Jurnal Riset Informatika dan Teknologi Informasi (JRITI)
Jejaring Penelitian dan Pengabdian Masyarakat (JPPM)
Vol. 3, No. 2, Desember 2025 – Maret 2026, Hal. 338 - 353
Web : https://ejurnal.jejaringppm.org/index.php/jriti
ISSN (Print) : 3024-8167
E-ISSN : 3109-8959

Penerapan Kecerdasan Buatan dalam Keamanan Siber pada Infrastruktur Kritis: Tinjauan Sistematis terhadap Ancaman, Solusi, dan Tantangan

Mufid Athooyaa[1], Satria Krisna Prabantara[2], Dwi Alvin Hidayat[3], Shaviraj Samad Shaikh[4], dan Arief Arfriandi[5]
Universitas Negeri Semarang[1,2,3,4,5]
Email: mufidathooyaa13@students.unnes.ac.id, satriaprabantara19@students.unnes.ac.id, dwihidayat285@students.unnes.ac.id, samadshaviraj2@students.unnes.ac.id, arfriandi@mail.unnes.ac.id

ABSTRAKSI
Infrastruktur kritis seperti sistem energi, jaringan distribusi air, fasilitas kesehatan, dan sistem transportasi merupakan tulang punggung operasional masyarakat modern yang sangat bergantung pada teknologi digital dan sistem *cyber-physical*. Konvergensi teknologi informasi (IT) dan teknologi operasional (OT) dalam *Industrial Control Systems* (ICS) dan SCADA telah meningkatkan efisiensi operasional, namun secara bersamaan memperluas permukaan serangan yang dapat dieksploitasi oleh aktor jahat. Artikel ini menyajikan tinjauan sistematis terhadap 89 publikasi periode 2020-2025 untuk menganalisis perkembangan penerapan kecerdasan buatan (AI) dalam keamanan siber infrastruktur kritis. Hasil kajian mengidentifikasi enam kategori utama serangan siber, diantaranya serangan terhadap jaringan dan komunikasi industri, manipulasi data dan injeksi perintah, Advanced Persistent Threats (APT), malware dan ransomware, insider threats, serta ancaman terhadap sistem berbasis AI. Penelitian ini menunjukkan bahwa algoritma AI, termasuk deep learning (CNN, LSTM, Transformer), machine learning klasik (Random Forest, SVM), Generative Adversarial Networks (GAN), Reinforcement Learning, dan Federated Learning memberikan kontribusi signifikan dalam deteksi dini anomali, respons adaptif, dan pemulihan sistem dengan tingkat akurasi mencapai 96-99%. Namun, implementasi AI menghadapi tantangan berupa kompleksitas komputasi tinggi, keterbatasan dataset, kerentanan terhadap adversarial attacks, serta kebutuhan transparansi dan interpretabilitas. Artikel ini merekomendasikan pengembangan model hybrid yang efisien, integrasi explainable AI, penerapan federated learning lintas sektor, dan pembentukan kerangka kolaboratif untuk membangun sistem keamanan siber yang tangguh dan berkelanjutan.

Kata Kunci: *Keamanan Siber, Infrastruktur Kritis, Kecedasan Buatan, Sistem Kontrol Industri, Tinjauan Pustaka Sistematis*

ABSTRACT
Critical infrastructure such as energy systems, water distribution networks, healthcare facilities, and transportation systems constitute the operational backbone of modern society, which heavily relies on digital technology and cyber-physical systems. The convergence of information technology (IT) and operational technology (OT) in Industrial Control Systems (ICS) and SCADA has enhanced operational efficiency, yet simultaneously expanded the attack surface that can be exploited by malicious actors. This article presents a systematic review of 89 publications from the 2020-2025 period to analyze the development of artificial intelligence (AI) applications in critical infrastructure cybersecurity. The review identifies six main categories of cyber attacks, such as attacks on industrial networks and communications, data manipulation and command injection, Advanced Persistent Threats (APT), malware and ransomware, insider threats, and threats against AI-based systems. This research demonstrates that AI algorithms, including deep learning (CNN, LSTM, Transformer), classical machine learning (Random Forest, SVM), Generative Adversarial Networks (GAN), Reinforcement Learning, and Federated Learning, contribute significantly to early anomaly detection, adaptive response, and system recovery with accuracy rates reaching 96-99%. However, AI implementation faces challenges including high computational complexity, dataset limitations, vulnerability to adversarial attacks, and the need for transparency and interpretability. This article recommends the development of efficient hybrid models, integration of explainable AI, cross-sector implementation of federated learning, and establishment of collaborative frameworks to build robust and sustainable cybersecurity systems.

Keywords: *Cybersecurity, Critical Infrastructure, Artificial Intelligence (AI), Industrial Control System (ICS), Systematic Literature Review (SLR)*

Penulis Korespondensi
Mufid Athooyaa

Tanggal Submit : 14/11/2025
Tanggal Diterima : 30/11/2025
Tanggal Terbit : 23/12/2025

I. PENDAHULUAN

Infrastruktur kritis seperti sistem energi, jaringan distribusi air, fasilitas kesehatan, sistem transportasi, dan jaringan telekomunikasi merupakan tulang punggung operasional masyarakat modern yang sangat bergantung pada teknologi digital dan sistem cyber-physical (CPS). Konvergensi antara teknologi informasi (IT) dan teknologi operasional (OT) dalam Industrial Control Systems (ICS) dan Supervisory Control and Data Acquisition (SCADA) telah meningkatkan efisiensi operasional, namun secara bersamaan memperluas permukaan serangan (attack surface) yang dapat dieksploitasi oleh aktor jahat [1], [2], [3]. Serangan siber terhadap infrastruktur kritis tidak hanya mengancam integritas data dan ketersediaan layanan, tetapi juga dapat menimbulkan konsekuensi fisik yang berbahaya, kerugian ekonomi masif, dan bahkan mengancam keselamatan publik [4], [5], [6].

Kompleksitas ancaman siber terhadap infrastruktur kritis terus meningkat dengan munculnya serangan canggih seperti Advanced Persistent Threats (APT), False Data Injection Attacks (FDIA), zero-day exploits, dan serangan tersembunyi (stealthy attacks) yang dirancang untuk menghindari deteksi sistem keamanan konvensional [7], [8], [9]. Metode deteksi intrusi tradisional berbasis signature dan rule-based systems menghadapi keterbatasan signifikan dalam mengenali pola serangan baru yang tidak ada dalam database tanda tangan, serta tidak mampu beradaptasi dengan lanskap ancaman yang dinamis [4], [10], [11]. Ketidakseimbangan data antara lalu lintas normal dan aktivitas berbahaya dalam dataset keamanan siber mempersulit pengembangan model deteksi yang efektif [12], [13], [14].

Dalam konteks ini, kecerdasan buatan (Artificial Intelligence/AI) dan pembelajaran mesin (Machine Learning/ML) telah muncul sebagai pendekatan transformatif untuk meningkatkan keamanan siber infrastruktur kritis. Algoritma AI menawarkan kemampuan untuk mendeteksi anomali secara otomatis, mempelajari pola serangan yang kompleks, beradaptasi dengan ancaman baru, dan memberikan respons yang cerdas tanpa memerlukan intervensi manual berkelanjutan [15], [16], [17]. Berbagai paradigma pembelajaran—mulai dari supervised learning, unsupervised learning, hingga reinforcement learning—telah diterapkan dengan tingkat keberhasilan yang bervariasi tergantung pada karakteristik sistem dan jenis ancaman yang dihadapi [18], [19], [20].

Namun demikian, implementasi AI dalam keamanan siber juga menghadirkan tantangan baru. Model deep learning yang kompleks rentan terhadap adversarial attacks di mana penyerang dapat memanipulasi input dengan perubahan minimal untuk mengelabui sistem deteksi [21], [22], [23]. Kebutuhan sumber daya komputasi tinggi membatasi penerapan di edge devices dan sistem embedded [24], [25], [26]. Keterbatasan dataset pelatihan, terutama data serangan nyata yang sering diklasifikasikan sebagai rahasia, menjadi hambatan dalam pengembangan model yang robust [17], [27], [28]

Artikel review ini bertujuan mengulas perkembangan riset terkini mengenai penerapan AI dalam keamanan siber untuk ICS/CI. Selain itu, artikel ini akan membahas tantangan-tantangan yang ada, baik dari aspek teknis, data, regulasi, maupun etika. Dengan menganalisis secara komprehensif, review ini juga akan memberikan wawasan mengenai arah riset masa depan, termasuk pengembangan *explainable* AI, *federated learning*, peningkatan ketahanan siber *(cyber resilience)*, dan strategi pertahanan adaptif berbasis AI.

II. PENELITIAN YANG TERKAIT

Penelitian mengenai keamanan siber pada infrastruktur kritis dan sistem kontrol industri telah berkembang pesat seiring meningkatnya integrasi teknologi digital dan *cyber-physical systems* (CPS). Kajian terdahulu menunjukkan bahwa ancaman terhadap sistem industri modern tidak hanya meningkat dari sisi jumlah, tetapi juga dari kompleksitas dan kemampuan adaptif serangan. Oleh karena itu, berbagai pendekatan berbasis kecerdasan buatan, teori permainan, serta teknologi baru seperti *federated learning* dan *large language models* (LLMs) dikembangkan untuk memperkuat ketahanan sistem terhadap ancaman siber.

Ayachi et al. (2022) dalam penelitian menyoroti pentingnya kualitas dataset dalam pengembangan *Intrusion Detection System* (IDS). Mereka mengusulkan penggunaan *Information Security and Object Technology–Cloud Intrusion Dataset* (ISOT-CID), yang lebih representatif dibanding dataset lama seperti NSL-KDD. Melalui pengujian beberapa algoritma *machine learning* seperti *Random Forest*, *Gradient Boosting*, dan *Artificial Neural Network*, penelitian ini menunjukkan bahwa keseimbangan dan relevansi data berpengaruh lebih besar terhadap akurasi deteksi dibanding kompleksitas model semata [29].

Alharthi et al. (2025) dalam studi memperluas pendekatan tersebut dengan membandingkan model *machine learning* (ML) dan *deep learning* (DL) pada dua dataset industri (SDN dan IEC 60870-5-104). Hasil penelitian menunjukkan bahwa model klasik seperti *Random Forest* dan *XGBoost* tetap unggul pada klasifikasi multikelas, mengungguli CNN dan LSTM dalam efisiensi dan stabilitas. Studi ini juga menekankan pentingnya pertimbangan *computational cost* dan ketidakseimbangan data dalam desain IDS yang dapat diterapkan pada sistem industri nyata [30].

Pendekatan konseptual yang berbeda ditawarkan oleh Mejdi et al. (2024) melalui kajian yang menerapkan *game theory* untuk menganalisis interaksi antara penyerang dan pembela pada CPS. Dengan meninjau lebih dari 800 publikasi, penelitian ini mengkategorikan lima kerangka permainan utama—*zero-sum*, *non-zero-sum*, *stochastic*, *differential*, dan *Stackelberg games*—dan menunjukkan bahwa model *Stackelberg* efektif untuk strategi pertahanan proaktif, sementara *stochastic games* lebih cocok untuk menghadapi ketidakpastian tinggi [31].

