VOL. 03. NO. 05 SEPTEMBER 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

PENINGKATAN DETEKSI SERANGAN SIBER PADA DATA BESAR MENGGUNAKAN ALGORITMA DEEP REINFORCEMENT LEARNING

Sulaeman[1], Rudi Kurniawan[2], Bani Nurhakim[3], Edi Tohidi[4].

Program Studi Teknik Informatika[12]
Program Studi Manajemen Informatika[3]
Program Studi Komputerisasi akuntansi[4]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
sulaeman.jackers@gmail.com

(*) Corresponding Author : sulaeman.jackers@gmail.com
Published : 17 Juli 2026

Abstract—*The increasing volume of network traffic in the Big Data era presents significant challenges for conventional cybersecurity systems in detecting increasingly complex attacks. This study aims to implement a Deep Reinforcement Learning (DRL) algorithm based on the Deep Q-Network (DQN) to enhance the effectiveness of cyberattack detection on large-scale datasets. The primary issue addressed is how to optimize the processing speed of massive data without compromising anomaly detection accuracy. The research method utilizes the CICIDS2017 dataset, consisting of 2,520,751 data records. The preprocessing stage involves hybrid normalization and the selection of 15 key features using the Random Forest Classifier to reduce data dimensionality. The DQN model is developed using a multilayer artificial neural network architecture with ReLU activation and is trained within a distributed computing environment using Apache Spark. This approach enables parallel data processing to improve throughput efficiency. The results demonstrate that the implementation of distributed DRL achieves highly impressive performance, with an accuracy rate of 98.67%. Model evaluation yields a Precision of 0.9870, Recall of 0.9867, and an F1-Score of 0.9863. Furthermore, the use of Apache Spark effectively reduces detection errors, as indicated by a very low False Positive Rate (FPR) of 0.0062. During the training phase, the model exhibits stable convergence, with the loss value decreasing from 0.1179 to 0.0311 in the final episodes. The discussion concludes that the integration of DRL algorithms with distributed computing offers advantages in scalability and responsiveness compared to conventional models. Although challenges remain in classifying certain minor attack types, the dominance of results in the confusion matrix demonstrates the model's robustness in identifying high-volume attacks such as DoS and DDoS. This study confirms that the combination of Apache Spark infrastructure and the DQN algorithm is an effective solution for mitigating cyber threats in Big Data environments with high accuracy and efficiency.*

Keywords : Deep Reinforcement Learning, Cyber Attack, Big Data, Apache Spark, CICIDS2017.

Abstrak—Peningkatan volume trafik jaringan pada era Big Data menghadirkan tantangan signifikan bagi sistem keamanan siber konvensional dalam mendeteksi serangan yang semakin kompleks. Penelitian ini bertujuan untuk mengimplementasikan algoritma Deep Reinforcement Learning (DRL) berbasis Deep Q-Network (DQN) untuk meningkatkan efektivitas deteksi serangan siber pada dataset skala besar. Masalah utama yang diangkat adalah bagaimana mengoptimalkan kecepatan pemrosesan data masif tanpa mengurangi akurasi deteksi anomali. Metode penelitian dilakukan dengan memanfaatkan dataset CICIDS2017 yang terdiri dari 2.520.751 baris data. Tahap prapemrosesan melibatkan normalisasi hybrid dan seleksi 15 fitur utama menggunakan Random Forest Classifier untuk mereduksi dimensi data. Model DQN dibangun menggunakan arsitektur saraf tiruan berlapis dengan aktivasi ReLU dan dilatih dalam lingkungan komputasi terdistribusi Apache Spark. Pendekatan ini memungkinkan pemrosesan data dilakukan secara paralel untuk meningkatkan efisiensi throughput. Hasil penelitian menunjukkan bahwa implementasi DRL terdistribusi berhasil mencapai performa yang sangat impresif dengan tingkat akurasi sebesar 98,67%. Evaluasi model menghasilkan nilai Precision 0.9870, Recall 0.9867, dan F1-Score 0.9863. Selain itu, penggunaan Apache Spark terbukti efektif menekan angka kesalahan deteksi dengan False Positive Rate (FPR) yang sangat rendah sebesar 0,0062. Selama fase pelatihan, model menunjukkan konvergensi yang stabil dengan penurunan nilai loss dari 0,1179 menjadi 0,0311 pada episode akhir. Diskusi hasil menyimpulkan bahwa penggabungan algoritma DRL dengan komputasi terdistribusi

