บทความวิจัย

การเพิ่มประสิทธิภาพการตรวจจับภัยคุกคามภายในโดยใช้แบบจำลองภัยคุกคามสไตรด์ร่วมกับอัลกอริทึมเชิงวิวัฒนาการ

ทรงพล นคเรศเรืองศักดิ์*
สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีดิจิทัลและนวัตกรรม มหาวิทยาลัยเซาธ์อีสท์บางกอก

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08 3946 1514 อีเมล: songpon@southeast.ac.th DOI: 10.14416/j.kmutnb.2026.06.003
รับเมื่อ 28 กันยายน 2568 แก้ไขเมื่อ 5 กุมภาพันธ์ 2569 ตอบรับเมื่อ 26 กุมภาพันธ์ 2569 เผยแพร่ออนไลน์ 12 มิถุนายน 2569

บทคัดย่อ

ภัยคุกคามภายในเป็นหนึ่งในความเสี่ยงสำคัญที่ส่งผลกระทบต่อความมั่นคงปลอดภัยของระบบสารสนเทศ งานวิจัยนี้มุ่งพัฒนาวิธีการตรวจจับภัยคุกคามดังกล่าวโดยใช้การสร้างแบบจำลองภัยคุกคามตามกรอบแนวคิดสไตรด์ร่วมกับขั้นตอนวิธีเชิงวิวัฒนาการ ได้แก่ ขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง เพื่อปรับแต่งกฎการตรวจจับให้เหมาะสมกับพฤติกรรมที่ปรากฏในข้อมูลจริง โดยใช้ชุดข้อมูล LANL จำนวน 50,000 แถว ซึ่งมีสัดส่วนภัยคุกคามร้อยละ 15 และทำการเปรียบเทียบกับแบบจำลองการเรียนรู้ของเครื่อง ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน แบบจำลองป่าสุ่ม และโครงข่ายประสาทเทียมหลายชั้น ผลการทดลองพบว่าสไตรด์ (STRIDE) พื้นฐานให้ค่าความถูกต้อง 0.823 และพื้นที่ใต้เส้นโค้งอาร์โอซี 0.832 แต่เมื่อผสานกับขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง ทำให้มีประสิทธิภาพเพิ่มขึ้นเป็นค่าความถูกต้อง 0.878 พื้นที่ใต้เส้นโค้งอาร์โอซี 0.894 และค่าความถูกต้อง 0.889 พื้นที่ใต้เส้นโค้งอาร์โอซี 0.902 ตามลำดับ ซึ่งสูงกว่าซัพพอร์ตเวกเตอร์แมชชีนที่มีพื้นที่ใต้เส้นโค้งอาร์โอซี 0.875 และแบบจำลองป่าสุ่มที่มีพื้นที่ใต้เส้นโค้งอาร์โอซี 0.891 รวมทั้งยังใกล้เคียงกับโครงข่ายประสาทเทียมหลายชั้นที่มีพื้นที่ใต้เส้นโค้งอาร์โอซี 0.895 โดยยังคงความสามารถในการตีความเชิงกฎได้ดีกว่า สะท้อนศักยภาพของการบูรณาการสไตรด์เข้ากับขั้นตอนวิธีเชิงวิวัฒนาการในการตรวจจับภัยคุกคามภายในอย่างมีประสิทธิภาพ ยืดหยุ่น และมีความเป็นไปได้ต่อการใช้งานจริงในองค์กร

คำสำคัญ: การตรวจจับภัยคุกคามภายใน แบบจำลองภัยคุกคาม สไตรด์ ขั้นตอนวิธีเชิงวิวัฒนาการ

การอ้างอิงบทความ: ทรงพล นคเรศเรืองศักดิ์, "การเพิ่มประสิทธิภาพการตรวจจับภัยคุกคามภายในโดยใช้แบบจำลองภัยคุกคามสไตรด์ร่วมกับอัลกอริทึมเชิงวิวัฒนาการ," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 36, ฉบับที่ 3, หน้า 1–15, ก.ค.–ก.ย. 2569, เลขที่บทความ 263-8228, doi: 10.14416/j.kmutnb.2026.06.003.

Research Article

Enhancing Insider Threat Detection Using STRIDE Threat Modeling with Evolutionary Algorithms

Songpon Nakharacruangsak*
Department of Information Technology, Faculty of Digital Technology and Innovation, Southeast Bangkok University, Bangkok, Thailand

* Corresponding Author, Tel. 08 3946 1514, E-mail: songpon@southeast.ac.th
DOI: 10.14416/j.kmutnb.2026.06.003
Received 28 September 2025; Revised 5 February 2026; Accepted 26 February 2026; Published online: 12 June 2026

Abstract

Insider threats are among the most critical risks affecting the security of information systems, as insiders often possess legitimate access that makes detection by traditional methods challenging. This study aims to develop an effective detection approach by integrating the STRIDE threat modeling framework with evolutionary optimization techniques, namely the Genetic Algorithm (GA) and Differential Evolution (DE), to refine detection rules based on real behavioral data. Using the LANL dataset comprising 50,000 records with 15% insider threat instances, the proposed models were evaluated against widely used machine learning classifiers, including the Support Vector Machine (SVM), Random Forest (RF), and Multi-Layer Perceptron (MLP). Experimental results show that the baseline STRIDE model achieved an accuracy of 0.823 and an AUC of 0.832, while STRIDE combined with GA and DE significantly improved performance, reaching an accuracy of 0.878 with an AUC of 0.894, and an accuracy of 0.889 with an AUC of 0.902, respectively. These results outperform SVM (AUC = 0.875) and RF (AUC = 0.891), and are comparable to MLP (AUC = 0.895), while maintaining superior interpretability through rule-based modeling. The findings highlight the potential of integrating STRIDE with evolutionary optimization techniques to achieve accurate, flexible, and interpretable insider threat detection that is practical for real-world organizational environments.

Keywords: Insider Threat Detection, Threat Modeling, STRIDE, Evolutionary Algorithm

Please cite this article as: S. Nakharacruangsak, "Enhancing insider threat detection using STRIDE threat modeling with evolutionary algorithms," *The Journal of KMUTNB*, vol. 36, no. 3, pp. 1–15, Jul.–Sep. 2026 (in Thai), Art. no. 263-8228, doi: 10.14416/j.kmutnb.2026.06.003.

1. บทนำ

ในยุคดิจิทัลองค์กรต่าง ๆ พึ่งพาเทคโนโลยีสารสนเทศและระบบเครือข่ายเป็นกลไกหลักในการขับเคลื่อนธุรกิจ ไม่ว่าจะเป็นการจัดเก็บข้อมูลลูกค้า การประมวลผลคำสั่งซื้อ การสื่อสารภายในองค์กร หรือการบริหารจัดการโครงสร้างพื้นฐานของระบบ โดยเฉพาะอย่างยิ่งในช่วงที่การเปลี่ยนผ่านสู่ระบบคลาวด์ (Cloud Computing) อินเทอร์เน็ตของสรรพสิ่ง (IoT) และการประมวลผลที่ขอบเครือข่าย (Edge Computing) เกิดขึ้นอย่างรวดเร็ว ส่งผลให้ประเด็นด้านความมั่นคงปลอดภัยไซเบอร์ (Cybersecurity) มีความสำคัญเพิ่มขึ้นอย่างต่อเนื่อง [1], [2] การโจมตีทางไซเบอร์ในปัจจุบันมีความซับซ้อน และแฝงตัวมากขึ้น ครอบคลุมทั้งภัยคุกคามจากภายนอก และภัยคุกคามจากบุคคลภายใน (Insider Threats) ซึ่งมีแนวโน้มทวีความรุนแรง เนื่องจากผู้ก่อเหตุมักเป็นบุคลากรที่มีสิทธิ์เข้าถึงระบบโดยชอบธรรม [3]–[5] ภัยคุกคามจากบุคคลภายในมีความท้าทายมากกว่าภัยจากภายนอก เนื่องจากพฤติกรรมของผู้ใช้ภายในมักไม่ก่อให้เกิดความผิดปกติที่ชัดเจน เช่น การเข้าสู่ระบบนอกเวลา หรือการล็อกอินจากเครื่องใหม่ ซึ่งอาจยังอยู่ในขอบเขตที่ระบบรักษาความปลอดภัยแบบดั้งเดิม เช่น การตรวจจับแบบอิงลายเซ็นและแบบอิงกฎเกณฑ์ มองว่าเป็นพฤติกรรมปกติ [1], [6] เพื่อศึกษาปัญหานี้ในเชิงพฤติกรรมห้องปฏิบัติการแห่งชาติลอสอาลามอส (Los Alamos National Laboratory; LANL) ได้ชุดข้อมูลเหตุการณ์การพิสูจน์ตัวตนของ LANL (LANL Authentication Dataset) ซึ่งเป็นข้อมูลเหตุการณ์การเข้าสู่ระบบจริงเป็นระยะเวลา 58 วัน โดยมีการฝังเหตุการณ์ภัยคุกคามในสัดส่วนที่สมจริง [7], [8] แนวคิดด้านแบบจำลองภัยคุกคาม (Threat Modeling) จึงถูกนำมาใช้เพื่อวิเคราะห์และจำแนกภัยคุกคามตั้งแต่ระดับสถาปัตยกรรมของระบบ

