสิทธิภาพดีที่สุดเมื่อฝึกด้วยข้อมูล ICU-ISU และทดสอบด้วยข้อมูล n2c2 มีค่าคะแนน F1 ร้อยละ 0.78 และฝึกด้วยข้อมูล n2c2 และทดสอบด้วยข้อมูล ICU-ISU มีค่าคะแนน F1 ร้อยละ 0.74

ตารางที่ 1 (Table 1)

งานวิจัยที่เกี่ยวข้อง (A Literature Review)

งานวิจัย (Research)	เทคนิค (Techniques)				รูปแบบข้อมูล (Data format)		ภาษา (Language)		ยา (Drug)
	TML	DL	NLP	LM	Structured	Unstructured	Thai	English	
Kunakorntham et al. (2022)	/	/			/		/		Statin
Wu et al. (2022)	/	/			/			/	Chinese herbal injections containing Panax notoginseng saponin
Tan et al. (2022)	/	/	/			/		/	Multiple types/ Not specified
Ishikawa et al. (2022)		/	/	/		/		/	Diabetes
McMaster et al. (2023)		/	/	/		/		/	Multiple types/ Not specified
Chaichulee et al. (2022)	/	/	/	/		/	/	/	Multiple types/ Not specified
Zitu et al. (2023)	/	/	/	/		/		/	Immune Checkpoint Inhibitors

หมายเหตุ: เทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิม (TML) เทคนิคการเรียนรู้เชิงลึก (DL) การประมวลผลภาษาธรรมชาติ (NLP), แบบจำลองทางภาษา (LM)

จากการศึกษางานวิจัยที่ผ่านมา พบว่างานวิจัยที่ใช้ข้อมูลภาษาไทยมีการศึกษาค่อนข้างน้อย เนื่องจากมีความท้าทายหลายประการ (Chaichulee et al., 2022; Kunakorntham et al., 2022) อาทิ ความแตกต่างระหว่างภาษาพูดและภาษาเขียน มีคำศัพท์เฉพาะทาง และความหลากหลายของศัพท์อาการในภาษาไทยที่ยังไม่เป็นมาตรฐาน อีกทั้งยังไม่มีแบบจำลองที่ถูกฝึกอบรมล่วงหน้าด้วยข้อมูลขนาดใหญ่ที่เฉพาะเจาะจงสำหรับข้อมูลอาการไม่พึงประสงค์จากการใช้ยาในภาษาไทย นอกจากนี้ได้ศึกษางานวิจัยทางด้านอาการไม่พึงประสงค์จากการใช้ยาอีกหลายงานวิจัย พบข้อด้อยทางด้านเทคนิคซึ่งสามารถสรุปได้ 3 ประเด็น 1) ข้อมูล ปัญหาขนาดตัวอย่างที่ใช้สร้างแบบจำลองมีจำนวนน้อย และขาดรายละเอียดทางการแพทย์ งานวิจัยหลายงานพบปัญหาการขาดข้อมูลจำเพาะทาง เช่น การวินิจฉัยจากแพทย์ หรือข้อมูลผู้ป่วยที่เป็นรายบุคคล ส่งผลให้ความแม่นยำของแบบจำลองลดลงเมื่อใช้งานกับข้อมูลจริงหรือนำไปประยุกต์ใช้กับบริบทอื่น ๆ ได้ยาก (Nafea et al., 2024; Nishioka et al., 2022; Roosan et al., 2022; Wu et al., 2022) งานวิจัยที่ใช้ชุดข้อมูลเก่า ซึ่งอาจไม่สอดคล้องกับสถานการณ์ปัจจุบันของ (Tan et al., 2022) งานวิจัยที่ใช้ข้อมูลเพียงแหล่งเดียว มักมีปัญหาที่ข้อมูลมาจากโรงพยาบาลแห่งเดียวในประเทศไทย ทำให้ความชุกของข้อมูลกรณีที่ศึกษาต่ำและไม่สามารถใช้แบบจำลองกับสถาบันที่มีโครงสร้างข้อมูลแตกต่างออกไปได้ (Chaichulee et al., 2022; Kunakorntham et al., 2022) 2) การประมวลผลและทรัพยากร งานวิจัยหลายงานประสบปัญหาด้านการประมวลผลและทรัพยากรที่จำเป็นต่อการวิเคราะห์ข้อมูลเชิงลึก มีการใช้ทรัพยากรในการประมวลผลสูง เนื่องจากความซับซ้อนของข้อมูลและการปรับแต่งพารามิเตอร์ (Ahmad et al., 2022; Nithinsha et al., 2023) มีงานวิจัยชี้ว่าการเรียนรู้เชิงลึกต้องใช้ทรัพยากรและเวลาสูงซึ่งเพิ่มต้นทุนในการวิจัย (Dongre et al., 2023) ในด้านกระบวนการเตรียมข้อมูลและตรวจสอบข้อมูลที่จำเป็นโดยเฉพาะจากเภสัชกรมีความซับซ้อนและใช้เวลานาน (Kunakorntham et al., 2022; Roosan et al., 2022) อีกทั้งพบว่างานวิจัยหลายงานยังไม่มีทำการเปรียบเทียบประเด็นด้านเวลาที่ใช้ในการประมวลผลกับวิธีการอื่น ๆ ที่มีอยู่ และประเด็นที่ 3) เทคนิคที่ใช้ในแบบจำลอง มีงานวิจัยพบว่าประสิทธิภาพในการฝึกสอนลดลง โดยเฉพาะในกรณีของการรู้จำอาการไม่พึงประสงค์ที่มีคำที่สื่อความหมายเหมือนกันแต่อยู่ในรูปแบบต่างกัน (Wei et al., 2023) ส่วนเทคนิคการสกัดคุณลักษณะแบบดั้งเดิมอย่างเทคนิคหาความถี่คำ-ความถี่เอกสารผกผัน (TF-IDF) และการวิเคราะห์ความหมายแฝง (Latent Semantic Analysis : LSA) ยังมีข้อจำกัด จึงควรใช้เทคนิคการฝังคำที่ทันสมัยกว่า เช่น Word2Vec, GloVe และ BERT (Nafea et al., 2024; Rawat et al., 2022) สำหรับการใช้แบบจำลองเดี่ยวมักพบปัญหาความไม่เสถียร เช่น TextCNN แก้ไขโดยใช้เทคนิคการรวมผลลัพธ์จากหลายแบบจำลอง (Rawat et al., 2022) การใช้ DCNN เพียงอย่างเดียวยังไม่เพียงพอ ควรผสมผสานเทคนิคการเรียนรู้เชิงลึกอื่น ๆ เช่น LSTM และ Transformer (Spandana et al., 2024) นอกจากนี้ยังพบปัญหาขาดการเปรียบเทียบกับเทคนิคการเรียนรู้เชิงลึกขั้นสูงการทดสอบกับเทคนิคขั้นสูงที่มีประสิทธิภาพ เช่น CNN, RNN หรือ Transformer (Nafea et al., 2024)

ดังนั้น งานวิจัยนี้จึงมุ่งพัฒนาแบบจำลองที่สามารถจำแนกอาการไม่พึงประสงค์จากการใช้ยารักษาโรคทางจิตเวชจากข้อมูลเวชระเบียนอิเล็กทรอนิกส์ของผู้ป่วยในรูปแบบที่ไม่มีโครงสร้างที่เป็นภาษาไทยปนภาษาอังกฤษ และใช้เทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิม เทคนิคการเรียนรู้เชิงลึก ร่วมกับเทคนิคการ

ประมวลผลภาษาธรรมชาติในภาษาไทยและภาษาอังกฤษเข้ามาช่วยจำแนกอาการไม่พึงประสงค์จากการใช้ยา ได้แก่ เทคนิคเกาส์เซียนนาอีฟเบย์ (Gaussian Naïve Bay) เทคนิคเคเนียร์เรสเนเบอร์ (K-Nearest Neighbours) เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น (Linear Support Vector Machine) เทคนิคการถดถอยโลจิสติค (Logistic Regression) และเทคนิคแรนดอมฟอเรสต์ (Random Forest) ร่วมกับการสกัดคุณลักษณะ (Feature Extraction) ด้วยเทคนิคหาความถี่คำ-ความถี่เอกสารผกผัน (TF-IDF) และการเรียนรู้เชิงลึก 2 เทคนิค ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชั่น (Convolutional Neural Network : CNN) และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง (Bidirectional Long Short-Term Memory : Bi-LSTM) เพื่อนำเสนอการเปรียบเทียบประสิทธิภาพและความเร็วของเทคนิคที่ใช้ บนสมมติฐานที่ว่าแบบจำลองที่พัฒนาขึ้นสามารถจำแนกว่าอาการที่เกิดขึ้นเป็นอาการไม่พึงประสงค์จากการใช้ยาได้อย่างถูกต้อง ช่วยเฝ้าระวังสัญญาณความเสี่ยงจากอาการไม่พึงประสงค์หลังการใช้ยาที่แม่นยำและทันต่อเหตุการณ์ โดยมีค่าประสิทธิภาพโดยรวมในระดับที่ยอมรับได้

วัตถุประสงค์ของการวิจัย

เพื่อนำเสนอเทคนิคที่เหมาะสมสำหรับจำแนกอาการไม่พึงประสงค์จากการใช้ยารักษาโรคทางจิตเวชจากบันทึกรายงานอาการไม่พึงประสงค์จากการใช้ยาในรูปแบบข้อความโดยเปรียบเทียบเทคนิคการเรียนรู้ของเครื่อง

กระบวนการวิจัย

ชุดข้อมูล (Dataset)

การศึกษาครั้งนี้เก็บข้อมูลรายงานเหตุการณ์ไม่พึงประสงค์จากการใช้ยาของผู้ป่วยจิตเวชจากแผนกเภสัชกรรม ตั้งแต่เดือน ตุลาคม พ.ศ. 2562 ถึงเดือน กันยายน พ.ศ. 2566 จากบันทึกรายงานโดยบุคลากรทางการแพทย์ ได้แก่ พยาบาลและเภสัชกร ประกอบด้วยข้อมูลจำนวน 6 คอลัมน์ ได้แก่ อาการไม่พึงประสงค์ที่เกิดขึ้น (ADE) ยาที่มีประวัติแพ้ (Drug_allergy) ยาที่ทำให้เกิดอาการไม่พึงประสงค์ (Drug_ADR) ประเภทอาการไม่พึงประสงค์ตามการป้องกัน (Preventable_type) ระดับความเกี่ยวข้อง (Possibility) และระดับความรุนแรง (Severity) ทั้งสิ้น 1,652 แถวข้อมูล ตัวอย่างข้อมูลดังตารางที่ 2 (Table 2) โดยข้อมูลทั้งหมดผ่านการรับรองจริยธรรมการวิจัยจากคณะกรรมการจริยธรรมการวิจัยในมนุษย์และได้รับอนุญาตการนำข้อมูลมาใช้ในเชิงวิจัยโดยไม่เปิดเผยตัวบุคคล ข้อมูลทุกระเบียนถูกลบข้อมูลส่วนบุคคล ได้แก่ ประจำตัวผู้ป่วย (HN) และแทนที่ด้วยรหัสผู้ป่วยเพื่อให้ไม่สามารถระบุตัวบุคคลได้ตามมาตรฐานการคุ้มครองข้อมูลส่วนบุคคล

ตารางที่ 2 (Table 2)

