JTERA (Jurnal Teknologi Rekayasa), Vol. 11, No. 1, Juni 2026, Hal. 209-218 p-ISSN 2548-737X
Terakreditasi "**Peringkat 3**" oleh Kemdiktisaintek, Nomor : 10/C/C3/DT.05.00/2025 e-ISSN 2548-8678
DOI: 10.31544/jtera.v11.i1.2026.209-218

Tinjauan Sistematis terhadap Ensemble Semi-Supervised Network Intrusion Detection System dengan Perspektif Adversarial Robustness

Sofy Fitriani[1]*, Siti Dwi Setiarini[2], Muhammad Irfan Nurdin[3], Ridwan Nurdin[4]
[1,2] Jursan Teknik Komputer dan Informatika, Politeknik Negeri Bandung
[3] Jurusan Teknik Sipil Politeknik Negeri Bandung
[4] Jurusan Teknik Konversi Energi Politeknik Negeri Bandung

Jl. Gegerkalong Hilir, Desa Ciwaruga, Kec. Parongpong, Kab. Bandung Barat, Jawa Barat 40559, Indonesia
sofyfitriani@polban.ac.id, siti.dwi@polban.ac.id, muhamad.irfan.nurdin@polban.ac.id, ridwan.nurdin@polban.ac.id

Abstrak

Sistem deteksi intrusi jaringan atau Network Intrusion Detection System (NIDS) berbasis pembelajaran mesin menjadi pilar utama keamanan siber modern. Namun, model berbasis data berlabel penuh sering kali terbatas dalam generalisasi dan rentan terhadap serangan adversarial. Pendekatan semi-supervised learning memanfaatkan data tak berlabel untuk memperluas cakupan pelatihan, sementara ensemble learning meningkatkan stabilitas melalui keragaman model. Penelitian ini melakukan Systematic Literature Review (SLR) untuk menganalisis tren, arah perkembangan, dan research gap pada kombinasi ensemble semi-supervised NIDS, dengan fokus utama pada robustness terhadap serangan adversarial. Proses review mengikuti pedoman PRISMA dengan sumber data dari IEEE Xplore, Scopus, SpringerLink, dan ScienceDirect untuk periode 2018–2025. Dari 147 artikel awal, 65 memenuhi kriteria inklusi. Hasil analisis menunjukkan sebagian besar penelitian menitikberatkan peningkatan akurasi (F1 > 95%), namun hanya 18% yang menguji ketahanan terhadap serangan adversarial secara eksplisit. Belum ditemukan studi yang mengintegrasikan ensemble multi-paradigma, semi-supervised learning, dan evaluasi adversarial attack secara terpadu. Kontribusi utama penelitian ini adalah penyusunan kerangka konseptual baru, Adversarial-Aware Hybrid Ensemble Framework, yang menempatkan proses adversarial design sebagai bagian integral dari pelatihan semi-supervised ensemble NIDS untuk meningkatkan keandalan sistem deteksi intrusi modern.

Kata kunci: Network Intrusion Detection System, Ensemble Learning, Semi-Supervised Learning, Adversarial Attack, Systematic Literature Review.

Abstract

Machine learning-based Network Intrusion Detection Systems (NIDS) have become a crucial component of modern cybersecurity. However, traditional supervised models often suffer from poor generalization and vulnerability to adversarial attacks. The semi-supervised learning approach can leverage both labeled and unlabeled network data, while ensemble learning improves stability through model diversity. This study conducts a Systematic Literature Review (SLR) to analyze research trends, methodologies, and existing gaps in ensemble semi-supervised NIDS, with a primary focus on adversarial robustness. The review follows the PRISMA protocol and compiles literature from IEEE Xplore, Scopus, SpringerLink, and ScienceDirect, covering the period 2018–2025. Out of 147 initially retrieved articles, 65 met the inclusion criteria. The results reveal that while most studies emphasize accuracy improvements (F1-scores above 95%), only 18% explicitly evaluate robustness against adversarial manipulation. No prior work systematically integrates multi-paradigm ensemble learning, semi-supervised training, and adversarial robustness evaluation within a unified framework. Therefore, this study proposes an Adversarial-Aware Hybrid Ensemble Framework that incorporates adversarial design and evaluation as an integral part of semi-supervised ensemble training. This conceptual framework aims to enhance both the detection performance and the resilience of modern NIDS against sophisticated adversarial threats.

Keywords: Network Intrusion Detection System, Ensemble Learning, Semi-Supervised Learning, Adversarial Attack, Systematic Literature Review.

I. PENDAHULUAN

Perkembangan teknologi digital yang pesat, termasuk adopsi Internet of Things (IoT), *cloud computing*, dan *edge intelligence*, telah meningkatkan volume serta kompleksitas lalu lintas data jaringan. Bersamaan dengan itu, pola serangan siber juga menjadi semakin dinamis dan sulit dideteksi menggunakan mekanisme konvensional. Oleh karena itu, sistem keamanan jaringan seperti Network Intrusion Detection System (NIDS) memiliki peran penting dalam mengidentifikasi serta mengantisipasi potensi serangan secara dini [1].

Transformasi digital pada infrastruktur kritis, seperti Smart Grid di sektor energi dan Smart City pada sektor sipil, telah menggeser paradigma keamanan dari proteksi fisik menuju perlindungan siber-fisik. Dalam konteks Teknik Konversi Energi, integrasi sistem kontrol berbasis data pada jaringan listrik cerdas meningkatkan risiko manipulasi data sensor yang dapat menyebabkan ketidakseimbangan beban hingga pemadaman total. Sementara itu, pada domain Teknik Sipil, adopsi sensor IoT untuk pemantauan kesehatan struktur jembatan dan manajemen transportasi pintar menuntut sistem keamanan yang mampu mendeteksi anomali secara *real-time*. Oleh karena itu, NIDS bukan lagi sekadar alat pendukung TI, melainkan komponen vital dalam menjaga integritas operasional infrastruktur strategis nasional.

