 doi.org/10.56706/ik.v20i2.159

Analisis Komparatif IDS dan WAF terhadap SQL *Injection* dan *Brute Force*: Tinjauan Literatur Sistematis

Hermawan Setiawan[1)], Ahmad Muflih Izfatara[2,*)], Rahadian Ronggo Kusumo[3)], Zamir Achmad Sachio[4)], Moch Radhit Julian Randito[5)]

1) Program Studi Rekayasa Kriptografi, Politeknik Siber dan Sandi Negara, hermawan.setiawan@poltekssn.ac.id

2) Program Studi Rekayasa Kriptografi, Politeknik Siber dan Sandi Negara, ahmad.muflih@student.poltekssn.ac.id

3) Program Studi Rekayasa Keamanan Siber, Politeknik Siber dan Sandi Negara, rahadian.ronggo@student.poltekssn.ac.id

4) Program Studi Rekayasa Perangkat Keras Kriptografi, Politeknik Siber dan Sandi Negara, zamir.achmad@student.poltekssn.ac.id

5) Program Studi Teknik Sipil, Sekolah Tinggi Teknologi Pekerjaan Umum Jakarta, julianradhit@gmail.com

Riwayat Artikel

Dikirim 26 Feb 2026
Diterima 27 Agu 2026
Diterbitkan 25 Sep 2026

Kata kunci:

Firewall Aplikasi Web
Injeksi SQL
Serangan Brute Force
Sistem Deteksi Intrusi
Tinjauan Literatur Sistematis

Keywords:

Brute Force
Intrusion Detection System
SQL Injection
Systematic Literature Review
Web Application Firewall

Abstrak

Aplikasi web merupakan sasaran utama serangan SQL *Injection* (SQLi) dan *Brute Force*, sementara *Intrusion Detection System* (IDS) dan *Web Application Firewall* (WAF) berbasis tanda tangan terbatas dalam menghadapi serangan tersamar maupun serangan *zero-day*. Tinjauan Literatur Sistematis ini disusun mengikuti protokol PRISMA 2020 dan kerangka PICOC untuk memetakan tantangan penerapan, metode deteksi yang dominan, serta membandingkan metrik performa IDS dan WAF terhadap SQLi dan *Brute Force*. Pencarian pada tujuh basis data untuk publikasi 2021–2025, dengan penyaringan oleh dua penelaah independen dan penilaian kualitas lima butir, menghasilkan 14 studi primer yang dianalisis melalui sintesis tematik. Hasil tinjauan mengindikasikan bahwa perbandingan performa kedua sistem secara langsung rentan terhadap bias karena perbedaan lapisan operasional, unit analisis, dan *dataset* evaluasi. IDS berbasis *Deep Learning* hibrida melaporkan akurasi biner 99,84%–99,98% pada *dataset* arus jaringan yang hampir tidak memuat muatan SQLi, sedangkan WAF berbasis *Machine Learning* melaporkan akurasi 89,34%–97,57% pada muatan HTTP, dan pendekatan AI generatif memblokir 99% serangan SQLi yang sebelumnya lolos dengan 23 aturan tambahan. Penelitian ini mengusulkan kerangka normalisasi perbandingan berbasis strata, analisis *base rate*, dan daftar periksa pelaporan minimum delapan butir, serta menyimpulkan bahwa IDS dan WAF bersifat komplementer, bukan substitutif.

Abstract

Web applications are a primary target of SQL Injection (SQLi) and Brute Force attacks, while signature-based Intrusion Detection Systems (IDS) and Web Application Firewalls (WAF) are limited against obfuscated and zero-day attacks. This Systematic Literature Review follows the PRISMA 2020 protocol and the PICOC framework to map deployment challenges, dominant detection methods, and performance-metric comparisons between IDS and WAF against SQLi and Brute Force. Searches across seven databases for publications from 2021 to 2025, with screening by two independent reviewers

and a five-item quality appraisal, yielded 14 primary studies analyzed through thematic synthesis. The findings indicate that direct performance comparisons between the two systems are prone to bias because they differ in operational layer, unit of analysis, and evaluation dataset. Hybrid Deep Learning IDS report binary accuracies of 99.84%–99.98% on network flow datasets that barely contain SQLi payloads, whereas Machine Learning-based WAFs report accuracies of 89.34%–97.57% on HTTP payloads, and a generative AI approach blocks 99% of previously successful SQLi attacks with 23 additional rules. The study proposes a stratified comparison framework, a base rate analysis, and an eight-item minimum reporting checklist, and concludes that IDS and WAF are complementary rather than substitutive.

1. PENDAHULUAN

Aplikasi web telah menjadi komponen fundamental dalam infrastruktur digital, namun sekaligus menjadi sasaran utama eksploitasi siber [1]. Dua serangan yang paling persisten terhadapnya adalah SQL *Injection* (SQLi) dan *Brute Force* [2]. SQLi menargetkan kelemahan validasi masukan dengan memanipulasi kueri SQL untuk mengakses, memodifikasi, atau mengekstraksi data sensitif [3], sedangkan *Brute Force* mengeksploitasi mekanisme autentikasi melalui penebakan kredensial secara massal sehingga menghasilkan lalu lintas bervolume tinggi [4]. Perbedaan karakteristik ini berimplikasi langsung pada lapisan pertahanan yang relevan untuk masing-masing serangan.

Intrusion Detection System (IDS) memantau lalu lintas jaringan atau peristiwa sistem untuk mendeteksi anomali keamanan [4], [5], sedangkan Web Application Firewall (WAF) beroperasi pada lapisan aplikasi (Layer 7) sebagai *reverse proxy* yang menyaring lalu lintas HTTP/HTTPS yang berbahaya [6]. Keduanya secara tradisional bergantung pada deteksi berbasis tanda tangan (*signature-based*) yang sulit mendeteksi serangan *zero-day* maupun SQLi dengan teknik penyandian (*obfuscation*) [3], [7], [8], sehingga penelitian bergerak menuju integrasi *Machine Learning* (ML) dan *Deep Learning* (DL) untuk deteksi berbasis anomali [2], [4].