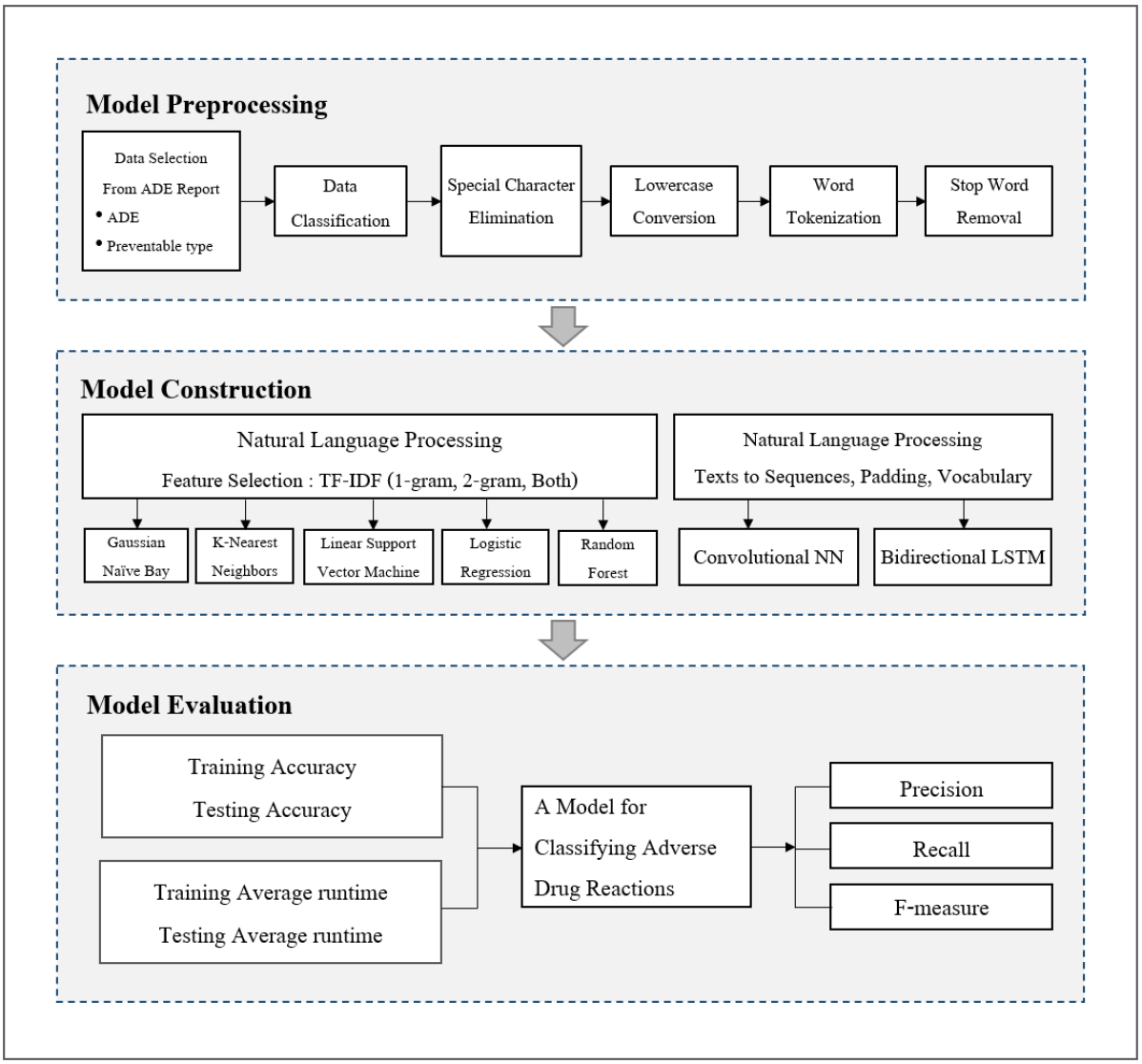
ตัวอย่างข้อมูลจากรายงานเหตุการณ์ไม่พึงประสงค์ (Sample Data from Adverse Event Report)

อาการไม่พึงประสงค์ที่เกิดขึ้น (Adverse Drug Event)	**การแพ้ยา** (Drug_allergy)	**ยาที่มีประวัติแพ้** (Drug_ADR)	**ประเภทอาการไม่พึงประสงค์ตามการป้องกัน** (Preventable_type)	**ระดับความเกี่ยวข้อง** (Possibility)	**ระดับความรุนแรง** (Severity)
น้ำนมไหลเวลาบีบ สีขาวขุ่น ถ้าไม่บีบไม่ไหล ประจำเดือนมาปกติ อาการอื่น ๆ ปกติ (Galactorrhea on expression, no spontaneous discharge, Regular periods, Others Regular)	Penicillin G Procaine	Risperidone	ไม่สามารถป้องกันได้ (Non-preventable ADR)	อาจจะใช่ (Possible)	D
ผู้ป่วยมีพูดเรื่อยเปื่อย ลิ้นแข็ง พูดไม่ค่อยชัด มีน้ำลายไหลมากกลางคืน (Patient presents with rambling speech and dysarthria secondary to lingual rigidity, with nocturnal sialorrhea)	None	Quetiapine	สามารถป้องกันได้ (Preventable ADR)	น่าจะใช่ (Probable)	E
พูดไม่ชัด คอบิดเกร็ง ตาเหลือบมองบน (Dysarthria, torticollis, and upward gaze deviation)	None	Clozapine, Haloperidol IM	ไม่สามารถป้องกันได้ (Non-preventable ADR)	น่าจะใช่ (Probable)	F

การเตรียมข้อมูลก่อนนำเข้าแบบจำลอง (Model Preprocessing)

การดำเนินการในหัวข้อนี้ ประกอบด้วย 3 ขั้นตอนย่อย ได้แก่ กระบวนการประมวลผลข้อมูล (Model Preprocessing) การสร้างแบบจำลอง (Model Construction) และการประเมินแบบจำลอง (Model Evaluation) แสดงดังภาพที่ 1 (Figure 1) มีรายละเอียดดังนี้

ภาพที่ 1 (Figure 1) *ภาพรวมของกระบวนการวิจัย (Overview of Research Methodology)*



ขั้นตอนที่ 1 กระบวนการเลือกข้อมูล (Data Selection) ในการวิจัยครั้งนี้มุ่งเน้นการพัฒนาแบบจำลองจำแนกข้อความจากคำบรรยายอาการเป็นหลัก จึงใช้ข้อมูลจำนวน 2 คอลัมน์ ได้แก่ คอลัมน์เหตุการณ์ไม่พึงประสงค์ที่เกิดขึ้นจากยา (ADE) เป็นคอลัมน์ที่อยู่ในรูปแบบข้อความบรรยายลักษณะอาการไม่พึงประสงค์จากการใช้ยา และคอลัมน์การแบ่งประเภทตามการป้องกัน (Preventable type) เป็นคอลัมน์ป้ายกำกับ (Label) สำหรับจำแนกว่าเป็น ADR หรือไม่ ซึ่งเป็นผลจากการวิเคราะห์ตามแนวปฏิบัติของโรงพยาบาล จากนั้นเลือกเฉพาะแถวข้อมูลที่สมบูรณ์ กล่าวคือ ต้องมีรายละเอียดอาการไม่พึงประสงค์ที่เกิดขึ้น และผลการประเมินประเภทการป้องกันที่ถูกบันทึกจากบุคลากรทางการแพทย์เท่านั้น พิจารณาตัดแถวที่มีข้อมูลไม่สมบูรณ์ (Missing Value) ได้แก่ ข้อมูลในคอลัมน์ ADE และ Preventable type ที่มีค่า NaN หรือ ค่าว่างออก เหลือแถวข้อมูลจำนวน 1,278 แถว จากทั้งหมด 1,652 แถวข้อมูล

ขั้นตอนที่ 2 การแบ่งกลุ่มข้อมูล (Data Classification) โดยพิจารณาข้อมูลในคอลัมน์ Preventable Type หากข้อมูลในคอลัมน์ดังกล่าว มีผลการประเมินของเภสัชกรอยู่ใน 2 กลุ่มนี้ กลุ่มที่ 1 ADR ที่ไม่สามารถป้องกันได้ (Non-Preventable ADR) ข้อมูลมีค่าเป็น “Non-preventable ADR“ หรือ “Not preventable ADR” หรือกลุ่มที่ 2 ADR ที่สามารถป้องกันได้ (Preventable ADR) ข้อมูลมีค่าเป็น "ขนาดยา วิถี ขนาด ความถี่ของการบริหารยาไม่เหมาะสม", "Preventable ADR", "มีประวัติแพ้ยาหรือ ADR จากยานี้มาก่อน", "ระดับยาในเลือดสูงกว่าระดับการรักษา หรือ non-adherence", "Drug interaction", "ไม่เหมาะสมต่อโรค", "เกี่ยวกับยา", "ไม่มีการตรวจวัดระดับยาในเลือด", "มีความถี่ในการตรวจไม่เหมาะสม" ให้กำหนดค่าข้อมูล Class ใหม่เป็น ADR หรือ Class 0 แต่หากค่าในคอลัมน์ Preventable type มีค่าผลการประเมินเป็น “Not ADR” หรือ “ไม่สงสัย ADR” ให้กำหนดค่าข้อมูล Class ใหม่เป็น “Not ADR” หรือ Class 1 แต่ละ Class มีรายละเอียด ดังนี้

ADR ให้ Class เท่ากับ 0 มีจำนวน 1,055 แถว

Not ADR ให้ Class เท่ากับ 1 หรือ มีจำนวน 223 แถว

ขั้นตอนที่ 3 การกำจัดอักขระพิเศษและตัวเลข (Special Character Elimination) นำข้อมูลในคอลัมน์ ADE มาลบอักขระพิเศษและตัวเลข แทนที่ลงไปด้วยช่องว่าง หลังดำเนินการข้อมูลคอลัมน์ ADE มีความยาวตัวอักษรเฉลี่ย 39 ตัวอักษร จำนวนคำโดยเฉลี่ย 4 คำ ประโยคที่มีจำนวนคำมากที่สุด 21 คำ

ขั้นตอนที่ 4 การแปลงอักษรภาษาอังกฤษเป็นตัวเล็ก (Lowercase Conversion) แปลงคอลัมน์ ADE ที่มีคำภาษาอังกฤษทั้งหมดเป็นตัวเล็ก เพื่อให้แบบจำลองทราบว่าเป็นคำเดียวกัน

ขั้นตอนที่ 5 การตัดคำ (Word Tokenization) ดำเนินการตัดคำข้อมูลข้อความในคอลัมน์ ADE โดยใช้ไลบรารี PyThaiNLP เพื่อแยกข้อความเป็นหน่วยย่อยสำหรับการวิเคราะห์และประมวลผลภาษาธรรมชาติ (NLP)

ขั้นตอนที่ 6 การกำจัดคำหยุด (Stop Word Removal) ลบคำหยุดหรือคำทั่วไปที่พบบ่อยในประโยคภาษาไทยและภาษาอังกฤษ เพื่อเพิ่มประสิทธิภาพของแบบจำลองและลดสัญญาณรบกวน เนื่องจากคำเหล่านี้ไม่ช่วยบ่งบอกอาการไม่พึงประสงค์จากการใช้ยาโดยตรงและพบได้ในทุกข้อความ ช่วยให้แบบจำลองมุ่งเน้นไปที่คำสำคัญและลดการเรียนรู้ข้อมูลผิดพลาด ลบคำหยุดในภาษาไทยด้วย Library PythaiNLP อาทิ “ความ” “การ” “มี” “ไม่” “ช้า” “ไว” “เยอะ” เป็นต้น นอกจากนี้ได้ลบคำที่พบบ่อยในบริบททางการแพทย์ภาษาไทยและคำที่ไม่บ่งบอกถึงอาการไม่พึงประสงค์จากการใช้ยา อาทิ คำว่า “ผป” (ย่อมาจากผู้ป่วย) “ผู้ป่วย”, “อาการ”, “แพทย์”, “พยาบาล”, “เภสัช”, “คนไข้”, “ยา”, “โรงพยาบาล” เป็นต้น และลบคำหยุดหรือคำทั่วไปที่พบบ่อยในประโยคภาษาอังกฤษ ด้วยไลบราลีของ nltk อาทิ “cpk”, “mg”, “bpm”, “kg”, “hr”, “oct” เป็นต้น หลังดำเนินการพบว่า คอลัมน์ ADE มีความยาวเฉลี่ยจำนวน 29 ตัวอักษร มีคำเฉลี่ยจำนวน 5 คำ และประโยคที่ยาวที่สุดมีจำนวนคำ 32 คำ