Dalam ranah deteksi intrusi berbasis AI pada penelitian, Canonico dan Sperlì (2023) mengklasifikasikan serangan CPS industri berdasarkan metodologi dan jenis pertahanan. Mereka menegaskan bahwa *deep learning*—khususnya CNN dan LSTM—memiliki keunggulan dalam menangani data berdimensi tinggi dan mengenali pola temporal jaringan [32]. Holdbrook et al. (2024) pada studi menambahkan bahwa keterbatasan dataset publik seperti NSL-KDD dan SWaT menjadi hambatan utama bagi efektivitas IDS industri. Mereka merekomendasikan pengembangan dataset baru yang merepresentasikan lalu lintas spesifik industri, protokol, dan skenario operasional seperti *smart grids* atau *automated factories* [33], [34].

Dalam konteks infrastruktur militer, Semenenko et al. (2024) menunjukkan bahwa integrasi *renewable energy*, *intelligent energy management systems*, dan *microgrids* meningkatkan efisiensi energi dan otonomi fasilitas, tetapi memerlukan sistem *cybersecurity* berlapis untuk melindungi jaringan kontrol otomatis dari serangan [35], [36].

Sementara itu, Yigit et al. (2025) meninjau penerapan *Generative AI* dan *Large Language Models* untuk *Critical Infrastructure Protection* (CIP) [37]. Mereka memperkenalkan dataset evaluasi seperti SECURE dan NetEval, serta menekankan pentingnya *federated learning* untuk menjaga privasi data tanpa mengurangi kemampuan deteksi ancaman.

Secara keseluruhan, literatur terdahulu menunjukkan tiga arah utama perkembangan riset keamanan siber untuk infrastruktur kritis, yaitu (1) peningkatan efektivitas IDS berbasis AI dan dataset yang representatif, (2) penerapan *game theory* untuk strategi pertahanan adaptif, dan (3) integrasi teknologi baru seperti LLMs, *digital twins*, dan *federated learning*. Meski kemajuan signifikan telah dicapai, tantangan utama masih meliputi keterbatasan dataset industri, deteksi serangan tersembunyi (*stealthy attacks*), serta integrasi AI dengan sistem *legacy*. Oleh karena itu, sinergi antara pembelajaran mesin, teori permainan, dan teknologi cerdas menjadi arah strategis penelitian di masa depan untuk meningkatkan keamanan dan resiliensi sistem industri modern.

III. METODE PENELITIAN

Artikel review ini menggunakan pendekatan tinjauan literatur sistematis (*Systematic Literatur Review*/*SLR*). Pendekatan sistematis ini bertujuan untuk menelusuri serta mengidentifikasi seluruh bukti empiris yang memenuhi kriteria inklusi yang telah ditetapkan sebelumnya guna menjawab pertanyaan penelitian yang dirumuskan oleh peneliti, [Literatur review as a research method]. Dalam melakukan SLR, terdapat empat langkah yang dilakukan, yaitu (1) perencanaan tinjauan, (2) melaksanakan tinjauan, (3) analisis, dan (4) menulis tinjauan, [38].

Langkah pertama yang dilakukan pada perencanaan tinjauan adalah penetapan protokol. Protokol menjelaskan prosedur yang dilakukan oleh peneliti dalam melakukan tinjauan agar tinjauan dapat direplikasi, [39]. Dalam melakukan perencanaan tinjauan, peneliti menggunakan prosedur PICOC (Population, Intervention, Comparison, Outcome, dan Context) [39] untuk menguraikan tujuan SLR dalam kata kunci yang teridentifikasi sehingga dapat membantu penyusunan pertanyaan penelitian (*Research Question/RQ*) [40].

A. Perencanaan

Prosedur PICOC digunakan untuk mendefinisikan objek yang akan dikaji menjadi kata kunci dan mendefinisikan pertanyaan peneliti. Dalam penjabaran elemen dari PICOC, sinonim dari masing-masing penjabaran dibutuhkan untuk membuat kueri pencarian pada pencarian database, [39]. Elemen-elemen dari PICOC yang terpapar pada Tabel 1 digunakan untuk membuat RQ. Pada penelitian ini, peneliti mengajukan dua RQ berdasarkan elemen PICOC yang telah ditetapkan, yaitu:

1. Apa saja jenis ancaman siber pada Infrastruktur Kritis?
2. Bagaimana algoritma-algoritma AI yang berbeda berkontribusi terhadap peningkatan keamanan siber?

Tabel 1. Elemen PICOC

Elemen	Kata Kunci	Sinonim	Deskripsi
P	*Cyber Threat*	*Cyber Attack, Cybercrime, Information Security Threat*	Fokus pada tren ancaman siber modern yang semakin kompleks dengan adanya Kecerdasan Buatan (AI).
I	*Artificial Intelligence*	*AI, Machine Learning, Deep Learning, Generative AI*	Pemanfaatan Kecerdasan Buatan (AI) baik untuk memperkuat serangan siber (offensif) maupun untuk membangun keamanan siber (defensif)
C	-	-	Tidak ada perbandingan. Penelitian berfokus pada pemetaan peran AI dalam keamanan siber
O	*Cyber Security*	*Information Technology Security*	Upaya menjaga kerahasiaan, integritas, dan ketersediaan informasi dalam menghadapi ancaman berbasis AI.
C	*Critical Infra-structure*	*Industrial Control System*	Lingkungan organisasi/industri yang menggunakan sistem informasi digital dan rentan terhadap serangan siber.

Tabel 2. Kriteria Inklusi

Kriteria	Deskripsi	Inklusi
Area	Kriteria ini memastikan seluruh artikel yang disertakan berfokus pada bidang ilmu komputer.	Computer Science
Periode	Rentang waktu ini dipilih untuk menjamin bahwa penelitian yang dikaji bersifat terbaru dan relevan dengan perkembangan teknologi terkini.	2020 - 2025
Bahasa	Hanya artikel berbahasa Inggris yang disertakan untuk menjaga konsistensi terminologi ilmiah dan memudahkan proses analisis literatur internasional.	Inggris
Tipe Dokumen	Artikel yang dipilih merupakan hasil penelitian asli bukan review, editorial, laporan teknis, atau prosiding konferensi.	Artikel Penelitian
Tipe Sumber	Kriteria ini membatasi sumber publikasi pada jurnal ilmiah yang terakreditasi bereputasi. Jurnal dipilih karena memiliki proses *peer review* yang menjamin kualitas, validitas, dan kredibilitas informasi yang disajikan.	Jurnal
Aksesibilitas	Hanya artikel yang dapat diakses secara terbuka dan gratis yang disertakan. Hal ini memastikan bahwa seluruh referensi dapat diverifikasi, diakses kembali, dan dimanfaatkan oleh peneliti lain tanpa hambatan berlangganan atau pembatasan akses.	Akses Terbuka

B. Pelaksanaan Tinjauan

Langkah kedua yang dilakukan adalah melaksanakan tinjauan. Tinjauan dilakukan dengan menggunakan protokol PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta Analysis*). Dalam artikel [41], metode yang dapat diterapkan dalam peninjauan artikel berupa penentuan (1) kriteria kelayakan, (2) sumber informasi, (3) strategi pencarian, (4) proses pengumpulan data.

C. Kriteria Kelayakan

Kriteria kelayakan merupakan pedoman yang digunakan peneliti untuk menentukan artikel yang diinklusi dan dikecualikan dalam proses tinjauan, [41]. Dalam penelitian ini, peneliti menetapkan kriteria kelayakan sebagaimana terpapar pada Tabel 2.

D. Sumber Informasi

Sumber database yang digunakan pada penelitian ini adalah SCOPUS. Saat ini, belum terdapat pedoman baku yang secara khusus mengatur penggunaan database, seperti database mana yang terbaik, dan berapa banyak yang harus dicari. Namun, secara umum disarankan untuk melakukan pencarian literatur secara luas untuk memperoleh hasil tinjauan yang komprehensif, [42]. SCOPUS merupakan salah satu database bibliografi terbesar dan paling bereputasi di dunia yang dikelola oleh Elsevier. Database ini mencakup berbagai disiplin ilmu, seperti teknik, ilmu komputer, kedokteran, ilmu sosial dan humaniora, serta menyediakan akses ke artikel jurnal, prosiding konferensi, dan dokumen ilmiah lainnya yang telah melalui proses *peer-review*. Sumber database SCOPUS dapat diakses melalui laman resmi https://www.scopus.com/.

E. Strategi Pencarian

Pencarian literatur yang diulas dalam penelitian ini menggunakan database SCOPUS dengan menerapkan pencarian lanjutan (*advanced search*) melalui penggunaan string pencarian. String pencarian tersebut disusun berdasarkan elemen-elemen PICOC yang terpapar pada Tabel 1. Dalam penyusunannya, string pencarian memperhatikan setiap elemen dan sinonim dari komponen PICOC, di mana setiap sinonim dipisahkan menggunakan operator OR, sedangkan setiap elemen PICOC dipisahkan menggunakan tanda kurung beserta operator AND [39]. String pencarian yang digunakan pada pencarian perpustakaan database sebagaiaman berikut.

TITLE-ABS-KEY(("Cyber Threat" OR "Cyber Attack" OR "Cybercrime" OR "Information Security Threat" OR "Digital Threat") AND ("Artificial Intelligence" OR

"AI" OR "Machine Learning" OR "Deep Learning" OR "Generative AI") AND ("Critical Infrastructure" OR "Industrial Control System" OR "ICS"))