Literatur yang ada umumnya membahas peningkatan IDS atau WAF secara terpisah, sementara analisis komparatif keduanya dalam konteks SQLi dan *Brute Force* masih terbatas [1], [3], [9]. Kesenjangan ini berdampak nyata karena perbandingan literatur saat ini belum sepenuhnya seimbang. Penjajaran metrik lintas domain ini berisiko menciptakan ilusi keamanan. Secara praktis, organisasi dengan anggaran terbatas cenderung memilih investasi berdasarkan angka akurasi tunggal yang berasal dari kondisi pengujian yang berbeda-beda. Secara metodologis, metrik IDS dilaporkan pada *dataset* arus jaringan seperti CICIDS2017 atau UNSW-NB15, sedangkan metrik WAF dilaporkan pada muatan HTTP, sehingga perbandingan akurasi secara langsung berpotensi menyesatkan. Selain itu, sejumlah studi WAF tidak menguji *Brute Force* [1], sementara *dataset* IDS hanya memuat sedikit sampel SQLi, sehingga terdapat area ancaman yang kurang dievaluasi pada kedua lini penelitian.

Penelitian ini disusun sebagai Tinjauan Literatur Sistematis (*Systematic Literature Review*, SLR) mengikuti PRISMA 2020 [10] dan pedoman Kitchenham dan Charters [11] dengan lima kontribusi, yaitu (1) sintesis komparatif berbasis bukti per artikel antara IDS dan WAF berbasis ML/DL untuk SQLi dan *Brute Force*; (2) kerangka normalisasi perbandingan berbasis strata; (3) tabel konseptual perbandingan IDS dan WAF; (4) identifikasi wilayah ancaman yang kurang terevaluasi beserta agenda penelitian lanjutan; dan (5) daftar periksa pelaporan minimum berisi delapan butir. Penelitian ini tidak merancang sistem deteksi baru. Kontribusi utama makalah ini berfokus pada kritik metodologis, yaitu mengevaluasi bias metrik akibat kondisi data yang tidak seimbang (*base rate fallacy*) serta menetapkan standar perbandingan yang valid.

2. LANDASAN TEORI

2.1. IDS dan WAF: Definisi dan Perbandingan Konseptual

IDS didefinisikan sebagai sistem yang memantau peristiwa pada jaringan atau *host* dan memberikan peringatan atas indikasi pelanggaran kebijakan keamanan tanpa harus melakukan pemblokiran [5], [12], sedangkan WAF adalah kontrol keamanan pada lapisan aplikasi yang memeriksa dan dapat memblokir permintaan HTTP/HTTPS sebelum mencapai aplikasi [13], [14]. WAF tradisional seperti *ModSecurity* dengan OWASP *Core Rule* Set memerlukan pemutakhiran aturan secara berkala dan berpotensi gagal terhadap serangan termutasi [7], [8]. Integrasi kecerdasan buatan dilaporkan meningkatkan akurasi sekaligus menekan positif palsu [15], meskipun dalam kondisi eksperimental yang belum tentu bertahan dalam lalu lintas produksi.

Tabel 1 merangkum perbedaan konseptual antara kedua sistem. Karena unit analisisnya berbeda, metrik keduanya berada pada ruang pengukuran yang berbeda; IDS yang tidak mendeteksi SQLi tidak berarti gagal, melainkan beroperasi di luar cakupan rancangannya. Analoginya seperti membandingkan tensimeter dengan termometer: keduanya sahih dalam ranahnya, tetapi angkanya tidak dapat dibandingkan secara langsung.

Tabel 1. Perbandingan konseptual IDS dan WAF

Dimensi	*Intrusion Detection System* (IDS)	*Web Application Firewall* (WAF)
Lapisan dan posisi	Lapisan 3–4; pasif pada segmen jaringan (*span port*/*tap*) atau *host*	Lapisan 7; *in-line* sebagai *reverse proxy* di depan server web
Unit analisis	Arus (*flow*) atau paket jaringan	Permintaan HTTP/HTTPS (URL, *header*, *payload*)
Mode operasi	Deteksi dan peringatan; pemblokiran hanya pada varian IPS	Deteksi sekaligus pencegahan (*blocking*)
Sumber fitur	Durasi arus, jumlah paket, *flag* TCP, *port*, alamat IP	Struktur kueri, karakter khusus, token SQL, entropi *payload*
Serangan yang tertangkap baik	Anomali volumetrik: *Brute Force*, DoS/DDoS, pemindaian *port*	Injeksi berbasis muatan: SQLi, XSS, *path traversal*
Dataset tipikal	CICIDS2017, CSE-CIC-IDS2018, UNSW-NB15, X-IIoTID, NF-BoT-IoT-v2	*Payload* SQLi/XSS, log HTTP, CSIC 2010, *dataset* simulasi
Metrik lazim	Akurasi, *detection rate*, *false positive rate* per kelas	*Precision*, *recall*, *block rate*, *bypass rate*
Keterbatasan inheren	Ketidakseimbangan kelas, beban komputasi, buta muatan terenkripsi	*Overhead* inspeksi, sensitif variasi sintaksis antar-DBMS

2.2. Deteksi SQL *Injection* dan *Brute Force*

SQLi memanipulasi kueri basis data untuk mencuri atau mengubah data sensitif [3]. Tinjauan Alghawazi et al. [3] menunjukkan dominasi *supervised learning* seperti *Support Vector Machine*, *Decision Tree*, dan *Random Forest*, serta efektivitas model hibrida CNN–LSTM pada tingkat *payload*. Untuk menghadapi teknik penyandian yang menghindari aturan WAF, GenSQLi memanfaatkan *Large Language Model* untuk menghasilkan varian SQLi sekaligus menurunkan aturan keamanan [8], sedangkan PROGESI memanfaatkan *Context-Free Grammar* pada proksi Nginx untuk menggeneralisasi pola injeksi [6].