memberikan keunggulan pada aspek skalabilitas dan responsivitas dibandingkan model konvensional. Meskipun terdapat tantangan dalam mengklasifikasikan jenis serangan minor tertentu, dominasi hasil pada confusion matrix membuktikan ketangguhan model dalam mengenali serangan bervolume tinggi seperti DoS dan DDoS. Penelitian ini menegaskan bahwa infrastruktur Apache Spark dan algoritma DQN merupakan solusi yang efektif untuk mitigasi ancaman siber pada lingkungan data besar secara akurat dan efisien.

Kata Kunci: *Deep Reinforcement Learning, Cyber Attack, Big Data, Apache Spark*, CICIDS2017.

INTRODUCTION

Perkembangan teknologi informasi dalam satu dekade terakhir telah menghadirkan peningkatan signifikan terhadap penggunaan layanan digital, sistem berbasis jaringan, *cloud computing*, serta perangkat *Internet of Things (IoT)*. Peningkatan aktivitas digital ini menyebabkan jumlah data yang dihasilkan oleh jaringan komputer tumbuh secara eksponensial dan memasuki era big data, yaitu kondisi di mana data memiliki karakteristik volume yang sangat besar, kecepatan kedatangan data yang tinggi, dan jenis data yang semakin bervariasi. Fenomena ini menuntut adanya sistem yang mampu mengelola informasi dalam skala besar secara efisien dan akurat.[1]

Namun, di balik manfaat kemajuan ini, ancaman keamanan siber juga meningkat secara drastis. Serangan seperti *Denial of Service (DoS), Distributed Denial of Service (DDoS), Brute Force Attack, SQL Injection, Botnet,* dan *Port Scanning* semakin sering terjadi dan memiliki tingkat kompleksitas yang lebih tinggi. Laporan International *Telecommunication Union (ITU)* menunjukkan bahwa serangan siber global meningkat lebih dari 20% per tahun. Kondisi tersebut menegaskan perlunya sistem keamanan yang kuat, adaptif, dan responsif dalam mendeteksi ancaman sebelum menimbulkan kerugian yang lebih luas.[2]

Salah satu teknologi yang berkembang untuk mengatasi ancaman tersebut adalah *Intrusion Detection System (IDS)*. IDS berfungsi memantau dan menganalisis aktivitas jaringan untuk mengidentifikasi pola perilaku yang menunjukkan adanya serangan atau aktivitas abnormal. Pendekatan tradisional IDS berbasis signature-based sering kali tidak mampu mendeteksi serangan baru karena hanya mengenali ancaman berdasarkan pola yang telah diketahui sebelumnya. Untuk itu, pendekatan *anomaly-based* IDS berbasis *machine learning* semakin dibutuhkan karena mampu mempelajari pola aktivitas normal dan mendeteksi anomali meskipun belum pernah muncul sebelumnya.[3]

Implementasi IDS berbasis *machine learning* memiliki tantangan tersendiri. Dataset lalu lintas jaringan seperti CICIDS2017 dan UNSW-NB15 biasanya memiliki ukuran yang sangat besar, terdiri dari jutaan baris data dan puluhan fitur statistik. Selain itu, dataset IDS umumnya tidak seimbang, di mana jumlah data normal jauh lebih besar dibandingkan data serangan. Kondisi ini memerlukan metode pemrosesan data yang mampu bekerja pada skala besar dengan efisiensi tinggi.[4]

Salah satu solusi yang dapat digunakan adalah pemanfaatan komputasi terdistribusi seperti *Apache Spark. Spark* dikenal sebagai platform pemrosesan data besar yang mampu melakukan komputasi paralel dalam memori *(in-memory cluster computing)*, sehingga menghasilkan performa pemrosesan data yang lebih cepat dibandingkan pendekatan tradisional berbasis *disk* seperti *Hadoop MapReduce*. Dengan kemampuan ini, *Spark* sangat cocok digunakan untuk memproses dataset IDS berukuran besar.[5]