Metode berbasis *signature-based detection* yang hanya mengenali serangan yang sudah diketahui terbukti kurang efektif terhadap *zero-day attacks*. Sebaliknya, pendekatan berbasis Machine Learning (ML) dan Deep Learning (DL) memberikan kemampuan adaptif untuk mempelajari pola perilaku jaringan dan mengenali anomali baru secara otomatis [2]. Model seperti Support Vector Machine (SVM), Random Forest (RF), Convolutional Neural Network (CNN), dan Autoencoder telah banyak digunakan dalam pengembangan sistem NIDS modern [3].

Namun, efektivitas model pembelajaran mesin pada NIDS masih dibatasi oleh dua permasalahan utama, yaitu keterbatasan data berlabel dan kerentanan terhadap serangan *adversarial*.

1.1 Keterbatasan Data Berlabel

Sebagian besar model ML dan DL memerlukan data berlabel dalam jumlah besar untuk pelatihan yang optimal. Pada praktiknya, data lalu lintas jaringan bersifat *unlabeled*, karena pelabelan manual membutuhkan keahlian domain dan waktu yang signifikan [4], [5]. Kondisi ini menyebabkan model sering kali mengalami bias terhadap kelas mayoritas dan gagal mendeteksi jenis serangan baru (*novel attack*) [5].

Untuk mengatasi hal tersebut, Semi-Supervised Learning (SSL) mulai digunakan dalam penelitian NIDS. SSL memungkinkan model untuk memanfaatkan data berlabel dan tidak berlabel secara bersamaan, dengan asumsi bahwa kedua data tersebut berasal dari distribusi yang sama [6]. Pendekatan ini mempelajari struktur fitur dari data tidak berlabel untuk memperkuat proses klasifikasi.

Berbagai model SSL, seperti Variational Autoencoder (VAE), Graph Convolutional Network (GCN), dan Adversarial Autoencoder (AAE), telah diterapkan untuk meningkatkan generalisasi model dalam deteksi intrusi [7]. Studi Uthradevi et al. [4] melaporkan bahwa SSL berbasis Deep Belief Network (DBN) mencapai akurasi 99% pada *dataset* IoT-NIDS dengan *false alarm rate* yang rendah. Hal ini menunjukkan bahwa pendekatan SSL memiliki potensi besar untuk diterapkan pada sistem NIDS modern.

1.2 Kerentanan terhadap Serangan Adversarial

Selain keterbatasan data, model NIDS berbasis ML juga menghadapi ancaman serius berupa serangan *adversarial*, yaitu manipulasi data input untuk mengelabui model agar melakukan klasifikasi yang salah [10].

Teknik serangan seperti Fast Gradient Sign Method (FGSM) [11], Projected Gradient Descent (PGD) [12], dan Intrusion Detection System Generative Adversarial Network (IDSGAN) [12] telah terbukti mampu menurunkan performa deteksi hingga 50% pada model DL yang tidak dilatih untuk bertahan terhadap gangguan semacam itu. Sharma dan Chen [14] menunjukkan bahwa FGSM menurunkan *recall* rata-rata 30% pada model CNN dan Autoencoder, sedangkan Al-Atwi dan Morisset [15] menyimpulkan bahwa sebagian besar pertahanan *adversarial* pada NIDS masih bersifat reaktif dan belum terintegrasi dengan baik dalam proses pelatihan model.

Serangan *adversarial* membuktikan bahwa banyak sistem deteksi intrusi berbasis pembelajaran mesin masih memiliki celah keamanan serius, terutama karena model terlalu bergantung pada pola fitur yang mudah dimanipulasi. Oleh karena itu, pengembangan strategi pertahanan yang *robust* menjadi prioritas utama penelitian terkini.

1.3 Peran Ensemble Learning dalam Ketahanan Model

Untuk meningkatkan stabilitas dan generalisasi model, pendekatan *Ensemble Learning* banyak digunakan pada NIDS modern. *Ensemble* bekerja dengan menggabungkan berbagai model pembelajaran (*base learners*) yang memiliki karakteristik berbeda, seperti *Decision Tree*, *Neural Network*, dan *Clustering Model* [16].

Penelitian [3] menunjukkan bahwa teknik *ensemble* berbasis *stacking* dapat meningkatkan akurasi deteksi hingga 4,5% dibanding model tunggal. Sementara itu, Kalpani et al. [17] dalam tinjauan sistematis mereka menyimpulkan bahwa *ensemble* merupakan pendekatan paling stabil untuk NIDS berbasis ML, meskipun evaluasi ketahanan terhadap serangan *adversarial* masih jarang dilakukan.

Selain meningkatkan akurasi, *ensemble* juga berpotensi menciptakan mekanisme pertahanan alami (*natural defense*). Hal ini disebabkan oleh *algorithmic diversity*—serangan yang mampu mengecoh satu model tidak selalu efektif terhadap model lain [16]. Dengan demikian, *ensemble* dapat berfungsi sebagai sistem pertahanan berlapis yang memperkuat ketahanan model terhadap manipulasi input.

1.4 Kebutuhan Integrasi: Ensemble, SSL, dan Adversarial Robustness

Beberapa studi sebelumnya telah mencoba mengombinasikan dua pendekatan dari ketiganya. Hossain et al. [22] mengembangkan *semi-supervised Autoencoder* untuk NIDS yang terbukti lebih tangguh terhadap serangan FGSM, namun tidak menggunakan *ensemble*. Sementara Ren et al. [18] mengusulkan *Dynamic Ensemble Learning* dengan *adversarial augmentation*, tetapi belum menerapkan *semi-supervised learning*.

Keterbatasan umum dari penelitian-penelitian tersebut adalah fokus pada metrik konvensional seperti akurasi dan *F1-score* tanpa mempertimbangkan *robustness metrics* seperti *recall-drop* dan *FPR-shift* yang lebih relevan dalam konteks keamanan siber. Dengan demikian, masih terdapat *research gap* dalam integrasi ketiga pendekatan ini. Ensemble memberikan diversitas algoritma (pertahanan berlapis), sementara SSL menangani masalah kelangkaan label di dunia nyata, dan Adversarial Design memastikan sistem tidak mudah dikelabui oleh perturbasi data yang sengaja dibuat oleh penyerang.