Brute Force menghasilkan volume lalu lintas yang tinggi dan pola yang berulang pada tingkat arus jaringan. IDS berbasis ML/DL umumnya dievaluasi pada CICIDS2017 [16], UNSW-NB15 [17], dan NF-BoT-IoT-v2 [15], [18], dengan *Brute Force* pada CICIDS2017 direpresentasikan melalui kelas *FTP-Patator*, *SSH-Patator*, dan *Web Attack–Brute Force* [4], [16]. Model DL dilaporkan unggul dalam mengenali pola tersebut [4], [19]. Namun, kelas *Brute Force* merupakan kelas minoritas sehingga akurasi agregat tidak dapat langsung dibaca sebagai kemampuan deteksi *Brute Force*.

2.3. Evaluasi Deteksi pada Kelas Tidak Seimbang

Axelsson [20] menunjukkan bahwa efektivitas praktis sistem deteksi intrusi ditentukan oleh *base rate*, yaitu prevalensi serangan dalam lalu lintas yang diamati. Berdasarkan *teorema Bayes*, presisi sistem dinyatakan dalam persamaan (1).

$$P(\text{serangan} \mid \text{alarm}) = \frac{\text{TPR} \times p}{\text{TPR} \times p + \text{FPR} \times (1 - p)} \quad (1)$$

dengan P (serangan | alarm) adalah presisi atau peluang sebuah alarm merupakan serangan, p adalah prevalensi kelas serangan, TPR adalah *true positive rate*, dan FPR adalah *false positive rate*. Pada p yang sangat kecil, sistem dengan TPR mendekati sempurna tetap menghasilkan mayoritas alarm palsu apabila FPR tidak ditekan jauh di bawah p; kekeliruan pembacaan inilah yang membedakan performa laboratorium dan performa operasional [21]. Pada distribusi timpang, kurva *precision-recall* lebih informatif daripada ROC [22], sehingga metrik per kelas beserta prevalensinya merupakan prasyarat penafsiran. Arp et al. [23] juga mengidentifikasi jebakan metodologis seperti bias sampel dan pengabaian *base rate*, sedangkan Engelen et al. [24] dan Lanvin et al. [25] menemukan kekeliruan pelabelan pada CICIDS2017 yang secara signifikan memengaruhi performa deteksi.

3. METODE PENELITIAN

3.1. Pertanyaan Penelitian dan Strategi Pencarian

Penelitian ini berpedoman pada tiga pertanyaan penelitian (RQ), yaitu RQ1: apa tantangan dan keterbatasan penerapan IDS dibandingkan dengan WAF di lingkungan industri; RQ2: metode deteksi apa yang dominan pada IDS dan WAF untuk SQLi dan *Brute Force*; dan RQ3: bagaimana perbandingan akurasi dan *false positive rate* keduanya serta sejauh mana perbandingan tersebut sahih secara metodologis. Kerangka PICOC ditetapkan dengan *Population* aplikasi web, *Intervention* IDS, *Comparison* WAF, *Outcome* efektivitas deteksi SQLi dan *Brute Force*, serta *Context* lingkungan industri dan sistem produksi.