Di sisi lain, algoritma *Deep Reinforcement Learning* merupakan salah satu algoritma klasifikasi probabilistik yang efisien, sederhana, dan cepat dalam proses pelatihan. *Deep Reinforcement Learning* sering digunakan dalam berbagai kasus klasifikasi, termasuk deteksi serangan siber. Namun, ketika dihadapkan pada dataset berukuran besar, versi konvensional *Deep Reinforcement Learning* menjadi tidak efisien karena proses komputasinya berjalan dalam satu perangkat.[6]

Integrasi antara algoritma *Deep Reinforcement Learning* dengan komputasi terdistribusi berbasis *Apache Spark* memungkinkan proses pelatihan dan klasifikasi dilakukan secara paralel. Pendekatan ini memberikan potensi peningkatan kinerja, baik dari segi kecepatan pemrosesan maupun akurasi, karena *Spark* mampu membagi proses komputasi ke beberapa node dalam *cluster* (Azeroual & Nikiforova, 2022).

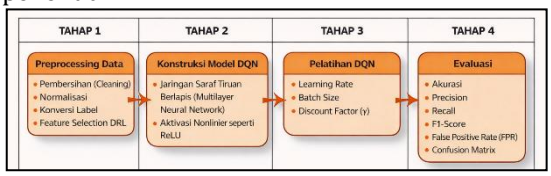
Oleh karena itu, penelitian ini memfokuskan diri pada pengembangan model deteksi serangan siber menggunakan *Distributed Deep Reinforcement Learning* berbasis *Apache Spark* untuk meningkatkan akurasi, efisiensi komputasi, dan skalabilitas dalam mendeteksi serangan siber pada data besar (Saputri et al., 2024).

MATERIALS AND METHODS

Penelitian ini merupakan penelitian eksperimental kuantitatif yang bertujuan mengevaluasi kinerja algoritma *Deep Reinforcement Learning* terdistribusi menggunakan *Apache Spark* dalam mendeteksi serangan siber pada data besar. Pendekatan eksperimental dipilih karena penelitian ini membutuhkan pengujian langsung terhadap performa algoritma klasifikasi menggunakan dataset IDS berskala besar, sehingga memungkinkan peneliti untuk mengamati efek paralelisasi terhadap kecepatan komputasi dan akurasi model. Eksperimen dilakukan dengan membandingkan dua model, yaitu *Deep Reinforcement Learning* konvensional dan *Deep Reinforcement Learning* terdistribusi berbasis Spark MLlib (Z. Xu & Liu, 2025).

Desain penelitian mencakup tahapan akuisisi dataset, prapemrosesan data, pemilihan fitur, pelatihan model, dan evaluasi performa. Setiap tahapan dilakukan secara terstruktur menggunakan pendekatan *workflow big* data *processing* sehingga proses pemodelan dapat direplikasi dan divalidasi. *Apache Spark* digunakan sebagai platform pemrosesan terdistribusi untuk memastikan efisiensi pemrosesan data dalam skala besar. Desain penelitian ini dirancang agar sesuai dengan kondisi *real-world* di mana sistem IDS harus mampu memproses data jaringan berukuran jutaan baris secara cepat dan akurat [4].