สำหรับตัวอย่างข้อมูลที่ผ่านกระบวนการประมวลผลข้อมูล แสดงดังตารางที่ 3 (Table 3)

ตารางที่ 3 (Table 3)

ตัวอย่างของการเตรียมข้อมูลล่วงหน้า (Example of Preprocessing)

	ขั้นตอน (Step)	**ผลลัพธ์คอลัมน์** (Column Result)		**หมายเหตุ** (Remark)
		อาการไม่พึงประสงค์ที่เกิดขึ้น (Adverse Drug Event)	**คลาส** (Class)	
1	การคัดเลือกข้อมูล (Data Selection)	ผื่นคันทั้งตัว หลังให้ ceftri 3 วัน (มี hx แพ้ Amoxi แต่ทำ skin test ก่อนให้ ceftri neg) (generalized pruritic rash after ceftri administration for 3 days (patient has hx of Amoxi allergy but skin test prior to ceftri administration was neg))	ไม่สามารถป้องกันได้ (Not preventable)	การคัดเลือกข้อมูลจากคอลัมน์ Adverse Drug Event และประเภท Preventable (คอลัมน์ Class) (Selection of data from the Adverse Drug Event column and the Preventable type (Class column).
2	การจัดหมวดหมู่ข้อมูล (Data Classification)	ผื่นคันทั้งตัว หลังให้ ceftri 3 วัน (มี hx แพ้ Amoxi แต่ทำ skin test ก่อนให้ ceftri neg) (generalized pruritic rash after ceftri administration for 3 days (patient has hx of Amoxi allergy but did skin test prior to ceftri administration was neg))	0	การนิยามค่าใน Class (ADR = 0, ไม่ใช่ ADR = 1) (Definition of Class values.(ADR = 0, Not ADR = 1))
3	การลบอักขระพิเศษ (Special Character Elimination)	ผื่นคันทั้งตัว หลังให้ ceftri วัน มี hx แพ้ Amoxi แต่ทำ skin test ก่อนให้ ceftri neg (generalized pruritic rash after ceftri administration for days patient has hx of Amoxi allergy but skin test prior to ceftri administration was neg)	0	ลบเลข 3 และเครื่องหมายวงเล็บออก (Remove the number 3 and the parentheses.)

ตารางที่ 3 (Table 3) (Cont.)

ขั้นตอน (Step)		ผลลัพธ์คอลัมน์ (Column Result)		หมายเหตุ (Remark)
		อาการไม่พึงประสงค์ที่เกิดขึ้น (Adverse Drug Event)	คลาส (Class)	
4	การแปลงข้อความเป็นตัวพิมพ์เล็ก (Lowercase Conversion)	ผื่นคันทั้งตัว หลังให้ ceftri วัน มี hx แพ้ amoxi แต่ทำ skin test ก่อนให้ ceftri neg (generalized pruritic rash after ceftri administration for days patient has hx of amoxi allergy but skin test prior to ceftri administration was neg)	0	แปลงตัวอักษร 'A' ตัวใหญ่เป็น 'a' ตัวเล็ก (Convert the uppercase letter ‘A’ to lowercase ‘a’.)
5	การตัดแบ่งคำ (Word Tokenization)	ผื่นคัน, ทั้งตัว, หลัง, ให้, ceftri, วัน, มี, hx, แพ้, amoxi, แต่, ทำ, skin, test, ก่อน, ให้, ceftri, neg (pruritic rash, generalized, after, administration, ceftri, days, has, hx, allergy, amoxi, but, did, skin, test, prior to, administration, ceftri, neg)	0	แยกประโยคออกเป็นคำ (Split the sentence into words.)
6	การลบคำที่ไม่สำคัญ (Stop Word Removal)	ผื่นคัน, ceftri, hx, แพ้, amoxi, skin, test, ceftri, neg (pruritic rash, ceftri, hx, allergy, amoxi, skin, test, ceftri, neg)	0	นำคำที่ไม่จำเป็นออก ได้แก่ ‘ทั้งตัว’, ‘หลัง’, ‘ให้’, ‘วัน’, ‘มี’, ‘ก่อน’, ‘ให้’ (Remove stop words, including: ‘generalized’, ‘after’, ‘administration’, ‘days’, ‘has’, ‘prior to’, ‘administration’)

กระบวนการสร้างแบบจำลอง (Model Construction)

การวิจัยครั้งนี้ได้พัฒนาแบบจำลองสำหรับจำแนกอาการไม่พึงประสงค์จากการใช้ยา โดยทดสอบ 7 เทคนิค ได้แก่ การเรียนรู้ของเครื่องจำนวน 5 เทคนิคเกาส์เซียนนาอีฟเบย์ เทคนิคเคเนียร์เรสเนเบอร์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น เทคนิคการถดถอยโลจิสติค และเทคนิคแรนดอมฟอเรสต์ และการเรียนรู้เชิงลึก 2 เทคนิค ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชั่น และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง และใช้ข้อมูลที่ผ่านกระบวนการประมวลผลข้อมูลแล้วมาแบ่งข้อมูล

เป็นข้อมูลชุดฝึกสอน 70% และข้อมูลชุดทดสอบ 30% ซึ่งแบบจำลองได้รับการฝึกอบรมบนข้อมูลชุดเดียวกันและทดสอบกับชุดข้อมูลที่แตกต่างกัน มีกระบวนการดังนี้

การประมวลผลภาษาธรรมชาติและการเรียนรู้ของเครื่อง (Natural Language Processing and Machine Learning)

นำข้อความรายงานเหตุการณ์ไม่พึงประสงค์ (ADE) และคลาส (Class) ที่ผ่านกระบวนการประมวลผลข้อมูล ทั้งหมด 1,278 เอกสาร มาแบ่งเป็นชุดข้อมูลเรียนรู้ (Training set) 70% จำนวน 894 ชุดเอกสาร และข้อมูลทดสอบ 30% จำนวน 384 เอกสาร (Test set) เพื่อสกัดคุณลักษณะ (Feature Extraction) จากชุดข้อมูลฝึกด้วยเทคนิคหาความถี่คำ-ความถี่เอกสารผกผัน (TF-IDF) เพื่อค้นหาคำ (Term) ที่สำคัญในเอกสาร (Document) โดยดูจากเนื้อหาของเอกสารทั้งหมด เป็นขั้นตอนที่ต้องดำเนินการก่อนนำข้อมูลเข้าแบบจำลอง เพื่อประเมินความถี่ของคำนั้นต่อเอกสารทั้งชุดเอกสาร มีสมการคำนวณดังสมการที่ 1

$$TF - IDF\ (t,d)\ =\ TF(t,d)\ \times\ IDF(t) \tag{1}$$

โดยที่ *TF* คือ ความถี่ของคำ ๆ *Term* (t) ปรากฏในเอกสาร *Document* (d)

IDF คือ จำนวนเอกสารที่มีคำที่สนใจเกิดขึ้น หากคำนั้นพบได้ทั่วไปในเอกสารต่าง ๆ บ่อย คำนั้นจะมีคะแนน *IDF* ต่ำ มีสูตรการคำนวณดังสมการที่ 2

$$IDF(t) = log\frac{n}{DF(t)} + 1 \tag{2}$$

โดยที่ *n* คือ จำนวนเอกสารทั้งหมด

DF คือ จำนวนเอกสารที่มีคำที่สนใจเกิดขึ้น

นำค่า *DF-TDF (V)* ที่คำนวณได้จากสมการที่ 1 มาทำ *Normalization* ดังสมการที่ 3

$$V_{norm} = \frac{V}{\sqrt{V_1^2+V_2^2+\cdots V_n^2}} \tag{3}$$

ขั้นตอนเริ่มตั้งแต่การนำเข้า TfidfVectorizer ด้วยไลบรารี่ sklearn จากนั้นนำข้อมูลชุดทดสอบ 70% มาคำนวณหาคะแนนความถี่คำ-ความถี่เอกสารผกผัน (TF-IDF) พบคุณลักษณะที่มีความถี่สูงในแต่ละคลาสเรียงลำดับคำที่มีความถี่สูงจากมากไปน้อย ดังแสดงในตารางที่ 4 (Table 4) และเมื่อกำหนดพารามิเตอร์รูปแบบของลำดับของจำนวนคำที่สนใจ หรือ N-Gram ทั้ง 3 แบบ แบบที่ 1 Unigram เช่นคำว่า ลิ้นแข็ง, น้ำนม เป็นต้น แบบที่ 2 Bigram เช่นคำว่า น้ำนม ไหล, กลืน ลำบาก เป็นต้น และแบบที่ 3 เป็นการนำคำทั้งใน Unigram และ Bigram รวมกัน เช่นคำว่า ลิ้นแข็ง, น้ำนม, น้ำนม ไหล, กลืน ลำบาก เป็นต้น พบว่า แต่ละรูปแบบมีจำนวนคุณลักษณะที่แตกต่างกันดังตารางที่ 5 (Table 5) จากนั้นพัฒนาแบบจำลองการเรียนรู้ของ

เครื่องด้วยเทคนิคเกาส์เซียนนาอีฟเบย์ เทคนิคเคเนียร์เรสเนเบอร์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น เทคนิคการถดถอยโลจิสติค และเทคนิคแรนดอมฟอเรสต์ กับ N-Gram ทั้ง 3 รูปแบบ ใช้ชุดข้อมูลฝึกสอนจำนวน 70% ให้แบบจำลองเรียนรู้ และเปรียบเทียบความถูกต้องในการฝึกแบบจำลอง จากนั้นใช้ชุดข้อมูลทดสอบ 30% ทดสอบแบบจำลองและเปรียบเทียบความถูกต้องของชุดข้อมูลทดสอบ กำหนดพารามิเตอร์โดยใช้เทคนิคการค้นหาไฮเปอร์พารามิเตอร์ที่เหมาะสมด้วยเทคนิคการตรวจสอบข้ามแบบกริด (Grid Search Cross-Validation) เริ่มจากกำหนดช่วงค่าของไฮเปอร์พารามิเตอร์ที่ต้องการทดสอบ จากนั้นทดสอบแบบจำลองด้วยไฮเปอร์พารามิเตอร์แต่ละชุด โดยใช้การตรวจสอบข้ามพบว่าที่ไฮเปอร์พารามิเตอร์ให้ผลลัพธ์ดีที่สุดมีดังนี้

```
models = [
        GaussianNB(priors=None, var_smoothing = 1e-09),
        neighbors.KNeighborsClassifier(algorithm = 'ball_tree',leaf_size = 40, metric=
            'euclidean' ,  n_neighbors = 11 , p = 1, weights = 'distance'),
        LinearSVC(C=1, class_weight = 'balanced', dual= False, loss = 'squared_hinge',
            max_iter = 1000, multi_class = 'ovr', penalty = 'l1',  random_state=0),
        LogisticRegression(C = 1, class_weight='balanced', multi_class = 'auto', max_iter=500,
            penalty = 'l2' , solver='liblinear', random_state=0),
        RandomForestClassifier(class_weight = 'balanced', criterion="entropy", max_depth =
            100, max_features= 'sqrt', min_samples_split= 6,
            n_estimators=50,random_state=0),
]
```