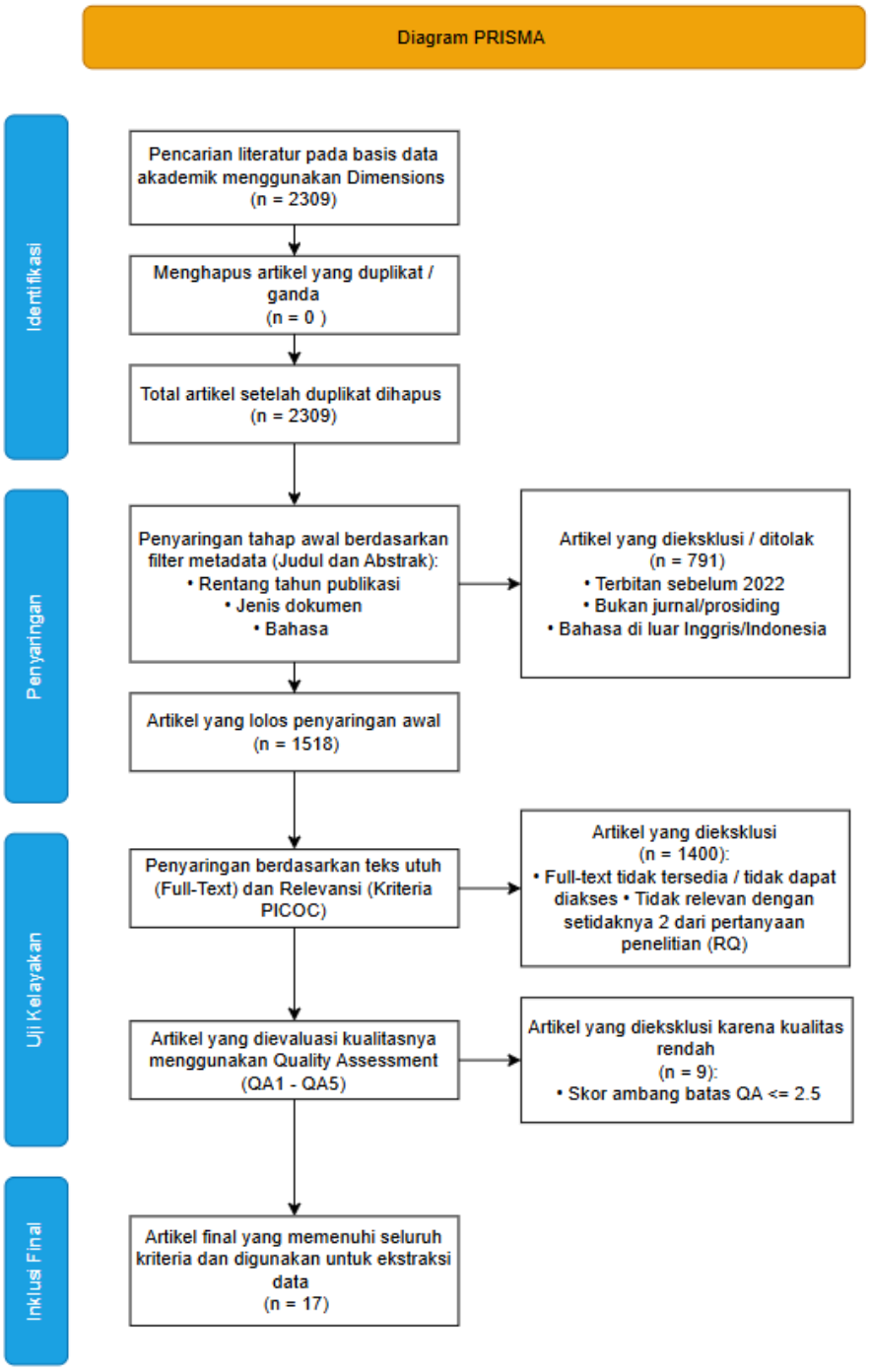
Pencarian dilakukan pada Scopus, IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, Dimensions, dan arXiv pada 12 Februari 2026 dan dimutakhirkan pada 10 Agustus 2026 menggunakan rangkaian *("Intrusion Detection System" OR "IDS") AND ("Web Application Firewall" OR "WAF") AND ("SQL Injection" OR "Brute Force") AND ("Machine Learning" OR "Deep Learning")* yang disesuaikan dengan sintaksis tiap basis data. Pencarian dilengkapi dengan *backward snowballing* sesuai dengan pedoman Wohlin [26].

3.2. Seleksi, Penilaian Kualitas, dan Ekstraksi Data

Kriteria inklusi meliputi artikel jurnal atau prosiding *peer-reviewed* terbitan 1 Januari 2021–31 Desember 2025, berbahasa Inggris atau Indonesia, teks penuh tersedia, menjawab sekurang-kurangnya dua RQ, serta melaporkan metrik kuantitatif beserta *dataset*. *Grey literature*, jurnal predator, dan sumber nonakademis dikeluarkan. Deduplikasi dilakukan melalui pencocokan DOI dan judul yang ternormalisasi, dilanjutkan dengan pemeriksaan manual, dengan versi jurnal yang dipertahankan pada tahap pra-cetak. Dua penelaah menyaring secara independen, dan kesepakatan diukur menggunakan kappa Cohen [27]; perbedaan diselesaikan melalui diskusi atau penelaah ketiga.

Kualitas studi dinilai dengan lima butir, yaitu kejelasan metode deteksi (QA1), data metrik kuantitatif beserta *dataset* (QA2), konteks implementasi industri (QA3), mitigasi spesifik SQLi atau *Brute Force* (QA4), serta jenis tugas klasifikasi dan distribusi kelas (QA5), dengan skor Ya = 1; Sebagian = 0,5; Tidak = 0 dan ambang penerimaan di atas 2,5. Proses seleksi disajikan pada Gambar 1. Rujukan fundamental di luar rentang 2021–2025, seperti NIST SP 800-31, tetap digunakan sebagai landasan teori, tetapi tidak dihitung sebagai bagian dari korpus, sehingga diperoleh 14 studi primer.

Data diekstraksi ke dalam formulir terstruktur yang memuat penulis, tahun, sistem, metode, *dataset*, jumlah sampel, jenis tugas klasifikasi, kelas serangan, metrik, konteks implementasi, dan keterbatasan, kemudian diverifikasi oleh penelaah kedua. Sintesis tematik dilakukan melalui pengodean baris demi baris, pembentukan tema deskriptif, dan pembentukan tema analitis yang menjawab RQ. Karena heterogenitas *dataset* tinggi, meta-analisis kuantitatif tidak dilakukan.



Gambar 1. *Diagram alir PRISMA proses seleksi studi*

3.3. Kerangka Normalisasi Perbandingan

Untuk menjawab RQ3 secara sahih, dua metrik hanya dinyatakan dapat diperbandingkan apabila memenuhi empat prasyarat sekaligus, yaitu (1) kesetaraan lapisan OSI; (2) kesetaraan unit analisis, yaitu sama-sama arus jaringan atau sama-sama permintaan HTTP; (3) kesetaraan jenis tugas, yaitu sama-sama biner atau multi-kelas; dan (4) kesetaraan prevalensi kelas, atau metrik dilaporkan per kelas. Apabila salah satu prasyarat tidak terpenuhi, perbandingan hanya dilakukan secara kualitatif atas cakupan ancaman, bukan secara kuantitatif atas besaran metrik. Kerangka ini diterapkan secara konsisten pada Subbab 4.3.

4. HASIL DAN PEMBAHASAN

4.1. Karakteristik Studi dan Bukti per Artikel

Tabel 2 menyajikan ekstraksi bukti dari 14 studi primer. Kolom tugas dan kelas serangan ditampilkan untuk menentukan strata perbandingan, sedangkan angka yang tidak dilaporkan studi asal tidak diperkirakan.