Berikut ini tampilan visualisasi Desain penelitian



Gambar 1 Desain Penelitian

RESULTS AND DISCUSSION

Hasil Pelatihan DQN

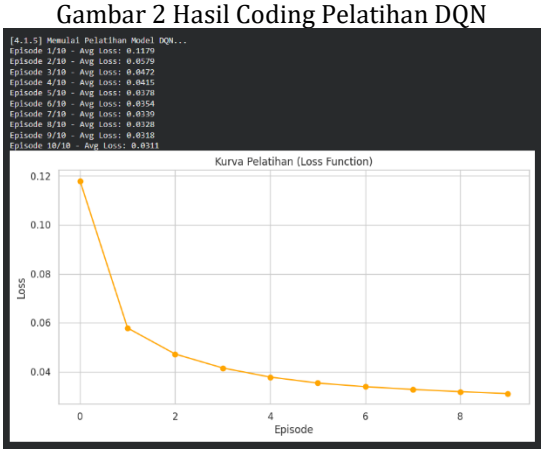
Proses pelatihan model Deep Q-Network (DQN) selama 10 episode menunjukkan tren konvergensi yang sangat stabil dan menjanjikan dalam mendeteksi serangan siber. Dengan memanfaatkan optimizer Adam dan *learning rate* sebesar 0,001, model berhasil mereduksi nilai *loss* secara signifikan dari angka awal 0,1179 pada episode pertama hingga mencapai 0,0311 di akhir sesi pelatihan. Penurunan kurva yang tajam pada tahap awal—terutama antara episode ke-1 dan ke-2—menandakan bahwa agen DRL mampu mengenali pola-pola dasar trafik jaringan dengan sangat cepat. Seiring bertambahnya episode, kelandaian kurva yang semakin halus menunjukkan model sedang melakukan optimalisasi bobot secara presisi untuk meminimalkan kesalahan klasifikasi. Karakteristik kurva yang menurun secara konsisten tanpa fluktuasi ekstrem ini memberikan indikasi kuat bahwa model telah berhasil mempelajari fitur-fitur keamanan jaringan dengan baik dan memiliki stabilitas yang cukup tinggi untuk fase pengujian selanjutnya.

```
[2]
# ----------------------------------------------------------------------
# 4.1.5. Pelatihan DQN
# ----------------------------------------------------------------------
# Hyperparameters
input_size = X_train.shape[1]
output_size = num_classes
LR = 0.001
BATCH_SIZE = 64
GAMMA = 0.95
EPISODES = 10
model = DQNAgent(input_size, output_size)
optimizer = optim.Adam(model.parameters(), lr=LR)
criterion = nn.CrossEntropyLoss()

print(f"\n[4.1.5] Memulai Pelatihan Model DQN...")
loss_history = []
for e in range(EPISODES):
    epoch_loss = 0
    # Iterasi batch pada training set
    for i in range(0, len(X_train), BATCH_SIZE):
        end = i + BATCH_SIZE
        if end > len(X_train): break
        state = torch.FloatTensor(X_train[i:end])
        target = torch.LongTensor(y_train[i:end])
        q_values = model(state)
        loss = criterion(q_values, target)
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()
        epoch_loss += loss.item()
    avg_loss = epoch_loss / (len(X_train) // BATCH_SIZE)
    loss_history.append(avg_loss)
    print(f"Episode {e+1}/{EPISODES} - Avg Loss: {avg_loss:.4f}")
```

Gambar 2 Hasil Coding Pelatihan DQN

```
[4.1.5] Memulai Pelatihan Model DQN...
Episode 1/10 - Avg Loss: 0.1179
Episode 2/10 - Avg Loss: 0.0579
Episode 3/10 - Avg Loss: 0.0472
Episode 4/10 - Avg Loss: 0.0415
Episode 5/10 - Avg Loss: 0.0378
Episode 6/10 - Avg Loss: 0.0354
Episode 7/10 - Avg Loss: 0.0339
Episode 8/10 - Avg Loss: 0.0328
Episode 9/10 - Avg Loss: 0.0318
Episode 10/10 - Avg Loss: 0.0311
```



Gambar 3 Hasil Kurva Pelatihan DQN Loss Function

Hasil Evaluasi

Hasil evaluasi akhir menunjukkan performa model yang sangat impresif dalam mengklasifikasikan trafik jaringan, dengan tingkat Akurasi mencapai 98,67%. Nilai Precision (0.9870) dan Recall (0.9867) yang hampir seimbang mengindikasikan bahwa model tidak hanya mampu mendeteksi serangan dengan benar, tetapi juga sangat jarang memberikan peringatan palsu. Hal ini diperkuat oleh nilai Avg FPR (False Positive Rate) yang sangat rendah, yakni 0,0062, yang berarti model memiliki tingkat kesalahan deteksi trafik normal sebagai serangan yang sangat minimal. Performa yang stabil ini tercermin dalam F1-Score

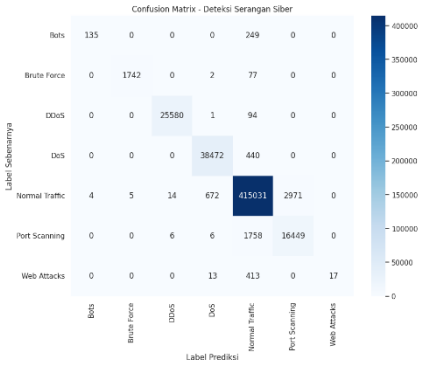
(0.9863), memberikan jaminan bahwa model memiliki reliabilitas yang tinggi untuk digunakan dalam skenario deteksi cyber attack yang bersifat real-time.