โดยหนึ่งในกรอบแนวคิดที่ได้รับความนิยมคือสไตรด์ (STRIDE) ซึ่งถูกพัฒนาและนำเสนอโดย กระบวนการพัฒนาซอฟต์แวร์ด้านความมั่นคงปลอดภัยของไมโครซอฟท์ (Microsoft Security Development Lifecycle; SDL) และสามารถจำแนกภัยคุกคามออกเป็น 6 ประเภท คือ การปลอมแปลงตัวตน (Spoofing) การแก้ไขหรือดัดแปลงข้อมูล (Tampering) การปฏิเสธความรับผิดชอบ (Repudiation) การเปิดเผยข้อมูลโดยมิชอบ (Information Disclosure) การปฏิเสธการให้บริการ (Denial of Service) และการยกระดับสิทธิ์ (Elevation of Privilege) [2], [9] กรอบสไตรด์ได้รับการประยุกต์ใช้ในหลากหลายบริบท เช่น อินเทอร์เน็ตของสรรพสิ่ง [1], [10] ระบบบ้านอัจฉริยะ (Smart Home) [6] ระบบเครือข่ายยานยนต์ (VANET) [11] รวมทั้งระบบคลาวด์ และการประมวลผลที่ขอบเครือข่าย [12], [13] อย่างไรก็ตามในทางปฏิบัติสไตรด์มักถูกใช้งานในรูปแบบการวิเคราะห์ด้วยตนเองหรือแบบจำลองแบบอิงกฎ ซึ่งขาดความสามารถในการปรับตัวต่อพฤติกรรมผู้ใช้ที่เปลี่ยนแปลงอย่างต่อเนื่อง จากข้อจำกัดของการใช้แบบจำลองแบบอิงกฎ (Rule-based Models) งานวิจัยจำนวนมากจึงหันไปพึ่งพาเทคนิคการเรียนรู้ของเครื่อง (Machine Learning; ML) และการเรียนรู้เชิงลึก (Deep Learning; DL) ในการตรวจจับพฤติกรรมที่ผิดปกติ เช่น ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine; SVM) แบบจำลองป่าสุ่ม (Random Forest; RF) โครงข่ายประสาทเทียม (Neural Networks) และโครงข่ายประสาทเทียมแบบเวียนกลับ (Recurrent Neural Networks; RNN) ซึ่งสามารถวิเคราะห์พฤติกรรมเชิงลึกได้จากข้อมูลจำนวนมาก และมีความสามารถในการตรวจจับพฤติกรรมแฝงหรือความเบี่ยงเบนได้ดี [4], [14]–[16] ถึงแม้ว่าแบบจำลองเหล่านี้จะให้ผลลัพธ์เชิงตัวเลขที่ดี แต่ยังขาดความสามารถในการตีความ ซึ่งสอดคล้องกับผลการศึกษาก่อนหน้าโดยเฉพาะในบริบทองค์กรที่ต้องการความโปร่งใสและการตรวจสอบได้ [17], [18]

เพื่อลดช่องว่างดังกล่าวนักวิจัยหลายท่านจึงพัฒนาแนวทางแบบจำลองเชิงผสมผสาน (Hybrid Models) โดยนำสไตรด์มาผนวกรวมกันกับขั้นตอนวิธีเชิงวิวัฒนาการ (Evolutionary Algorithm) เช่น ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm; GA) และ การวิวัฒนาการเชิงผลต่าง (Differential Evolution; DE) [19]–[21] เพื่อปรับแต่งกฎการตรวจจับให้เหมาะสมกับข้อมูล อย่างไรก็ตามงานวิจัยที่

ผ่านมายังขาดการประเมินเชิงเปรียบเทียบอย่างเป็นระบบในบริบทภัยคุกคามจากบุคคลภายในและยังไม่ชี้ให้เห็นศักยภาพด้านการตีความของกฎที่ได้อย่างชัดเจน

งานวิจัยนี้นำเสนอกรอบการตรวจจับภัยคุกคามจากบุคคลภายในโดยผสานสไตรด์เข้ากับขั้นตอนวิธีเชิงวิวัฒนาการ (ขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง) ในลักษณะเชิงปริมาณ พร้อมการประเมินเชิงเปรียบเทียบกับแบบจำลองการเรียนรู้ของเครื่องมาตรฐาน ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน แบบจำลองป่าสุ่ม และโครงข่ายประสาทเทียมหลายชั้น โดยมุ่งเน้นทั้งความแม่นยำและความสามารถในการอธิบายกฎการตัดสินใจได้อย่างเป็นเหตุเป็นผล โดยแบบจำลองที่พัฒนาขึ้นถูกประเมินด้วยตัวชี้วัดทางประสิทธิภาพ ได้แก่ ความถูกต้อง (Accuracy; ACC) ความแม่นยำ (Precision; PRE) ความไว (Recall; REC) ค่าความกลมกลืน (F1-score; F1) และพื้นที่ใต้เส้นโค้งอาร์โอซี (Area Under the ROC Curve; AUC-ROC) เพื่อสะท้อนศักยภาพในการนำไปใช้งานจริงในระบบเครือข่ายขององค์กร

2. วิธีการทดลอง

งานวิจัยนี้ได้ออกแบบการทดลองสำหรับประเมินประสิทธิภาพของแบบจำลองตรวจจับภัยคุกคามภายในโดยกำหนดสภาพแวดล้อมการทดลองตั้งแต่ การเตรียมชุดข้อมูล พัฒนาแบบจำลองสไตรด์ที่ผสานเข้ากับขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง พร้อมเปรียบเทียบกับแบบจำลองมาตรฐาน และประเมินผลด้วยตัวชี้วัดที่เป็นสากล โดยสามารถสรุปได้ดังนี้

2.1 ซอฟต์แวร์และฮาร์ดแวร์

งานวิจัยนี้ใช้ไลบรารีประกอบด้วย ไซคิต-เลิร์น (Scikit-learn) เวอร์ชัน 1.2.2 [5] สำหรับการสร้างและประเมินแบบจำลองการเรียนรู้ของเครื่อง ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน แบบจำลองป่าสุ่ม และโครงข่ายประสาทเทียมหลายชั้น ร่วมกับเครื่องมือสำหรับการขั้นตอนวิธีเชิงวิวัฒนาการ ได้แก่ ขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง [19], [21] โดยทั้งหมดถูกพัฒนาด้วยภาษาไพธอน (Python) เวอร์ชัน 3.9.13 ซึ่งได้รับความนิยมอย่างแพร่หลายในการพัฒนาและทดลองแบบจำลองเชิงข้อมูลในงานวิจัยก่อนหน้า และทำการทดลองบนเครื่องคอมพิวเตอร์ที่ใช้หน่วยประมวลผลกลาง (Central Processing Unit; CPU) Intel(R) Core i7-10750H @ 2.60 GHz หน่วยความจำหลัก (Random Access Memory; RAM) ขนาด 16GB ระบบปฏิบัติการ (Operating System; OS) เป็น Windows 10 Pro 64-bit โดยไม่ได้ใช้หน่วยประมวลผลด้านกราฟิก (GPU) เพิ่มเติม เนื่องจากขั้นตอนการประมวลผลของอัลกอริทึมขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง และแบบจำลองการเรียนรู้ของเครื่องที่เลือกใช้สามารถทำงานได้อย่างมีประสิทธิภาพบนหน่วยประมวลผลกลาง

2.2 แหล่งข้อมูลและชุดข้อมูล

งานวิจัยนี้ใช้ข้อมูลเหตุการณ์การเข้าสู่ระบบจากชุดข้อมูลเหตุการณ์การพิสูจน์ตัวตนของ LANL ซึ่งเป็นข้อมูลจริงจากระบบเครือข่ายภายในองค์กรและถูกใช้อย่างแพร่หลายในงานวิจัยด้านการตรวจจับภัยคุกคามจากบุคคลภายใน โดยคัดเลือกข้อมูลจำนวน 50,000 แถว ซึ่งแต่ละแถวแทนเหตุการณ์การตรวจสอบสิทธิ์ของผู้ใช้งาน ข้อมูลถูกจัดกลุ่มออกเป็น 2 กลุ่ม ได้แก่ กลุ่มเหตุการณ์ภัยคุกคาม (Positive) ซึ่งเป็นเหตุการณ์ที่เกี่ยวข้องกับภัยคุกคามจากบุคคลภายใน และกลุ่มเหตุการณ์ปกติ (Negative) ซึ่งเป็นเหตุการณ์การใช้งานปกติ โดยในการทดลองหลักกำหนดสัดส่วนของกลุ่มเหตุการณ์ภัยคุกคามเท่ากับ ร้อยละ 15 ของข้อมูลทั้งหมด เพื่อให้มีจำนวนเหตุการณ์เพียงพอต่อการเรียนรู้และการประเมินประสิทธิภาพของแบบจำลอง ทั้งนี้ ชุดข้อมูลดังกล่าวถูกใช้เป็นฐานสำหรับการทดลองเพิ่มเติมภายใต้สัดส่วนข้อมูลที่แตกต่างกัน โดยรายละเอียดของกระบวนการเตรียมข้อมูลและการปรับสัดส่วนจะอธิบายในหัวข้อถัดไป

2.3 แบบจำลองและอัลกอริทึม

งานวิจัยนี้นำทั้งวิธีการแบบอิงกฎ และการเรียนรู้ของเครื่อง มาประยุกต์ใช้ร่วมกับการเพิ่มประสิทธิภาพเชิง