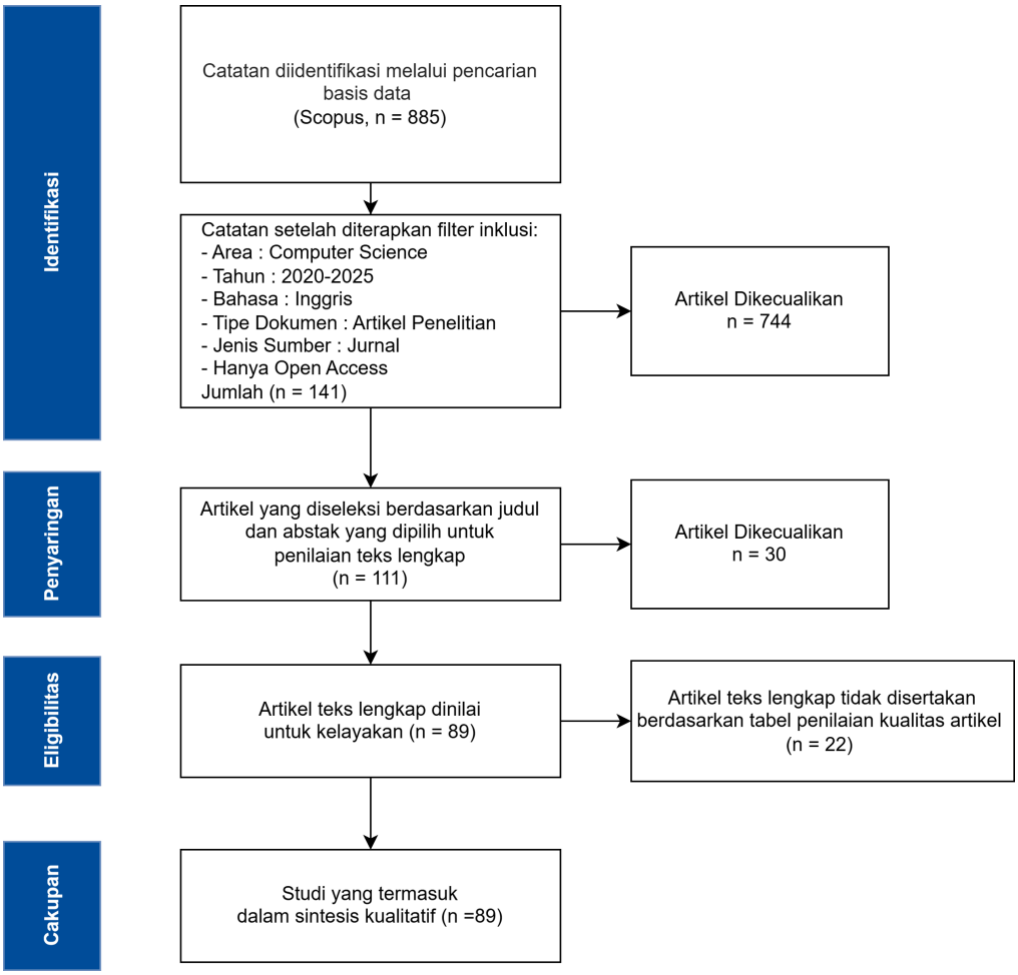
F. Proses Pengumpulan Data

Proses pengumpulan data diawali dengan pencarian literatur pada database yang telah ditentukan menggunakan string pencarian berdasarkan elemen PICOC. Penerapan string pencarian pada database SCOPUS menghasilkan sebanyak 885 artikel. Selanjutnya, prosedur inkulsi dan eksklusi diterapkan untuk mendapatkan literatur yang relevan dengan topik kajian. Sebanyak 141 artikel diperoleh dari hasil penerapan kriteria. Artikel-artikel tersebut kemudian disintesis berdasarkan penilaian kualitas artikel [41].

Penilaian kualitas artikel dilakukan menggunakan skala penilaian yang telah ditetapkan oleh peneliti, [39]. Setiap artikel dinilai berdasarkan kriteria pada Tabel 3. Proses penilaian artikel dilakukan melalui pembacaan menyeluruh terhadap abstrak dan teks lengkap (*full text*) dari masing-masing artikel. Dalam proses penilaian, para peneliti membagi tugas penelaahan kepada empat auhor. Author pertama bwetanggung jawab mengulas artikel yang diterbitkan pada tahun 2025, author kedua untuk tahun 2024, author ketiga untuk periode tahun 2022 – 2023, dan author keempat untuk periode tahun 2020 – 2021. Hanya artikel yang memperoleh skor sintesis ≥ 3,5 yang dianggap memenuhi kriteria kualitas dan layak untuk dianalis lebih lanjut dalam proses tinjauan sistematis.

Tabel 3. Aspek Penilaian Kualitas Artikel

Aspek	Penilaian
Apakah literatur menjawab kedua pertanyaan penelitian yang diajukan peneliti?	1 – Tidak (Skor: 0) 2 – Sebagian (Skor: 0.5) 3 – Ya (Skor: 1)
Apakah literatur yang tersaring berupa artikel *review* (bukan *primary study*)?	1 – Tidak (Skor: 1) 2 – Ya (Skor: 0)
Apakah artikel yang tersaring relevan dengan topik yang diangkat? (Infrastruktur Kritis, Sistem Kontrol Industri, dan teknologi AI dalam keamanan siber) Apakah literatur menyajikan *framework* dan eksperimen yang jelas?	1 – Tidak (Skor: 0) 2 – Ya (Skor: 1) 1 – Tidak (Skor: 0) 2 – Sebagian (Skor: 0.5) 3 – Ya (Skor: 1)

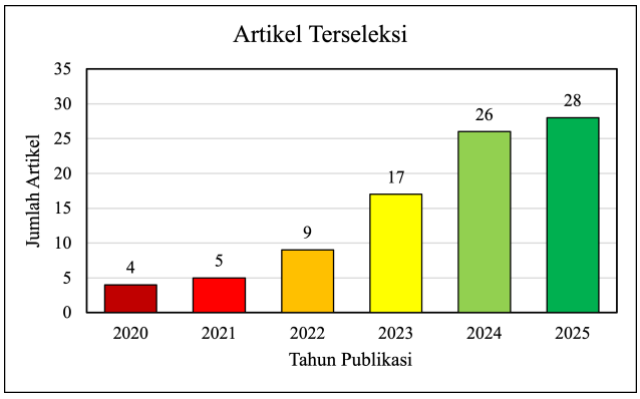


Gambar. 1 Alur Kerja PRISMA.

Berdasarkan hasil pengulasan artikel, didapatkan 89 artikel yang layak untuk dianalisis lebih lanjut. Artikel-artikel tersebut digunakan untuk menjawab RQ yang telah ditetapkan oleh peneliti. Proses seleksi literatur secara keseluruhan digambarkan pada Gambar 1, yang menunjukkan identifikasi, penyaringan, penilaian kelayakan, hingga inklusi artikel akhir.

IV. HASIL DAN PEMBAHASAN

Sebanyak 92 artikel telah terseleksi pada bagian metode penelitian. Artikel-artikel tersebut diterbitkan pada selang periode tahun 2020 – 2025. Gambar 2 menampilkan distribusi artikel yang digunakan pada artikel tinjauan ini. Terlihat adanya tren peningkatan tiap tahunnya yang menunjukkan topik ini masih menjadi fokus penelitian hingga saat ini.



Gambar. 2 Distribusi Artikel dalam Kajian Literatur.

A. Klasifikasi Serangan Siber

Hasil kajian dari artikel-artikel terpilih menunjukkan bahwa infrastruktur kritis menghadapi beragam jenis serangan siber yang memiliki karakteristik berbeda. Serangan tersebut dapat dikategorikan ke dalam beberapa kelompok utama sebagaimana berikut.

1. Serangan terhadap Jaringan dan Komunikasi Industri

Ketersediaan jaringan dan komunikasi adalah pilar utama operasi infrastruktur kritis, dan penyerang secara agresif menargetkan lapisan ini. Denial-of-Service (DoS) dan Distributed Denial-of-Service (DDoS) merupakan ancaman yang paling banyak disoroti, bertujuan melumpuhkan layanan atau jaringan dengan membanjiri lalu lintas komunikasi ICS atau *Software-Defined Networking* (SDN) Controller [7], [10], [13], [15], [18], [27], [43], [44], [45], [46], [47], [48], [49]. Varian serangan ini mencakup SYN, UDP, ICMP, dan HTTP Flooding [13], [27], [44], [50]. serta serangan spesifik seperti DoS GoldenEye dan DoS Hulk [2], [13], [44].

Ancaman lain yang menargetkan integritas komunikasi adalah serangan intervensi jalur. Man-in-the-Middle (MITM) Attacks memungkinkan penyerang mencegat dan memodifikasi data atau perintah yang dikirim antara *Supervisory Control and Data Acquisition* (SCADA) dan *Remote Terminal Unit* (RTU) [5], [7], [10], [12], [13], [27], [43], [50], [51], [52], [53], [54], [55], [56], [49]. Jurnal-jurnal juga mencatat varian MITM seperti ARP Poisoning dan DNS Spoofing [18], [43], [57]. Sementara itu, Serangan *Spoofing* diidentifikasi secara independen, berfokus pada pemalsuan identitas perangkat sah untuk mengirimkan informasi palsu [19], [24], [46], [58], [59], [60], [61].

2. Manipulasi Data dan Injeksi Perintah pada Sistem Siber-Fisik

Serangan canggih bergeser dari gangguan jaringan ke manipulasi langsung pada sistem kendali, yang melibatkan pengambilalihan kendali sistem ICS [1], [62].

- Manipulasi dan Injeksi Data (False Data Injection Attack/FDIA)

  Ini adalah salah satu ancaman yang paling banyak dibahas [3], [7], [9], [10], [11], [14], [15], [56], [59], [62], [63], [64], [65], [66], [67]. FDIA, termasuk Serangan Injeksi Data Palsu pada sensor [1], [10], [15], [20], [63], [68] dan Perturbation Attacks [15], bertujuan menipu sistem *state estimation* agar operator atau kontrol otomatis membuat keputusan salah, seperti mematikan generator [10] atau mengubah parameter kimia air [1].
- Command Injection Attack

  Penyerang menyuntikkan perintah ilegal langsung ke perangkat kontrol seperti *Programmable Logic Controller* (PLC) atau RTU melalui protokol industri (seperti Modbus/DNP3) [3], [7], [12], [14], [50], [51], [53], [57], [61], [62], [69], [70]. Serangan ini dapat memanipulasi aktuator [54], [62] atau mengirim perintah *Remote Tripping* ke pemutus arus [22], [71].
- Replay Attack

  Penyerang merekam dan memutar ulang data sensor atau perintah kontrol yang sah untuk menipu sistem agar menganggap kondisi berjalan normal padahal sedang disabotase [3], [7], [8], [11], [15], [51], [61], [62], [64], [70], [72], [49].

3. Advanced Persistent Threats (APT) dan Serangan Berbasis Spionase

Ancaman ini dicirikan oleh operasi jangka panjang dan target yang sangat spesifik [10], [57], [73], [74], [75], [76], [77], [78]. APT, sering kali *state-sponsored* atau peretas yang disponsori oleh negara [1], [54]. Ancaman ini menargetkan seluruh rantai serangan dari pengintaian hingga eksekusi [10], [73]. Serangan Reconnaissance [5], [12], [13], [18], [25], [43], [52], [53], [57], [77], [79], seperti *PortScan* [13], [18], [43], [44], [69], [77], [79], [80], [81] dan *Vulnerability Scanning* [43], [50], adalah fase krusial dalam persiapan APT [77].

4. Malware dan Ransomware Attacks

Malware industri menargetkan *firmware* dan logika kontrol, menimbulkan kerugian fisik yang ekstrem [4], [10], [12], [13], [26], [43], [51], [82], [71]. Insiden bersejarah yang secara konsisten diidentifikasi sebagai studi kasus penting meliputi:

- Stuxnet
  Malware mani yang memanipulasi PLC Siemens untuk merusak sentrifugal nuklir [10], [11], [19], [51], [55], [63], [69], [83], [84].
- BlackEnergy
  Malware yang digunakan dalam serangan pemadaman listrik di Ukraina [63], [83], [85] .
- Triton (Trisis)
  Malware yang secara eksplisit menargetkan *Safety Instrumented Systems* (SIS), sistem pertahanan terakhir terhadap bencana industri [55], [63].

Selain malware yang spesifik untuk ICS, serangan umum lainnya seperti Ransomware [12], [17], [28], [43], [59], [67], [86], [87], [88] dan Trojan/Backdoor [12], [18], [26], [43], [80] juga menjadi ancaman serius, seringkali menargetkan sistem medis [87] atau rantai pasok [88].

5. Insider Threats dan Kesalahan Manusia

Ancaman yang berasal dari dalam organisasi (Insider Threats) diakui sebagai vektor serangan yang sulit dideteksi [2], [13], [14], [17], [19], [54], [61], [63], [76], [78], [89]. Ancaman ini dapat melibatkan bebepa hal, seperti berikut.