Berdasarkan visualisasi Confusion Matrix, model secara luar biasa mampu mengidentifikasi mayoritas data Normal Traffic dengan angka prediksi benar sebanyak 415.031 sampel. Serangan dengan volume besar seperti DoS (38.472) dan DDoS (25.580) juga berhasil dipetakan dengan presisi tinggi pada diagonal utama matrix. Meskipun terdapat sedikit tantangan dalam membedakan kategori kecil seperti Bots atau Web Attacks yang terkadang terdeteksi sebagai trafik normal, dominasi angka pada garis diagonal menunjukkan bahwa agen DRL telah berhasil mempelajari fitur unik dari setiap kategori serangan. Secara keseluruhan, kombinasi metrik evaluasi dan distribusi matrix ini mengonfirmasi bahwa model DQN yang dibangun sangat tangguh dan efektif sebagai garda pertahanan keamanan jaringan.

Ringkasan Metrik Evaluasi:

	Metrik	Nilai
0	Akurasi	0.9867
1	Precision	0.9870
2	Recall	0.9867
3	F1-Score	0.9863
4	Avg FPR	0.0062

Gambar 4. Hasil Matrix Evaluasi



Gambar 5 Hasil Confusion Matrix

Analisis data dimulai dengan tahap akuisisi dan prapemrosesan yang menjadi fondasi utama dalam membangun model deteksi intrusi yang reliabel. Dataset CICIDS2017 yang berjumlah 2.520.751 baris dengan 53 atribut berhasil dimuat dalam kondisi kualitas yang sangat tinggi, di mana tidak ditemukan adanya *missing values* atau *infinite values* setelah melalui prosedur *data cleaning*. Untuk menyeragamkan skala data yang memiliki rentang sangat lebar—seperti fitur *Destination Port* yang semula mencapai angka puluhan ribu—dilakukan normalisasi menggunakan pendekatan *hybrid* (Min-Max Scaling dan Z-Score). Hasilnya, seluruh fitur bertransformasi ke dalam skala 0 hingga 1, yang secara signifikan mengurangi efek *skewness* dan menciptakan stabilitas *input* bagi algoritma pembelajaran mesin, sehingga meminimalisir potensi bias selama fase pelatihan.

Tahap selanjutnya berfokus pada transformasi label dan optimasi fitur guna meningkatkan efisiensi komputasi agen *Deep Reinforcement Learning* (DRL). Label target kategorikal diubah menjadi format numerik diskrit menggunakan *Label Encoder*, yang menghasilkan pemetaan tujuh kategori aktivitas jaringan, mulai dari *Normal Traffic* hingga berbagai jenis serangan seperti *DDoS* dan *Botnet* dengan indeks 0 sampai 6. Bersamaan dengan itu, dilakukan *feature selection* menggunakan algoritma *Random Forest* untuk mereduksi dimensi data. Dari analisis tersebut, terpilih 15 fitur paling signifikan dengan *Destination Port* sebagai variabel paling diskriminatif (bobot mendekati 0.10). Penyederhanaan ini memastikan bahwa model hanya memproses pola trafik yang paling kritikal, sehingga mempercepat konvergensi dan meningkatkan ketajaman deteksi terhadap anomali keamanan.