วิวัฒนาการ (Evolutionary Optimization) เพื่อทำการเปรียบเทียบประสิทธิภาพในการตรวจจับภัยคุกคามภายในโดยเฉพาะการผสานแบบจำลองสไตรด์ [19], [21] เข้ากับขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง ซึ่งมีความสามารถในการค้นหาคำตอบที่เหมาะสมในพื้นที่ค้นหาที่มีความซับซ้อนสูง ทั้งนี้ เพื่อให้การเปรียบเทียบระหว่างอัลกอริทึมเป็นไปอย่างเป็นธรรม (Fair Comparison) ได้กำหนดให้ทั้งขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่างทำงานบนโครงสร้างการเข้ารหัสกฎ (Encoded Rule Representation) ในรูปแบบเดียวกัน

2.3.1 ขั้นตอนวิธีเชิงพันธุกรรม

ขั้นตอนวิธีเชิงพันธุกรรมจำลองหลักการคัดเลือกโดยธรรมชาติผ่านกระบวนการเลือก (Selection) การสับเปลี่ยน (Crossover) และการกลายพันธุ์ (Mutation) โดยใช้ฟังก์ชันวัดความเหมาะสม (Fitness Function) ดังสมการที่ (1)

$$Finess(x) = \alpha \cdot Accuracy(x) + \beta \cdot Interpretability(x) \quad (1)$$

โดยที่ α และ β เป็นค่าน้ำหนักที่สะท้อนความสำคัญของแต่ละตัวชี้วัด

2.3.2 การวิวัฒนาการเชิงผลต่าง

การวิวัฒนาการเชิงผลต่างใช้ความแตกต่างระหว่างสมาชิกในประชากรเพื่อสร้างคำตอบใหม่ โดยนิยามเวกเตอร์ทดลอง (Trial Vector) ดังสมการ (2)

$$v_i = x_{r1} + F \cdot (x_{r2} - x_{r3}) \quad (2)$$

โดยที่ x_{r1}, x_{r2}, x_{r3} เป็นสมาชิกที่สุ่มเลือกจากประชากร และ F คือค่า Scaling Factor

2.3.3 ซัพพอร์ตเวกเตอร์แมชชีน

ซัพพอร์ตเวกเตอร์แมชชีนพยายามหาขอบเขตเชิงเส้นที่ดีที่สุด (Optimal Hyperplane) สำหรับจำแนกข้อมูล ดังสมการที่ (3)

$$f(x) = w^T x + b \quad (3)$$

โดยแก้ปัญหา Optimization:

$$\min_{w,b} \frac{1}{2} \| w \|^2 \text{ subject to } y_i(w^T x_i + b) \geq 1 \quad (4)$$

2.3.4 แบบจำลองป่าสุ่ม

แบบจำลองป่าสุ่ม ใช้การรวมผลลัพธ์จากหลายต้นไม้ตัดสินใจ (Decision Trees) ผ่านการโหวตแบบเสียงข้างมาก ดังสมการที่ (5)

$$\hat{y} = majority\_vote\ (h_1(x), h_2(x), \ldots, h_t(x)) \quad (5)$$

โดยที่ $h_t(x)$ แทนฟังก์ชันการจำแนกจากต้นไม้ลำดับที่ t

2.3.5. โครงข่ายประสาทเทียมหลายชั้น (MLP)

โครงข่ายประสาทเทียมหลายชั้น เป็นโครงข่ายประสาทเทียมที่ประกอบด้วยหลายชั้น (Input, Hidden, Output) โดยแต่ละชั้นคำนวณดังสมการที่ (6)

$$a^{(l)} = f(w^{(l)} \cdot a^{(l-1)} + b^{(l)}) \quad (6)$$

โดยที่ $f(\cdot)$ คือ Activation Function เช่น ReLU หรือ Sigmoid

แบบจำลองสไตรด์แบบอิงกฎ (STRIDE-Rules) ใช้เป็นจุดตั้งต้น ขณะที่แบบจำลองสไตรด์ที่ผสานกับขั้นตอนวิธีเชิงพันธุกรรม (STRIDE+GA) และแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง (STRIDE+DE) ถูกออกแบบเพื่อเพิ่มความแม่นยำเชิงปริมาณและคงไว้ซึ่งความสามารถในการตีความ ส่วนซัพพอร์ตเวกเตอร์แมชชีน แบบจำลองป่าสุ่ม และโครงข่ายประสาทเทียมหลายชั้น ถูกใช้เป็นตัวเปรียบเทียบเชิงมาตรฐาน (Benchmark Models) ตามแนวทางของงานวิจัยที่เกี่ยวข้อง [4], [5], [17], [21]

2.4 วิธีการวัดและประเมินผล

ในการประเมินประสิทธิภาพของแบบจำลองตรวจจับภัยคุกคามภายใน งานวิจัยนี้เลือกใช้ตารางความสับสนของผลการจำแนก (Confusion Matrix) เป็นพื้นฐานในการคำนวณ

ตัวชี้วัดที่สำคัญหลายประการ ตารางความสับสนของผลการจำแนกประกอบด้วยสี่องค์ประกอบ ได้แก่

ผลบวกจริง (True Positive; TP) จำนวนกรณีที่แบบจำลองทำนายว่าเป็นภัยคุกคาม และเป็นภัยคุกคามจริง

ผลลบจริง (True Negative; TN) จำนวนกรณีที่แบบจำลองทำนายว่าไม่เป็นภัยคุกคาม และไม่เป็นภัยคุกคามจริง

ผลบวกลวง (False Positive; FP) จำนวนกรณีที่แบบจำลองทำนายว่าเป็นภัยคุกคาม แต่ไม่เป็นภัยคุกคามจริง

ผลลบลวง (False Negative; FN) จำนวนกรณีที่แบบจำลองทำนายว่าไม่เป็นภัยคุกคาม แต่เป็นภัยคุกคามจริง

จากตารางความสับสนของผลการจำแนกดังกล่าว สามารถคำนวณตัวชี้วัดที่สำคัญในการประเมินผลได้ดังนี้

ประการแรก คือค่าความถูกต้อง (Accuracy) ซึ่งสะท้อนถึงความแม่นยำโดยรวมของแบบจำลอง โดยคำนวณจากสัดส่วนของกรณีที่ทำนายถูกทั้งหมดเมื่อเทียบกับจำนวนข้อมูลทั้งหมด ดังสมการที่ (7)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

อย่างไรก็ตามการใช้ค่าความถูกต้องเพียงอย่างเดียวอาจไม่เพียงพอในกรณีที่ข้อมูลไม่สมดุล จึงต้องพิจารณาตัวชี้วัดอื่นเพิ่มเติม ได้แก่ ค่าความแม่นยำ (Precision) และค่าความไว (Recall) โดยค่าความแม่นยำมุ่งวัดสัดส่วนของกรณีที่แบบจำลองทำนายว่าเป็นภัยคุกคามและเป็นภัยคุกคามจริงจากจำนวนที่ทำนายว่าเป็นภัยทั้งหมด ซึ่งคำนวณได้จาก สมการที่ (8)

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

ในขณะที่ ค่าความไว วัดความสามารถของแบบจำลองในการตรวจจับภัยคุกคามที่เกิดขึ้นจริง โดยคำนวณจากสมการที่ (9)

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

และเพื่อสร้างความสมดุลระหว่าง ค่าความแม่นยำและค่าความไว งานวิจัยนี้ยังใช้ตัวชี้วัดค่าความกลมกลืน (F1-Score) ซึ่งเป็นค่าเฉลี่ยเชิงฮาร์มอนิกระหว่างสองค่าดังกล่าว โดยคำนวณจากสมการที่ (10)

$$\text{F1-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

นอกจากนี้เพื่อประเมินความสามารถของแบบจำลองในการแยกแยะข้อมูลระหว่างกลุ่มภัยคุกคามและไม่เป็นภัยคุกคาม งานวิจัยนี้ใช้ค่าพื้นที่ใต้เส้นโค้งอาร์โอซี โดยกราฟเส้นโค้งอาร์โอซี (Receiver Operating Characteristic; ROC Curve) เป็นกราฟที่แสดงความสัมพันธ์ระหว่าง ค่าอัตราการตรวจพบจริง (True Positive Rate; TPR หรือ Recall) กับค่าอัตราการแจ้งเตือนผิดพลาด (False Positive Rate; FPR) และค่าพื้นที่ใต้เส้นโค้งอาร์โอซี คือพื้นที่ใต้กราฟเส้นโค้งอาร์โอซี (ROC) ซึ่งสะท้อนความสามารถโดยรวมของแบบจำลองในการจำแนกกลุ่มข้อมูลทั้งสองประเภท โดยคำนวณจากสมการที่ (11)

$$AUC = \int_0^1 TPR(FPR)\,dFPR \tag{11}$$

โดยที่

TPR คือสัดส่วนของภัยคุกคามจริงที่ถูกตรวจพบ

FPR คือสัดส่วนของข้อมูลปกติที่ถูกระบุว่าเป็นภัยคุกคาม

$dFPR$ คือค่าการเปลี่ยนแปลงของ FPR

2.5 วิธีการวิจัย

งานวิจัยนี้ใช้รูปแบบเชิงทดลอง (Experimental Research) โดยใช้สถานการณ์ภัยคุกคามจากภายในผ่านชุดข้อมูล LANL และทดสอบประสิทธิภาพของแต่ละแบบจำลอง โดยแบ่งเป็นขั้นตอนดังต่อไปนี้