ตารางที่ 4 (Table 4)

ตัวอย่างคุณลักษณะของแต่ละคลาส (Sample Features for Each Class)

คลาส (Class)	**คุณลักษณะ** (Features)
0 หรือ ADR (0 or ADR)	ลิ้นแข็ง (Tardive dyskinesia), น้ำลายไหล (Hypersalivation), สั่น (Tremor), มือ (Hand), มือ สั่น (Hand tremor), ง่วง (Drowsiness), น้ำลาย (Drooling), ไหล (Drooling), ตัว (Body), ปาก (Mouth), ชัด (Clear Speech), เกร็ง (Rigidity), น้ำนม (Breast milk), น้ำนม ไหล (Galactorrhea), เดิน (Walking), แข็ง (Stiffness), ลำบาก (Difficulty), กลืน (Swallowing), หน้ามืด (Syncope), ลิ้น (Tongue), ตัว แข็ง (Muscle rigidity), คอแห้ง (Dry Throat), ลิ้นแข็ง ชัด (Dysarthria), ปาก แห้ง (Dry Mouth), แห้ง (Dryness), pt (Patient), แห้ง คอแห้ง (Dryness of throat), กลืน ลำบาก (Dysphagia), ต่ำ (Hypotension), เวียน (Vertigo), เซ (Unsteady gait), น้ำ (Fluid), eps (Extrapyramidal Symptoms), เข็ง (Stiffness), แข็ง (Rigidity), pulse (Pulse Rate), ลิ้นแข็ง กลืน (Dysphagia with Dysarthria), กิน (Eating), ซึม (Somnolence)

ตารางที่ 4 (Table 4) (Cont.)

คลาส (Class)	คุณลักษณะ (Features)
1 หรือ Not ADR (1 or Not ADR)	ผื่น (Rash), แจ้ง (Report), pulse (Pulse), ประเมิน (Assessment), ผื่นคัน (Rash), ปาก (Mouth), เวียน (Dizziness), กิน (Eating), เวียน หน้า (Dizziness), ลิ้นแข็ง (Tongue rigidity), คัน (Itching), เดิน (Walk), หน้า (Face), สงสัย (Suspected), pt (Patient), ทาน (Taking), ดี (Well), adr (Adverse Drug Reaction), มีค่า (Significant), อาเจียน (Vomiting), น้ำ (Fluid), เคี้ยว ปาก (Tardive dyskinesia), เคี้ยว (Chewing), eps (Extrapyramidal Symptoms), คลื่นไส้ (Nausea), แห้ง (Dryness), ปกติ (Normal), ปาก แห้ง (Dry Mouth), กลืน (Swallowing), ดีขึ้น (Improved), ลำบาก (Difficulty), ขา (Leg), กลืน ลำบาก (Dysphagia), ทางจิต (Psychiatric), ที่อ (Stiff), เวลา (Time), ตัว (Body), งง (Confusion),น้ำลายไหล (Sialorrhea)

ตารางที่ 5 (Table 5)

พารามิเตอร์ของ N-Gram และจำนวนฟีเจอร์ (Parameters of N-Gram and Number of Features)

ชุดคำต่อเนื่อง N คำ (N-Gram)	พารามิเตอร์ (Parameter)	จำนวนคุณลักษณะ (Features Number)
Unigram (1-gram)	ngram_range = (1, 1)	1,139
Bigram (2-gram)	ngram_range = (2, 2)	3,327
Unigram และ Bigram (Both)	ngram_range = (1, 2)	4,465

การประมวลผลภาษาธรรมชาติและการเรียนรู้เชิงลึก (Natural Language Processing and Deep Learning)

ก่อนนำข้อมูลเข้าไปฝึกในแบบจำลองการเรียนรู้เชิงลึก 2 เทคนิค ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชั่น และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง ได้ดำเนินการเตรียมข้อมูลข้อความรายงานเหตุการณ์ไม่พึงประสงค์ (ADE) ให้อยู่ในรูปแบบที่แบบจำลองสามารถประมวลผลได้ (Tokenize) โดยใช้ไลบราลีของ Keras กับชุดข้อมูลฝึก จำนวน 894 ชุดเอกสาร และข้อมูลทดสอบ จำนวน 384 เอกสาร ทำการแบ่งประโยคออกเป็นคำ (Texts to Sequences) แปลงแต่ละคำให้อยู่ในรูปแบบ Sequence หรือตัวเลขดัชนี (Index) กำหนดใช้ทุกคำในคลังคำศัพท์สำหรับข้อมูลการฝึก โดยกำหนด num_words=None จากคลังคำศัพท์ทั้งหมด 952 คำ (vocab_size) จากนั้นทำ pad_sequences ให้ทุกประโยคมีจำนวน Sequence ที่เท่ากัน เมื่อพิจารณาจากเอกสารที่มีความยาวสูงสุด 32 คำ และค่าเฉลี่ยจำนวนคำจากเอกสารทั้งหมดเท่ากับ 5 คำ จึงเติมค่า 0 ต่อท้ายใน Array โดยกำหนด padding=post และ maxlen=32 (จำนวนความยาวเมื่อเติม 0 แต่ละประโยคแล้ว) ผลลัพธ์ดังตารางที่ 6 (Table 6) ตัวอย่างการทำ Texts to Sequences และ Pad Sequences

ตารางที่ 6 (Table 6)

การแปลงข้อความเป็นลำดับตัวเลขและการเติมลำดับคำให้มีความยาวเท่ากัน (Texts to Sequences and Pad Sequences)

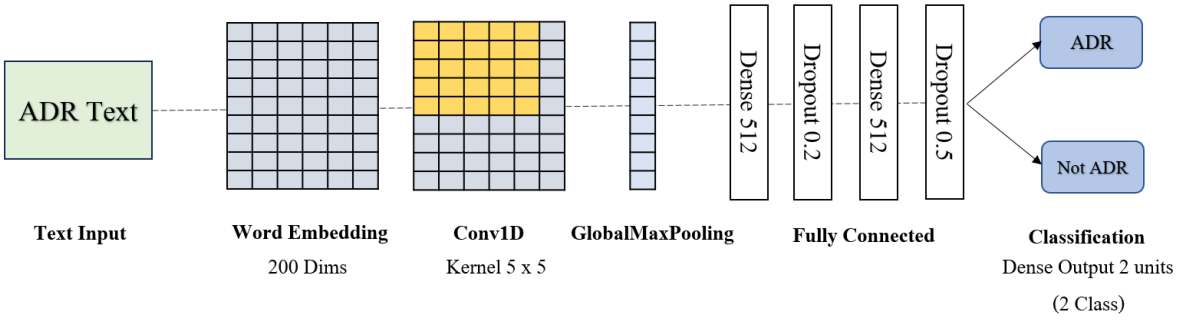
ขั้นตอน (Step)	พารามิเตอร์ (Parameter)	ผลลัพธ์ (Result)
1) ข้อมูลเหตุการณ์ไม่พึงประสงค์ต้นฉบับ (Original Adverse Drug Event Data)	-	ปาก (Mouth) แห้ง (Dry) คอแห้ง (Dry throat) ปวดหัว (Headache) ท้ายทอย (Back of neck) กลืน (Swallow) ข้าว (Rice) ลำบาก (Difficult) ฝืดคอ (Throat tightness)
2) แปลงข้อความเป็นลำดับตัวเลข (Texts to Sequences)	num_words= None, vocab_size=952	[5 18 20 107 322 15 108 14 241]
3) เติมลำดับให้มีความยาวเท่ากัน (Pad sequences)	padding='post', maxlen = 32	[5 18 20 107 322 15 108 14 241 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

จากนั้นพัฒนาแบบจำลองด้วยเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง และประเมินประสิทธิภาพแบบจำลองด้วยการเปรียบเทียบความถูกต้องในการฝึก (Training Accuracy) และความถูกต้องในทดสอบ (Test Accuracy) ดังนี้

แบบจำลองที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น มีโครงสร้างตามภาพที่ 2 (Figure 2) โดยเริ่มจากชั้น Input รับข้อมูลที่มีความยาว 32 คำ ตามด้วยชั้น Embedding ประมวลผลชุดคำจำนวน 952 คำ ผลลัพธ์จำนวน 200 มิติ เพื่อแปลงคำเป็นเวกเตอร์มีความหมาย ถัดมาคือชั้น Conv1D ซึ่งมี Filters จำนวน 256 และ Kernel ขนาด 5x5 กำหนดชั้นการกระตุ้น (Activation) เป็น ‘ReLU’ (Rectified Linear Unit) เพื่อสกัดคุณลักษณะ และ GlobalMaxPooling1D เพื่อลดมิติและคัดเลือกข้อมูลสำคัญ ต่อด้วยชั้น Dense ขนาด 512 หน่วย พร้อมฟังก์ชัน ‘relu’ ตามด้วย Dropout อัตรา 0.2 และ Dense อีกชั้น ขนาด 512 หน่วย พร้อม Dropout อัตรา 0.5 ปิดท้ายด้วยชั้น Dense ขนาด 2 หน่วย มีฟังก์ชัน ‘softmax’ สำหรับจำแนกผลลัพธ์ 2 คลาส แบบจำลองถูกฝึกด้วยช่วยปรับน้ำหนักของแบบจำลอง ‘adam’ ฟังก์ชันการสูญเสียเป็น ‘sparse_categorical_crossentropy’ และประเมินผลด้วย ‘accuracy’ เป็นตัวชี้วัด

ภาพที่ 2 (Figure 2)

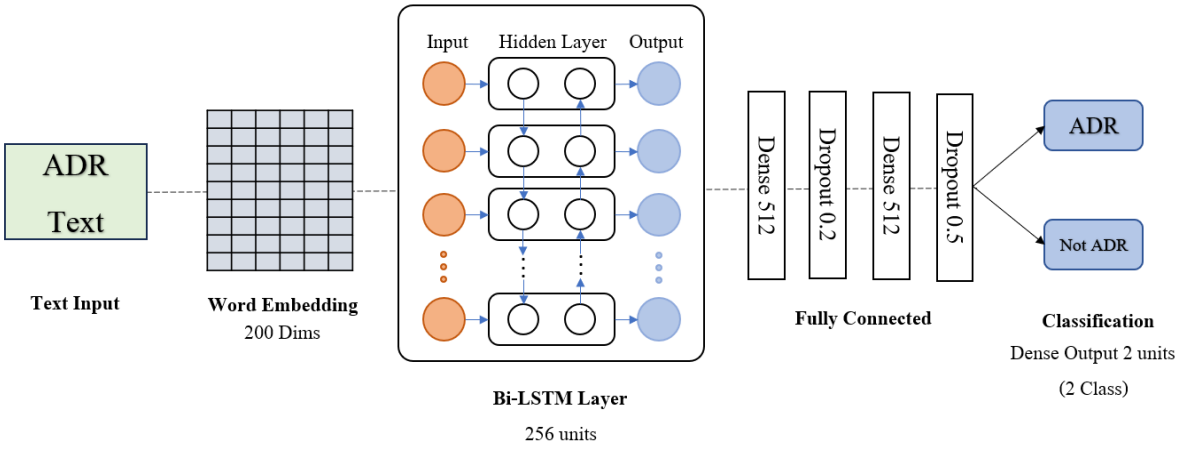
สถาปัตยกรรมของแบบจำลองโดยใช้เทคนิค CNN (Architecture of the Model Using CNN Technique)