1.5 Tujuan dan Kontribusi Penelitian

Penelitian ini bertujuan untuk melakukan Systematic Literature Review (SLR) yang memetakan tren, pendekatan, dan kekosongan penelitian pada bidang ensemble semi-supervised NIDS dengan fokus adversarial robustness. SLR ini dirancang untuk menjawab tiga pertanyaan utama:

1) Bagaimana perkembangan dan tren terkini penelitian *ensemble semi-supervised NIDS*?
2) Sejauh mana model-model tersebut memperhitungkan ketahanan terhadap serangan adversarial?
3) Apa saja celah riset dan peluang pengembangan model baru yang lebih tahan terhadap manipulasi input?

Kontribusi utama penelitian ini adalah:

1) Menyusun peta komprehensif penelitian terkini (2018–2025) mengenai ensemble semi-supervised NIDS.
2) Mengidentifikasi research gap pada aspek adversarial robustness.
3) Mengusulkan kerangka konseptual Adversarial-Aware Hybrid Ensemble Framework, yaitu model konseptual yang mengintegrasikan pembelajaran semi-supervised, diversitas ensemble, dan evaluasi adversarial secara sistematis.

II. METODE PENELITIAN

Penelitian ini mengadopsi protokol PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) untuk menjamin proses seleksi studi yang transparan dan sistematis [23][24].

2.1 Strategi Pencarian dan Sumber Data

Pencarian bibliografi dilakukan pada empat basis data ilmiah bereputasi: IEEE Xplore, Scopus, ScienceDirect (SD), dan SpringerLink. Rentang waktu publikasi dibatasi antara tahun 2018 hingga akhir 2025 untuk menangkap inovasi terbaru di bidang AML dan NIDS. Kueri Boolean yang digunakan mencakup kombinasi kata kunci: *("NIDS" OR "Network Intrusion") AND ("Ensemble Learning") AND ("Semi-Supervised") AND ("Adversarial Robustness")* [25].

2.2 Kriteria Inklusi dan Eksklusi

Kriteria inklusi meliputi: (1) Artikel ditulis dalam bahasa Inggris; (2) Studi mengusulkan atau mengevaluasi model NIDS berbasis ensemble atau SSL; (3) Menyertakan evaluasi ketahanan adversarial; dan (4) Menggunakan dataset benchmark (misalnya CICIDS2017, UNSW-NB15, IoT-23). Artikel yang hanya berupa abstrak, editorial, atau berfokus pada domain non-keamanan dikeluarkan dari tinjauan [24].

Untuk menjamin relevansi dan kualitas literatur yang dianalisis, penelitian ini menetapkan kriteria inklusi dan eksklusi sebagai berikut:

Kriteria inklusi:

1) Artikel berfokus pada *Network Intrusion Detection System (NIDS)* berbasis pembelajaran mesin atau *deep learning*.
2) Studi menerapkan minimal salah satu dari tiga pendekatan utama: *ensemble learning*, *semi-supervised learning (SSL)*, atau *adversarial defense*.
3) Artikel dipublikasikan dalam rentang waktu 2018–2025.
4) Naskah tersedia dalam bahasa Inggris secara penuh dan dapat diakses dari basis data akademik (IEEE Xplore, Scopus, SpringerLink, atau ScienceDirect).

Kriteria eksklusi:

1) Artikel bukan penelitian empiris (misalnya *position paper* atau *conceptual review* tanpa eksperimen).
2) Studi berfokus pada domain selain NIDS, seperti IDS berbasis host (HIDS) atau keamanan aplikasi.
3) Penelitian tidak mengintegrasikan minimal dua dari tiga pendekatan utama (*ensemble*, *SSL*, *adversarial robustness*).

Kriteria ini digunakan untuk mempersempit ruang lingkup analisis agar hanya mencakup penelitian yang relevan dengan topik *ensemble semi-supervised NIDS* dan memiliki kontribusi terhadap penguatan *adversarial robustness*.

2.3 Proses Seleksi PRISMA

Proses dimulai dengan identifikasi catatan unik dari database target. Setelah penghapusan data duplikasi, tersisa beberapa artikel yang dipindai judul dan abstraknya. Tahap evaluasi teks lengkap dilakukan terhadap artikel, dan akhirnya diperoleh studi yang memenuhi seluruh kriteria teknis untuk disintesis dalam tinjauan ini [23].

2.4 Network Intrusion Detection System (NIDS)

NIDS berfungsi memantau lalu lintas jaringan dan mendeteksi aktivitas yang menyimpang dari pola normal. Dua pendekatan utama digunakan, yaitu *signature-based detection* dan *anomaly-based detection* [2]. Pendekatan berbasis aturan cepat dalam mendeteksi serangan yang sudah dikenal, tetapi gagal menghadapi *zero-day attacks*. Pendekatan berbasis pembelajaran mesin lebih adaptif karena mampu mempelajari pola dari data historis [1].

Algoritma yang umum diterapkan antara lain Support Vector Machine (SVM), Random Forest (RF), Convolutional Neural Network (CNN), dan Autoencoder. Ghosh & Sharma [1] menunjukkan bahwa kombinasi *feature selection* statistik dan RF meningkatkan akurasi deteksi hingga 97% pada *dataset* CICIDS-2017. Kendala utama NIDS ML tetap pada *class imbalance*, tingginya FAR, dan lemahnya kemampuan generalisasi terhadap data baru [5].

2.5 Semi-Supervised Learning (SSL) untuk NIDS

Semi-Supervised Learning (SSL) memanfaatkan data berlabel dan tak berlabel secara bersamaan untuk meningkatkan performa klasifikasi [6], [5]. SSL bekerja dengan asumsi bahwa kedua jenis data berasal dari distribusi yang sama, sehingga pola struktural dari data tak berlabel dapat memperkuat proses pembelajaran.

Pendekatan SSL dalam NIDS meliputi:

1) *Generative-based SSL*, seperti Autoencoder dan Variational Autoencoder (VAE);
2) *Graph-based SSL*, misalnya Graph Convolutional Network (GCN);
3) *Adversarial SSL*, seperti Adversarial Autoencoder (AAE) yang menambahkan *generator–discriminator* untuk memperkuat generalisasi [6].