2.5.1 การเตรียมข้อมูล (Data Preparation)

งานวิจัยนี้คัดเลือกเฉพาะบันทึกการตรวจสอบสิทธิ์ (Authentication Logs) จากชุดข้อมูลเหตุการณ์การพิสูจน์ตัวตน

ของ LANL และกรองข้อมูลที่ไม่สมบูรณ์ออก จากนั้นแปลงให้อยู่ในรูปแบบตารางเพื่อความสะดวกในการวิเคราะห์จากนั้นได้ปรับสัดส่วนของกลุ่มเหตุการณ์ภัยคุกคาม เป็นร้อยละ 15 10 และ 5 ผ่านกระบวนการสุ่มคัดเลือกข้อมูลแบบแบ่งชั้นอย่างสุ่ม (Stratified Random Sub-sampling) โดยไม่ทำการแทนที่ข้อมูล ขณะที่ยังคงโครงสร้างของกลุ่มเหตุการณ์ปกติ (Negative) ไว้ตามเดิม พร้อมดำเนินการเลือกคุณลักษณะ (Feature Engineering) ที่สะท้อนพฤติกรรมผู้ใช้งาน ได้แก่ ความถี่ในการเข้าสู่ระบบ จำนวนครั้งที่เข้าสู่ระบบล้มเหลว ความหลากหลายของโฮสต์ที่เข้าถึง และความผิดปกติของช่วงเวลาการใช้งาน เพื่อใช้เป็นตัวแปรในการเรียนรู้และการตรวจจับภัยคุกคามของแบบจำลอง โดยรายละเอียดแสดงไว้ในตารางที่ 1

ตารางที่ 1 โครงสร้างข้อมูลของตารางที่ใช้ในการวิจัย

ฟิลด์ในชุดข้อมูลแลนแอล	ความหมาย
timestamp	วันที่-เวลา ของเหตุการณ์ login
source_user	ผู้ใช้งานที่พยายาม login
dest_computer	เครื่องปลายทางที่เข้าถึง
auth_type	วิธีการ authentication เช่น Kerberos, LDAP
success	0 = สำเร็จ / 1 = ล้มเหลว
logon_type	Interactive / Remote / Network
login_success	login สำเร็จหรือไม่ (True / False)
login_hour	ชั่วโมงของวัน (ใช้หาพฤติกรรมผิดเวลา)
is_sensitive_server	เป็น server สำคัญหรือไม่ (1/0)
rare_host	เป็น host ที่มีการเข้าถึงน้อย (1/0)
failed_last_5min	จำนวน login fail 5 นาทีที่ผ่านมา
user_risk_score	คะแนนพฤติกรรมเสี่ยงจำลอง (0–100)
label	1 = Insider Threat, 0 = ปกติ

2.5.2 การวิเคราะห์แบบดั้งเดิม (Baseline Modeling)

เพื่อกำหนดกรอบการวิเคราะห์เบื้องต้น งานวิจัยนี้ได้นำแนวคิดการสร้างแบบจำลองภัยคุกคามตามกรอบสไตรด์มาจัดประเภทภัยคุกคามและสร้างแบบจำลองเชิงกฎ (Rule-based Model) เพื่ออธิบายความเชื่อมโยงระหว่างผู้ใช้งาน (User) และเครื่องที่เข้าถึง (Host) แสดงดังตารางที่ 2 วิธีการดังกล่าวทำหน้าที่เป็นแบบจำลองพื้นฐาน (Baseline) โดยสะท้อนข้อจำกัดของการวิเคราะห์แบบอิงกฎ (Rule-based) ที่แม้จะตีความได้ง่าย แต่ยังขาดประสิทธิภาพเชิงปริมาณเมื่อเทียบกับแบบจำลองเชิงข้อมูล

ตารางที่ 2 ตัวอย่างของกฎ

หมวดสไตรด์	จำนวนกฎที่ใช้	ตัวอย่างกฎ	พฤติกรรมที่ตรวจจับได้
Spoofing	3	หาก user ใช้งาน host ที่ไม่เคย login มาก่อน → แจ้งเตือน	สวมรอยบัญชีผู้ใช้
Tampering	3	หากมี logon เข้า host แล้วทำ logoff ภายในเวลา < 10 วินาที → ผิดปกติ	พยายามเข้าถึง/เปลี่ยนแปลงข้อมูล
Repudiation	2	หากมี log ที่ไม่สามารถยืนยันความถูกต้องได้ → แจ้งเตือน	การปฏิเสธความรับผิด
Information Disclosure	2	ใช้เครื่องนอกเวลาทำการ และเข้า host ใหม่ → พฤติกรรมรั่วไหล	การเข้าถึงข้อมูลลับ
Denial of Service	2	พยายาม login ผิด > 10 ครั้งในเวลาอันสั้น → อาจโจมตีระบบ	พฤติกรรมบั่นทอนระบบ
Elevation of Privilege	3	เข้าสู่ระบบแบบ admin จาก user ปกติ โดยไม่มีประวัติ → แจ้งเตือน	การยกระดับสิทธิ์โดยมิชอบ

2.5.3 การจำลองด้วยการเรียนรู้ของเครื่อง (ML Modeling)

เพื่อประเมินศักยภาพของเทคนิคการเรียนรู้ของเครื่องงานวิจัยนี้ได้สร้างซัพพอร์ตเวกเตอร์แมชชีน แบบจำลองป่าสุ่ม และโครงข่ายประสาทเทียมหลายชั้น โดยแบ่งข้อมูลออกเป็นชุดฝึกสอน (80%) และชุดทดสอบ (20%) ผลลัพธ์ถูกประเมินด้วยตัวชี้วัดมาตรฐาน ได้แก่ ความถูกต้อง ความแม่นยำ ความไว ค่าความกลมกลืน และพื้นที่ใต้เส้นโค้งอาร์โอซี เพื่อสะท้อนความแม่นยำและความสามารถในการแยกแยะภัยคุกคาม

จากกิจกรรมปกติ

2.5.4 การพัฒนาแบบจำลองสไตรด์ที่ผสานอัลกอริทึมวิวัฒนาการ (Proposed STRIDE+GA/DE)

จุดเด่นของงานวิจัยนี้คือ การปรับปรุงการสร้างแบบจำลองภัยคุกคามตามกรอบสไตรด์ด้วยการผสานขั้นตอนวิธีวิวัฒนาการ (Evolutionary Optimization Techniques) ได้แก่ ขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง เพื่อค้นหากฎการตรวจจับที่เหมาะสมที่สุดกับข้อมูลจริง วิธีการดังกล่าวช่วยลดข้อจำกัดของการพึ่งพากฎคงที่ เพิ่มความยืดหยุ่นต่อข้อมูลที่มีความซับซ้อน และสามารถปรับตัวเข้ากับพฤติกรรมที่เปลี่ยนแปลงได้อย่างต่อเนื่อง แต่เพื่อให้เหมาะสมจึงต้องเปลี่ยนชุดพารามิเตอร์ของกฎให้กลายเป็นเว็กเตอร์ที่สามารถเข้ารหัสในขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง ได้ดังแสดงในตารางที่ 3

ตารางที่ 3 การกำหนดชุดพารามิตเตอร์ของกฎ

กฎ	พารามิเตอร์	ความหมาย
Failed login threshold	3	จำนวนครั้งที่ถือว่าเป็นผิดปกติ
Time window	5 min	ช่วงเวลาสำหรับนับพฤติกรรม
Rare host access	Yes / No	เปิดหรือปิดฟีเจอร์นี้
Sensitive host access weight	0–1	น้ำหนักความเสี่ยง

สามารถแปลงให้กลายเป็นเว็กเตอร์ที่สามารถเข้ารหัส (Encode) ใน GA/DE เช่น [3, 5, 1, 0.7] เป็นต้น

2.5.5 การเปรียบเทียบผลลัพธ์ (Result Comparison)

งานวิจัยได้ทำการเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองสไตรด์แบบอิงกฎ แบบจำลอง สไตรด์ที่ผสานกับขั้นตอนวิธีเชิงพันธุกรรม แบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง และแบบจำลองการเรียนรู้ของเครื่อง (RF, SVM, MLP) ผ่านตารางผลลัพธ์และกราฟเส้นโค้งอาร์โอซี โดยพิจารณาตัวชี้วัดประสิทธิภาพ (Accuracy, Precision, Recall, F1-score, AUC) เพื่อสะท้อนความเหมาะสมของแต่ละวิธีในบริบทการใช้งานจริง

3. ผลการทดลองและอภิปรายผลการทดลอง

3.1 ผลการทดลอง

การทดลองนี้ผู้วิจัยได้ใช้ตัวชี้วัดมาตรฐาน ได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความไว ค่าความกลมกลืน และพื้นที่ใต้เส้นโค้งอาร์โอซี เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองสไตรด์ที่ผสานกับขั้นตอนวิธีวิวัฒนาการ (ขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง) กับแบบจำลองมาตรฐาน ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน แบบจำลองป่าสุ่ม และโครงข่ายประสาทเทียมหลายชั้น โดยผลการทดลองภายใต้สัดส่วนข้อมูลกลุ่มเหตุการณ์ภัยคุกคาม ร้อยละ 15 แสดงไว้ในตารางที่ 4