แบบจำลองที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง มีโครงสร้างตามภาพที่ 3 (Figure 3) โดยเริ่มจากชั้นนำเข้า (Input) รับข้อมูลที่มีความยาว 32 คำ ตามด้วยชั้น Embedding ประมวลผลชุดคำจำนวน 952 คำ ผลลัพธ์จำนวน 200 มิติ เพื่อแปลงคำเป็นเวกเตอร์มีความหมาย ถัดมาคือชั้น Bidirectional(LSTM(256)) ซึ่งเป็นโครงข่าย LSTM แบบสองทิศทาง ที่มีขนาด 256 หน่วย ทำให้สามารถวิเคราะห์ความสัมพันธ์ของข้อมูลได้ทั้งจากซ้ายไปขวาและขวาไปซ้าย จึงจับบริบทของข้อมูลได้อย่างครบถ้วน เหมาะสำหรับงานที่ต้องการพิจารณาความเชื่อมโยงของข้อมูลในทุกมิติ ต่อด้วยชั้น Dense ขนาด 512 หน่วย พร้อมฟังก์ชัน ‘ReLU’ ตามด้วย Dropout อัตรา 0.2 และ Dense อีกชั้นขนาด 512 หน่วย พร้อม Dropout อัตรา 0.5 ปิดท้ายด้วยชั้น Dense ขนาด 2 หน่วย มีฟังก์ชัน ‘softmax’ สำหรับจำแนกผลลัพธ์ 2 คลาส แบบจำลองถูกฝึกด้วยช่วยปรับน้ำหนักของแบบจำลอง ‘adam’ ฟังก์ชันการสูญเสียเป็น ‘sparse_categorical_crossentropy’ และประเมินผลด้วย ‘accuracy’ เป็นตัวชี้วัด

ภาพที่ 3 (Figure 3)

สถาปัตยกรรมของแบบจำลองโดยใช้เทคนิค Bi-LSTM (Architecture of the Model Using Bi-LSTM Technique)



จากนั้นฝึกสอนแบบจำลองโดยใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง ด้วยชุดข้อมูลฝึกสอน (Training set) กำหนดรอบการฝึกสอน (epochs=20, batch_size=20) และใช้ชุดข้อมูลทดสอบ (Test set) เพื่อประเมินประสิทธิภาพแต่ละแบบจำลอง

การประเมินแบบจำลอง (Model Evaluation)

งานวิจัยนี้เปรียบเทียบประสิทธิภาพแบบจำลองที่พัฒนาด้วยเทคนิคทั้ง 7 เทคนิค โดยนำชุดข้อมูลฝึกไปฝึกสอนแบบจำลองและทดสอบแบบจำลองด้วยชุดข้อมูลทดสอบเพื่อเปรียบเทียบความถูกต้องในการฝึก (Training Accuracy) ความถูกต้องในการทดสอบ (Test Accuracy) เวลาในการสร้างแบบจำลอง

(Training Runtime) เวลาในการทดสอบแบบจำลอง (Testing Runtime) ค่าความแม่นยำ (Precision) ค่าระลึก (Recall) และค่าถ่วงดุล (F1-score) ดังสมการที่ 4-7

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \times 100 \quad (4)$$

$$Precision = \frac{TP}{(TP + FP)} \times 100 \quad (5)$$

$$Recall = \frac{TP}{(TP + FN)} \times 100 \quad (6)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{recision + Recall} \quad (7)$$

โดยที่

True Positive (*TP*) คือ สิ่งที่แบบจำลองทำนายว่า “จริง” และมีค่าเป็น “จริง”

True Negative (*TN*) คือ สิ่งที่แบบจำลองทำนายว่า “ไม่จริง” และมีค่า “ไม่จริง”

False Positive (*FP*) คือ สิ่งที่แบบจำลองทำนายว่า “จริง” แต่มีค่าเป็น “ไม่จริง”

False Negative (*FN*) คือ สิ่งที่แบบจำลองทำนายว่า “ไม่จริง” แต่มีค่าเป็น “จริง”

และเนื่องจากจำนวนข้อมูลในแต่ละคลาสไม่เท่ากันจึงหาค่าประสิทธิภาพโดยรวมแบบถ่วงน้ำหนัก (Weighted Average F1-Score) ดังสมการที่ 8

$$Weighted\ Avg\ F1\ Score\ = \sum_{i=1}^{N} w_i \times F1\ Score_i \quad (8)$$

โดยที่ N คือ 2 (จำนวนคลาส) W_i คือ จำนวนข้อมูลในแต่ละคลาส I และ $F1\ Score_i$ คือ ค่าประสิทธิภาพโดยรวมของแต่ละคลาส i

ผลการศึกษา

ผลการพัฒนาแบบจำลองสำหรับจำแนกอาการไม่พึงประสงค์จากการใช้ยา โดยทดสอบกับเทคนิคการเรียนรู้ของเครื่อง และเทคนิคการเรียนรู้เชิงลึก และใช้ข้อมูลที่ผ่านกระบวนการประมวลผลข้อมูลแล้วมาแบ่งข้อมูลเป็นข้อมูลชุดฝึกสอน (Training set) ร้อยละ 70 และข้อมูลชุดทดสอบ (Test set) ร้อยละ 30 ดังนี้

ผลการเปรียบเทียบความถูกต้องของเทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิม

ตารางที่ 7 (Table 7) นำเสนอผลการเปรียบเทียบความถูกต้อง (Accuracy) ของแบบจำลองการเรียนรู้ของเครื่อง 5 เทคนิค ได้แก่ เทคนิคเกาส์เซียนนาอีฟเบย์ เทคนิคเคเนียร์เรสเนเบอร์ เทคนิคซัพพอร์ตเวกเตอร์

แมชชีนเชิงเส้น เทคนิคการถดถอยโลจิสติค และเทคนิคแรนดอมฟอเรสต์ โดยใช้คุณลักษณะที่ถูกสกัดด้วยเทคนิคหาความถี่ของคำ-ส่วนกลับความถี่ของเอกสารและทดสอบกับชุดข้อมูลที่มีการกำหนดค่า *n*-gram range ที่แตกต่างกัน 3 รูปแบบ ได้แก่ 1-gram, 2-gram, และ Both (1-gram และ 2-gram) การประเมินความถูกต้องแบ่งออกเป็น 2 ส่วน คือ ความถูกต้องบนชุดข้อมูลฝึก (Training Accuracy) และความถูกต้องบนชุดข้อมูลทดสอบ (Testing Accuracy)

เมื่อดูค่าเฉลี่ยของทุกเทคนิคบนผลลัพธ์บนชุดข้อมูลฝึก พบว่า *n*-gram ที่ให้ค่าความถูกต้องสูงสุดคือ 2-gram มีค่าเฉลี่ยความถูกต้องที่ร้อยละ 93.69 รองลงมาคือการใช้ทั้ง 1-gram และ 2-gram (Both) ที่ร้อยละ 93.31 ซึ่งมีความใกล้เคียงกัน ส่วน 1-gram มีค่าเฉลี่ยความแม่นยำต่ำสุดที่ร้อยละ 87.65 เทคนิคที่ให้ค่าความถูกต้องในการฝึกสูงที่สุดคือ เทคนิคเคเนียร์เรสเนเบอร์ร่วมกับการใช้ทั้ง 1-gram และ 2-gram (Both) มีค่าความถูกต้องสูงสุดที่ร้อยละ 98.09 รองลงมาคือเทคนิคแรนดอมฟอเรสต์กับการใช้ทั้ง 1-gram และ 2-gram (Both) มีค่าความถูกต้องอยู่ที่ร้อยละ 97.54 ส่วนค่าเฉลี่ยของทุกเทคนิคบนผลลัพธ์บนชุดข้อมูลทดสอบ พบว่า *n*-gram ที่ให้ค่าความถูกต้องสูงสุดคือเทคนิคแรนดอมฟอเรสต์ กับการใช้ทั้ง 1-gram และ 2-gram (Both) มีค่าความถูกต้องสูงสุดที่ร้อยละ 84.64 รองลงมาคือเทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น กับ 2-gram ที่ร้อยละ 83.59 จากผลวิจัยพบว่าเทคนิคแรนดอมฟอเรสต์ให้ผลลัพธ์ดีที่สุดสำหรับ 1-gram และ 2-gram (Both) ด้วยความถูกต้องร้อยละ 84.64 ส่วนเทคนิคเกาส์เซียนนาอีฟเบย์ ยังคงให้ผลลัพธ์ต่ำที่สุดในการทดสอบโดยเฉพาะกับ 1-gram (ร้อยละ 46.09) การใช้ 1-gram และ 2-gram (Both) จึงให้ผลลัพธ์ดีกว่าการใช้ 1-gram เพียงอย่างเดียวโดยเฉพาะในการใช้เทคนิคเรนดอมฟอเรสต์

ตารางที่ 7 (Table 7)

การเปรียบเทียบความถูกต้องของแบบจำลองโดยใช้เทคนิคการเรียนรู้ของเครื่อง 5 เทคนิคร่วมกับการสกัดคุณลักษณะ 3 แบบ (Accuracy Comparison of Models Using Five Machine Learning Techniques Combined with Three Feature Extraction)

เทคนิคการเรียนรู้ของเครื่อง (Machine Learning Techniques)	**ความถูกต้อง (%)** (Accuracy (%))					
	การฝึกฝน (Training)			**การทดสอบ** (Testing)		
ngram_range	1-gram	2-gram	Both	1-gram	2-gram	Both
GaussianNB	64.65	83.56	85.01	46.09	50.52	60.16
K-Nearest Neighbors	97.98	96.87	98.09	82.29	82.03	82.29
LinearSVC	90.60	96.09	93.18	77.08	83.59	80.21
Logistic Regression	87.92	95.52	92.73	72.92	77.86	74.48
Random Forest	97.09	96.42	97.54	82.81	82.03	84.64
ค่าเฉลี่ย (Average)	**87.65**	**93.69**	**93.31**	**72.24**	**75.21**	**76.35**