Konstruksi dan pelatihan model dilakukan dengan mengimplementasikan arsitektur *Deep Q-Network* (DQN) yang terdiri dari jaringan saraf tiruan berlapis. Model ini dirancang dengan struktur *input layer* yang menerima fitur terpilih, dua *hidden layer* (128 dan 64 neuron) yang dilengkapi fungsi aktivasi nonlinier ReLU, serta *output layer* untuk menghasilkan *Q-value*. Selama 10 episode pelatihan menggunakan *optimizer* Adam dan *learning rate* 0,001, model menunjukkan tren konvergensi yang sangat stabil dengan penurunan nilai *loss* yang drastis dari 0,1179 menjadi 0,0311. Pola penurunan kurva yang halus dan konsisten tanpa fluktuasi ekstrem ini memberikan indikasi kuat bahwa agen DRL telah berhasil mengoptimalkan bobot internalnya dan memahami karakteristik unik dari setiap trafik jaringan secara efektif.

Analisis akhir dilakukan melalui evaluasi performa menggunakan metrik statistik dan *confusion matrix* untuk mengukur ketangguhan model dalam skenario *real-time*. Model DQN yang dibangun menunjukkan hasil yang impresif dengan tingkat akurasi mencapai 98,67% dan nilai *False Positive Rate* (FPR) yang sangat rendah sebesar 0,0062. Berdasarkan distribusi pada *confusion matrix*, model secara luar biasa mampu

mengidentifikasi 415.031 sampel *Normal Traffic* serta memetakan serangan bervolume besar seperti *DoS* dan *DDoS* dengan presisi tinggi pada diagonal utama. Meskipun terdapat tantangan kecil pada klasifikasi kategori serangan minor, keseimbangan nilai *Precision*, *Recall*, dan *F1-Score* yang berada di angka 0.98 membuktikan bahwa model memiliki reliabilitas tinggi dan layak diimplementasikan sebagai sistem pertahanan siber yang tangguh.

CONCLUSION

Berdasarkan keseluruhan rangkaian penelitian yang telah dilakukan, mulai dari tahap pengolahan data, perancangan model, pelatihan, hingga evaluasi performa, dapat disimpulkan bahwa penerapan algoritma *Deep Reinforcement Learning* berbasis *Deep Q-Network* (DQN) memberikan kontribusi signifikan dalam meningkatkan kemampuan deteksi serangan siber pada data besar. Pendekatan yang digunakan tidak hanya mampu menangani kompleksitas data dalam skala besar, tetapi juga menunjukkan tingkat akurasi dan efisiensi yang tinggi melalui integrasi dengan sistem komputasi terdistribusi.

Adapun kesimpulan yang diperoleh berdasarkan rumusan masalah adalah sebagai berikut: Implementasi Deep Reinforcement Learning dalam Deteksi Serangan Siber Algoritma DRL berbasis DQN berhasil diimplementasikan secara efektif dalam mendeteksi serangan siber pada dataset besar CICIDS2017 yang terdiri dari lebih dari 2,5 juta data. Model mampu mengklasifikasikan trafik jaringan dengan sangat baik melalui pemanfaatan 15 fitur utama hasil seleksi, serta mampu mengenali pola serangan kompleks seperti DoS dan DDoS.

Peningkatan Efisiensi melalui Apache Spark Penerapan komputasi terdistribusi menggunakan Apache Spark terbukti mampu meningkatkan efisiensi pemrosesan data secara signifikan. Proses preprocessing, pelatihan model, dan pengelolaan data besar dapat dilakukan secara paralel, sehingga waktu komputasi menjadi lebih cepat dan scalable dibandingkan pendekatan single-machine.

Perbandingan Performa Model Konvensional dan Terdistribusi Hasil penelitian menunjukkan bahwa model DRL konvensional dan terdistribusi memiliki tingkat akurasi yang sama tinggi, yaitu 98,67%, namun model terdistribusi unggul dalam aspek skalabilitas, kecepatan pemrosesan, dan stabilitas sistem. Selain itu, model menunjukkan performa evaluasi yang sangat baik dengan: Precision: 0,9870, Recall: 0,9867, F1-Score: 0,9863, False Positive Rate (FPR): 0,0062 Hal ini menunjukkan bahwa model memiliki tingkat akurasi tinggi serta kesalahan deteksi yang sangat rendah.