- Penyalahgunaan hak akses (*privilege misuse*) oleh karyawan yang memiliki kredensial sah [89].
- Akses tidak sah (*unauthorized access*) dan eskalasi hak (*Privilege Escalation*) [11], [13], [20], [66], [69], [90], seringkali dicapai melalui Brute Force Attack [2], [13], [18], [79], [81], [91], [92], [93] atau Phishing [2], [13], [17], [28], [76], [78], [94] yang menargetkan operator manusia [17].

6. Ancaman Baru terhadap Sistem Berbasis AI dan Machine Learning

Integrasi AI/ML ke dalam sistem infrastruktur kritis dan sistem deteksi intrusi (IDS) telah menciptakan permukaan serangan baru [6], [90]. Ancaman ini bersifat meta-level, menargetkan model AI itu sendiri:

- Adversarial Attacks
  Serangan yang didesain untuk menipu model AI agar salah mengklasifikasikan data berbahaya sebagai normal [6], [21], [22], [23], [59], [82], [95], seringkali dengan menambahkan *noise* kecil (Adversarial Perturbation) [21].
- Data Poisoning Attacks
  Penyerang memanipulasi data pelatihan AI untuk merusak model, yang dapat menyebabkan keputusan yang salah pada sistem kontrol [6], [59], [65], [96], termasuk menciptakan *backdoor* pada model [21], [96].
- Ancaman AI Generatif
  Penggunaan *Generative Adversarial Networks* (GANs) untuk menciptakan data sensor atau citra radar palsu (*deepfake*) yang tampak realistis, menyesatkan sistem navigasi atau kontrol [95], [97].

B. Dampak Serangan Siber pada Infrastruktur Kritis

Setiap kategori serangan tersebut memiliki dampak signifikan terhadap keberlangsungan operasional sistem, termasuk gangguan layanan publik, kerugian finansial, hingga infrastruktur kritis. Serangan siber terhadap infrastruktur kritis, yang sering dikategorikan sebagai Cyber Disaster [1], menghasilkan dampak kaskade yang melampaui batas kerugian finansial [1].

1. Dampak Fisik dan Keselamatan Publik

Ancaman langsung terhadap keselamatan dan kesehatan manusia merupakan kekhawatiran terbesar. Serangan injeksi data pada *water treatment plant* dapat menyebabkan kontaminasi atau overdosis bahan kimia [1]. Serangan terhadap sistem listrik dapat menimbulkan pemadaman massal (blackout) [9], [10], [21], [70], sementara serangan pada sistem medis (IoMT) dapat membahayakan perangkat medis vital [80] atau mengganggu layanan kesehatan [87]. Penargetan sistem keselamatan industri (SIS) melalui malware seperti Triton secara eksplisit menunjukkan niat untuk menyebabkan bencana industri [55].

2. Stabilitas Operasional dan Ekonomi

Serangan pada ICS dan jaringan komunikasi (seperti DoS/DDoS) menyebabkan gangguan operasional dan downtime signifikan [1], [47], [69]. Manipulasi data pada sistem energi atau PLTA dapat mengakibatkan penurunan produksi energi [66], ketidakseimbangan daya [21], dan kerugian finansial besar [14].

3. Stabilitas Sosial dan Keamanan Nasional

Serangan yang disponsori negara (*state-sponsored*) menargetkan stabilitas sosial dan keamanan nasional [1], [10], [54]. Ancaman juga muncul pada fase *pra-serangan* melalui Serangan Berbasis Media Sosial seperti ujaran kebencian dan propaganda yang mengindikasikan rencana sabotase [94]. Serangan rantai pasok (*supply chain attack*) juga menyoroti kerentanan sistem di fase pengadaan perangkat lunak atau *firmware* [17], [88], [90], [98].

C. Kontribusi Algoritma AI terhadap Peningkatan Keamanan Siber pada Infrastruktur Kritis

Algoritma-algoritma kecerdasan buatan (AI) memberikan kontribusi signifikan dalam meningkatkan keamanan siber infrastruktur kritis melalui tiga dimensi utama: deteksi dini ancaman, respons adaptif terhadap serangan, dan pemulihan sistem. Setiap kategori algoritma menunjukkan karakteristik dan keunggulan

spesifik dalam menangani kompleksitas ancaman siber modern.

1. Algoritma Deep Learning untuk Deteksi Anomali dan Klasifikasi Serangan

Algoritma deep learning menunjukkan superioritas dalam mengekstraksi fitur kompleks dan mengenali pola serangan yang sulit diidentifikasi oleh metode konvensional. Recurrent Neural Network (RNN) dan variannya, khususnya Long Short-Term Memory (LSTM), terbukti sangat efektif dalam menganalisis data time-series dari sensor Industrial Control Systems (ICS) [48]. Penelitian menunjukkan bahwa RNN mampu memodelkan dependensi jangka panjang pada data sensor ICS, memungkinkan deteksi pola abnormal secara dini sebelum terjadi kerusakan signifikan [1]. Implementasi Bidirectional LSTM (BiLSTM) dalam arsitektur autoencoder mencapai tingkat deteksi (TPR) sebesar 96,63% dengan false positive rate hanya 0,07% pada jaringan Manufacturing Message Specification (MMS) [85][99].

Convolutional Neural Network (CNN) memberikan kontribusi penting dalam ekstraksi fitur spasial dari data multivariat. Arsitektur 1-Dimensional CNN terbukti efektif menangkap korelasi antar-sensor dalam sistem ICS, mengurangi kemungkinan false negative pada deteksi serangan [1], [48]. Kombinasi CNN dengan BiLSTM menghasilkan model hybrid yang mencapai akurasi 97,7% pada dataset CICIDS-2017 dan 85,5% pada dataset Natural Gas Pipeline, dengan keunggulan dalam mendeteksi pola spasial lokal sekaligus memahami dinamika temporal [14].

Autoencoder, sebagai arsitektur unsupervised learning, menunjukkan efektivitas luar biasa dalam deteksi anomali tanpa memerlukan data serangan berlabel. Improved Autoencoder (ImpAE) yang dioptimalkan dengan regularisasi adaptif mencapai akurasi 96,2% pada dataset Secure Water Treatment (SWaT) dengan false positive rate hanya 1,8% [62]. Selain itu, model multiple unsupervised Stacked Autoencoders (SAE) mencapai akurasi 95,86% untuk dataset Gas Pipeline dan 99,67% untuk SWaT dataset [71]. Keunggulan utama autoencoder terletak pada kemampuannya mendeteksi serangan zero-day melalui perbandingan reconstruction error antara data normal dan anomali [4], [7]. Model autoencoder yang dilatih eksklusif pada data normal mampu menghasilkan kesalahan rekonstruksi tinggi ketika dihadapkan pada pola serangan, mencapai tingkat deteksi rata-rata 99% pada dataset IEC 104 dan 97% pada dataset IEC 61850 [4].

Arsitektur Transformer dengan mekanisme self-attention memberikan kontribusi revolusioner dalam pemahaman konteks temporal dan dependensi jangka panjang. Framework Transformer-GAN-AE yang dioptimalkan dengan Improved Chimp Optimization Algorithm (IChOA) mencapai akurasi 98,92% pada dataset TON_IoT dan 97,86% pada WUSTL-IIoT-2021, dengan konvergensi tercepat mencapai RMSE di bawah 2,5 dalam 128 detik [12]. Mekanisme Causal Window Attention dalam arsitektur Transformer memungkinkan analisis multi-scale untuk menangkap hubungan kausal lokal dan dependensi global dalam strategi serangan [73].

2. Algoritma Machine Learning Klasik dan Metode Ensemble

Algoritma *machine learning* klasik berkontribusi signifikan terhadap peningkatan keamanan siber pada infrastruktur kritis melalui kemampuan deteksi anomali yang cepat, efisien, dan dapat diinterpretasikan. Dalam studi [49], lima algoritma klasik KNN diantaranya LDA, QDA, Decision Tree, dan Bernoulli Naive Bayes diterapkan pada dataset SCADA WUSTL-IIOT-2018 dan WUSTL-IIOT-2021 untuk mengidentifikasi berbagai serangan seperti reconnaissance, DoS, dan command injection. KNN dan Decision Tree menunjukkan kinerja terbaik dengan akurasi tinggi serta *false positive rate* rendah, menjadikannya efektif dalam memisahkan trafik normal dan berbahaya. LDA dan QDA tetap berguna untuk pemetaan pola awal meskipun menghasilkan FPR lebih tinggi, sementara BNB menawarkan efisiensi komputasi meskipun performanya paling rendah. Secara keseluruhan, algoritma ML klasik memberikan fondasi yang kuat bagi sistem deteksi intrusi SCADA karena ringan, cepat, dan mampu bekerja dengan baik pada data industri yang tidak seimbang serta mudah diintegrasikan ke perangkat lapangan.

Random Forest (RF) menunjukkan performa konsisten tinggi dengan akurasi mencapai 99,9% pada dataset CICIDS2017 dan 93-95% pada CIC Modbus [2]. Keunggulan RF terletak pada ketahanannya terhadap overfitting, kemampuan menangani data berdimensi tinggi, dan efisiensi komputasi yang menjadikannya cocok untuk implementasi real-time [2], [70]. Dalam konteks deteksi ancaman ICS, RF menunjukkan keseimbangan, keandalan, dan tingkat alarm palsu rendah, menjadikannya pilihan kuat untuk aplikasi industri [2].

Support Vector Machine (SVM), khususnya One-Class SVM, efektif untuk deteksi anomali pada data tanpa label. Namun, penelitian menunjukkan performa yang bervariasi tergantung kompleksitas dataset—akurasi hanya 94,2% pada CICIDS2017 dengan presisi rendah dan tingkat alarm palsu tinggi [2]. Pendekatan Ellipsoid-based Process Anomaly and Stealthy Attack Detection (EPASAD) yang menggunakan decision boundary berbasis ellipsoid mencapai akurasi 98,7% dengan false positive rendah sekitar 2%, mengungguli Autoencoder dan Principal Component Analysis (PCA) dalam mendeteksi serangan tersembunyi [8].

Isolation Forest menunjukkan efektivitas luar biasa dalam deteksi anomali unsupervised, mencapai akurasi 100% dan AUC 1.0 pada eksperimen smart grid [10].

Algoritma ini bekerja dengan mengisolasi data anomali melalui pembangunan banyak pohon keputusan acak, efektif mendeteksi perilaku menyimpang pada fase pre-attack tanpa memerlukan label data [10].

3. Generative Adversarial Networks untuk Augmentasi Data dan Deteksi Adversarial

Generative Adversarial Networks (GAN) memberikan kontribusi inovatif dalam mengatasi ketidakseimbangan data dan meningkatkan ketahanan model terhadap perturbation attacks. Wasserstein GAN (WGAN) dalam framework SPARK menciptakan sampel serangan sintetis realistis yang memperkaya dataset pelatihan, memungkinkan deteksi zero-day attacks dengan akurasi 99% [13]. Arsitektur Bidirectional GAN (BiGAN) yang dikombinasikan dengan RNN dan 1-D CNN tidak hanya melakukan deteksi dini melalui pembelajaran perilaku normal ICS, tetapi juga mendukung anomaly localization untuk respons yang lebih cepat dan tepat sasaran [1].