ตารางที่ 8 (Table 8) แสดงเวลาในการประมวลผลแบบจำลอง (Testing Runtime) ประยุกต์กับเทคนิค 5 ประเภท ได้แก่ เทคนิคเกาส์เซียนนาอีฟเบย์ เทคนิคเคเนียร์เรสเนเบอร์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น เทคนิคการถดถอยโลจิสติค และเทคนิคแรนดอมฟอเรสต์ โดยแบ่งการวัดผลเป็นสองส่วน คือข้อมูลฝึกสอน และข้อมูลทดสอบ โดยทดสอบกับค่า ngram_range ที่แตกต่างกัน 3 รูปแบบคือ 1-gram, 2-gram, และ Both (1-gram และ 2-gram) พบว่าในการฝึกแบบจำลองโดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น และเทคนิคการถดถอยโลจิสติค มีเวลาการประมวลผลที่เร็วที่สุด โดยใช้เวลาประมาณ 18-35 มิลลิวินาที เทคนิคเกาส์เซียนนาอีฟเบย์ใช้เวลามากขึ้นอย่างมีนัยสำคัญ โดยเฉพาะเมื่อประมวลผลรูปแบบ Both (1-gram และ 2-gram) ใช้เวลาประมาณ 1,491 มิลลิวินาที ส่วนเทคนิคเคเนียร์เรสเนเบอร์ ใช้เวลาในการประมวลผลที่สูงที่สุด โดยเฉพาะเมื่อใช้รูปแบบ Both (1-gram และ 2-gram) ใช้เวลาประมาณ 13,596 มิลลิวินาที การทดสอบแบบจำลองโดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น และเทคนิคการถดถอยโลจิสติค ยังคงมีประสิทธิภาพดีที่สุดกับรูปแบบ 1-gram หรือ 2-gram ใช้เวลาประมาณ 4-5 มิลลิวินาที แต่เมื่อใช้กับรูปแบบ Both (1-gram และ 2-gram) เวลาการทดสอบเพิ่มขึ้นเป็น 7-9 มิลลิวินาที เทคนิคเคเนียร์เรสเนเบอร์ มีเวลาการทดสอบที่สูงที่สุดในทุกกรณี โดยเฉพาะเมื่อใช้กับรูปแบบ Both (1-gram และ 2-gram) ใช้เวลาประมาณ 6,973 มิลลิวินาที การใช้เทคนิคเกาส์เซียนนาอีฟเบย์ มีเวลาการทดสอบที่ค่อนข้างดี ใช้เวลาประมาณ 12-36 มิลลิวินาที ส่วนเทคนิคแรนดอมฟอเรสต์ มีเวลาการทดสอบที่สูงกว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น เทคนิคการถดถอยโลจิสติค และเทคนิคเกาส์เซียนนาอีฟเบย์ แต่ยังน้อยกว่าเทคนิคเคเนียร์เรสเนเบอร์ โดยเฉพาะเมื่อใช้รูปแบบ Both (1-gram และ 2-gram) โดยใช้เวลาประมาณ 423 มิลลิวินาที

ตารางที่ 8 (Table 8)

การเปรียบเทียบเวลาในการประมวลผลของแบบจำลองโดยใช้เทคนิคการเรียนรู้ของเครื่อง 5 เทคนิคร่วมกับการสกัดคุณลักษณะ 3 แบบ (Runtime Comparison of Models Using Five Machine Learning Techniques Combined with Three Feature Extraction)

เทคนิคการเรียนรู้ของเครื่อง (Machine Learning Techniques)	**เวลาในการประมวลผล (มิลลิวินาที)** (Runtime (Milliseconds))					
	การฝึกฝน (Training)			**การทดสอบ** (Testing)		
พารามิเตอร์ที่ใช้กำหนดขนาดของกลุ่มคำ (ngram_range)	1-gram	2-gram	Both	1-gram	2-gram	Both
GaussianNB	43	513	1,491	13	21	36
K-Nearest Neighbors	3,090	9,595	13,596	1,242	3,629	6,973
LinearSVC	19	23	35	4	6	7
Logistic Regression	18	23	33	4	6	10
Random Forest	1,112	2,838	3,334	23	42	423
ค่าเฉลี่ย (Average)	**856**	**2,598**	**3,698**	**257**	**741**	**1,490**

ผลการเปรียบเทียบความถูกต้องของเทคนิคการเรียนรู้เชิงลึก

ผลการพัฒนาแบบจำลองด้วยเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง แสดงดังตารางที่ 9 (Table 9) แบบจำลองที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น มีประสิทธิภาพสูงกว่าโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง ทั้งการฝึกสอนด้วยข้อมูลฝึก การใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น พบว่ามีความถูกต้องเมื่อใช้ชุดข้อมูลฝึกฝนร้อยละ 98 และความถูกต้องในชุดข้อมูลทดสอบร้อยละ 83 ส่วนเทคนิคโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง ความถูกต้องเมื่อใช้ชุดข้อมูลฝึกฝนร้อยละ 98 และความถูกต้องในชุดข้อมูลทดสอบร้อยละ 82 นอกจากนี้ ประสิทธิภาพด้านเวลาประมวลผลของเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น ยังดีกว่าทั้งการฝึกและการทดสอบแบบจำลอง โดยใช้เวลาในการฝึก 641,214 มิลลิวินาที เทียบกับ 1,548,725 มิลลิวินาทีของเทคนิคโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง และใช้เวลาทดสอบ 13,097 มิลลิวินาที เทียบกับ 28,064 มิลลิวินาทีของโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทางเช่นเดียวกัน

ตารางที่ 9 (Table 9)

การเปรียบเทียบความแม่นยำและระยะเวลาการทำงานของโมเดลโดยใช้เทคนิคการเรียนรู้เชิงลึก (Comparison of Accuracy and Runtime of Models Using Deep Learning Techniques)

เทคนิคการเรียนรู้เชิงลึก (Deep Learning Techniques)	**ความถูกต้อง (%)** (Accuracy (%))		**เวลาในการประมวลผล (มิลลิวินาที)** (Runtime (Milliseconds))	
	การฝึกฝน (Training)	**การทดสอบ** (Testing)	**การฝึกฝน** (Training)	**การทดสอบ** (Testing)
CNN	98.10	82.55	641,214	13,097
Bi-LSTM	97.65	81.77	1,548,725	28,064

ผลการเปรียบเทียบประสิทธิภาพของเทคนิคการจำแนกประเภท 7 เทคนิค โดยพิจารณาจากค่าความแม่นยำ ความระลึก และค่าถ่วงดุล บนชุดข้อมูลทดสอบ ดังตารางที่ 10 (Table 10) พบว่าเทคนิคแรนดอมฟอเรสต์ (Both) ให้ประสิทธิภาพดีที่สุดในด้านความแม่นยำ คิดเป็นร้อยละ 82.58 ตามด้วย เทคนิคการถดถอยโลจิสติก (Both) ที่ร้อยละ 81.93 และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น (Both) ที่ร้อยละ 80.72

ในส่วนของค่าความระลึก เทคนิคแรนดอมฟอเรสต์ (Both) ยังคงให้ผลลัพธ์สูงสุดที่ 84.64 รองลงมาคือเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน ร้อยละ 82.55 และเทคนิคเคเนียร์เรสเนเบอร์ (Both) ร้อยละ 82.29

เมื่อพิจารณาค่าถ่วงดุลระหว่างค่าความแม่นยำและความระลึก พบว่า เทคนิคแรนดอมฟอเรสต์ (Both) ยังคงมีประสิทธิภาพสูงสุดที่ร้อยละ 83.15 ตามด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชัน ร้อยละ 81.40 และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง ร้อยละ 81.10 ตามลำดับ

ผลการทดลองชี้ให้เห็นว่าเทคนิคแรนดอมฟอเรสต์ (Both) มีความสมดุลและประสิทธิภาพโดยรวมที่ดีที่สุดในการจำแนกประเภทข้อมูลชุดนี้ ในขณะที่เทคนิคเกาส์เซียนนาอีฟเบย์ (Both) ให้ประสิทธิภาพต่ำที่สุดในทุกตัวชี้วัด

ตารางที่ 10 (Table 10)

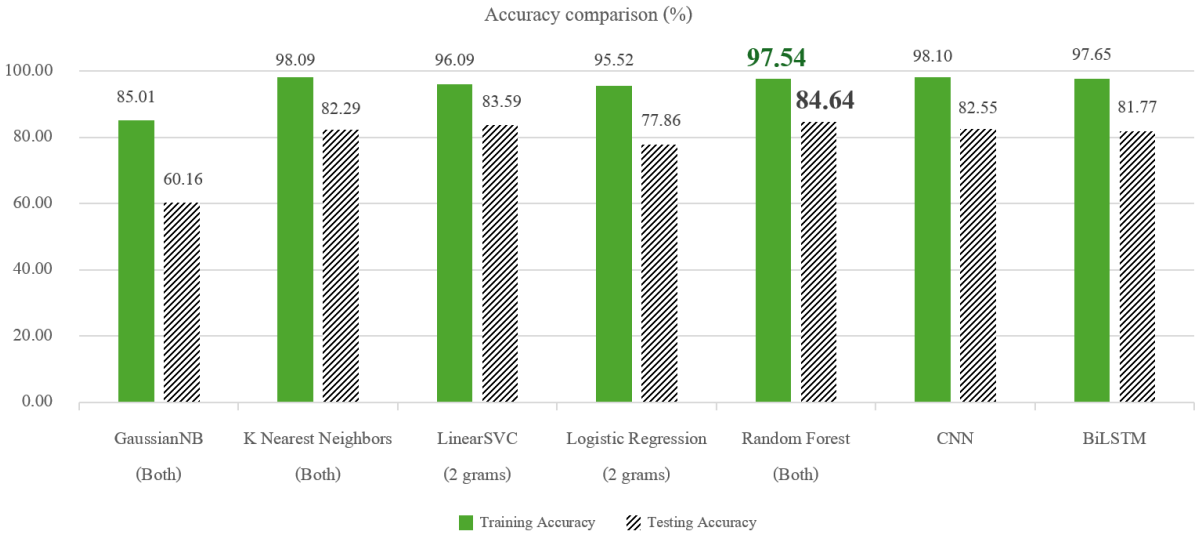
การเปรียบเทียบค่าความแม่นยำ ความระลึก และค่าถ่วงดุลบนชุดข้อมูลทดสอบด้วย 7 เทคนิค (Comparison of Accuracy, Recall, and F1-Score on the Test Dataset Using Seven Techniques)

เทคนิค (Techniques)	ความแม่นยำ (%) (Precision (%))	ความระลึก (%) (Recall (%))	ค่าถ่วงดุล (F1-score (%))
GaussianNB (Both)	74.58	60.16	65.20
K-Nearest Neighbors (Both)	75.32	82.29	77.31
LinearSVC (Both)	80.72	80.21	80.46
Logistic Regression (Both)	81.93	74.48	77.07
Random Forest (Both)	**82.58**	**84.64**	**83.15**
CNN	80.29	82.55	81.40
Bi-LSTM	80.44	81.77	81.10

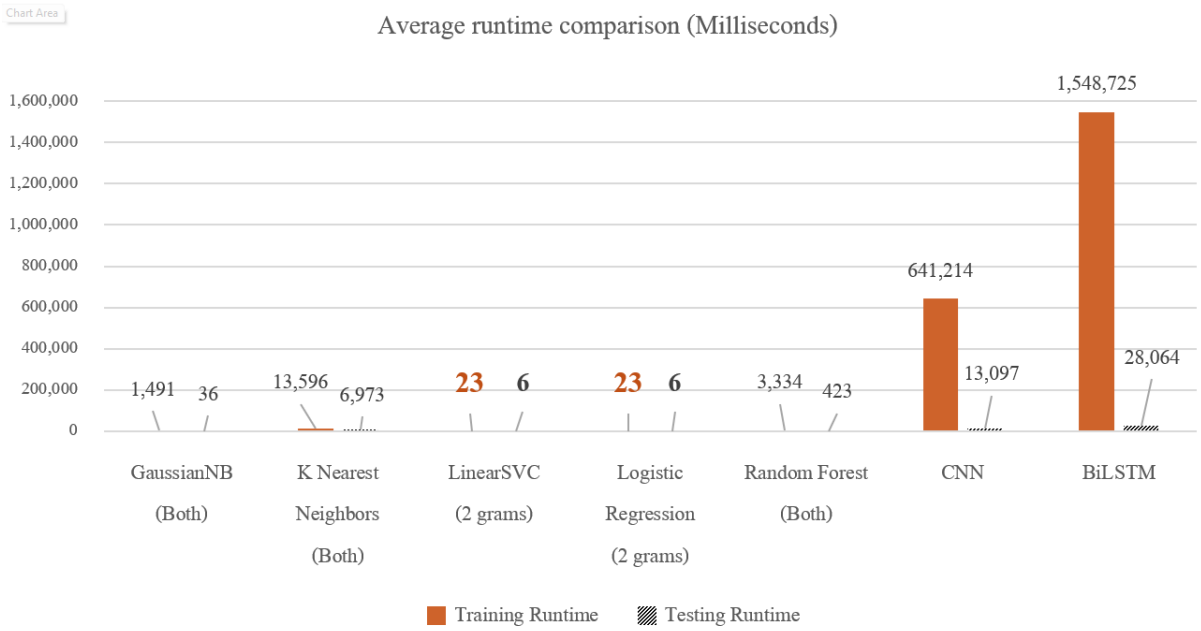
การเปรียบเทียบประสิทธิภาพแบบจำลองที่ประยุกต์ใช้เทคนิค 7 แบบที่แตกต่างกัน ในมิติความถูกต้องมีผลลัพธ์ดังภาพที่ 4 (Figure 4) ซึ่งแสดงให้เห็นว่าแบบจำลองที่พัฒนาด้วยเทคนิคแรนดอมฟอเรสต์มีความถูกต้องสูงที่สุด และผลการเปรียบเทียบแบบจำลองในมิติด้านเวลา ภาพที่ 5 (Figure 5) แสดงให้เห็นว่า แบบจำลองที่พัฒนาด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น และเทคนิคการถดถอยโลจิสติค มีความเร็วในการประมวลผลดีที่สุด