Tabel 2. Ekstraksi bukti per artikel

Studi (sistem)	Metode	*Dataset*	Tugas dan kelas relevan	Metrik yang dilaporkan
Kamal & Mashaly [4] (IDS)	*Autoencoder*–CNN; *Transformer*–DNN; ADASYN-SMOTE+ENN	CICIDS2017; NF-BoT-IoT-v2	Biner dan multi-kelas; *Patator*, *Web Attack*–*Brute Force*, SQLi	99,90%/99,92% biner dan 99,95%/99,96% multi-kelas (CICIDS2017); 99,98% biner (NF-BoT-IoT-v2)
Altunay & Albayrak [19] (IDS)	CNN, LSTM, CNN+LSTM	UNSW-NB15; X-IIoTID	Biner dan multi-kelas; tanpa kelas *Brute Force* eksplisit	93,21%/92,9% (UNSW-NB15); 99,84%/99,80% (X-IIoTID)
Wahab et al. [28] (IDS)	DL dan *Transformer*	DDoS jaringan IoT	Multi-kelas; DDoS	Di luar cakupan RQ; konteks metode
Talukder et al. [15] (IDS)	*Oversampling, stacking feature embedding*	CICIDS2017; UNSW-NB15	Biner dan multi-kelas; termasuk *Brute Force*	Akurasi agregat tinggi; metrik per kelas tidak tersedia
Yoshimura et al. [29] (IDS)	DOC-IDS: CNN 1D dan *Autoencoder*	Lalu lintas jaringan	Deteksi anomali tanpa label	Tidak dapat ditempatkan pada strata
Aldhaheri et al. [2] (IDS)	Tinjauan DL ancaman IoT	Beragam	*Botnet*, *password cracking*	Sintesis; tanpa metrik tunggal
Aslam et al. [30] (IDS)	ML adaptif SDN-IoT	DDoS pada SDN	Biner; DDoS	Di luar cakupan RQ
Dawadi et al. [1] (WAF)	Arsitektur berlapis LSTM	ISCX, CICDDoS, simulasi	Biner per lapisan; DDoS, XSS/SQLi; *Brute Force* di luar cakupan	97,57% (DDoS); 89,34% (XSS/SQLi)
Babaey & Ravindran [8] (WAF)	GenSQLi: LLM pembangkit serangan dan aturan	514 *payload* SQLi; 1.596 normal, 475 serangan	Evaluasi aturan; SQLi termutasi	92,5% *payload* valid; 99% serangan lolos terblokir dengan 23 aturan; *bypass* Gemini-Pro 56,6%
Coscia et al. [6] (WAF)	PROGESI: CFG pada proksi Nginx	*Payload* SQLi termutasi	Berbasis aturan; SQLi	Efektif pada *payload* termutasi (strata S3)
Shaheed & Kurdy [7] (WAF)	ML dengan rekayasa fitur	Permintaan HTTP	Biner; termasuk SQLi	Metrik muatan HTTP (strata S2)
Alghawazi et al. [3] (SQLi)	SLR teknik ML	Beragam	SQLi	Sintesis; *supervised learning* dominan
Alotaibi [31] (lintas)	Survei keamanan IIoT	Beragam	Beragam	Sintesis; tanpa metrik tunggal
Shang et al. [9] (lintas)	Tinjauan pertahanan *edge*	Beragam	Beragam	Sintesis; tanpa metrik tunggal

Tiga hal terbaca dari Tabel 2. Pertama, tidak ada satu pun studi yang menguji IDS dan WAF pada *dataset* yang sama. Kedua, studi IDS didominasi oleh *dataset* arus jaringan, sedangkan studi WAF didominasi oleh muatan HTTP. Ketiga, sebagian besar studi hanya melaporkan metrik agregat, bukan metrik per kelas serangan.

4.2. Tantangan Penerapan dan Metode Deteksi Dominan (RQ1 dan RQ2)

Tantangan utama IDS adalah mengelola lalu lintas berskala besar tanpa mengorbankan waktu respons. *Dataset* yang sangat tidak seimbang membuat model rentan terhadap *overfitting* dan *false negative* pada kelas minoritas [4], [15], sehingga digunakan SMOTE [32], ADASYN [33], atau kombinasinya dengan *Edited Nearest Neighbors* [4], sementara kompleksitas DL menimbulkan persoalan latensi pada perangkat tepi [9], [31]. Pada WAF, pendekatan tanda tangan rentan terhadap varian baru [7], [8]; *payload* SQLi hasil LLM terbukti melewati *ModSecurity* dengan OWASP CRS sebelum aturan tambahan diturunkan [8]; perbedaan sintaksis antar-DBMS menyulitkan

generalisasi aturan [6], [8]; dan inspeksi mendalam menambah *overhead* meskipun dampaknya tidak signifikan pada pengujian Dawadi et al. [1].

Dari sisi metode, IDS pada 2021–2025 didominasi oleh arsitektur hibrida spasial-temporal seperti CNN+LSTM [19], *Autoencoder*–CNN, dan *Transformer*–DNN [4], sementara *ensemble* klasik tetap kompetitif untuk metadata jaringan [15]. WAF berorientasi pada analisis muatan melalui LSTM dan Bi-LSTM [1], [7], tata bahasa formal [6], serta AI generatif untuk menghasilkan varian serangan dan menurunkan aturan [8]. Dominasi ini merupakan dominasi publikasi, bukan bukti superioritas produksi, dan pendekatan generatif yang memperkuat aturan tidak setara metriknya dengan pendekatan klasifikasi yang memprediksi label.