Pendekatan hybrid GAN-Transformer-Autoencoder mengatasi masalah class imbalance dengan menghasilkan sampel data serangan sintetis yang meningkatkan kemampuan model mengenali pola serangan tanpa bias terhadap data normal [12]. Framework ini mencapai kinerja superior dengan akurasi 98,92% dan recall 99,52%, dengan stabilitas tinggi (varians <0,0001) yang mengonfirmasi keandalan untuk sistem keamanan kritis [12].

4. Reinforcement Learning untuk Respons Adaptif dan Mitigasi

Reinforcement Learning (RL) menunjukkan kontribusi signifikan dalam pengambilan keputusan mitigasi yang cerdas dan cost-aware. Pendekatan Thompson Sampling dalam framework berbasis Modbus/TCP memungkinkan sistem memilih strategi mitigasi optimal dari tiga opsi: isolasi aset sepenuhnya, dropping lalu lintas berbahaya parsial, atau eskalasi ke administrator manusia, berdasarkan unit cost yang meminimalkan kerugian finansial dan operasional [16]. Deep Reinforcement Learning (DRL) dengan Deep Q-Network (DQN) mencapai akurasi deteksi 98,79% pada sistem SCADA, dengan kemampuan beradaptasi terhadap serangan baru berkat mekanisme experience replay dan target network [20].

Agent-based Dynamic Thresholding (ADT) yang menggabungkan DQN dengan autoencoder menunjukkan kemampuan luar biasa dalam menyesuaikan threshold deteksi secara dinamis berdasarkan kondisi lingkungan [65]. Q-learning untuk respons ancaman otomatis mencapai tingkat keberhasilan 96,5% setelah 500 episode pelatihan, menunjukkan pembelajaran efektif dalam memilih tindakan mitigasi yang tepat untuk berbagai skenario ancaman [93].

5. Federated Learning untuk Keamanan Terdistribusi dengan Preservasi Privasi

Federated Learning (FL) menghadirkan paradigma revolusioner dalam pelatihan model keamanan terdistribusi tanpa kompromi privasi data. Framework Federated Residual Convolutional Network (FedRCNet) yang dioptimalkan dengan Enhanced Levy Flight Optimization (ELFO) mencapai akurasi 99,2% pada dataset CICIoT2023 dengan AUC 0,99-1,0, mengungguli metode machine learning individual seperti SVM (87,33%) dan MLP (89,31%) [18]. Pendekatan FL memungkinkan setiap node industri melatih model lokal menggunakan datanya sendiri, dengan hanya pembaruan model yang dikirim ke server pusat untuk agregasi, menjaga privasi data sensitif [18], [24].

AI4FIDS (AI for Federated Intrusion Detection System) menerapkan deteksi multimodal dengan menganalisis empat jenis data secara paralel: statistik aliran jaringan, log sistem, data operasional, dan visualisasi lalu lintas [95]. Evaluasi berbagai strategi agregasi (FedAvg, FedProx, FedAdam) menunjukkan bahwa tidak ada satu strategi terbaik untuk semua kasus—FedProx mencapai akurasi terbaik 82,29% pada TON_IoT dan 86,73% pada CSE CIC IDS 2018, sementara FedAvg optimal pada CIC IoT 2022 dengan akurasi 97,6% [95].

6. Explainable AI untuk Transparansi dan Interpretabilitas

Explainable AI (XAI) memberikan kontribusi krusial dalam membangun kepercayaan operator dan memfasilitasi pengambilan keputusan berbasis bukti. Framework berbasis Random Forest dengan Explainable AI methods memberikan interpretasi transparan tentang fitur yang paling berpengaruh dalam deteksi serangan, mendukung respons adaptif melalui human-in-the-loop [15]. Pendekatan shapelet extraction dengan Transformer dan Random Forest mencapai akurasi 99,4% dengan kemampuan memberikan penjelasan berbasis aturan yang intuitif, memenuhi tiga kriteria XAI: penjelasan global, identifikasi pentingnya fitur, dan penjelasan penyebab anomali [100].

SHapley Additive exPlanations (SHAP) menunjukkan efektivitas luar biasa dalam mengidentifikasi parameter kritis yang terpengaruh serangan. Analisis SHAP pada sistem hidroelektrik mengungkapkan bahwa parameter seperti Nozzle2 Aperture (nilai SHAP 0,3642), Water Flow (0,3575), dan Mechanical Power (0,3439) memiliki pengaruh signifikan, sementara Gravity (0,0001) dan Water Density (0,0018) hampir tidak berpengaruh, memungkinkan optimalisasi pemantauan dengan fokus pada variabel kritis [66].

7. Algoritma Optimasi Bio-Inspired untuk Hyperparameter Tuning

Algoritma optimasi metaheuristik memberikan kontribusi penting dalam menemukan konfigurasi parameter optimal untuk model deep learning. Improved Chimp Optimization Algorithm (IChOA) berhasil mengoptimalkan hiperparameter Transformer-GAN-AE, menghasilkan konvergensi tercepat dan stabilitas tinggi dengan varians minimal [12]. Hippopotamus Optimization Algorithm (HOA) untuk penyetelan CNN-BiLSTM mencapai akurasi 99,50% pada ToN-IoT dengan waktu pemrosesan tercepat 10,67 detik [86].

Binary Grey Wolf Optimizer (BGWO) untuk seleksi fitur dan Archimedes Optimization Algorithm (AOA) untuk penyetelan hiperparameter Enhanced Elman Spike Neural Network (EESNN) mencapai akurasi 99,12% pada NSLKDD2015 dan 99,36% pada CICIDS2017, dengan efisiensi waktu training hanya 0,30 menit [80], [101]. Snow Ablation Optimization (SAO) untuk seleksi fitur mengurangi kompleksitas komputasi, menghilangkan noise, dan meningkatkan akurasi model klasifikasi utama [86].

8. Hybrid Models dan Ensemble Approaches

Model hybrid yang menggabungkan kekuatan berbagai algoritma menunjukkan performa superior konsisten. Ensemble Intelligent System Model (EISM) yang mengombinasikan Random Forest, SVM, Gradient Boosting, dan XGBoost dengan Bayesian Optimization mencapai akurasi 99,67%, melampaui metode IDS konvensional dan model tunggal, dengan efisiensi pelatihan 25% lebih cepat dibanding Grid Search [81]. Segmented Neural Network (SNN) yang membagi analisis menjadi DNN untuk fitur statis, LSTM untuk fitur temporal, dan RandNN untuk payload mencapai akurasi hampir sempurna 99,86% pada dataset DNP3 [57].

Arsitektur CNN-GRU dalam framework federated learning menunjukkan akurasi 98,84% pada NSL-KDD, lebih tinggi dibanding CNN (97,07%), GRU (96,65%), dan LSTM (94,11%), dengan latensi rendah sekitar 0,04 detik per paket [53]. Model hybrid CNN-LSTM untuk deteksi serangan web mencapai akurasi 99,12%, precision 98,87%, dan recall 99,03%, menandakan peningkatan signifikan dibandingkan pendekatan konvensional [92].

9. Natural Language Processing untuk Threat Intelligence

Natural Language Processing (NLP) memberikan kontribusi unik dalam menganalisis data tidak terstruktur untuk ekstraksi intelijen ancaman. Organization-Specific Threat Intelligence System (OSTIS) menggunakan deep learning berbasis NLP untuk mengekstraksi informasi relevan dari blog keamanan, laporan insiden, dan forum cyber, dengan F1-score klasifikasi domain mencapai 0,93, entity identification 0,95, dan relation identification 0,89 [76]. Pendekatan TF-IDF untuk mengubah paket data MMS menjadi vektor numerik, dikombinasikan dengan Truncated SVD untuk reduksi dimensi, memungkinkan deteksi anomali efisien dengan Bidirectional LSTM Autoencoder [85].

Framework ML-NLP hybrid untuk smart grid meningkatkan akurasi deteksi serangan hingga 12% dibandingkan metode ML konvensional, dengan kemampuan menganalisis baik data terstruktur (log jaringan, telemetri sensor) maupun tidak terstruktur (laporan ancaman, komunikasi operator) [17]. Sentiment analysis berbasis Bidirectional LSTM pada data media sosial mencapai akurasi 78,52% dalam mengenali sinyal niat serangan siber dari percakapan daring, memberikan early warning untuk pencegahan preventif [94].

D. Kontribusi terhadap Deteksi Dini, Respons Adaptif, dan Pemulihan Sistem

Secara kolektif, algoritma AI berkontribusi dalam tiga fase kritis keamanan siber. Untuk deteksi dini, model deep learning seperti LSTM, CNN, dan Transformer menunjukkan kemampuan mengenali pola anomali pada tahap awal sebelum dampak signifikan terjadi, dengan tingkat deteksi mencapai 96-99% [1], [4], [12], [62], [85]. Autoencoder dan model unsupervised lainnya memungkinkan deteksi zero-day attacks tanpa memerlukan data serangan sebelumnya [4], [7], [10].

Untuk respons adaptif, Reinforcement Learning dan algoritma optimasi bio-inspired memungkinkan sistem menyesuaikan strategi pertahanan secara dinamis berdasarkan kondisi lingkungan yang berubah [16], [20], [65], [93]. Federated Learning mendukung pembelajaran terdistribusi yang beradaptasi dengan pola ancaman baru tanpa kompromi privasi [18], [24], [95]. Framework Explainable AI memfasilitasi human-in-the-loop decision making dengan memberikan justifikasi transparan untuk setiap deteksi [15], [66], [100].

Untuk pemulihan sistem, model AI mendukung anomaly localization untuk identifikasi komponen yang diserang [1], rekonstruksi data yang dimanipulasi [68], dan pembelajaran berkelanjutan dari insiden untuk meningkatkan resiliensi terhadap serangan berulang [43], [44]. Framework adaptive learning dengan drift detection methods seperti ADWIN memungkinkan sistem beradaptasi terhadap perubahan pola serangan tanpa pelatihan ulang penuh [19].

E. Implikasi dan Tantangan

Meskipun menunjukkan efektivitas tinggi, implementasi algoritma AI dalam keamanan siber infrastruktur kritis menghadapi beberapa tantangan. Kompleksitas komputasi model deep learning yang tinggi membatasi penerapan pada perangkat dengan sumber daya terbatas [18], [53], [81]. Ketergantungan

pada kualitas dan ketersediaan dataset pelatihan menjadi kendala, terutama untuk data serangan yang bersifat rahasia [17], [28]. Kerentanan model terhadap adversarial attacks memerlukan pengembangan teknik pertahanan yang robust [17], [21], [22].

Penelitian menunjukkan pentingnya pendekatan hybrid dan ensemble yang menggabungkan kekuatan berbagai algoritma untuk mencapai keseimbangan antara akurasi, efisiensi, dan interpretabilitas [12], [53], [57], [81]. Integrasi Explainable AI menjadi krusial untuk membangun kepercayaan operator dan memfasilitasi adopsi sistem AI dalam lingkungan kritis [15], [66], [100]. Federated Learning menawarkan solusi promising untuk pelatihan kolaboratif dengan preservasi privasi, namun memerlukan strategi agregasi yang disesuaikan dengan karakteristik dataset spesifik [95].