ตารางที่ 4 การเปรียบเทียบประสิทธิภาพของแบบจำลองที่กลุ่มเหตุการณ์ภัยคุกคาม ร้อยละ 15

แบบจำลอง	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าความไว	ค่าความกลมกลืน	พื้นที่ใต้เส้นโค้ง
STRIDE	0.823	0.793	0.765	0.779	0.832
STRIDE+GA	0.878	0.861	0.847	0.854	0.894
STRIDE+DE	0.889	0.872	0.859	0.865	0.902
SVM	0.861	0.838	0.826	0.832	0.875
Random	0.876	0.854	0.842	0.848	0.891
MLP	0.883	0.865	0.854	0.859	0.895

จากตารางที่ 4 แสดงให้เห็นว่าแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง ให้ค่าประสิทธิภาพสูงที่สุดในทุกตัวชี้วัด โดยเฉพาะค่าพื้นที่ใต้เส้นโค้งอาร์โอซี (AUC) เท่ากับ 0.902 ซึ่งสะท้อนถึงความสามารถในการแยกแยะระหว่างเหตุการณ์ปกติและเหตุการณ์ที่เป็นภัยคุกคามได้อย่างมีประสิทธิภาพสูง เมื่อเปรียบเทียบกับแบบจำลองสไตรด์แบบอิงกฎ แบบดั้งเดิมซึ่งให้ค่าต่ำที่สุดในทุกตัวชี้วัด (Accuracy = 0.823 และ AUC = 0.832) ผลลัพธ์ดังกล่าวชี้ให้เห็นข้อจำกัดของการใช้กฎคงที่ในการตรวจจับภัยคุกคามที่มีความซับซ้อนและเปลี่ยนแปลงตามพฤติกรรมผู้ใช้

เพื่อศึกษาผลกระทบของความไม่สมดุลของข้อมูล (Class Imbalance) ซึ่งพบได้บ่อยในบริบทของภัยคุกคาม

จากบุคคลภายใน งานวิจัยนี้ได้ดำเนินการทดลองเพิ่มเติม โดยปรับสัดส่วนของกลุ่มเหตุการณ์ภัยคุกคาม เป็นร้อยละ 10 และร้อยละ 5 ผ่านกระบวนการสุ่มคัดเลือกข้อมูลแบบแบ่งชั้นอย่างสุ่ม ผลการทดลองภายใต้สัดส่วนดังกล่าวแสดงไว้ในตารางที่ 5 และตารางที่ 6 ตามลำดับ

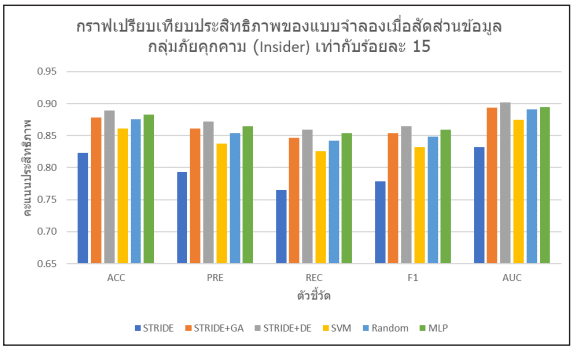
ตารางที่ 5 การเปรียบเทียบประสิทธิภาพของแบบจำลองที่กลุ่มเหตุการณ์ภัยคุกคาม ร้อยละ 10

แบบจำลอง	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าความไว	ค่าความกลมกลืน	พื้นที่ใต้เส้นโค้ง
STRIDE	0.801	0.765	0.712	0.737	0.810
STRIDE+GA	0.869	0.845	0.812	0.828	0.883
STRIDE+DE	0.873	0.852	0.825	0.838	0.891
SVM	0.854	0.826	0.793	0.809	0.866
Random	0.868	0.841	0.801	0.820	0.879
MLP	0.871	0.848	0.814	0.831	0.884

ตารางที่ 6 การเปรียบเทียบประสิทธิภาพของแบบจำลองที่กลุ่มเหตุการณ์ภัยคุกคาม ร้อยละ 5

แบบจำลอง	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าความไว	ค่าความกลมกลืน	พื้นที่ใต้เส้นโค้ง
STRIDE	0.774	0.732	0.661	0.695	0.782
STRIDE+GA	0.851	0.812	0.742	0.775	0.872
STRIDE+DE	0.856	0.821	0.758	0.788	0.879
SVM	0.842	0.798	0.721	0.758	0.858
Random	0.848	0.803	0.721	0.760	0.865
MLP	0.852	0.811	0.739	0.773	0.871

จากผลการทดลองพบว่าเมื่อสัดส่วนของข้อมูลกลุ่มเหตุการณ์ภัยคุกคามลดลงจากร้อยละ 15 เป็นร้อยละ 10 และร้อยละ 5 ค่าตัวชี้วัดของทุกแบบจำลองมีแนวโน้มลดลง โดยเฉพาะค่าความไว ค่าความกลมกลืน ซึ่งสะท้อนถึงความท้าทายในการตรวจจับภัยคุกคามภายใต้สภาวะข้อมูลไม่สมดุล อย่างไรก็ตามแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่างยังคงให้ผลการทดลองที่เหนือกว่าแบบ



รูปที่ 1 กราฟเปรียบเทียบประสิทธิภาพของแบบจำลองเมื่อสัดส่วนข้อมูลกลุ่มภัยคุกคาม (Insider) เท่ากับร้อยละ 15

จำลองอื่นอย่างสม่ำเสมอในทุกสัดส่วนข้อมูล แสดงถึงความคงทนและความเสถียรของแนวทางที่ผสานการวิเคราะห์เชิงความหมายของแบบจำลองสไตรด์ เข้ากับกระบวนการปรับแต่งกฎแบบอัตโนมัติ ขณะที่แบบจำลองการเรียนรู้ของเครื่อง เช่น แบบจำลองป่าสุ่ม และโครงข่ายประสาทเทียมหลายชั้น แม้จะให้ค่าความถูกต้อง และค่าพื้นที่ใต้เส้นโค้งอาร์โอซีอยู่ในระดับสูง แต่ประสิทธิภาพลดลงชัดเจนเมื่อข้อมูลมีความไม่สมดุลมากขึ้น และยังขาดความสามารถในการตีความเชิงเหตุผลเมื่อเทียบกับแนวทางที่นำเสนอในงานวิจัยนี้

เมื่อแสดงผลลัพธ์ในรูปแบบกราฟแท่งสำหรับการทดลองหลักที่มีสัดส่วนกลุ่มเหตุการณ์ภัยคุกคาม ร้อยละ 15 แสดงดังรูปที่ 1 พบว่าแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่างมีค่าประสิทธิภาพสูงกว่าแบบจำลองอื่นในทุกตัวชี้วัดอย่างชัดเจน ขณะที่แบบจำลองสไตรด์แบบอิงกฎให้ผลต่ำสุดตลอดการเปรียบเทียบ ส่วนแบบจำลองสไตรด์ที่ผสานกับขั้นตอนวิธีเชิงพันธุกรรมให้ผลลัพธ์ใกล้เคียงกับแบบจำลองการเรียนรู้ของเครื่อง ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน แบบจำลองป่าสุ่ม และโครงข่ายประสาทเทียมหลายชั้น โดยเฉพาะในด้าน ความถูกต้อง ค่าความกลมกลืน อย่างไรก็ตามแม้แบบจำลองการเรียนรู้ของเครื่อง จะมีความแม่นยำสูง แต่ยังขาดความสามารถในการตีความ ซึ่งเป็นข้อจำกัดสำคัญต่อการนำไปใช้งานจริงในบริบทขององค์กร

แม้ว่าขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่างจะเป็นขั้นตอนวิธีเชิงวิวัฒนาการที่นิยมใช้ในการปรับปรุงกฎการตัดสินใจ แต่ผลการทดลองพบว่าการวิวัฒนาการเชิงผลต่างให้ประสิทธิภาพสูงกว่าขั้นตอนวิธีเชิงพันธุกรรมอย่างสม่ำเสมอทั้งในด้าน ความถูกต้อง ค่าความกลมกลืน และค่าพื้นที่ใต้เส้นโค้งอาร์โอซี รวมถึงใช้เวลาลู่เข้าน้อยกว่า ความได้เปรียบนี้เกิดจากกลไกการค้นหาของการวิวัฒนาการเชิงผลต่างที่ปรับคำตอบผ่านความแตกต่างของเวกเตอร์ ทำให้การค้นหาเป็นไปอย่างต่อเนื่องและมีทิศทาง โดยในบริบทของงานวิจัยนี้ กฎของสไตรด์ สามารถแทนในรูปของเวกเตอร์พารามิเตอร์ที่ประกอบด้วยค่าเกณฑ์การตัดสินใจและน้ำหนักของเงื่อนไขต่าง ๆ กลไกของการวิวัฒนาการเชิงผลต่าง จึงสามารถปรับค่าพารามิเตอร์เหล่านี้ได้อย่างค่อยเป็นค่อยไปโดยไม่ทำลายโครงสร้างของกฎเดิม ในทางตรงกันข้าม กระบวนการสับเปลี่ยน และการกลายพันธุ์ของขั้นตอนวิธีเชิงพันธุกรรมอาจก่อให้เกิดการเปลี่ยนแปลงของกฎแบบไม่ต่อเนื่องและรบกวนโครงสร้างกฎที่มีประสิทธิภาพ ส่งผลให้การลู่เข้าไม่เสถียร ด้วยเหตุนี้ แบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง จึงเหมาะสมกับปัญหาการปรับแต่งกฎอย่างเหมาะสม (Rule Optimization) ที่ต้องการทั้งความแม่นยำและความสามารถในการตีความ