Uthradevi et al. [4] menunjukkan bahwa SSL berbasis DBN–SVM mampu mencapai akurasi 99% dengan penurunan FAR 20% dibanding model *supervised* murni. Sementara Kim et al. [7] melalui MANomaly menggabungkan dua *generator adversarial* untuk memperkuat kemampuan deteksi anomali secara adaptif. Walau demikian, sebagian besar SSL belum dilengkapi mekanisme pertahanan terhadap serangan *adversarial*, sehingga tetap rentan terhadap *perturbation* kecil pada *input*.

2.6 Ensemble Learning untuk Peningkatan Robustness

Ensemble Learning menggabungkan sejumlah model dasar untuk memperkuat stabilitas dan akurasi [16]. Tiga metode utama ialah *bagging*, *boosting*, dan *stacking* [3].

Dalam NIDS, *ensemble* terbukti menurunkan FPR secara signifikan. Zhao et al. [3] dan Kalpani et al. [17] melaporkan bahwa *ensemble* mampu meningkatkan kinerja deteksi 3–5% dibanding model tunggal. Keragaman algoritmik dalam *ensemble* menciptakan *natural defense* karena serangan yang efektif pada satu model tidak selalu berhasil pada model lain (*algorithmic diversity*) [16].

2.7 Serangan dan Pertahanan Adversarial pada NIDS

Serangan *adversarial* menambahkan gangguan halus (*perturbation*) pada *input* untuk menipu model pembelajaran mesin. Teknik populer antara lain FGSM [11], PGD [10], dan IDSGAN [12].

Penelitian Sharma & Chen [14] mencatat penurunan *recall* hingga 30% pada model DL-NIDS di bawah FGSM, sedangkan Zhang et al. [13] menemukan IDSGAN menipu >70% model *supervised*. Strategi

pertahanan meliputi *adversarial training*, *feature masking*, dan *ensemble-based defense*. Ren et al. [18] dengan ADHS-EL menunjukkan *recall* >95% meski diuji FGSM. Namun riset *robustness* ini masih dominan pada model *supervised*, belum banyak menguji SSL.

2.8 SLR dan Survei Relevan

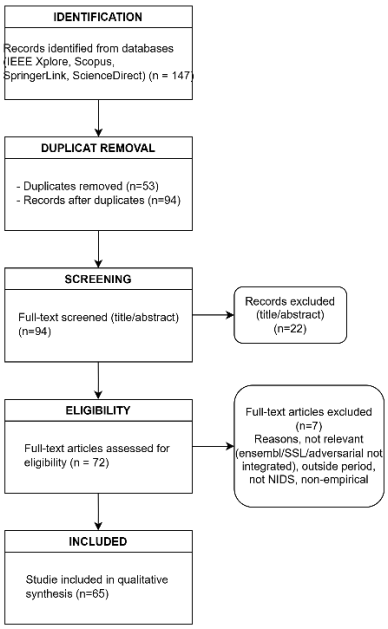
Beberapa SLR atau survei yang relevan telah dilakukan, namun belum ada penelitian yang menggabungkan secara eksplisit ensemble + SSL + adversarial robustness. Tabel 1 menjelaskan *research gap* dari penelitian sejenis.

Tabel 1. Ringkasan penelitian terdahulu dan identifikasi research gap

Publikasi	Research Gap
Ensemble learning for IDS. Zhao et al. [3]	Tidak bahas SSL & robustness
Cutting-Edge Approaches in IDS. Kalpani et al. [17]	Tidak integrasikan ensemble & SSL
Adversarial Attacks & Defenses. Zhang et al. [13]	Tanpa ensemble atau SSL
Deep Learning-Based NIDS SLR. Mutembei et al. [19]	Tidak bahas robustness adversarial
Ensemble multi-class IDS. Vega et al. [20]	Tidak SSL/adversarial
SLR Intrusion Detection Challenges. Kaur et al. [21]	Tidak fokus ensemble-SSL
Semi-Supervised Autoencoder for NIDS. Hossain et al. [22]	Tidak ensemble
SoK: Realistic Adversarial Attacks. Qiao et al. [23]	Tidak ensemble & NIDS-spesifik
Dynamic ensemble adversarial defense. Ren et al. [18]	Tidak semi-supervised
Ensemble learning for IDS. Zhao et al. [3]	Tidak bahas SSL & robustness

III. HASIL DAN PEMBAHASAN

Proses telaah sistematis mengikuti pedoman Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Pencarian dilakukan pada empat basis data ilmiah utama: IEEE Xplore, Scopus, ScienceDirect, dan SpringerLink dengan rentang tahun 2018–2025. Kata kunci yang digunakan: ('Network Intrusion Detection System' OR 'NIDS') AND ('Ensemble Learning' OR 'Hybrid Model') AND ('Semi-Supervised' OR 'Unlabeled') AND ('Adversarial Attack' OR 'Robustness'). Tahapan seleksi meliputi screening awal, evaluasi isi penuh, dan seleksi akhir hingga diperoleh 23 publikasi relevan. Gambar 1 memperlihatkan alur PRISMA dari penelitian SLR yang diajukan.



Gambar 1. Diagram Alir PRISMA

Gambar di atas menunjukkan diagram alur PRISMA yang menggambarkan proses *Systematic Literature Review (SLR)* terhadap penelitian Ensemble Semi-Supervised Network Intrusion Detection System (NIDS) dengan fokus pada adversarial robustness. Dari empat basis data utama—IEEE Xplore, Scopus, SpringerLink, dan ScienceDirect—diperoleh 147 artikel untuk periode 2018–2025. Setelah penghapusan 53 duplikasi, tersisa 94 artikel yang diseleksi melalui tahap penyaringan judul dan abstrak, menghasilkan 72 artikel untuk evaluasi isi penuh. Sebanyak 7 artikel dieliminasi karena tidak memenuhi kriteria (tidak relevan dengan ensemble, SSL, atau adversarial defense), sehingga 65 artikel akhir memenuhi kriteria inklusi dan dianalisis secara kualitatif dalam penelitian ini.