Secara keseluruhan, algoritma AI telah terbukti memberikan kontribusi transformatif dalam meningkatkan keamanan siber infrastruktur kritis, dengan kemampuan deteksi yang superior, respons yang adaptif, dan dukungan pemulihan yang efektif. Keberhasilan implementasi bergantung pada pemilihan algoritma yang tepat sesuai konteks aplikasi, optimasi hiperparameter yang cermat, dan integrasi dengan mekanisme human oversight untuk keputusan kritis.

V. KESIMPULAN

Berdasarkan hasil tinjauan sistematis terhadap 89 artikel periode 2020-2025, penerapan kecerdasan buatan (AI) terbukti memberikan kontribusi besar terhadap peningkatan keamanan siber pada infrastruktur kritis. Algoritma seperti CNN, LTSM, Autoencoder, dan Transformer menunjukkan performa tinggi dalam mendeteksi anomali, mengidentifikasi serangan zero-day, serta memahami pola serangan kompleks dengan akurasi 96-99%. Pendekatan Reinforcement Learning dan Federated Learning juga memperkuat kemampuan sistem untuk beradaptasi terhadap ancaman baru tanpa mengorbankan privasi data. Model Machine Learning Klasik seperti Random Forest dan SVM tetap relevan karena efisiensi komputasi dan kemampuan interpretasi yang baik, terutama untuk implementasi pada sistem dengan sumber daya terbatas.

Kajian ini mengidentifikasi enam kategori utama ancaman siber terhadap infrastruktur kritis: serangan terhadap jaringan dan komunikasi industri (DoS/DDoS, MITM), manipulasi data dan injeksi perintah (FDIA, command injection, replay attacks), Advanced Persistent Threats (APT), malware dan ransomware (Stuxnet, BlackEnergy, Triton), insider threats dan kesalahan manusia, serta ancaman baru terhadap sistem berbasis AI (adversarial attacks, data poisoning). Dampak serangan tersebut tidak hanya mengancam keamanan data, tetapi juga menimbulkan konsekuensi fisik berbahaya, kerugian ekonomi masif, gangguan layanan publik, dan ancaman terhadap keselamatan dan keamanan nasional.

Namun demikian, implementasi AI dalam keamanan siber infrastruktur kritis menghadapi sejumlah tantangan signifikan yang perlu diatasi. Model deep learning yang kompleks memerlukan sumber daya komputasi tinggi yang membatasi penerapan pada edge devices dan sistem embedded. Keterbatasan ketersediaan dataset serangan nyata yang sering diklasifikasikan sebagai rahasia menjadi hambatan dalam pengembangan model yang robust dan representatif. Sistem berbasis AI masih rentan terhadap adversarial attacks dan data poisoning yang dapat memanipulasi model untuk menghasilkan keputusan yang salah. Ketidakseimbangan data antara lalu lintas normal dan aktivitas berbahaya, serta kebutuhan akan transparansi dan interpretabilitas dalam lingkungan industri yang menuntut keandalan tinggi, menjadi pertimbangan krusial dalam desain dan implementasi sistem deteksi intrusi berbasis AI.

Untuk penelitian selanjutnya, disarankan beberapa arah strategis yang mencakup aspek teknis, kolaboratif, dan regulasi. Pertama, pengembangan model hybrid dan ensemble yang efisien untuk mencapai keseimbangan optimal antara akurasi, efisiensi komputasi, dan interpretabilitas. Kedua, integrasi explainable AI secara *real-time* dalam sistem operasional untuk memfasilitasi dan membangun kepercayaan operator. Ketiga, penerapan federated learning lintas sektor industri untuk pembelajaran kolaboratif dengan preservasi privasi. Keempat, pengembangan dataset serangan representatif yang mencakup protokol industri spesifik dan skenario operasional nyata. Kelima, penelitian tentang pertahanan terhadap *adversarial attacks* untuk meningkatkan ketahanan model AI. Keenam, eksplorasi teknologi emerging seperti Large Language Models untuk *threat intelligence* dan digital twins untuk simulasi serangan.

DAFTAR PUSTAKA

[1] N. Neshenko, E. Bou-Harb, B. Furht, and M. Baghersad, "A deep learning-based adaptive cyber disaster management framework," *J Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01241-3.

[2] B.-V. Vasilică, F.-D. Anton, R. Pietraru, S.-O. Anton, and B.-N. Chiriac, "Enhancing Security in Smart Robot Digital Twins Through Intrusion Detection Systems," *Applied Sciences (Switzerland)*, vol. 15, no. 9, 2025, doi: 10.3390/app15094596.

[3] A. Dehlaghi-Ghadim, A. Balador, M. H. Moghadam, H. Hansson, and M. Conti, "ICSSIM — A framework for building industrial control systems security testbeds," *Comput Ind*, vol. 148, 2023, doi: 10.1016/j.compind.2023.103906.

[4] G. Amritha, M. G. Nair, and F. Granelli, "Machine Learning-Enriched Cybersecurity in Smart Grids," *IEEE Access*, vol. 13, pp. 99303–99313, 2025, doi: 10.1109/ACCESS.2025.3576497.

[5] N. Muller, K. Bao, J. Matthes, and K. Heussen, "CyPhERS: A cyber-physical event reasoning system providing real-time situational awareness for attack and fault response," *Comput Ind*, vol. 151, 2023, doi: 10.1016/j.compind.2023.103982.

[6] L. SAMBUCCI and E.-A. PARASCHIV, "The accelerated integration of artificial intelligence systems and its potential to expand the vulnerability of the critical infrastructure," *Revista Română de Informatică și Automatică*, vol. 34, no. 3, pp. 113–148, 2024, doi: 10.33436/v34i3y202410.

[7] M. Barbhaya, P. R. Dasari, S. K. Damarla, R. Srinivasan, and B. Huang, "A deep learning framework for cyberattack detection and classification in Industrial Control Systems," *Comput Chem Eng*, vol. 202, 2025, doi: 10.1016/j.compchemeng.2025.109278.

[8] V. Maurya, R. Agarwal, S. Kumar, and S. Shukla, "EPASAD: ellipsoid decision boundary based Process-Aware Stealthy Attack Detector," *Cybersecurity*, vol. 6, no. 1, 2023, doi: 10.1186/s42400-023-00162-z.

[9] S. Mokhtari and K. K. Yen, "False Data Injection Attack Detection, Isolation, and Identification in Industrial Control Systems Based on Machine Learning: Application in Load Frequency Control," *Electronics (Switzerland)*, vol. 13, no. 16, 2024, doi: 10.3390/electronics13163239.

[10] S. Kabir, N. Hannan, A. Shufian, and M. S. Rahman Zishan, "Proactive detection of cyber-physical grid attacks: A pre-attack phase identification and analysis using anomaly-based machine learning models," *Array*, vol. 27, 2025, doi: 10.1016/j.array.2025.100441.

[11] A. M. Awaad, K. M. Ali Alheeti, and A. K. A. H. Najem, "Anomaly-Based IDS (Intrusion Detection System) for Cyber-Physical Systems," *Mesopotamian Journal of Big Data*, vol. 2024, pp. 230–240, 2024, doi: 10.58496/MJBD/2024/017.

[12] A. Salehiyan, P. S. Moghaddam, and M. Kaveh, "An Optimized Transformer–GAN–AE for Intrusion Detection in Edge and IIoT Systems: Experimental Insights from WUSTL-IIoT-2021, EdgeIIoTset, and TON_IoT Datasets," *Future Internet*, vol. 17, no. 7, 2025, doi: 10.3390/fi17070279.

[13] R. Bhukya, S. A. Moeed, A. Medavaka, A. O. Khadidos, A. O. Khadidos, and S. Selvarajan, "SPARK and SAD: Leading-edge deep learning frameworks for robust and effective intrusion detection in SCADA systems," *International Journal of Critical Infrastructure Protection*, vol. 49, 2025, doi: 10.1016/j.ijcip.2025.100759.

[14] J. Wang, C. Si, Z. Wang, and Q. Fu, "A New Industrial Intrusion Detection Method Based on CNN-BiLSTM," *Computers, Materials and Continua*, vol. 79, no. 3, pp. 4297–4318, 2024, doi: 10.32604/cmc.2024.050223.

[15] U. U. Izuazu, C. I. Nwakanma, D.-S. Kim, and J. M. Lee, "Explainable and perturbation-resilient model for cyber-threat detection in industrial control systems Networks," *Discover Internet of Things*, vol. 5, no. 1, 2025, doi: 10.1007/s43926-025-00100-0.

[16] T. Kotsiopoulos, P. Radoglou-Grammatikis, Z. Lekka, V. Mladenov, and P. Sarigiannidis, "Defending industrial internet of things against Modbus/TCP threats: A combined AI-based detection and SDN-based mitigation solution," *Int J Inf Secur*, vol. 24, no. 4, 2025, doi: 10.1007/s10207-025-01076-2.

[17] R. K. Jha, "Strengthening Smart Grid Cybersecurity: An In-Depth Investigation into the Fusion of Machine Learning and Natural Language Processing," *Journal of Trends in Computer Science and Smart Technology*, vol. 5, no. 3, pp. 284–301, 2023, doi: 10.36548/jtcsst.2023.3.005.

[18] N. Manogaran *et al.*, "Federated Learning and EEL-Levy Optimization in CPS ShieldNet Fusion: A New Paradigm for Cyber–Physical Security," *Sensors*, vol. 25, no. 12, 2025, doi: 10.3390/s25123617.

[19] G. Ahmadi-Assalemi, H. al-Khateeb, V. Benson, B. Adamyk, and M. Ammi, "Adaptive learning anomaly detection and classification model for cyber and physical threats in industrial control systems," *IET Cyber-Physical Systems: Theory and Applications*, vol. 10, no. 1, 2025, doi: 10.1049/cps2.70004.

[20] F. Mesadieu, D. Torre, and A. Chennamaneni, "Leveraging Deep Reinforcement Learning Technique for Intrusion Detection in SCADA Infrastructure," *IEEE Access*, vol. 12, pp. 63381–63399, 2024, doi: 10.1109/ACCESS.2024.3390722.

[21] G. Sampedro, S. Ojo, M. Krichen, M. A. Alamro, A. Mihoub, and V. Karovič, "Defending AI Models Against Adversarial Attacks in Smart Grids Using Deep Learning," *IEEE Access*, vol. 12, pp. 157408–157417, 2024, doi: 10.1109/ACCESS.2024.3473531.

[22] E. Anthi, L. Williams, M. Rhode, P. Burnap, and A. Wedgbury, "Adversarial attacks on machine learning cybersecurity defences in Industrial Control Systems," *Journal of Information Security and Applications*, vol. 58, 2021, doi: 10.1016/j.jisa.2020.102717.

[23] A. L. Perales Gomez, L. F. Fernandez-Maimo, F. J. G. Clemente, J. A. M. Maroto, A. H. Huertas Celdrán, and G. Bovet, "A Methodology for Evaluating the Robustness of Anomaly Detectors to Adversarial Attacks in Industrial Scenarios," *IEEE Access*, vol. 10, pp. 124582–124594, 2022, doi: 10.1109/ACCESS.2022.3224930.

[24] T. Rehman, N. Tariq, F. A. Khan, and S. U. Rehman, "FFL-IDS: A Fog-Enabled Federated Learning-Based Intrusion Detection System to Counter Jamming and Spoofing Attacks for the Industrial Internet of Things," *Sensors*, vol. 25, no. 1, 2025, doi: 10.3390/s25010010.