นอกจากการประเมินเชิงตัวเลขแล้ว งานวิจัยนี้ยังวิเคราะห์กฎการตรวจจับขั้นสุดท้าย (Final Rules) ที่ได้จากขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง เพื่อยืนยันว่าแบบจำลองสามารถให้ผลลัพธ์ที่ตีความได้และเชื่อมโยงกับหมวดภัยคุกคามตามกรอบสไตรด์อย่างชัดเจน ซึ่งจะอภิปรายในหัวข้อถัดไป

3.2 การวิเคราะห์ความสามารถในการตีความของแบบจำลอง

แม้ว่าผลการทดลองเชิงปริมาณจะแสดงให้เห็นว่าแบบจำลองสไตรด์ที่ผสานกับอัลกอริทึมเชิงวิวัฒนาการ โดยเฉพาะแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่างมีประสิทธิภาพสูงสุดในทุกตัวชี้วัด อย่างไรก็ตามหนึ่งในวัตถุประสงค์สำคัญของงานวิจัยนี้คือการพัฒนาแบบจำลองที่ไม่เพียงมีความแม่นยำสูง แต่ยังสามารถอธิบายเหตุผลของการตัดสินใจได้อย่างชัดเจน เพื่อให้เหมาะสมต่อการนำไปใช้งานจริงในบริบทขององค์กร งานวิจัยนี้จึงนำเสนอชุดกฎการตรวจจับขั้นสุดท้ายที่ได้จากกระบวนการปรับแต่งด้วยขั้นตอนวิธีเชิงพันธุกรรม และการวิวัฒนาการเชิงผลต่าง เพื่อแสดงให้เห็นถึงความสามารถในการตีความของแบบจำลองอย่างเป็นรูปธรรม

ตารางที่ 7 ตัวอย่างกฎการตรวจจับภัยคุกคามภายในขั้นสุดท้ายที่ได้จากแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง

ลำดับกฎ	เงื่อนไขของกฎ	พฤติกรรมที่สังเกตุได้	กลุ่มของสไตรด์
R1	login_fail ≥ 3 ∧ login_success = 1 ∧ new_host = TRUE	พยายามล็อกอินซ้ำหลายครั้งก่อนสำเร็จจากเครื่องใหม่	Spoofing
R2	unique_host ≥ 5 ∧ session_time < threshold	เข้าถึงหลายโฮสต์ในช่วงเวลาสั้นผิดปกติ	Information Disclosure
R3	access_frequency > $\mu + 2\sigma$ ∧ user_role = normal	ผู้ใช้ทั่วไปมีความถี่การใช้งานสูงผิดปกติ	Elevation of Privilege
R4	login_time ∉ working_hours ∧ privileged_host = TRUE	เข้าถึงโฮสต์สำคัญนอกเวลาทำงาน	Elevation of Privilege
R5	unique_host ≥ 4 ∧ login_fail ≥ 2	ลักษณะการสแกนหรือเคลื่อนไหวข้ามระบบ	Spoofing
R6	session_time ≥ τ ∧ access_frequency ≥ high	การใช้งานยาวนานและต่อเนื่องผิดปกติ	Information Disclosure

ตารางที่ 7 แสดงตัวอย่างกฎการตรวจจับขั้นสุดท้ายที่ได้จากแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง ซึ่งเป็นกฎที่มีค่า ความเหมาะสม (Fitness) สูงสุดหลังจากกระบวนการเรียนรู้เชิงวิวัฒนาการ กฎเหล่านี้ถูกจัดหมวดหมู่ตามแนวคิดสไตรด์และสะท้อนพฤติกรรมที่มีความเสี่ยงต่อ

การเป็นภัยคุกคามจากบุคคลภายในอย่างชัดเจน เช่น ความถี่ในการเข้าสู่ระบบที่ผิดปกติ การเข้าถึงโฮสต์จำนวนมากในช่วงเวลาสั้น หรือการใช้งานนอกช่วงเวลาปกติของผู้ใช้

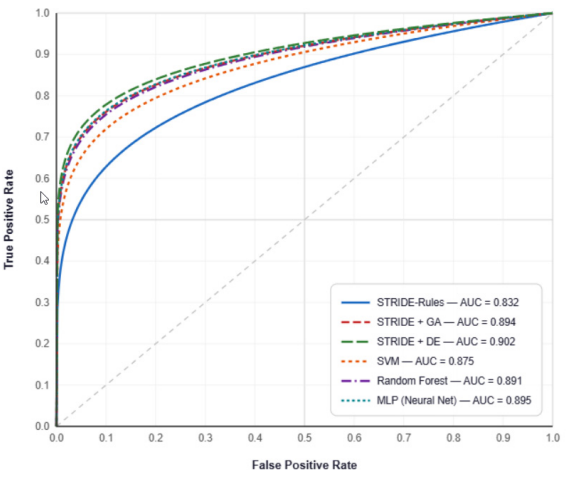
จากการวิเคราะห์กฎการตรวจจับขั้นสุดท้ายที่ได้จากแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง พบว่ากฎส่วนใหญ่สามารถจัดอยู่ในหมวดการปลอมแปลงตัวตน (Spoofing) การยกระดับสิทธิ์ (Elevation of Privilege) และการเปิดเผยข้อมูลโดยมิชอบ (Information Disclosure) ตามกรอบสไตรด์ โดยพิจารณาจากลักษณะพฤติกรรมในเงื่อนไขของกฎ เช่น การเข้าสู่ระบบจากหลายโฮสต์ภายในช่วงเวลาสั้น การเข้าถึงทรัพยากรเกินบทบาทปกติ หรือการถ่ายโอนข้อมูลจำนวนมากในช่วงเวลาที่ผิดปกติ ดังแสดงในตารางที่ 7 ผลลัพธ์นี้แสดงให้เห็นว่าแบบจำลองไม่ได้ทำงานในลักษณะกล่องดำ แต่ให้ผลลัพธ์ในรูปของกฎที่สามารถตีความ ตรวจสอบ และเชื่อมโยงกับแนวคิดแบบจำลองภัยคุกคามได้อย่างชัดเจน อีกทั้งกฎที่ได้ยังสอดคล้องกับลักษณะภัยคุกคามจากบุคคลภายในที่พบบ่อยในบริบทองค์กรจริง สะท้อนว่าแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่างให้ทั้งความแม่นยำเชิงปริมาณและความสามารถในการอธิบายผลลัพธ์ได้อย่างเป็นเหตุเป็นผล

3.3 การอภิปรายผลข้อผิดพลาดของแบบจำลอง

แม้ว่าแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่างจะให้ประสิทธิภาพโดยรวมอยู่ในระดับสูง แต่การวิเคราะห์เชิงลึกยังพบข้อผิดพลาดในรูปแบบของผลบวกลวงและผลลบลวง ซึ่งสะท้อนข้อจำกัดด้านการทำความเข้าใจบริบทของผู้ใช้งาน โดยผลบวกลวงมักเกิดจากพฤติกรรมที่เบี่ยงเบนจากรูปแบบทั่วไปแต่ไม่ใช่ภัยคุกคามจริง เช่น การเข้าสู่ระบบนอกเวลางานหรือการเข้าถึงหลายโฮสต์ของผู้ดูแลระบบ ขณะที่ผลลบลวงพบในกรณีผู้ใช้งานที่มีสิทธิ์ระดับสูงซึ่งสามารถดำเนินกิจกรรมที่มีความเสี่ยงได้โดยไม่แสดงความผิดปกติอย่างชัดเจน อย่างไรก็ตามการมีชุดกฎจากแบบจำลองสไตรด์ที่ผสานกับขั้นตอนวิธีเชิงวิวัฒนาการ (STRIDE+GA/DE) ที่สามารถตีความได้ช่วยให้ผู้เชี่ยวชาญสามารถวิเคราะห์และปรับแต่งเงื่อนไขเชิงบริบทเพิ่มเติมได้ ซึ่งเป็นแนวทางสำคัญในการลดข้อผิดพลาดและเพิ่มความเหมาะสมในการนำแบบจำลองไปใช้งานจริงในระบบเครือข่ายองค์กร

เพื่อให้แนวทางการประมวลผลสามารถนำไปทำซ้ำได้อย่างชัดเจน งานวิจัยนี้กำหนดค่าพารามิเตอร์ของในการทดลอง ขั้นตอนวิธีเชิงพันธุกรรมถูกตั้งค่าพารามิเตอร์ดังนี้ ขนาดประชากร 100 อัตราการสับเปลี่ยน 0.8 และอัตราการกลายพันธุ์ 0.05 โดยใช้การคัดเลือกแบบทัวร์นาเมนต์ (Tournament Selection) และการสับเปลี่ยนพันธุกรรมแบบจุดเดียว (Single-Point Crossover) ส่วนการวิวัฒนาการเชิงผลต่างใช้กำหนดค่าตัวประกอบการกลายพันธุ์ (Mutation Factor; F) เท่ากับ 0.5 และ อัตราการสับเปลี่ยน (Crossover Rate; CR) เท่ากับ 0.7 โดยใช้กลยุทธ์ DE/rand/1/bin ทั้งสองอัลกอริทึมทำงานสูงสุด 100 รอบ หรือหยุดเมื่อค่าความเหมาะสมคงที่เกิน 20 รอบ โดยการวิวัฒนาการเชิงผลต่างใช้เวลาเฉลี่ยในการลู่เข้า 112.7 วินาที ซึ่งน้อยกว่าขั้นตอนวิธีเชิงพันธุกรรมที่ใช้เวลา 145.3 วินาที แสดงถึงประสิทธิภาพและความเสถียรที่สูงกว่า