4.3. Perbandingan Performa dan Batas Keterbandingannya (RQ3)

Penerapan kerangka pada Subbab 3.3 menunjukkan bahwa perbandingan langsung antara akurasi IDS dan WAF tidak memenuhi prasyarat komparabilitas. Pertama, unit analisisnya berbeda: akurasi 99,92% pada klasifikasi biner CICIDS2017 [4] menyatakan proporsi arus jaringan yang diklasifikasikan dengan benar, sedangkan akurasi 89,34% [1] menyatakan proporsi permintaan HTTP yang diklasifikasikan sebagai XSS/SQLi. Kedua, prevalensi kelas tidak setara: kelas Web *Attack*–SQL *Injection* pada CICIDS2017 hanya menyisakan sembilan sampel setelah pembersihan pencilan dan peningkatan sintetis melalui SMOTE [4], sehingga akurasi multikelas sebesar 99,96% hampir tidak mencerminkan kemampuan deteksi SQLi yang sebenarnya. Ketiga, cakupan ancaman tidak setara: Dawadi et al. [1] mengecualikan *Brute Force*, sedangkan akurasi 99,84% [19] diperoleh pada X-IIoTID yang tidak secara eksplisit mencakup kelas *Brute Force*.

Dengan demikian, akurasi tertinggi yang sahih untuk IDS pada korpus ini adalah 99,98% (biner, NF-BoT-IoT-v2) dan 99,96% (multi-kelas, CICIDS2017) [4], sedangkan WAF melaporkan 89,34%–97,57% pada muatan HTTP [1], dan pendekatan generatif menyatakan capaian dalam satuan yang berbeda, yaitu 99% serangan terblokir dengan 23 aturan [8]. Tabel 3 menyajikan perbandingan yang dinormalisasi; perbandingan hanya dilakukan dalam setiap strata.

Tabel 3. Perbandingan performa IDS dan WAF setelah normalisasi berbasis strata

Strata	Lapisan dan unit analisis	Sistem	Serangan	Ringkasan bukti
S1	Jaringan; arus (*flow*)	IDS berbasis ML/DL	*Brute Force (Patator)*	Akurasi biner 99,90%–99,98% [4]; metrik per kelas jarang dilaporkan
S2	Aplikasi; permintaan HTTP	WAF berbasis ML/DL	SQLi dan XSS	89,34% untuk XSS/SQLi gabungan [1]; presisi tinggi pada fitur *payload* [7]
S3	Aplikasi; aturan pengamanan	WAF aturan dan AI generatif	SQLi termutasi	99% serangan lolos terblokir dengan 23 aturan [8]; generalisasi CFG [6]
S4	Aplikasi; permintaan HTTP	WAF	*Brute Force*	Tidak ada studi dalam korpus; dinyatakan di luar cakupan [1]
S5	Jaringan; arus	IDS	SQLi	Kelas minoritas ekstrem pada CICIDS2017; bukti tidak memadai

Strata S4 dan S5 merupakan temuan yang tersembunyi di balik perbandingan agregat, yaitu *Brute Force* pada lapisan aplikasi dan SQLi pada lapisan jaringan sebagai wilayah ancaman yang secara sistematis kurang dievaluasi.

4.4. Analisis Base Rate

Tabel 4 menyajikan presisi teoretis berdasarkan persamaan (1) pada TPR 99% untuk berbagai prevalensi dan FPR. Prevalensi kelas Web *Attack*–SQL *Injection* pada CICIDS2017 berada pada kisaran 10^{-5} [16], sehingga pengklasifikasi dengan TPR 99% dan FPR 0,1% hanya mencapai presisi sekitar 0,98%; artinya sekitar 99 dari 100 alarm SQLi merupakan alarm palsu. Akurasi 99,96% [4] karenanya tidak bertentangan dengan belum memadainya deteksi SQLi pada lapisan jaringan, karena keduanya mengukur hal yang berbeda, seperti alat uji saring dengan sensitivitas 99% yang tetap didominasi oleh positif palsu untuk penyakit yang diderita oleh satu dari seratus ribu orang.

Tabel 4. Presisi teoretis pada TPR 99% menurut prevalensi kelas serangan dan FPR

Prevalensi (p)	Contoh konteks empiris	FPR 1%	FPR 0,1%	FPR 0,01%
10^{-1}	Kelas hasil penyeimbangan sintetis (SMOTE/ADASYN)	91,67%	99,10%	99,91%
10^{-2}	*Brute Force* gabungan (*Patator*) pada CICIDS2017	50,00%	90,91%	99,01%
10^{-3}	Serangan web pada lalu lintas kampus atau korporat	9,02%	49,77%	90,83%
10^{-4}	Injeksi pada lalu lintas produksi bervolume tinggi	0,98%	9,01%	49,75%
10^{-5}	Web *Attack*–SQL *Injection* pada CICIDS2017	0,10%	0,98%	9,01%

SMOTE [32] dan ADASYN [33] meningkatkan prevalensi kelas minoritas pada data latih, yaitu dengan memindahkan kondisi pengujian dari baris terbawah ke baris teratas pada Tabel 4. Prosedur ini sahih untuk mengatasi bias pembelajaran, tetapi ketika diterapkan pada himpunan uji, presisi yang dilaporkan mencerminkan distribusi sintetis dan tidak dapat digeneralisasi ke lalu lintas produksi [23]. Hal yang sama berlaku bagi *false positive rate* rendah yang dilaporkan setelah penyeimbangan dan pembuangan pencilan [4], yang merupakan batas atas performa pada data bersih, bukan estimasi performa di lingkungan produksi.

4.5. Daftar Periksa Pelaporan Minimum dan Implikasi Arsitektural

Kerangka normalisasi hanya dapat diterapkan apabila studi melaporkan informasi yang diperlukan. Tabel 5 menyajikan daftar periksa pelaporan minimum yang diturunkan dari Subbab 4.3 dan 4.4 serta rekomendasi Arp et al. [23] dan Sommer dan Paxson [21]. Daftar ini bersifat kumulatif; tidak ada satu pun studi dalam korpus yang memenuhi seluruh butir, dengan kelemahan yang terpusat pada MR4, MR6, dan MR7.