[25] A. Aldaej, T. A. Ahamed Ahanger, and I. Ullah, "Deep Learning-Inspired IoT-IDS Mechanism for Edge Computing Environments," *Sensors*, vol. 23, no. 24, 2023, doi: 10.3390/s23249869.

[26] I. Ahmad, Z. Wan, A. Ahmad, and S. S. Sajid Ullah, "A Hybrid Optimization Model for Efficient Detection and Classification of Malware in the Internet of Things," *Mathematics*, vol. 12, no. 10, 2024, doi: 10.3390/math12101437.

[27] J. H. Lee, I. H. Ji, S. H. Jeon, and J. T. Seo, "Generating ICS Anomaly Data Reflecting Cyber-Attack Based on Systematic Sampling and Linear Regression," *Sensors*, vol. 23, no. 24, 2023, doi: 10.3390/s23249855.

[28] A. Alqudhaibi, M. Albarrak, A. Aloseel, S. Jagtap, and K. Salonitis, "Predicting Cybersecurity Threats in Critical Infrastructure for Industry 4.0: A Proactive Approach Based on Attacker Motivations," *Sensors*, vol. 23, no. 9, 2023, doi: 10.3390/s23094539.

[29] Y. Ayachi, Y. Mellah, M. Saber, N. Rahmoun, I. Kerrakchou, and T. Bouchentouf, "A survey and analysis of intrusion detection models based on information security and object technology-cloud intrusion dataset," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 4, p. 1607, 2022, doi: 10.11591/ijai.v11.i4.pp1607-1614.

[30] A. Alharthi, M. Alaryani, and S. Kaddoura, "A comparative study of machine learning and deep learning models in binary and multiclass classification for intrusion detection systems," *Array*, vol. 26, p. 100406, 2025, doi: 10.1016/j.array.2025.100406.

[31] H. Mejdi, S. Elmadssia, M. Koubãa, and T. Ezzedine, "A Comprehensive Survey on Game Theory Applications in Cyber-Physical System Security: Attack Models, Security Analyses, and Machine Learning Classifications," *IEEE Access*, vol. 12, pp. 163638–163653, 2024, doi: 10.1109/ACCESS.2024.3491502.

[32] R. Canonico and G. Sperlì, "Industrial cyber-physical systems protection: A methodological review," *Comput Secur*, vol. 135, p. 103531, 2023, doi: 10.1016/j.cose.2023.103531.

[33] T. D. Ramotsoela, G. P. Hancke, and A. M. Abu-Mahfouz, "Behavioural Intrusion Detection in Water Distribution Systems Using Neural Networks," *IEEE Access*, vol. 8, pp. 190403–190416, 2020, doi: 10.1109/ACCESS.2020.3032251.

[34] R. Holdbrook, O. Odeyomi, S. Yi, and K. Roy, "Network-Based Intrusion Detection for Industrial and Robotics Systems: A Comprehensive Survey," *Electronics (Basel)*, vol. 13, no. 22, p. 4440, 2024, doi: 10.3390/electronics13224440.

[35] O. Semenenko, O. Nozdrachov, I. Chernyshova, A. Melnychenko, and D. Momot, "Innovative technologies to improve energy efficiency and security of military facilities," *Naukovij žurnal «Tehnika ta energetika»*, vol. 15, no. 4, pp. 147–156, 2024, doi: 10.31548/machinery/4.2024.147.

[36] K. Ayoub, H. Abdelmajid, T. Ayoub, and M. Soukaina, "Enhancing IoT RPL Protocol Security Against Black Hole Attacks with Deep Learning Techniques," *Journal of Computer Science*, vol. 20, no. 11, pp. 1530–1544, 2024, doi: 10.3844/jcssp.2024.1530.1544.

[37] Y. Yigit *et al.*, "Generative AI and LLMs for Critical Infrastructure Protection: Evaluation Benchmarks, Agentic AI, Challenges, and Opportunities," *Sensors*, vol. 25, no. 6, p. 1666, 2025, doi: 10.3390/s25061666.

[38] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *J Bus Res*, vol. 104, pp. 333–339, 2019, doi: 10.1016/j.jbusres.2019.07.039.

[39] A. Carrera-Rivera, W. Ochoa, F. Larrinaga, and G. Lasa, "How-to conduct a systematic literature review: A quick guide for computer science research," *MethodsX*, vol. 9, p. 101895, 2022, doi: 10.1016/j.mex.2022.101895.

[40] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf Softw Technol*, vol. 64, pp. 1–18, 2015, doi: 10.1016/j.infsof.2015.03.007.

[41] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, pp. n71–n71, 2021, doi: 10.1136/bmj.n71.

[42] A. H. Dehkordi, E. Mazaheri, H. A. Ibrahim, S. Dalvand, and R. Ghanei Gheshlagh, "How to Write a Systematic Review: A Narrative Review," *Int J Prev Med*, vol. 12, no. 1, 2021, doi: 10.4103/ijpvm.IJPVM_60_20.

[43] Y.-C. Chen, C.-H. Cheng, T.-W. Lin, and J.-S. Lee, "Diverse Machine Learning-Based Malicious Detection for Industrial Control System," *Electronics (Switzerland)*, vol. 14, no. 10, 2025, doi: 10.3390/electronics14101947.

[44] C.-S. Shieh, T.-L. Nguyen, T.-T. Nguyen, and M.-F. Horng, "Unknown DDoS Attack Detection with Sliced Iterative Normalizing Flows Technique," *Computers, Materials and Continua*, vol. 82, no. 3, pp. 4881–4912, 2025, doi: 10.32604/cmc.2025.061001.

[45] O. Polat *et al.*, "Multi-Stage Learning Framework Using Convolutional Neural Network and Decision Tree-Based Classification for Detection of DDoS Pandemic Attacks in SDN-Based SCADA Systems," *Sensors*, vol. 24, no. 3, 2024, doi: 10.3390/s24031040.

[46] W. Villegas-Ch, J. Govea, A. Maldonado Navarro, and P. Palacios Játiva, "Intrusion Detection in IoT Networks Using Dynamic Graph Modeling and Graph-Based Neural Networks," *IEEE Access*, vol. 13, pp. 65356–65375, 2025, doi: 10.1109/ACCESS.2025.3559325.

[47] S. Das, M. Ashrafuzzaman, F. T. Sheldon, and S. Shiva, "Ensembling Supervised and Unsupervised Machine Learning Algorithms for Detecting Distributed Denial of Service Attacks," *Algorithms*, vol. 17, no. 3, 2024, doi: 10.3390/a17030099.

[48] J. Zhao, P. Zeng, C. Chen, Z. Dong, and J. Han, "Deep learning anomaly detection based on hierarchical status-connection features in networked control systems," *Intelligent Automation and Soft Computing*, vol. 30, no. 1, pp. 337–350, 2021, doi: 10.32604/iasc.2021.016966.

[49] N. Yalçın, S. Çakir, and S. Ünaldı, "Attack Detection Using Artificial Intelligence Methods for SCADA Security," *IEEE Internet Things J*, vol. 11, no. 24, pp. 39550–39559, 2024, doi: 10.1109/JIOT.2024.3447876.

[50] E. Anthi, L. Williams, P. Burnap, and K. Jones, "A three-tiered intrusion detection system for industrial control systems," *J Cybersecur*, vol. 7, no. 1, 2021, doi: 10.1093/cybsec/tyab006.

[51] W. Choi, S. Pandey, and J. Kim, "Detecting Cybersecurity Threats for Industrial Control Systems Using Machine Learning," *IEEE Access*, vol. 12, pp. 153550–153563, 2024, doi: 10.1109/ACCESS.2024.3478830.

[52] M. Anwar, L. Lundberg, and A. Borg, "Improving anomaly detection in SCADA network communication with attribute extension," *Energy Informatics*, vol. 5, no. 1, 2022, doi: 10.1186/s42162-022-00252-1.

[53] S. Y. Diaba, M. Shafie-Khah, and M. Elmusrati, "On the performance metrics for cyber-physical attack detection in smart grid," *Soft comput*, vol. 26, no. 23, pp. 13109–13118, 2022, doi: 10.1007/s00500-022-06761-1.

[54] G. Raman Mr, N. Somu, and A. P. Mathur, "A multilayer perceptron model for anomaly detection in water treatment plants," *International Journal of Critical Infrastructure Protection*, vol. 31, 2020, doi: 10.1016/j.ijcip.2020.100393.

[55] S. Mubarak, M. H. Hadi Habaebi, M. R. Islam, F. D. A. Rahman, and M. Tahir, "Anomaly detection in ICS datasets with machine learning algorithms," *Computer Systems Science and Engineering*, vol. 37, no. 1, pp. 33–46, 2021, doi: 10.32604/CSSE.2021.014384.

[56] D. Agnew *et al.*, "Implementation Aspects of Smart Grids Cyber-Security Cross-Layered Framework for Critical Infrastructure Operation," *Applied Sciences (Switzerland)*, vol. 12, no. 14, 2022, doi: 10.3390/app12146868.

[57] S. Allah Bakhsh *et al.*, "Enhancing Security in DNP3 Communication for Smart Grids: A Segmented Neural Network Approach," *IEEE Access*, vol. 13, pp. 110436–110456, 2025, doi: 10.1109/ACCESS.2025.3580507.

[58] O. Polat, A. Ayid Ahmad, S. Oyucu, E. Algül, F. Doaan, and A. Aksoz, "Temporal-Spatial Feature Extraction in IoT-Based SCADA System Security: Hybrid CNN-LSTM and Attention-Based Architectures for Malware Classification and Attack Detection," *IEEE Access*, vol. 13, pp. 102109–102132, 2025, doi: 10.1109/ACCESS.2025.3577761.

[59] H. Benaddi, M. Jouhari, K. Ibrahimi, J. Ben Othman, and E. M. Amhoud, "Anomaly Detection in Industrial IoT Using Distributional Reinforcement Learning and Generative Adversarial Networks," *Sensors*, vol. 22, no. 21, 2022, doi: 10.3390/s22218085.

[60] C.-J. Huang, C.-J. Chi, and W.-T. Hung, "Hybrid-AI-Based iBeacon Indoor Positioning Cybersecurity: Attacks and Defenses," *Sensors*, vol. 23, no. 4, 2023, doi: 10.3390/s23042159.

[61] A. L. P. Perales Gomez, L. F. Fernandez-Maimo, A. H. Huertas Celdrán, and F. J. G. García Clemente, "MADICS: A methodology for anomaly detection in industrial control systems," *Symmetry (Basel)*, vol. 12, no. 10, 2020, doi: 10.3390/SYM12101583.

[62] M. M. Aslam, A. Tufail, L. C. De Silva, R. A. A. Haji Mohd Apong, and A. Namoun, "An improved autoencoder-based approach for anomaly detection in industrial control systems," *Systems Science and Control Engineering*, vol. 12, no. 1, 2024, doi: 10.1080/21642583.2024.2334303.

[63] B. Kim, M. A. Alawami, E. Kim, S. Oh, J. Park, and H. Kim, "A Comparative Study of Time Series Anomaly Detection Models for Industrial Control Systems," *Sensors*, vol. 23, no. 3, 2023, doi: 10.3390/s23031310.