ในส่วนของกราฟเส้นโค้งอาร์โอซีพบว่าเส้นโค้งของแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่างอยู่ใกล้มุมบนซ้ายมากที่สุด พร้อมค่าพื้นที่ใต้เส้นโค้งอาร์โอซีเท่ากับ 0.902 ซึ่งยืนยันถึงประสิทธิภาพในการแยกแยะระหว่างเหตุการณ์ปกติและภัยคุกคามได้ดีที่สุด รองลงมาคือโครงข่ายประสาทเทียมหลายชั้นที่ค่าพื้นที่ใต้เส้นโค้งอาร์โอซีเท่ากับ 0.895 และแบบจำลองสไตรด์ที่ผสานกับขั้นตอนวิธีเชิงพันธุกรรมที่ค่าพื้นที่ใต้เส้นโค้งอาร์โอซีเท่ากับ 0.894 ส่วนแบบจำลองป่าสุ่ม และซัพพอร์ตเวกเตอร์แมชชีน แม้จะแสดงผลในระดับที่น่าพอใจ ค่าพื้นที่ใต้เส้นโค้งอาร์โอซี เท่ากับ 0.891 และ 0.875 ตามลำดับแต่ยังด้อยกว่าแบบจำลองสไตรด์ที่ผสานกับขั้นตอนวิธีเชิงวิวัฒนาการอย่างสม่ำเสมอ ขณะที่แบบจำลองสไตรด์แบบอิงกฎมีค่าพื้นที่ใต้เส้นโค้งอาร์โอซีต่ำสุดที่ 0.832 ตอกย้ำข้อจำกัดของวิธีการ Rule-based ดั้งเดิม โดยเส้นโค้งอาร์โอซีถูกสร้างขึ้นจากการทำการตรวจสอบความถูกต้องแบบไขว้ชนิดแบ่งชั้น 10 ส่วน (10-Fold Stratified Cross-Validation) ร่วมกับการสุ่มตัวอย่างแบบ



รูปที่ 2 กราฟเส้นโค้งอาร์โอซีเปรียบเทียบประสิทธิภาพของแบบจำลอง

บูตสแตรป (Bootstrap Sampling) จำนวน 1,000 ครั้ง เพื่อประเมินค่าความเชื่อมั่น (95% Confidence Interval) แสดงดังรูปที่ 2

3.4 การอภิปรายผลเชิงเปรียบเทียบกับงานวิจัยก่อนหน้า

เมื่อพิจารณาเชิงเปรียบเทียบกับงานวิจัยก่อนหน้า ผลลัพธ์ของแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่างสอดคล้องกับข้อสังเกตของงานวิจัยหลายฉบับที่ชี้ว่าแบบจำลองการเรียนรู้ของเครื่อง และการเรียนรู้เชิงลึก แม้ให้ค่าความแม่นยำสูง แต่ยังมีข้อจำกัดด้านความสามารถในการตีความ ขณะที่แบบจำลองแบบอิงกฎมีความเข้าใจง่ายแต่ด้อยด้านประสิทธิภาพเชิงปริมาณ [4], [5], [17], [22] การผสานแบบจำลองสไตรด์เข้ากับขั้นตอนวิธีเชิงวิวัฒนาการ (GA/DE) จึงช่วยสร้างสมดุลระหว่างความแม่นยำและความสามารถในการอธิบายผลได้อย่างเป็นระบบ โดยค่าพื้นที่ใต้เส้นโค้งอาร์โอซีที่สูงของแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่างยังสะท้อนแนวคิดการออกแบบระบบตรวจจับเชิงรุกตามหลักแบบจำลองภัยคุกคาม [1], [2], [23] และสอดคล้องกับแนวทางที่เสนอใน [21] ที่ยืนยันว่าขั้นตอนเชิงวิวัฒนาการสามารถเพิ่มประสิทธิภาพการตรวจจับภัยคุกคามภายในได้อย่างมีประสิทธิผล

แต่เมื่อเปรียบเทียบกับงานร่วมสมัย เช่น งานของ Gong และคณะ [24] ซึ่งเป็นงานสำรวจด้านการตรวจจับภัยคุกคามภายในด้วยแนวทางแบบกราฟ (Graph-Based Approaches) พบว่าแม้แบบจำลองกราฟจะมีความสามารถในการเรียนรู้โครงสร้างความสัมพันธ์ของผู้ใช้ได้ดี แต่ยังมุ่งเน้นประสิทธิภาพเชิงปริมาณและความสามารถในการขยายขนาดระบบ (Scalability) เป็นหลัก โดยไม่ได้เน้นการตีความกฎในระดับนโยบายองค์กร ขณะที่งานของ Bin Sarhan และ Altwaijry [25] รายงานค่าความถูกต้องสูงบนชุดข้อมูล CERT แต่ยังคงอยู่ในลักษณะกล่องดำและต้องพึ่งพากระบวนการวิศวกรรมคุณลักษณะและการลดมิติข้อมูลที่ซับซ้อน ในทางตรงกันข้ามแนวทางของงานวิจัยนี้ให้ผลลัพธ์ในรูปของกฎที่สามารถเชื่อมโยงกับหมวดหมู่ภัยคุกคามตามแบบจำลองสไตรด์ได้อย่างชัดเจน ซึ่งสอดคล้องกับแนวโน้มงานปัญญาประดิษฐ์เชิงอธิบายได้ (Explainable Artificial Intelligence; XAI) ด้านความมั่นคงปลอดภัยไซเบอร์ที่ Rjoub และคณะ [26] ได้อภิปรายไว้ โดยเฉพาะในบริบทองค์กรที่ต้องการระบบสนับสนุนการตัดสินใจในศูนย์ปฏิบัติการความมั่นคงปลอดภัย

นอกจากนี้การประเมินเพิ่มเติมภายใต้สัดส่วนเหตุการณ์ผิดปกติที่ลดลง (10% และ 5%) แสดงให้เห็นว่าแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง ยังคงรักษาค่าพื้นที่ใต้เส้นโค้งอาร์โอซี และค่าความกลมกลืน ได้อย่างมีเสถียรภาพ สะท้อนความคงทนของแบบจำลองต่อปัญหาข้อมูลไม่สมดุล ซึ่งสอดคล้องกับประเด็นท้าทายที่ถูกอภิปรายไว้ในงานสำรวจ [24] โดยรวมแล้วแนวทางแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง จึงไม่เพียงแข่งขันได้กับแบบจำลองการเรียนรู้ของเครื่อง และแนวทางแบบกราฟในด้านประสิทธิภาพ พร้อมทั้งเพิ่มมิติของความสามารถในการอธิบายผลเชิงแบบจำลองภัยคุกคามอย่างเป็นระบบ

4. สรุปผลการทดลอง

ผลการทดลองแสดงให้เห็นว่าแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่างให้ประสิทธิภาพสูงสุดในการตรวจจับภัยคุกคามภายใน โดยมีค่าความถูกต้องเท่ากับ

0.889 และค่าพื้นที่ใต้เส้นโค้งอาร์โอซีเท่ากับ 0.902 ซึ่งมีแนวโน้มดีกว่าแบบจำลองสไตรด์แบบอิงกฎ และแบบจำลองการเรียนรู้ของเครื่องแบบมาตรฐานอย่างสม่ำเสมอ สะท้อนถึงศักยภาพในการจัดลำดับและจำแนกภัยคุกคามได้ด้วยความแม่นยำในระดับสูง อย่างไรก็ตามจากการวิเคราะห์เชิงลึกของผลลัพธ์ยังพบข้อผิดพลาดบางกรณีในรูปแบบของ ผลบวกลวง และผลลบลวงโดยผลบวกลวง มักเกิดในสถานการณ์ที่ผู้ใช้งานมีรูปแบบการใช้งานแตกต่างจากค่าเฉลี่ยทั่วไปแต่ไม่ใช่ภัยคุกคามจริง เช่น การเข้าสู่ระบบนอกเวลาทำงานของผู้ใช้งานที่ทำงานเป็นกะ หรือผู้ดูแลระบบที่ต้องเข้าถึงหลายโฮสต์ในช่วงเวลาสั้น ขณะที่ผลลบลวงมักพบในกรณีผู้ใช้งานที่มีสิทธิ์ระดับสูงซึ่งสามารถดำเนินกิจกรรมที่มีความเสี่ยงได้โดยไม่แสดงพฤติกรรมเบี่ยงเบนอย่างชัดเจน แนวโน้มดังกล่าวสะท้อนว่าข้อผิดพลาดของแบบจำลองไม่ได้เกิดขึ้นแบบสุ่ม แต่มีความสัมพันธ์กับบริบทด้านบทบาทและลักษณะงานของผู้ใช้ ซึ่งในเชิงปฏิบัติอาจส่งผลให้เกิดการแจ้งเตือนเกินความจำเป็นหรือการพลาดเหตุการณ์สำคัญในสภาพแวดล้อมของศูนย์ปฏิบัติการความมั่นคงปลอดภัยไซเบอร์ (SOC) และทีมตอบสนองเหตุการณ์ (CSIRT) ทั้งนี้ การมีชุดกฎที่สามารถตีความได้ช่วยให้ผู้เชี่ยวชาญสามารถนำผลลัพธ์ไปปรับแต่งนโยบายหรือเพิ่มเงื่อนไขเชิงบริบทเพิ่มเติม เพื่อบรรเทาผลกระทบของข้อผิดพลาดดังกล่าวในการใช้งานจริง