Tabel 5. Daftar periksa pelaporan minimum bagi studi evaluasi IDS dan WAF

Kode	Informasi yang wajib dilaporkan	Alasan metodologis
MR1	Lapisan operasional dan titik penempatan sensor	Menentukan strata dan mencegah penjajaran lintas lapisan
MR2	Unit analisis beserta alat ekstraksi fitur	Unit berbeda mengukur objek berbeda
MR3	Identitas, versi, dan praproses *dataset*	Versi CICIDS2017 berbeda menghasilkan performa berbeda [24], [25]
MR4	Jumlah sampel per kelas sebelum dan sesudah penyeimbangan	Membedakan distribusi asli dari distribusi sintetis
MR5	Jenis tugas klasifikasi dan definisi label	Akurasi biner dan multi-kelas tidak setara
MR6	Presisi, *recall*, dan F1 per kelas	Metrik agregat didominasi kelas mayoritas [22]
MR7	Prevalensi kelas uji dan asumsi prevalensi operasional	Presisi operasional memerlukan *base rate* [20]
MR8	Pembuangan pencilan dan skema pemisahan data (acak/temporal)	Pemisahan acak data berurutan menimbulkan kebocoran [23]

Secara arsitektural, IDS dan WAF tidak saling menggantikan: WAF ditempatkan di depan server web untuk menangani SQLi dan injeksi lain, sedangkan IDS pada segmen jaringan menangkap anomali volumetrik akibat *Brute Force* dan pemindaian. Rekomendasi ini bersifat inferensial karena tidak ada satu pun studi yang mengevaluasi arsitektur gabungan maupun korelasi peringatan antara kedua sistem secara empiris. Ancaman validitas penelitian ini meliputi keragaman definisi kelas serangan antar-*dataset* (validitas konstruk) yang dimitigasi dengan kerangka strata, bias seleksi penelaah (validitas internal) yang dimitigasi dengan penyaringan ganda dan kappa Cohen, jumlah studi primer yang terbatas (validitas eksternal), serta bias publikasi terhadap studi berperforma tinggi yang tidak dapat dimitigasi sepenuhnya.

5. KESIMPULAN

Tinjauan Literatur Sistematis atas 14 studi primer dari tujuh basis data (2021–2025) ini tidak bertujuan untuk mengklaim secara mutlak sistem mana yang lebih unggul. Bukti yang ada justru mengindikasikan bahwa pertanyaan tersebut rentan menghasilkan kesimpulan yang bias apabila dijawab melalui penjajaran metrik secara langsung, karena perbedaan lapisan operasional, unit analisis, jenis tugas klasifikasi, dan prevalensi kelas. Setelah dinormalisasi, IDS berbasis DL hibrida melaporkan akurasi biner 99,90%–99,98% pada *dataset* arus jaringan, WAF berbasis ML/DL

melaporkan 89,34%–97,57% pada muatan HTTP, dan pendekatan AI generatif memblokir 99% serangan SQLi yang sebelumnya lolos dengan 23 aturan tambahan; seluruh angka tersebut hanya berlaku pada strata masing-masing. Kontribusi penelitian ini meliputi ekstraksi bukti per artikel yang dapat ditelusuri, kerangka normalisasi berbasis empat prasyarat komparabilitas, identifikasi dua wilayah ancaman yang kurang dievaluasi, serta daftar periksa pelaporan dengan minimal delapan butir. Analisis *base rate* menunjukkan bahwa pada prevalensi SQL *Injection* 10^{-5}, sistem dengan TPR 99% dan FPR 0,1% hanya mencapai presisi sekitar 0,98%, sehingga akurasi agregat tidak layak dijadikan dasar kesimpulan untuk kelas dengan prevalensi rendah. Penelitian selanjutnya disarankan untuk menyusun tolok ukur bersama yang memuat lalu lintas jaringan dan muatan HTTP secara simultan, mengevaluasi arsitektur gabungan beserta korelasi peringatan antarlapisan, serta mengembangkan model yang efisien untuk komputasi tepi dengan pelaporan metrik per kelas.

REFERENSI

[1] B. R. Dawadi, B. Adhikari, and D. K. Srivastava, "Deep learning technique-enabled web application firewall for the detection of web attacks," Sensors, vol. 23, no. 4, art. 2073, 2023, doi: 10.3390/s23042073.

[2] A. Aldhaheri, F. Alwahedi, M. A. Ferrag, and A. Battah, "Deep learning for cyber threat detection in IoT networks: A review," Internet of Things and Cyber-Physical Systems, vol. 4, pp. 110–128, 2024, doi: 10.1016/j.iotcps.2023.09.003.

[3] M. Alghawazi, D. Alghazzawi, and S. Alarifi, "Detection of SQL injection attack using machine learning techniques: A systematic literature review," Journal of Cybersecurity and Privacy, vol. 2, no. 4, pp. 764–777, 2022, doi: 10.3390/jcp2040039.

[4] H. Kamal and M. Mashaly, "Enhanced hybrid deep learning models-based anomaly detection method for two-stage binary and multi-class classification of attacks in intrusion detection systems," Algorithms, vol. 18, no. 2, art. 69, 2025, doi: 10.3390/a18020069.

[5] R. Bace and P. Mell, "Intrusion detection systems," National Institute of Standards and Technology, NIST Special Publication 800-31, 2001, doi: 10.6028/NIST.SP.800-31.