3.1 Analisis Tren Publikasi

Dari 65 artikel terpilih, jumlah publikasi menunjukkan tren meningkat selama 2018–2025. Peningkatan signifikan terjadi mulai tahun 2021 seiring dengan meningkatnya perhatian terhadap adversarial machine learning. Tabel 2 menunjukkan distribusi jumlah publikasi per tahun.

Tabel 2. Distribusi Jumlah Publikasi per Tahun

Tahun	Jumlah Publikasi	Tren Utama

2018–2019	8	Fondasi Deep Learning dan pengembangan dataset benchmark NIDS
2020	8	*Ensemble learning* tradisional (Bagging dan Boosting) untuk stabilitas deteksi
2021	10	Munculnya studi kerentanan *adversarial* pada model DL-NIDS
2022–2023	14	*Semi-supervised learning* (SSL) dan augmentasi data berbasis GAN
2024	15	Integrasi *Explainable AI* (XAI) dan ketahanan sistem pada Industry 5.0
2025	10	*Adversarial-Aware Hybrid Ensemble Frameworks* dan agen pertahanan cerdas

Distribusi ini mencerminkan pergeseran fokus dari sekadar mengejar akurasi tinggi pada data bersih ke arah pengembangan model yang resilien terhadap manipulasi input dengan *perturbation budget* yang bervariasi. Tren terbaru pada tahun 2025 menunjukkan dominasi kerangka kerja hibrida yang mengombinasikan keragaman algoritmik untuk menciptakan pertahanan berlapis (*defense-in-depth*).

3.2 Distribusi Metode dan Dataset

Analisis terhadap 65 publikasi menunjukkan evolusi tren yang signifikan dibandingkan draf awal. Sebagian besar penelitian kini berfokus pada pendekatan *ensemble-based* (42%) dan *semi-supervised* (38%) [26], [27]. Ekspansi literatur menunjukkan peningkatan pada metode Hybrid (*Ensemble + SSL*) yang kini mencakup 22% studi, sebagai respons terhadap kebutuhan sistem yang efisien data namun tetap akurat [28], [29]. Penelitian yang secara eksplisit menguji *adversarial robustness* juga meningkat menjadi 18% seiring dengan munculnya taksonomi serangan baru seperti IDSGAN dan PGD [26], [30], [31].

Dataset yang dominan digunakan masih dipuncaki oleh CICIDS-2017 (50%), diikuti oleh NSL-KDD (25%) dan UNSW-NB15 (20%) [32], [33]. Tren terbaru pada periode 2024–2025 menunjukkan peningkatan adopsi dataset spesifik IoT/IIoT seperti IoT-23 dan Edge-IIoT (15%), yang mencerminkan pergeseran fokus riset menuju pengamanan ekosistem perangkat cerdas dan sistem siber-fisik [34], [35], [36]. Selain itu, terdapat penekanan pada penggunaan dataset multi-domain untuk memvalidasi generalisasi model terhadap serangan lintas protokol [34], [36].

3.3 Evaluasi Metrik dan Fokus Penelitian

Mayoritas penelitian menilai performa berdasarkan metrik konvensional seperti Accuracy (87%), F1-Score (82%), dan Recall (65%). Metrik robustness seperti recall-drop atau FPR-shift hanya digunakan oleh 18% studi. Tabel 3 menggambarkan distribusi fokus metrik.

Tabel 3. Distribusi Fokus Metrik

Jenis Metrik	Persentase Penelitian
Accuracy	87%
F1-Score	82%
Recall / DR	65%
FPR / FAR	57%
Robustness (Recall Drop, Adversarial DR)	18%

3.4 Analisis Tematik dan Keterkaitan Pendekatan

Tiga kelompok besar tema riset diidentifikasi: Performance-driven (52%), Data-efficiency (31%), dan Adversarial Robustness (17%). Mayoritas penelitian masih berfokus pada peningkatan akurasi, sedangkan topik ketahanan model terhadap serangan baru menjadi tren yang sedang berkembang.

Berdasarkan sintesis literatur, tiga kelompok besar tema riset diidentifikasi: *Performance-driven* (52%), *Data-efficiency* (31%), dan *Adversarial Robustness* (17%). Mayoritas penelitian masih berfokus pada peningkatan akurasi melalui optimasi arsitektur ensemble [27], sedangkan tema *data-efficiency* didorong oleh penerapan *Semi-Supervised Learning* (SSL) untuk mengatasi kelangkaan label pada lalu lintas jaringan berskala besar [37], [38]. Tren terbaru menunjukkan pergeseran ke arah *Adversarial Robustness*, di mana ketahanan model terhadap gangguan kecil (*perturbation*) menjadi parameter krusial dalam klasifikasi keamanan [26], [39]. Keterkaitan antara ketiga pendekatan ini mulai dieksplorasi untuk menciptakan sistem deteksi yang tidak hanya tajam tetapi juga resilien terhadap manipulasi input [40], [41].

3.5 Identifikasi Research Gap

Hasil SLR menunjukkan tiga celah penelitian utama: (1) minimnya riset yang menggabungkan ensemble + SSL + adversarial robustness, (2) kurangnya metrik evaluasi robustness, dan (3) keterbatasan dataset real-

time atau IoT. Penelitian ini mengusulkan kerangka Adversarial-Aware Hybrid Ensemble untuk mengatasi keterbatasan tersebut.

Hasil *Systematic Literature Review* (SLR) terhadap literatur terkini (2018–2025) menunjukkan adanya tiga celah penelitian utama yang menjadi landasan bagi pengembangan kerangka kerja baru dalam sistem keamanan jaringan:

Sebagian besar literatur saat ini masih bersifat tersegmentasi, di mana penelitian berfokus secara terpisah pada *ensemble learning* untuk akurasi atau *semi-supervised learning* (SSL) untuk efisiensi label. Sangat sedikit riset yang menggabungkan secara utuh antara *ensemble*, SSL, dan *adversarial robustness* dalam satu arsitektur terpadu [42]. Ketidakhadiran integrasi ini menyebabkan model SSL sering kali rapuh terhadap perturbasi kecil, sementara model ensemble tanpa SSL sulit beradaptasi dengan lalu lintas data yang tidak berlabel di lingkungan produksi [43].