ดังนั้นงานวิจัยนี้บรรลุวัตถุประสงค์ในการเพิ่มประสิทธิภาพการตรวจจับภัยคุกคามภายใน พร้อมทั้งนำเสนอกรอบการสร้างแบบจำลองภัยคุกคามแบบผสมผสานเชิงปริมาณ (Hybrid Threat Modeling) ที่ผสานสไตรด์เข้ากับอัลกอริทึมเชิงวิวัฒนาการ โดยเฉพาะแบบจำลองสไตรด์ที่ผสานกับการวิวัฒนาการเชิงผลต่าง ซึ่งไม่เพียงแข่งขันได้ในเชิงประสิทธิภาพกับแบบจำลองการเรียนรู้ของเครื่องและแนวทางแบบกราฟเท่านั้น แต่ยังยกระดับสไตรด์จากการวิเคราะห์เชิงแนวคิดไปสู่การประเมินเชิงตัวเลขที่สามารถตรวจสอบ เปรียบเทียบ และประเมินประสิทธิภาพได้อย่างเป็นระบบ พร้อมทั้งคงไว้ซึ่งความโปร่งใสและความสามารถในการตีความเชิงแบบจำลองภัยคุกคาม ซึ่งสอดคล้องกับแนวโน้มงานปัญญาประดิษฐ์เชิงอธิบายได้ ด้านความมั่นคงปลอดภัยไซเบอร์ในปัจจุบัน และเอื้อต่อการประยุกต์ใช้จริงในสภาพแวดล้อมองค์กรที่ต้องการความมั่นคงปลอดภัยระดับสูง

5. กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณห้องปฏิบัติการแห่งชาติลอสอาลามอส (Los Alamos National Laboratory; LANL) ได้เผยแพร่ชุดข้อมูลเหตุการณ์การพิสูจน์ตัวตนของ LANL (LANL Authentication Dataset) เพื่อใช้ในการศึกษาและทดสอบแบบจำลองตรวจจับภัยคุกคามภายในในงานวิจัยนี้

เอกสารอ้างอิง

[1] L. P. da Silva, B. S. Nascimento, R. A. M. P. Dias, and D. S. Mendonça, “A comprehensive approach for applying threat modeling to internet of things systems,” *IEEE 8th World Forum on Internet of Things (WF-IoT), Yokohama, Japan*, 2022, pp. 1–6, doi: 10.1109/WF-IoT 54382.2022.10152291.

[2] L. Nikolov and A. Aleksieva-Petrova, “Framework for integrating threat modeling into a devOps pipeline for enhanced software development,” in *Proceedings the International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Split, Croatia, 2024, pp. 1–5, doi: 10.23919/SoftCOM6-2040.2024.10721871.

[3] S. Preetam, M. Compastié, V. Daza, and S. Siddiqui, “An approach for intelligent behaviour-based threat modelling with explanations,” in *Proceedings the IEEE Conference on NFV-SDN, Dresden, Germany*, 2023, pp. 197–200, doi: 10.1109/NFV-SDN59219.2023.10329587.

[4] D. C. Le and N. Zincir-Heywood, “Anomaly detection for insider threats using unsupervised ensembles,” *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1152–

1164, 2021, doi: 10.1109/TNSM.2021.3071928.

[5] F. R. Alzaabi and A. Mehmood, "A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods," *IEEE Access*, vol. 12, pp. 30907–30927, 2024, doi: 10.1109/ACCESS.2024.3369906.

[6] R. Zhu, X. Wu, J. Sun, and Z. Li, "Research on smart home security threat modeling based on STRIDE-IAHP-BN," in *Proceedings DCABES, Nanning, China*, 2021, pp. 207–213, doi: 10.1109/DCABES52998.2021.00059.

[7] A. D. Kent, "Cyber security data sources for dynamic network research," in *Dynamic Networks and Cyber-Security, Singapore: World Scientific*, 2016, pp. 37–65, doi: 10.1142/9781786340757_0002.

[8] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," *arXiv preprint arXiv:1710.00811*, 2017, doi: 10.48550/arXiv.1710.00811.

[9] Microsoft, "The STRIDE threat model," *Microsoft Security Development Lifecycle (SDL)*. [Online]. Available: https://www.microsoft.com/en-us/securityengineering/sdl/threatmodeling.

[10] M. Shin, S. Dorbala, and D. Jang, "Threat modeling for security failure-tolerant requirements," in *Proceedings the International Conference on Social Computing*, pp. 594–599, doi: 10.1109/SocialCom.2013.89.

[11] E. R. Agustina, A. R. Hakim, and K. Ramli, "Modeling data security and privacy threats for VANET using STRIDE and LINDDUN," in *Proceedings ICoSEIT*, Bandung, Indonesia, 2024, pp. 114–119, doi: 10.1109/ICoSEIT60086.2024.10497513.

[12] S. Manzoor, H. Zhang, and N. Suri, "Threat modeling and analysis for the cloud ecosystem," in *Proceedings IEEE IC2E, Orlando, FL, USA*, 2018, pp. 278–281, doi: 10.1109/IC2E.2018.00056.

[13] M. Kravchik and A. Shabtai, "Detecting cyber attacks in industrial control systems using convolutional neural networks," in *Proceedings CPS-SPC*, 2018, pp. 72–83, doi: 10.1145/3264888.3264896.

[14] A. Magklaras and S. Furnell, "Insider threat prediction tool: Evaluating the probabilities of IT misuse," *Computers & Security*, vol. 21, no. 1, pp. 62–73, 2002, doi: 10.1016/S0167-4048(02)00109-8.

[15] A. Tosun, M. A. Andresen, and C. D. Jensen, "Threat modeling made simple: Method aimed at non-experts," in *Proceedings IST-Africa, Dublin, Ireland*, 2024, pp. 1–9, doi: 10.23919/IST-Africa63983.2024.10569804.

[16] Y. Wang and G. Tuerhong, "A Survey of interpretable machine learning methods," in *Proceedings VRHCIAI, Changsha, China*, 2022, pp. 232–237, doi: 10.1109/VRHCIAI57205.2022.00047.

[17] D. Sridevi, L. Kannagi, V. G, and S. Revathi, "Detecting insider threats in cybersecurity using machine learning and deep learning techniques," in *Proceedings ICCSAI, Greater Noida, India*, 2023, pp. 871–875, doi: 10.1109/ICCSAI59793.2023.10421133.

[18] N. Capuano, G. Fenza, V. Loia, and C. Stanzione,

"Explainable artificial intelligence in cybersecurity: A survey," *IEEE Access*, vol. 10, pp. 93575–93600, Sep. 2022, doi: 10.1109/ACCESS.2022.3204171.

[19] A. Jawad, J. Jaskolka, A. Matrawy, and M. Ibnkahla, "StrideSEA: A STRIDE-centric security evaluation approach," *arXiv:2503.19030*, 2025, doi: 10.48550/arXiv.2503.19030.

[20] N. Hu, P. G. Bradford, and J. Liu, "Applying role-based access control and genetic algorithms to insider threat detection," in *Proceedings ACMSE*, 2006, pp. 790–791, doi: 10.1145/1185448.1185638.

[21] D. C. Le, S. Khanchi, A. N. Zincir-Heywood, and M. I. Heywood, "Benchmarking evolutionary computation approaches to insider threat detection," in *Proceedings GECCO*, 2018, pp. 1286–1293, doi: 10.1145/3205455.3205612.

[22] A. Masood and A. Masood, "A taxonomy of insider threat in isolated (air-gapped) computer networks," in *Proceedings IBCAST, Islamabad, Pakistan*, 2021, pp. 678–685, doi: 10.1109/IBCAST51254.2021.9393281.

[23] V. Maheshwari and M. Prasanna, "Integrating risk assessment and threat modeling within SDLC process," in *Proceedings ICICT, Coimbatore, India*, 2016, pp. 1–5, doi: 10.1109/INVENTIVE.2016.7823275.

[24] Y. Gong, S. Cui, S. Liu, B. Jiang, C. Dong, and Z. Lu, "Graph-based insider threat detection: A survey," *Computer Networks*, vol. 254, 2024, Art. no. 110757. doi: 10.1016/j.comnet.2024.110757.

[25] B. B. Sarhan and N. Altwaijry, "Insider threat detection using machine learning approach," *Applied Sciences*, vol. 13, no. 1, 2023, Art. no. 259, doi: 10.3390/app13010259.

[26] G. Rjoub, J. Bentahar, O. A. Wahab, R. Mizouni, A. Song, R. Cohen, H. Otrok, and A. Mourad, "A survey on explainable artificial intelligence for cybersecurity," *IEEE Transactions on Network and Service Management*, vol. 20, no. 4, pp. 5115–5140, Dec. 2023, doi: 10.1109/TNSM.2023.3282740.