[6] A. Coscia, V. Dentamaro, S. Galantucci, A. Maci, and G. Pirlo, "PROGESI: A PROxy grammar to enhance web application firewall for SQL injection prevention," IEEE Access, vol. 12, pp. 107689–107703, 2024, doi: 10.1109/ACCESS.2024.3438092.

[7] A. Shaheed and M. H. D. B. Kurdy, "Web application firewall using machine learning and features engineering," Security and Communication Networks, vol. 2022, art. 5280158, 2022, doi: 10.1155/2022/5280158.

[8] V. Babaey and A. Ravindran, "GenSQLi: A generative artificial intelligence framework for automatically securing web application firewalls against structured query language injection attacks," Future Internet, vol. 17, no. 1, art. 8, 2025, doi: 10.3390/fi17010008.

[9] K. Shang, W. He, and S. Zhang, "Review on security defense technology research in edge computing environment," Chinese Journal of Electronics, vol. 33, no. 1, pp. 1–18, 2024, doi: 10.23919/cje.2022.00.170.

[10] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, art. n71, 2021, doi: 10.1136/bmj.n71.

[11] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," EBSE Technical Report EBSE-2007-01, Keele University and University of Durham, 2007.

[12] K. Scarfone and P. Mell, "Guide to intrusion detection and prevention systems (IDPS)," National Institute of Standards and Technology, NIST Special Publication 800-94, 2007, doi: 10.6028/NIST.SP.800-94.

[13] N. Gupta and A. Saikia, "Web application firewall," B.Tech Project Final Report, Dept. Comput. Sci. Eng., Indian Institute of Technology Kanpur, 2007.

[14] OWASP Foundation, "OWASP Top 10:2021 — The ten most critical web application security risks," 2021. [Daring]. Tersedia: https://owasp.org/Top10/

[15] M. A. Talukder et al., "Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction," Journal of Big Data, vol. 11, art. 33, 2024, doi: 10.1186/s40537-024-00886-w.

[16] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. 4th Int. Conf. Information Systems Security and Privacy (ICISSP), Funchal, Portugal, 2018, pp. 108–116, doi: 10.5220/0006639801080116.

[17] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in Proc. Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, 2015, pp. 1–6, doi: 10.1109/MilCIS.2015.7348942.

[18] M. Sarhan, S. Layeghy, and M. Portmann, "Towards a standard feature set for network intrusion detection system datasets," Mobile Networks and Applications, vol. 27, pp. 357–370, 2022, doi: 10.1007/s11036-021-01843-0.

[19] H. C. Altunay and Z. Albayrak, "A hybrid CNN+LSTM-based intrusion detection system for industrial IoT networks," Engineering Science and Technology, an International Journal, vol. 38, art. 101322, 2023, doi: 10.1016/j.jestch.2022.101322.

[20] S. Axelsson, "The base-rate fallacy and the difficulty of intrusion detection," ACM Transactions on Information and System Security, vol. 3, no. 3, pp. 186–205, 2000, doi: 10.1145/357830.357849.

[21] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in Proc. IEEE Symposium on Security and Privacy (S&P), Oakland, CA, USA, 2010, pp. 305–316, doi: 10.1109/SP.2010.25.

[22] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," PLOS ONE, vol. 10, no. 3, art. e0118432, 2015, doi: 10.1371/journal.pone.0118432.

[23] D. Arp, E. Quiring, F. Pendlebury, A. Warnecke, F. Pierazzi, C. Wressnegger, L. Cavallaro, and K. Rieck, "Dos and don'ts of machine learning in computer security," in Proc. 31st USENIX Security Symposium, Boston, MA, USA, 2022, pp. 3971–3988.

[24] G. Engelen, V. Rimmer, and W. Joosen, "Troubleshooting an intrusion detection dataset: The CICIDS2017 case study," in Proc. IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 2021, pp. 7–12, doi: 10.1109/SPW53761.2021.00009.

[25] M. Lanvin, P.-F. Gimenez, Y. Han, F. Majorczyk, L. Mé, and É. Totel, "Errors in the CICIDS2017 dataset and the significant differences in detection performances it makes," in Risks and Security of Internet and Systems (CRiSIS 2022), LNCS vol. 13857, Cham, Switzerland: Springer, 2023, pp. 18–33, doi: 10.1007/978-3-031-31108-6_2.

[26] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in Proc. 18th Int. Conf. Evaluation and Assessment in Software Engineering (EASE), London, UK, 2014, art. 38, doi: 10.1145/2601248.2601268.

[27] J. Cohen, "A coefficient of agreement for nominal scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37–46, 1960, doi: 10.1177/001316446002000104.

[28] S. A. Wahab, S. Sultana, N. Tariq, M. Mujahid, J. A. Khan, and A. Mylonas, "A multi-class intrusion detection system for DDoS attacks in IoT networks using deep learning and transformers," Sensors, vol. 25, no. 15, art. 4845, 2025, doi: 10.3390/s25154845.

[29] N. Yoshimura, H. Kuzuno, Y. Shiraishi, and M. Morii, "DOC-IDS: A deep learning-based method for feature extraction and anomaly detection in network traffic," Sensors, vol. 22, no. 12, art. 4405, 2022, doi: 10.3390/s22124405.

[30] M. Aslam et al., "Adaptive machine learning based distributed denial-of-services attacks detection and mitigation system for SDN-enabled IoT," Sensors, vol. 22, no. 7, art. 2697, 2022, doi: 10.3390/s22072697.

[31] B. Alotaibi, "A survey on industrial Internet of Things security: Requirements, attacks, AI-based solutions, and edge computing opportunities," Sensors, vol. 23, no. 17, art. 7470, 2023, doi: 10.3390/s23177470.

[32] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[33] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN), Hong Kong, 2008, pp. 1322–1328, doi: 10.1109/IJCNN.2008.4633969.