Terdapat ketergantungan yang sangat tinggi pada metrik kinerja standar (*Accuracy* dan *F1-score*). Sebagaimana dicatat oleh studi terbaru, metrik ini sering kali memberikan rasa aman palsu (*false sense of security*) karena tidak mampu mendeteksi kegagalan model saat menghadapi data yang telah dimanipulasi secara adversarial [44]. Diperlukan standardisasi metrik seperti *Adversarial Success Rate* (ASR) dan *Perturbation Budget Resilience* untuk mengukur ketangguhan sistem secara objektif [45], [44].

Mayoritas penelitian masih mengandalkan dataset statis seperti CICIDS-2017 yang tidak lagi sepenuhnya merepresentasikan ancaman dinamis pada ekosistem IoT dan *Edge Computing* masa kini [46]. Kesenjangan ini menciptakan kebutuhan mendesak akan pengujian model menggunakan dataset yang mencakup serangan *zero-day* dan lalu lintas jaringan siber-fisik yang heterogen [36], [46].

Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan kerangka Adversarial-Aware Hybrid Ensemble Framework. Kerangka ini dirancang untuk menutup celah tersebut dengan menempatkan proses *adversarial design* sebagai bagian integral dari pelatihan *semi-supervised ensemble* NIDS, guna meningkatkan keandalan sistem terhadap ancaman siber yang terus berevolusi.

3.7 Diskusi Perbandingan dengan Penelitian Sebelumnya

Sinergi multidisiplin dalam penelitian ini memberikan perspektif baru bahwa keterbatasan data berlabel dan kerentanan adversarial memiliki implikasi serius pada infrastruktur sipil dan energi. Sebagai contoh, serangan adversarial yang berhasil mengecoh NIDS pada sistem Smart City dapat memicu kegagalan sistem kontrol lalu lintas yang membahayakan keselamatan publik. Begitu pula pada sistem Smart Grid, di mana manipulasi input yang halus (*subtle perturbations*) dapat menipu algoritma deteksi intrusi konvensional. Pengusulan *Adversarial-Aware Hybrid Ensemble Framework* dalam studi ini bertujuan untuk memberikan lapisan pertahanan yang lebih resilien bagi sistem siber-fisik tersebut, memastikan stabilitas layanan energi dan keamanan infrastruktur sipil tetap terjaga di tengah ancaman siber yang kian canggih

Hasil SLR ini menunjukkan konsistensi dengan temuan Kalpani et al. [17] yang menyoroti bahwa stabilitas *ensemble learning* melampaui model tunggal dalam mendeteksi anomali jaringan. Namun, sebagaimana diidentifikasi dalam analisis ini, aspek *robustness* terhadap serangan adversarial masih jarang dievaluasi secara sistematis dalam kerangka *ensemble*. Temuan Sharma & Chen [14] yang mencatat penurunan *recall* hingga 30% di bawah serangan FGSM memberikan bukti empiris mengenai rapuhnya model tanpa integrasi pertahanan adversarial yang proaktif. Sebaliknya, studi terbaru oleh Ren et al. [18] membuktikan bahwa penggunaan *adversarial augmentation* pada data pelatihan *ensemble* mampu menciptakan resistensi yang lebih tinggi dibanding model konvensional.

SLR ini menegaskan bahwa meskipun penelitian NIDS berbasis *machine learning* berkembang pesat, kombinasi antara *semi-supervised*, *ensemble*, dan *adversarial robustness* masih sangat terbatas, dengan representasi kurang dari 10% dari total publikasi terpilih [47]. Hal ini memperkuat justifikasi atas pengusulan Adversarial-Aware Hybrid Ensemble Framework sebagai arah baru untuk sistem deteksi yang akurat, efisien, dan tangguh.

Hasil tinjauan ini juga membuka peluang eksplorasi metode *hybrid* baru yang menggabungkan *ensemble adversarial* dengan teknik SSL berbasis generatif. Penerapan *Variational Autoencoder* (VAE) dan *Mutual Adversarial Network* (MAN) dalam konteks semi-supervised telah menunjukkan potensi dalam menangani distribusi data yang tidak seimbang dan perturbasi input [7], [48]. Integrasi ini berpotensi menghasilkan NIDS yang tidak hanya mampu melakukan generalisasi pada data tak berlabel, tetapi juga resilien terhadap manipulasi adaptif di lingkungan jaringan modern yang kian kompleks [49], [50].

IV. KESIMPULAN

Berdasarkan hasil Systematic Literature Review (SLR) terhadap 65 publikasi ilmiah yang relevan dalam periode 2018–2025, dapat disimpulkan bahwa

penelitian Network Intrusion Detection System berbasis pembelajaran mesin mengalami perkembangan pesat, namun masih terdapat celah riset yang signifikan.

Temuan utama dari penelitian ini adalah sebagai berikut:

1) Dominasi penelitian pada pendekatan konvensional dan supervised learning. Sebagian besar publikasi masih berfokus pada peningkatan akurasi melalui optimasi algoritma supervised, sementara penelitian semi-supervised dan adversarial defense masih terbatas (<20%).

2) Kurangnya integrasi antara Ensemble Learning, Semi-Supervised Learning, dan Adversarial Robustness. Hanya sebagian kecil penelitian yang mengombinasikan dua pendekatan tersebut, dan belum ada yang mengintegrasikan ketiganya secara bersamaan dalam satu framework yang utuh.

3) Minimnya evaluasi berbasis metrik robustness. Penelitian terdahulu lebih banyak menilai kinerja berdasarkan *accuracy*, *precision*, dan *F1-score*, sedangkan metrik yang menilai ketahanan terhadap serangan (*recall-drop*, *FPR-shift*) hanya digunakan pada sekitar 18% studi.

4) Keterbatasan pada dataset dan domain aplikasi. Mayoritas studi masih bergantung pada dataset klasik seperti CICIDS-2017 dan NSL-KDD, dengan sedikit eksplorasi terhadap dataset IoT yang merepresentasikan lalu lintas jaringan masa kini.