ภาพที่ 4 (Figure 4)

การเปรียบเทียบความถูกต้อง (Accuracy Comparison)



ภาพที่ 5 (Figure 5) การเปรียบเทียบเวลาในการประมวลผล (Runtime Comparison)



อภิปรายผลและสรุปผลการวิจัย

การศึกษานี้มุ่งพัฒนาแบบจำลองสำหรับจำแนกอาการไม่พึงประสงค์จากการใช้ยา โดยวิเคราะห์ข้อมูลจากเวชระเบียนอิเล็กทรอนิกส์ที่อยู่ในรูปแบบข้อความภาษาไทยปนภาษาอังกฤษ ผลการศึกษาแสดงให้เห็นว่า การประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกหลายวิธีมีศักยภาพสูงในการวิเคราะห์และจำแนกรายงานอาการไม่พึงประสงค์จากการใช้ยา การสร้างแบบจำลองพื้นฐาน (Baseline Model) ในงานวิจัยครั้งนี้เน้นให้สามารถตีความได้ง่ายและปรับใช้ได้สะดวก เหมาะสมกับข้อมูลข้อความที่มีจำนวนน้อยและความยาวไม่มาก โดยทดสอบเทคนิคทั้ง 7 รูปแบบ ได้แก่ เทคนิคการเรียนรู้ของเครื่อง 5 แบบ ได้แก่ เทคนิคเกาส์เซียนนาอีฟเบย์ เทคนิคเคเนียร์เรสเนเบอร์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น เทคนิคการถดถอยโลจิสติค และเทคนิคแรนดอมฟอเรสต์ รวมถึงการเรียนรู้เชิงลึก 2 แบบ ได้แก่ เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทาง ผลการวิจัยพบว่า เมื่อใช้เทคนิคแรนดอมฟอเรสต์กับการสกัดคุณลักษณะด้วยเทคนิคหาความถี่คำ-ความถี่เอกสารผกผัน (TF-IDF) ร่วมกับการใช้ ngram_range แบบ Both (1-gram และ 2-gram) ให้ค่าความถูกต้องสูงที่สุด ร้อยละ 84.6 ค่าความแม่นยำ ร้อยละ 82.6 ความระลึก ร้อยละ 84.6 และค่าถ่วงดุลร้อยละ 83.2 แสดงให้เห็นว่าการพิจารณาคู่คำ (2-gram) มีความสำคัญในการจับความหมายและบริบทของเอกสารทางการแพทย์ภาษาไทย ซึ่งสอดคล้องกับงานวิจัยของ แทนและคณะ (Tan et al., 2022) ที่พบว่าเทคนิคแรนดอมฟอเรสต์เป็นหนึ่งในเทคนิคที่มีประสิทธิภาพสูงในการจับความสัมพันธ์ระหว่างยากับอาการไม่พึงประสงค์ อย่างไรก็ตาม ผลการวิจัยนี้มีความแตกต่างจากงานของ วูและคณะ (Wu et al., 2022) ซึ่งได้ประเมินเทคนิคการเรียนรู้ของเครื่อง 18 แบบ โดยพบว่าแบบจำลองที่มีประสิทธิภาพสูงสุดใช้เทคนิคการ

รวมแบบจำลอง (Ensemble) ที่ให้ค่า AUC สูงถึงร้อยละ 91.4 และค่าความถูกต้องร้อยละ 89.4 ซึ่งมีเทคนิคการถดถอยโลจิสติค และเทคนิคแรนดอมฟอเรสต์ อยู่ในกลุ่มเทคนิคที่ใช้ในการศึกษา แต่การทดสอบเพียงเทคนิคใดเทคนิคหนึ่งไม่ได้ให้ประสิทธิภาพดีที่สุด ความแตกต่างนี้เนื่องมาจากลักษณะของชุดข้อมูล วิธีการประมวลผล การเลือกคุณลักษณะ และบริบทของภาษาที่ใช้ในการสร้างแบบจำลอง โดยเฉพาะอย่างยิ่งงานวิจัยนี้มุ่งเน้นการวิเคราะห์เวชระเบียนภาษาไทยปนภาษาอังกฤษ ซึ่งมีความซับซ้อนด้านโครงสร้างและการตีความมากกว่าภาษาอังกฤษเพียงอย่างเดียว

เมื่อพิจารณาประสิทธิภาพของเทคนิคการเรียนรู้เชิงลึกในงานวิจัยนี้ พบว่าแบบจำลองที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น มีค่าความถูกต้องร้อยละ 82.6 ค่าความแม่นยำ ร้อยละ 80.3 ความระลึก ร้อยละ 82.6 และค่าถ่วงดุลร้อยละ 81.4 ซึ่งสูงกว่าเทคนิคโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทางที่ได้ค่าความถูกต้อง ร้อยละ 81.8 ค่าความแม่นยำ ร้อยละ 80.4 ความระลึก ร้อยละ 81.8 และค่าถ่วงดุลร้อยละ 81.1 แสดงให้เห็นว่าโครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชั่นสามารถจัดการกับข้อมูลข้อความในลักษณะของปัญหานี้ได้อย่างมีประสิทธิภาพมากกว่า นอกจากนี้ค่าความถูกต้องของแบบจำลองที่ใช้โครงข่ายประสาทเทียมแบบคอนโวลูชั่น ยังใกล้เคียงกับผลการศึกษาที่นำเสนอเทคนิคการเพิ่มข้อมูลแบบ self-attention และการจัดการปัญหาข้อมูลไม่สมดุล ส่งผลให้ค่าความถูกต้อง เพิ่มขึ้นถึงประมาณร้อยละ 80 (Ishikawa et al., 2022)

ประเด็นประสิทธิภาพด้านเวลาในการประมวลผล (Time Efficiency) ซึ่งเป็นมิติสำคัญสำหรับการนำไปประยุกต์ใช้จริงทางคลินิก พบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น และเทคนิคการถดถอยโลจิสติคมีประสิทธิภาพสูงสุด ใช้เวลาในการประมวลผลเพียง 18-35 มิลลิวินาทีสำหรับการฝึกฝน และ 4-9 มิลลิวินาทีสำหรับการทดสอบ ในขณะที่เทคนิคการเรียนรู้เชิงลึกอย่างเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น-ยาวสองทิศทางใช้เวลามากกว่าหลายเท่า สอดคล้องกับข้อสังเกตในงานวิจัยที่ผ่านมา ที่ระบุว่าการเรียนรู้เชิงลึกต้องใช้ทรัพยากรและเวลาสูงซึ่งเพิ่มต้นทุนในการวิจัยและการนำไปใช้งาน (Ahmad et al., 2022; Dongre et al., 2023; Nithinsha et al., 2023) ดังนั้น การเลือกเทคนิคที่เหมาะสมสำหรับการนำไปพัฒนาระบบเพื่อต่อยอด จึงต้องพิจารณาความสมดุลระหว่างประเด็นความแม่นยำและความเร็วในการประมวลผล โดยในกรณีที่ต้องการระบบที่ทำงานแบบเรียลไทม์ การใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้นอาจเป็นตัวเลือกที่เหมาะสมกว่า แม้จะมีความแม่นยำที่ต่ำกว่าเล็กน้อย

อย่างไรก็ตาม มีข้อจำกัดด้านเทคนิคที่ใช้ในงานวิจัยครั้งนี้ ได้แก่ การใช้เทคนิคหาความถี่คำ-ความถี่เอกสารผกผัน (TF-IDF) ในการจับความหมายเชิงบริบทและความสัมพันธ์ระหว่างคำได้ยังไม่ดีนัก ซึ่งสามารถพิจารณาใช้เทคนิคการฝังคำขั้นสูง เช่น Word2Vec, GloVe หรือใช้แบบจำลองภาษาสำเร็จรูปในตระกูล Transformer เช่น BERT หรือ XLM-R ซึ่งอาจปรับปรุงประสิทธิภาพของแบบจำลองได้ดียิ่งขึ้น โดยเฉพาะกับข้อมูลภาษาไทยที่มีความซับซ้อนด้านโครงสร้างและความหมาย ตามข้อเสนอแนะของงานวิจัยของของ Rawat et al. (2022) และ Nafea et al. (2024) นอกจากนี้ยังพบว่า กระบวนการการตัดคำภาษาไทยยังมีความผิดพลาด เช่น ข้อความที่มีศัพท์เทคนิคทางการแพทย์และที่มีทั้งภาษาอังกฤษปน

ภาษาไทย ทำให้การวิเคราะห์เพื่อจำแนกอาการไม่พึงประสงค์จากการใช้ยามีความถูกต้องลดลงส่งผลต่อการวัดประสิทธิภาพของแบบจำลอง อีกทั้งแบบจำลองยังไม่รองรับข้อมูลนำเข้าที่เป็นคำที่ผิดทั้งภาษาไทยและภาษาอังกฤษ เช่น “ยุ่” “pluse” “ศรีษะ” เป็นต้น