[64] V. Tkach, A. Kudin, V. R. Kebande, O. Baranovskyi, and I. Kudin, "Non-Pattern-Based Anomaly Detection in Time-Series," *Electronics (Switzerland)*, vol. 12, no. 3, 2023, doi: 10.3390/electronics12030721.

[65] X. Yang, E. Howley, and M. Schukat, "ADT: Time series anomaly detection for cyber-physical systems via deep reinforcement learning," *Comput Secur*, vol. 141, 2024, doi: 10.1016/j.cose.2024.103825.

[66] I. Erkek and E. Irmak, "Enhancing Cybersecurity of a Hydroelectric Power Plant Through Digital Twin Modeling and Explainable AI," *IEEE Access*, vol. 13, pp. 41887–41908, 2025, doi: 10.1109/ACCESS.2025.3547672.

[67] C. Atheeq, R. Sultana, S. A. Sabahath, and M. A. K. Mohammed, "Advancing IoT Cybersecurity: Adaptive Threat Identification with Deep Learning in Cyber-Physical Systems," *Engineering, Technology and Applied Science Research*, vol. 14, no. 2, pp. 13559–13566, 2024, doi: 10.48084/etasr.6969.

[68] A. Abughali, M. Alansari, and A. S. Al-Sumaiti, "Deep Learning Strategies for Detecting and Mitigating Cyber-Attacks Targeting Water-Energy Nexus," *IEEE Access*, vol. 12, pp. 129690–129704, 2024, doi: 10.1109/ACCESS.2024.3458788.

[69] P. Illy and G. Kaddoum, "A Collaborative DNN-Based Low-Latency IDPS for Mission-Critical Smart Factory Networks," *IEEE Access*, vol. 11, pp. 96317–96329, 2023, doi: 10.1109/ACCESS.2023.3311822.

[70] M. Zaman, D. Upadhyay, and C.-H. Lung, "Validation of a Machine Learning-Based IDS Design Framework Using ORNL Datasets for Power System with SCADA," *IEEE Access*, vol. 11, pp. 118414–118426, 2023, doi: 10.1109/ACCESS.2023.3326751.

[71] A. Al-Abassi, H. Karimipour, A. Dehghantanha, and R. M. Parizi, "An ensemble deep learning-based cyber-attack detection in industrial control system," *IEEE Access*, vol. 8, pp. 83965–83973, 2020, doi: 10.1109/ACCESS.2020.2992249.

[72] E. N. Yolaçan and H. Cavsi Zaim, "DCWM-LSTM: A Novel Attack Detection Framework for Robotic Arms," *IEEE Access*, vol. 13, pp. 20547–20560, 2025, doi: 10.1109/ACCESS.2025.3535225.

[73] S. Zhang, X. Xue, and X. Su, "DeepOP: A Hybrid Framework for MITRE ATT&CK Sequence Prediction via Deep Learning and Ontology," *Electronics (Switzerland)*, vol. 14, no. 2, 2025, doi: 10.3390/electronics14020257.

[74] M. Lozano, I. Pérez-Llopis, and M. Esteve Domingo, "Threat Hunting Architecture Using a Machine Learning Approach for Critical Infrastructures Protection," *Big Data and Cognitive Computing*, vol. 7, no. 2, 2023, doi: 10.3390/bdcc7020065.

[75] A. Shan and S. Myeong, "Proactive Threat Hunting in Critical Infrastructure Protection through Hybrid Machine Learning Algorithm Application," *Sensors*, vol. 24, no. 15, 2024, doi: 10.3390/s24154888.

[76] D. R. Arikkat *et al.*, "OSTIS: A novel Organization-Specific Threat Intelligence System," *Comput Secur*, vol. 145, 2024, doi: 10.1016/j.cose.2024.103990.

[77] X. Qin, F. Jiang, X. Qin, L. Ge, M. Lu, and R. Doss, "CGAN-based cyber deception framework against reconnaissance attacks in ICS," *Computer Networks*, vol. 251, 2024, doi: 10.1016/j.comnet.2024.110655.

[78] Q. Zhang, "Optimizing Threat Intelligence Strategies for Cybersecurity Awareness Using MADM and Hybrid GraphNet-Bipolar Fuzzy Rough Sets," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 11, pp. 1381–1392, 2024, doi: 10.14569/IJACSA.2024.01511136.

[79] D. S. Morozov, A. A. Yefimenko, T. M. Nikitchuk, R. O. Kolomiiets, and S. O. Semerikov, "The sweet taste of IoT deception: an adaptive honeypot framework for design and evaluation," *Journal of Edge Computing*, vol. 3, no. 2, pp. 207–223, 2024, doi: 10.55056/jec.607.

[80] S. S. Khatami, M. Shoeibi, A. E. Oskouei, D. Martín de Andrés, and M. K. Dashliboroun, "5DGWO-GAN: A Novel Five-Dimensional Gray Wolf Optimizer for Generative Adversarial Network-Enabled Intrusion Detection in IoT Systems," *Computers, Materials and Continua*, vol. 82, no. 1, pp. 881–911, 2025, doi: 10.32604/cmc.2024.059999.

[81] Z. A. Sheikh, Y. Singh, S. Tanwar, R. Sharma, F.-E. Țurcanu, and M. S. Raboaca, "EISM-CPS: An Enhanced Intelligent Security Methodology for Cyber-Physical Systems through Hyper-Parameter Optimization," *Mathematics*, vol. 11, no. 1, 2023, doi: 10.3390/math11010189.

[82] H. Tanyıldız, C. Şahin, and Ö. Batur Dinler, "Improving Deceptive Patch Solutions Using Novel Deep Learning-Based Time Analysis Model for Industrial Control Systems," *Applied Sciences (Switzerland)*, vol. 14, no. 20, 2024, doi: 10.3390/app14209287.

[83] W. Wang, F. Harrou, B. Bouyeddou, S.-M. Senouci, and Y. Sun, "A stacked deep learning approach to cyber-attacks detection in industrial systems: application to power system and gas pipeline systems," *Cluster Comput*, vol. 25, no. 1, pp. 561–578, 2022, doi: 10.1007/s10586-021-03426-w.

[84] D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper, "Lightweight Long Short-Term Memory Variational Auto-Encoder for Multivariate Time Series Anomaly Detection in Industrial Control Systems," *Sensors*, vol. 22, no. 8, 2022, doi: 10.3390/s22082886.

[85] J. Azar, M. Al Saleh, R. Couturier, and H. Noura, "Text Mining and Unsupervised Deep Learning for Intrusion Detection in Smart-Grid Communication Networks," *Internet of Things*, vol. 6, no. 2, 2025, doi: 10.3390/iot6020022.

[86] M. K. Ishak, "Mathematical Modeling of Cyberattack Defense Mechanism Using Hybrid Transfer Learning With Snow Ablation Optimization Algorithm in Critical Infrastructures," *IEEE Access*, vol. 13, pp. 13329–13340, 2025, doi: 10.1109/ACCESS.2025.3530931.

[87] H. Ünözkan, M. Ertem, and S. Bendak, "Using attack graphs to defend healthcare systems from cyberattacks: a longitudinal empirical study," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 11, no. 1, 2022, doi: 10.1007/s13721-022-00391-1.

[88] S. Li, "Comparative Analysis of Predicting Malware Attack Trends in Cyber Supply Chain Using Multiple Classification Models," *IEEE*

Access, 2024, doi: 10.1109/ACCESS.2024.3471802.
[89] A. Budzys, O. Kurasova, and V. Medvedev, "Deep learning-based authentication for insider threat detection in critical infrastructure," *Artif Intell Rev*, vol. 57, no. 10, 2024, doi: 10.1007/s10462-024-10893-1.
[90] K. A. Ali Abuhasel, "A Linear Probabilistic Resilience Model for Securing Critical Infrastructure in Industry 5.0," *IEEE Access*, vol. 11, pp. 80863–80873, 2023, doi: 10.1109/ACCESS.2023.3300650.
[91] P. Lanka, K. Gupta, and C. Varol, "Intelligent Threat Detection—AI-Driven Analysis of Honeypot Data to Counter Cyber Threats," *Electronics (Switzerland)*, vol. 13, no. 13, 2024, doi: 10.3390/electronics13132465.
[92] A. Salam, F. Ullah, F. Amin, and M. Abrar, "Deep Learning Techniques for Web-Based Attack Detection in Industry 5.0: A Novel Approach," *Technologies (Basel)*, vol. 11, no. 4, 2023, doi: 10.3390/technologies11040107.
[93] M. M. Saeed, "An AI-Driven Cybersecurity Framework for IoT: Integrating LSTM-Based Anomaly Detection, Reinforcement Learning, and Post-Quantum Encryption," *IEEE Access*, vol. 13, pp. 104027–104036, 2025, doi: 10.1109/ACCESS.2025.3576506.
[94] S. Lehominova, Y. Shchavinsky, T. Muzhanova, D. Rabchun, and M. Zaporozhchenko, "Application of Sentiment Analysis to Prevent Cyberattacks on Objects of Critical Information Infrastructure," *International Journal of Computing*, vol. 22, no. 4, pp. 534–540, 2023, doi: 10.47839/ijc.22.4.3362.
[95] P. Radoglou-Grammatikis *et al.*, "AI4FIDS: Multimodal Federated Intrusion Detection," *IEEE Trans Emerg Top Comput*, 2025, doi: 10.1109/TETC.2025.3562346.
[96] M. Kravchik, L. Demetrio, B. Biggio, and A. Shabtai, "Practical Evaluation of Poisoning Attacks on Online Anomaly Detectors in Industrial Control Systems," *Comput Secur*, vol. 122, 2022, doi: 10.1016/j.cose.2022.102901.
[97] A. H. Oveis, G. Meucci, F. Mancuso, F. Berizzi, and A. Cantelli-Forti, "Generative AI Threats to Maritime Navigation Using Deceptive ISAR Images," *IEEE Access*, vol. 12, pp. 173800–173809, 2024, doi: 10.1109/ACCESS.2024.3500774.
[98] G. Abbas, M. Ali, M. Ahmad, and A. Khan, "CIRA-Cyber Intelligent Risk Assessment Methodology for Industrial Internet of Things Based on Machine Learning," *IEEE Access*, vol. 13, pp. 77001–77016, 2025, doi: 10.1109/ACCESS.2025.3559617.
[99] J. Vávra, M. Hromada, L. Lukáš, and J. Dworzecki, "Adaptive anomaly detection system based on machine learning algorithms in an industrial control environment," *International Journal of Critical Infrastructure Protection*, vol. 34, 2021, doi: 10.1016/j.ijcip.2021.100446.
[100] K.-K. Kim, J.-S. Kim, and I.-C. Euom, "Explainable Anomaly Detection Based on Operational Sequences in Industrial Control Systems," *IEEE Access*, vol. 13, pp. 66170–66187, 2025, doi: 10.1109/ACCESS.2025.3560260.
[101] A. Al Mazroa, F. R. Albogamy, M. Ishak, and S. M. Mostafa, "Boosting Cyberattack Detection Using Binary Metaheuristics with Deep Learning on Cyber-Physical System Environment," *IEEE Access*, vol. 13, pp. 11280–11294, 2025, doi: 10.1109/ACCESS.2025.3526258.