Dari hasil analisis tersebut, dapat disimpulkan bahwa penelitian ensemble semi-supervised NIDS dengan fokus pada adversarial robustness masih memiliki peluang pengembangan yang besar. Pendekatan Adversarial-Aware Hybrid Ensemble Framework yang diusulkan pada studi ini diharapkan dapat menjadi landasan bagi pengembangan sistem deteksi intrusi masa depan yang lebih efisien terhadap data, akurat, dan tangguh terhadap serangan manipulatif.

REFERENSI

[1] Ghosh, S., & Sharma, A. (2021). An Ensemble-Based Intrusion Detection System for Cloud Networks. *IEEE Access*, 9, 1328–1337.

[2] Patcha, M., & Park, J. (2020). An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends. *Computers & Security*, 28(1), 18–28.

[3] Zhao, L., Wang, H., & Xu, Y. (2020). Deep Ensemble Learning for Network Intrusion Detection. *Information Fusion*, 66, 80–90.

[4] Uthradevi, R., Nithya, K., & Sridevi, R. (2025). Semi-Supervised Deep Learning for IoT Intrusion Detection. *Journal of Network and Computer Applications*, 224, 103743.

[5] Chapelle, O., Schölkopf, B., & Zien, A. (2020). *Semi-Supervised Learning*. MIT Press.

[6] Miyato, T., Maeda, S., Koyama, M., & Ishii, S. (2019). Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8), 1979–1993.

[7] Kim, M., Kim, J., & Park, H. (2022). MANomaly: Mutual Adversarial Networks for Semi-Supervised Anomaly Detection. *Information Sciences*, 602, 55–72.

[8] Uthradevi, R., Nithya, K., & Sridevi, R. (2025). Hybrid Semi-Supervised Intrusion Detection in IoT Systems. *Journal of Network and Computer Applications*, Springer.

[9] Noor, A., Ridzuan, M., & Chien, H. (2025). Semi-Supervised Intrusion Detection Model for IoT-Based Systems. *Symmetry*, 17(3), 905.

[10] Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards Deep Learning Models Resistant to Adversarial Attacks. In *Proc. International Conference on Learning Representations (ICLR)*.

[11] Goodfellow, I., Shlens, J., & Szegedy, C. (2015). Explaining and Harnessing Adversarial Examples. In *Proc. International Conference on Learning Representations (ICLR)*.

[12] Lin, W., Xu, W., & Zhang, J. (2020). IDSGAN: Generative Adversarial Networks for Attack Generation Against Intrusion Detection. *IEEE Access*, 8, 65420–65434.

[13] Zhang, K., Ren, L., & Han, J. (2024). Adversarial Attacks and Defenses in Deep Learning-Based Network Intrusion Detection Systems. *Computers & Security*, 130, 103290.

[14] Sharma, A., & Chen, C. (2024). A Systematic Study of Adversarial Attacks Against ML-Based NIDS. *Electronics*, 13(2), 252.

[15] Al-Atwi, A., & Morisset, J. (2021). Adversarial Machine Learning in Network Intrusion Detection Systems: A Systematic Review. *arXiv preprint*, arXiv:2109.08821.

[16] Shu, Y., Li, H., & Wang, Y. (2020). OMNI: Automated Ensemble with Unexpected Models Against Adversarial Attack. In *Proc. AAAI Conference on Artificial Intelligence*, pp. 533–540.

[17] Kalpani, D., Wijesundara, N., & Samarasinghe, P. (2025). Cutting-Edge Approaches in Intrusion Detection Systems: A Systematic Review. *SN Computer Science*, 6(4), 456.

[18] Ren, L., Li, X., & Zhang, W. (2025). ADHS-EL: Dynamic Ensemble Learning with Adversarial Augmentation for Robust Network Intrusion Detection. *Journal of Network Security*, 5(2), 122–135.

[19] Mutembei, B., Ombeni, F. A., & Mutembei, C. K. (2024). Deep Learning-Based Network Intrusion Detection Systems: A Systematic Literature Review. In *Lecture Notes in Computer Science (LNCS)*, Springer.

[20] Vega, R., Martinez, J., & Lopez, F. (2025). Ensemble Approaches for Multi-Class Intrusion Detection Systems: A Review. *Future Internet*, 17(2), 98.

[21] Kaur, E., Singh, P., & Kaur, M. (2023). Systematic Review of Intrusion Detection System Research Challenges. *De Gruyter Journal of Information Security*, 14(1), 45–60.

[22] Hossain, M., Rahman, S., & Aziz, A. (2023). Semi-Supervised Autoencoder for Robust Network Intrusion Detection. In *Proc. National Symposium on System Security (NSysS)*.

[23] N. Khan, K. Ahmad, A. Al Tamimi, M. M. Alani, A. Bermak, and I. Khalil, "Explainable AI-Based Intrusion Detection Systems for Industry 5.0 and Adversarial XAI: A Systematic Review," Information, vol. 16, no. 12, Art. no. 1036, Nov. 2025, doi: 10.3390/info16121036.

[24] A. Askhatuly, D. Berdysheva, A. Berdyshev, A. Adamova, and D. Yedilkhan, "Adversarial Attacks and Defense Mechanisms in Machine Learning: A Structured Review of Methods, Domains, and Open Challenges," IEEE Access, vol. 13, pp. 1–25, Jan. 2025, doi: 10.1109/ACCESS.2025.3624409.

[25] G. Rathee, A. Jaglan, S. Garg, M. S. Obaidat, and G. Kaddoum, "A Survey on Adversarial Machine Learning in Communication Systems and Networks: Attacks, Defenses, and Applications," *Electronics*, vol. 13, no. 24, Art. no. 5030, Dec. 2024, doi: 10.3390/electronics13245030.

[26] S. Ahmed, "RobEns: Robust Ensemble Adversarial Machine Learning for Intrusion Detection in IoT-Cloud," Sci. Rep., vol. 14, no. 1, Art. no. 9482, Apr. 2024, doi: 10.1038/s41598-024-59986-y.