ข้อเสนอแนะการพัฒนางานวิจัยในอนาคต ควรพัฒนาแบบจำลองด้วยเทคนิคการรวมแบบจำลอง (Ensemble) ที่รวมจุดแข็งของเทคนิคต่าง ๆ เข้าด้วยกัน อาจใช้การผสมผสานระหว่างเทคนิคแรนดอมฟอเรสต์ ที่มีความแม่นยำสูงกับเทคนิคการถดถอยโลจิสติค และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้นที่มีความเร็วในการประมวลผลสูง นอกจากนี้ในการสร้างและปรับปรุงคลังคำศัพท์อาการไม่พึงประสงค์ (ADR Lexicon) ให้ครอบคลุมคำที่มีความหมายคล้ายกันและรูปแบบการสะกดที่หลากหลาย เพื่อแก้ปัญหาที่เกิดจากการใช้คำย่อ คำทับศัพท์ และคำสะกดผิด ซึ่งพบได้บ่อยในข้อมูลเวชระเบียน รวมถึงความท้าทายของการตัดคำภาษาไทย (Tokenization Errors) ที่ไม่มีการเว้นวรรคระหว่างคำและการใช้ศัพท์ทางการแพทย์เฉพาะทาง (Domain-specific terms) ปัญหาเหล่านี้อาจทำให้แบบจำลองตีความผิดและมีความแม่นยำลดลง ดังนั้น การปรับปรุงระบบตัดคำ การสร้างแบบจำลองภาษาที่เรียนรู้จากข้อมูลสองภาษา ควบคู่กับการใช้เทคนิคฝังความหมายเชิงบริบทที่ทันสมัย เช่น Pre-trained Transformers จะช่วยเพิ่มประสิทธิภาพของแบบจำลองได้ ควรแก้ปัญหาข้อจำกัดในการจำแนกกลุ่ม Non-ADR อันเนื่องมาจากความไม่สมดุลของชุดข้อมูลด้วยวิธี อาทิ SMOTE และการสร้างข้อมูลเสริม (Data augmentation) อีกทั้ง ควรพัฒนาแบบจำลองภาษาทางการแพทย์เฉพาะทางภาษาไทย เพื่อรองรับข้อมูลเชิงลึกจากสาขาสุขภาพและเภสัชกรรมโดยตรง เพิ่มขนาดของข้อมูลโดยวิเคราะห์ข้อมูลทั้งเชิงโครงสร้างและไม่มีโครงสร้าง จากแหล่งต่าง ๆ เช่น เวชระเบียนอิเล็กทรอนิกส์หลายโรงพยาบาล รายงานอาการไม่พึงประสงค์ และข้อมูลจากเครือข่ายสังคมออนไลน์ เพื่อสร้างแบบจำลองที่มีความครอบคลุมและสามารถนำไปประยุกต์ใช้ได้อย่างกว้างขวาง ทำให้สามารถนำแบบจำลองไปใช้งานกับยากลุ่มอื่นหรือโรงพยาบาลอื่นได้

การศึกษาวิจัยนี้แสดงให้เห็นถึงศักยภาพในการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิมและเทคนิคการเรียนรู้เชิงลึกเพื่อการจำแนกอาการไม่พึงประสงค์จากการใช้ยา จากข้อมูลเวชระเบียนอิเล็กทรอนิกส์ภาษาไทย ผลการพัฒนาแบบจำลองสามารถเชื่อมโยงกับนโยบายด้านสาธารณสุขและการปฏิบัติทางคลินิกเพื่อพัฒนาระบบเฝ้าระวังอาการไม่พึงประสงค์จากการใช้ยาที่มีประสิทธิภาพ โดยเฉพาะในด้านการเพิ่มความรวดเร็วในการจำแนกและรายงานอาการไม่พึงประสงค์ การเพิ่มความแม่นยำในการตรวจจับอาการที่อาจถูกมองข้าม และการลดภาระงานของบุคลากรสาธารณสุขให้สามารถมุ่งเน้นการดูแลผู้ป่วยได้มากขึ้น การประยุกต์ใช้งานวิจัยนี้สอดคล้องกับแนวทางขององค์การอนามัยโลกและนโยบายสาธารณสุขดิจิทัลของประเทศไทย เพื่อพัฒนาต่อยอดเป็นระบบเฝ้าระวังความปลอดภัยของยาในระดับประเทศที่มีประสิทธิภาพและเหมาะสมกับบริบทไทยต่อไป

Author Contributions

SK: Conceptualization, Methodology, Data collection, Data analysis, Original draft preparation, Review and editing. AW: Data analysis consulting, Review and editing. TA: Validation, Funding Acquisition, Review and editing. JA: Conceptualization, Data analysis consulting, Validation, Supervision, Original draft preparation, Review and editing.

Declaration of the Use of Generative AI

During the preparation of this manuscript, the authors used Gemini 3 to assist with language editing and polishing of the draft. The tool was used solely to improve clarity, grammar, and overall readability. The authors reviewed and revised the output as necessary and take full responsibility for the content of the manuscript.

Ethics

This study received ethical approval from the Human Research Ethics Committee of Suranaree University of Technology (Project Code: EC-67-0227; COA No. 210/2567). The approval certificate was issued on January 8, 2025. All procedures followed relevant ethical guidelines. The requirement for informed consent was waived by the ethics committee because this study used retrospective data collection.

Funding None

References

Ahmad, F., Abbasi, A., Kitchens, B., Adjeroh, D., & Zeng, D. (2020). Deep learning for adverse event detection from web search. *IEEE Transactions on Knowledge and Data Engineering*, *34*(6), 2681-2695. https://doi.org/10.1109/TKDE.2020.3017786

Alomar, M. J. (2014). Factors affecting the development of adverse drug reactions. *Saudi pharmaceutical journal*, *22*(2), 83-94. https://doi.org/10.1016/j.jsps.2013.02.003

Chaichulee, S., Promchai, C., Kaewkomon, T., Kongkamol, C., Ingviya, T., & Sangsupawanich, P. (2022). Multi-label classification of symptom terms from free-text bilingual adverse drug reaction reports using natural language processing. *PLoS One*, *17*(8), e0270595. https://doi.org/10.1371/journal.pone.0270595

Dongre, S., & Agrawal, J. (2023). Deep-Learning-Based Drug Recommendation and ADR Detection Healthcare Model on Social Media. *IEEE Transactions on Computational Social Systems*, *10*(4), 1791–1799. https://doi.org/10.1109/TCSS.2022.3231701

Guo, K., Feng, Z., Chen, S., Yan, Z., Jiao, Z., & Feng, D. (2022). Safety Profile of Antipsychotic Drugs: Analysis Based on a Provincial Spontaneous Reporting Systems Database. *Frontiers in Pharmacology*, *13*, 848472. https://doi.org/10.3389/fphar.2022.848472

Iqbal, E. (2021). *Investigating adverse effects of psychiatric drugs through data-mining of electronic health records* [Doctoral dissertation]. King's College London.

Ishikawa, T., Yakoh, T., & Urushihara, H. (2022). An NLP-inspired data augmentation method for adverse event prediction using an imbalanced healthcare dataset. *IEEE Access*, *10*, 81166-81176. https://doi.org/10.1109/ACCESS.2022.3195212

Kobua, S., Wongkoblap, A., Angskun, T., & Angskun, J. (2025). A Model for Analyzing the Severity Level of Adverse Drug Reactions using Machine Learning. *Journal of Science and Technology Mahasarakham University*, *44*(1), 59-71. (In Thai)

Kunakorntham, P., Pattanaprateep, O., Dejthevaporn, C., Thammasudjarit, R., & Thakkinstian, A. (2022). Detection of statin-induced rhabdomyolysis and muscular related adverse events through data mining technique. *BMC Medical Informatics and Decision Making*, *22*(1), 233. https://doi.org/10.1186/s12911-022-01978-4

McMaster, C., Chan, J., Liew, D. F., Su, E., Frauman, A. G., Chapman, W. W., & Pires, D. E. (2023). Developing a deep learning natural language processing algorithm for automated reporting of adverse drug reactions. *Journal of biomedical informatics*, *137*104265. https://doi.org/10.1016/j.jbi.2022.104265

Nafea, A. A., Omar, N., & Al-qfail, Z. M. (2024). Artificial neural network and latent semantic analysis for adverse drug reaction detection. *Baghdad Science Journal*, *21*(1), 19. https://doi.org/10.21123/bsj.2023.7988

Naranjo, C. A., Busto, U., Sellers, E. M., Sandor, P., Ruiz, I., Roberts, E. A., ... & Greenblatt, D. J. (1981). A method for estimating the probability of adverse drug reactions. *Clinical Pharmacology & Therapeutics*, *30*(2), 239-245. https://doi.org/10.1038/clpt.1981.154

Nishioka, S., Watanabe, T., Asano, M., Yamamoto, T., Kawakami, K., Yada, S., Aramaki, E., Yajima, H., Kizaki, H., & Hori, S. (2022). Identification of hand-foot syndrome from cancer patients' blog posts: BERT-based deep-learning approach to detect potential adverse drug reaction symptoms. *PloS one*, *17*(5), e0267901. https://doi.org/10.1371/journal.pone.0267901

Nithinsha, S., & Anusuya, S. (2023). Robust Adverse Drug Reaction Prediction and Classification by Employing Deer Hunting Optimization Driven Deep Learning Approach. *International Journal of Electrical and Electronics Engineering*, *10*(5), 48–59. https://doi.org/10.14445/23488379/IJEEE-V10I5P105

Phianthai, B., & Vetchakun-anukun, K. (2020). The Association between Potential Factors and Seriousness Level of Adverse Drug Reactions (ADRs) at King Narai Hospital: a cross-sectional study. *Journal of Health and Environmental Education, 5*(2), 134-146. (In Thai)

Rawat, A., Wani, M. A., ElAffendi, M., Imran, A. S., Kastrati, Z., & Daudpota, S. M. (2022). Drug adverse event detection using text-based convolutional neural networks (TextCNN) technique. *Electronics, 11*(20), 3336. https://doi.org/10.3390/electronics11203336

Roosan, D., Law, A. V., Roosan, M. R., & Li, Y. (2022). Artificial Intelligent Context-Aware Machine-Learning Tool to Detect Adverse Drug Events from Social Media Platforms. *Journal of medical toxicology: official journal of the American College of Medical Toxicology*, *18*(4), 311–320. https://doi.org/10.1007/s13181-022-00906-2

Sisay, T., & Wami, R. (2021). Adverse drug reactions among major depressive disorders: patterns by age and gender. *Heliyon*, *7*(12), e08655. https://doi.org/10.1016/j.heliyon.2021.e08655

Songsiriphan, R. (2018). Factors Related to a Severity Level of Adverse Drug Reactions (ADRs). *The Southern College Network Journal of Nursing and Public Health, 5*(2), 46-56. (In Thai)

Spandana, S., & Prakash, R. V. (2024). Multiple features-based adverse drug reaction detection from social media using deep convolutional neural networks (DCNN). *Multimedia Tools and Applications*, *83*(26), 67779-67793. https://doi.org/10.1007/s11042-024-18144-9

Suwankesawong, W., Sriphiromya, P., Tragulpiankit, P., Phetcharat, C., & Sornsrivichai, V. (2016). Evaluation of Thai algorithm usage for adverse drug reaction monitoring. *Journal of Health Science of Thailand, 25(4), 673–682*. (In Thai)

Tan, H. X., Teo, C. H. D., Ang, P. S., Loke, W. P. C., Tham, M. Y., Tan, S. H., ... & Dorajoo, S. R. (2022). Combining machine learning with a rule-based algorithm to detect and identify related entities of documented adverse drug reactions on hospital discharge summaries. *Drug safety*, *45*(8), 853-862. https://doi.org/10.1007/s40264-022-01196-x

Tiamkaew, K. (2014). *Guidelines for dental treatment of psychiatric patients* (2nd ed., 1st printing). Srithanya Hospital, Ministry of Public Health. (In Thai)

World Health Organization. (2021). *The global patient safety action plan 2021-2030. Towards eliminating avoidable harm in health care*.

Wu, X. W., Zhang, J. Y., Chang, H., Song, X. W., Wen, Y. L., Long, E. W., & Tong, R. S. (2022). Develop an ADR prediction system of Chinese herbal injections containing Panax notoginseng saponin: a nested case–control study using machine learning. *BMJ Open*, *12*(9), e061457. https://doi.org/10.1136/bmjopen-2022-061457

Zitu, M. M., Zhang, S., Owen, D. H., Chiang, C., & Li, L. (2023). Generalizability of machine learning methods in detecting adverse drug events from clinical narratives in electronic medical records. *Frontiers in pharmacology*, *14*, 1218679. https://doi.org/10.3389/fphar.2023.1218679