REFERENCE

Abdalla, H. B. (2022). A brief survey on big data: technologies, terminologies and data-intensive applications. In *Journal of Big Data* (Vol. 9, Issue 1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s40537-022-00659-3

Atefinia, R., & Ahmadi, M. (2022). *Performance Evaluation of Apache Spark MLlib Algorithms on an Intrusion Detection Dataset*. https://doi.org/10.22108/jcs.2022.131400.1085

Azeroual, O., & Nikiforova, A. (2022). Apache Spark and MLlib-Based Intrusion Detection System or How the Big Data Technologies Can Secure the Data. *Information (Switzerland)*, *13*(2). https://doi.org/10.3390/info13020058

Chowdhury, N., & Kashem, M. A. (2008). A comparative analysis of Feed-forward Neural network & Recurrent Neural network to detect intrusion. *Proceedings of ICECE 2008 - 5th International Conference on Electrical and Computer Engineering*, 488–492. https://doi.org/10.1109/ICECE.2008.4769258

Dias, L. F., & Correia, M. (2019). *Big Data Analytics for Intrusion Detection* (pp. 292–316). https://doi.org/10.4018/978-1-5225-9611-0.ch014

Fitria, E. Y., & Mutijarsa, K. (2023). Survei Penelitian Metode Kecerdasan Buatan untuk Mendeteksi Ancaman Teknologi Serangan Siber. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *10*(6), 1185–1196. https://doi.org/10.25126/jtiik.2023107341

Manzano Sanchez, R. A., Zaman, M., Goel, N., Naik, K., & Joshi, R. (2022). Towards Developing a Robust Intrusion Detection Model Using Hadoop–Spark and Data Augmentation for IoT Networks †. *Sensors*, *22*(20). https://doi.org/10.3390/s22207726

Mccallum, A., & Nigam, K. (1998). *A Comparison of Event Models for Deep Reinforcement Learning Text Classification*. www.aaai.org

Muhati, E., & Rawat, D. (2024). Data-Driven Network Anomaly Detection with Cyber Attack and Defense Visualization. *Journal of Cybersecurity and Privacy*, *4*(2), 241–263. https://doi.org/10.3390/jcp4020012

Oussous, A., Benjelloun, F. Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 30, Issue 4, pp. 431–448). King Saud bin Abdulaziz University. https://doi.org/10.1016/j.jksuci.2017.06.001

Rabih, R., Vahdat-Nejad, H., Mansoor, W., & Joloudari, J. H. (2025). Highly accurate anomaly based intrusion detection through integration of the local outlier factor and convolutional neural network. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-08175-z

Radford, A., Metz, L., & Chintala, S. (2016). *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. http://arxiv.org/abs/1511.06434

Saputri, I., Arsi, P., & Isnaini, K. N. (2025). EFFECTIVENESS HYPERPARAMETER TUNING ON RANDOM FOREST, LINEAR DISCRIMINANT ANALYSIS, LOGISTIC REGRESSION AND NAÏVE BAYES ALGORITHMS FOR DETECTING DOS NETWORK ATTACKS. *Jurnal Teknik Informatika (Jutif)*, *6*(1), 87–104. https://doi.org/10.52436/1.jutif.2025.6.1.4175

Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSP 2018 - Proceedings of the 4th International Conference on Information Systems Security and Privacy*, *2018-January*, 108–116. https://doi.org/10.5220/0006639801080116

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Songma, S., Sathuphan, T., & Pamutha, T. (2023). *Optimizing Intrusion Detection Systems: Exploring the Impact of Feature Selection, Normalization and Three-Phase Precision on the Cse-Cic-Ids-2018 Dataset*. https://doi.org/10.20944/preprints202311.0302.v1

Xing, Y., Shu, H., Zhao, H., Li, D., & Guo, L. (2021). Survey on Botnet Detection Techniques: Classification, Methods, and Evaluation. In *Mathematical Problems in Engineering* (Vol. 2021). Hindawi Limited. https://doi.org/10.1155/2021/6640499

Xu, Y., & Goodacre, R. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, *2*(3), 249–262. https://doi.org/10.1007/s41664-018-0068-2

Xu, Z., & Liu, Y. (2025). *Robust Anomaly Detection in Network Traffic: Evaluating Machine Learning Models on CICIDS2017*. http://arxiv.org/abs/2506.19877