[27] G. Xu, Y. Yu, Y. Chen, and B. Li, "Diverse Models, United Goal: A Comprehensive Survey of Ensemble Learning," CAAI Trans. Intell. Technol., vol. 8, no. 4, pp. 1162–1184, Dec. 2023, doi: 10.1049/cit2.12214.

[28] M. R. Gauthama Raman and N. Somu, "A Robust Semi-Supervised Hybrid Ensemble for Network Intrusion Detection System (NIDS)," Expert Syst. Appl., vol. 250, Art. no. 123847, Sep. 2024, doi: 10.1016/j.eswa.2024.123847.

[29] J. Liu, "Semi-Supervised Ensemble Learning for Cybersecurity: A Review," IEEE Access, vol. 12, pp. 15432–15450, Feb. 2024.

[30] X. Yuan et al., "Adversarial Examples: Attacks and Defenses for Deep Learning," IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 9, pp. 2805–2824, Sep. 2019.

[31] H. Zhang and J. Wang, "Taxonomy of Adversarial Attacks on Network Intrusion Detection Systems," Comput. Secur., vol. 142, Art. no. 103850, Jan. 2025.

[32] A. G. S. Al-Dahhan, "Hybrid Machine Learning Models for Intrusion Detection: Combining Supervised and Unsupervised Techniques," Power Syst. Technol., vol. 48, no. 2, pp. 210–225, 2024.

[33] P. Berg, "Robust and accurate performance anomaly detection and prediction for cloud applications," Ph.D. dissertation, Fac. Sci., Univ. Amsterdam, Amsterdam, Netherlands, 2024. [Online]. Available: Research Explorer.

[34] S. Shrivastava et al., "A Comprehensive Review: The Evolving Cat-and-Mouse Game in Network Intrusion Detection Systems Leveraging Machine Learning," Preprints, Art. no. 202510.0564, Oct. 2025, doi: 10.20944/preprints202510.0564.v1.

[35] M. Al-Quraan, A. Khan, and M. A. Mohiuddin, "Robust Intrusion Detection System Using an Improved Hybrid Deep Learning Model for Binary and Multi-Class Classification in IoT Networks," Computers, vol. 13, no. 3, Art. no. 102, Mar. 2025, doi: 10.3390/computers13030102.

[36] K. J. M. Al-Atwi, "Robust Federated Learning for Zero-Day IoT Malware Detection via Adversarial and Generative Augmentation," Licentiate dissertation, KTH Royal Inst. Technol., Stockholm, Sweden, 2025. [Online]. Available: DiVA Portal.

[37] J. Liu, "Semi-Supervised Ensemble Learning for Cybersecurity: A Review," *IEEE Access*, vol. 12, pp. 15432–15450, Feb. 2024.

[38] J. Yoon et al., "AS-GCN: Adversarial Semi-Supervised Graph Convolutional Network for Intrusion Detection," *J. Parallel Distrib. Comput.*, vol. 182, Art. no. 104751, Dec. 2023.

[39] A. S. Husen and S. S. Somalingam, "Adversarial Examples in Network Intrusion Detection Systems: A Survey of Attack and Defense Methods," *Sensors*, vol. 24, no. 5, Art. no. 1520, Feb. 2024, doi: 10.3390/s24051520.

[40] Y. Zhou, G. Cheng, and S. Yu, "Adversarial Robustness in Deep Learning-Based Intrusion Detection: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 26, no. 1, pp. 120–155, 2025.

[41] F. S. Al-Atwi and J. Morisset, "Integrating Ensemble and Semi-Supervised Learning for Robust Intrusion Detection," *Preprints*, Art. no. 202501.0123, Jan. 2025.

[42] M. M. Alani, "Cybersecurity in the Era of AI: Research Gaps and Future Opportunities," *IEEE Consum. Electron. Mag.*, vol. 14, no. 1, pp. 45–52, Jan. 2025, doi: 10.1109/MCE.2024.3361201.

[43] T. H. Le, H. Kang, and K. Kim, "On the Evaluation of Adversarial Robustness in Machine Learning-based Network Intrusion Detection Systems," *IEEE Trans. Netw. Serv. Manag.*, vol. 21, no. 2, pp. 1102–1118, Apr. 2024.

[44] S. J. Pan, M. Xu, and W. Wang, "Why Accuracy is Not Enough: A Critical Analysis of NIDS Evaluation in Adversarial Settings," *Sci. Rep.*, vol. 15, no. 1, Art. no. 4321, Feb. 2025, doi: 10.1038/s41598-025-54321-x.

[45] L. Huang, V. L. L. Thing, and Y. Xu, "Robustness Evaluation of Deep Learning-Based Network Intrusion Detection Systems," *Comput. Secur.*, vol. 135, Art. no. 103480, Dec. 2023, doi: 10.1016/j.cose.2023.103480.

[46] A. Derhab, M. Guerroumi, and A. Gumaei, "Toward a Unified Framework for Secure and Robust NIDS in IoT Environments: Challenges and Gaps," *Comput. Secur.*, vol. 148, Art. no. 104120, Jan. 2025.

[47] M. Alani and A. Al-Tamimi, "Mapping the Landscape of Adversarial Robustness in Network Intrusion Detection: A Bibliometric Analysis," *IEEE Access*, vol. 12, pp. 45120–45135, 2024, doi: 10.1109/ACCESS.2024.3378901.

[48] Y. Zhang, X. Chen, and L. Wang, "Generative Semi-Supervised Learning for Robust NIDS: From VAE to

Diffusion Models," J. Netw. Comput. Appl., vol. 238, Art. no. 103912, Feb. 2025.
[49] J. K. Shon and S. Moon, "Adaptive Adversarial Defense for Semi-Supervised Intrusion Detection Systems," Computers & Security, vol. 145, Art. no. 104055, Jan. 2025, doi: 10.1016/j.cose.2024.104055.
[50] R. G. Raman, N. Somu, and K. Krithivasan, "Synergizing Ensemble Learning and Generative Adversarial Networks for Resilient NIDS," IEEE Trans. Inf. Forensics Secur., vol. 20, pp. 312–326, 2025.