



Machine Learning Techniques for Network Intrusion Detection: A Comparative Review of Algorithms, Datasets, and Performance Metrics (2020–2025)

Mustafa Almahdi Algaet ^{1*}, Mahmoud Sharnah ²

^{1,2} Department of Network, Faculty of Information Technology, Elmergib University, Al-Khums, Libya

تقنيات تعلم الآلة للكشف عن الاختراقات في الشبكات: مراجعة مقارنة للخوارزميات، ومجموعات البيانات، ومقاييس الأداء (2025–2020)

مصطفى المهدي القط^{1*}، محمود شرننة²

^{2,1} قسم الشبكات، كلية تقنية المعلومات، جامعة المرقب، الخمس، ليبيا

*Corresponding author: malgaet@elmergib.edu.ly

Received: August 16, 2026

Accepted: September 25, 2026

Published: October 21, 2026



Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract:

This paper presents a comparative review of machine learning techniques used in network intrusion detection systems by analyzing fourteen studies published between 2020 and 2025. The review was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, and the studies were identified using Google Scholar with a search string combining terms related to machine learning, intrusion detection, and performance evaluation. The studies were selected based on their application of at least one machine learning algorithm, the presence of quantitative performance metrics, and publication in English within the specified period. The reviewed studies employed algorithms that fall into three main categories: traditional models such as Random Forest, Decision Tree, and Support Vector Machine; deep learning models such as Long Short-Term Memory and Deep Belief Networks; and ensemble methods such as XGBoost and AdaBoost. Of the thirteen studies that reported a single summary accuracy value, the average accuracy was 96.42%, ranging from 83.27% to 99.99%. The Random Forest algorithm was the most frequently tested, appearing in ten out of fourteen studies. The NSL-KDD dataset was the most frequently used, appearing in six studies. However, nine out of fourteen studies did not report an F1-Score, thus limiting the possibility of direct comparison between the results. The results indicate that traditional models and ensemble methods remain highly competitive compared to deep learning methods when applied to standard datasets. Furthermore, the choice of dataset significantly impacts the reported performance levels. The study recommends that future research focus on using newer datasets that are more representative of real-world environments, adhering to consistent reporting of multiple performance metrics, and developing standardized evaluation frameworks that allow for more reliable comparisons between different studies.

Keywords: Machine Learning; Intrusion Detection System; Network Security; Random Forest; Deep Learning; Performance Evaluation; Comparative Review; Cybersecurity.

المخلص

تقدم هذه الورقة مراجعة مقارنة لتقنيات التعلم الآلي المستخدمة في أنظمة كشف التسلل عبر الشبكات، وذلك من خلال تحليل أربع عشرة دراسة نُشرت في الفترة ما بين عامي 2020 و 2025. أُجريت هذه المراجعة وفقاً لإرشادات "العناصر المفضلة

للمراجعات المنهجية وتحليل البيانات التجميعي (PRISMA)، حيث تم تحديد الدراسات باستخدام محرك البحث "جوجل سكولار" (Google Scholar) عبر سلسلة بحثية تجمع مصطلحات تتعلق بالتعلم الآلي، وكشف التسلسل، وتقييم الأداء. وقد وقع الاختيار على الدراسات بناءً على معايير محددة، منها: استخدام خوارزمية واحدة على الأقل من خوارزميات التعلم الآلي، وتوفر مقاييس أداء كمية، وأن تكون الدراسات منشورة باللغة الإنجليزية خلال الفترة الزمنية المذكورة. تضمنت الدراسات التي شملتها المراجعة خوارزميات تتدرج تحت ثلاث فئات رئيسية: النماذج التقليدية (مثل الغابة العشوائية "Random Forest"، وشجرة القرار "Decision Tree"، وآلة المتجهات الداعمة "Support Vector Machine")، ونماذج التعلم العميق (مثل الشبكات العصبية المتكررة "Long Short-Term Memory" وشبكات الاعتقاد العميق "Deep Belief Networks")، والأساليب التجميعية (مثل "XGBoost" و"AdaBoost"). ومن بين ثلاث عشرة دراسة قدمت قيمة إجمالية لدقة الأداء، بلغ متوسط الدقة 96.42%، تراوحت بين 83.27% و99.99%. وكانت خوارزمية "الغابة العشوائية" (Random Forest) هي الأكثر اختباراً، حيث ظهرت في عشر دراسات من أصل أربع عشرة، بينما كانت مجموعة بيانات "NSL-KDD" هي الأكثر استخداماً، إذ وردت في ست دراسات. ومع ذلك، لم تذكر تسع دراسات من أصل أربع عشرة مقياس "F1-Score"، مما حدّ من إمكانية إجراء مقارنة مباشرة بين النتائج. وتشير النتائج إلى أن النماذج التقليدية والأساليب التجميعية لا تزال تتمتع بقدرة تنافسية عالية مقارنة بأساليب التعلم العميق عند تطبيقها على مجموعات بيانات قياسية. علاوة على ذلك، تبين أن اختيار مجموعة البيانات يؤثر بشكل جوهري على مستويات الأداء المُبلغ عنها. وتوصي الدراسة بأن تركز الأبحاث المستقبلية على استخدام مجموعات بيانات أحدث وأكثر تمثيلاً لبيئات العمل الواقعية، مع الالتزام بتقديم تقارير متسقة تتضمن مقاييس أداء متعددة، وتطوير أطر تقييم موحدة تتيح إجراء مقارنات أكثر موثوقية بين الدراسات المختلفة.

الكلمات المفتاحية: التعلم الآلي؛ نظام كشف التسلسل؛ أمن الشبكات؛ الغابة العشوائية (Random Forest)؛ التعلم العميق؛ تقييم الأداء؛ مراجعة مقارنة؛ الأمن السيبراني.

1. Introduction

New opportunities have been created for cyber attackers due to the rapid expansion of network infrastructure, and this has made network security one of the most important challenges in the digital age [1]. For this reason, an Intrusion Detection System (IDS) is designed to detect malicious activities and monitor network traffic before any significant damage occurs [2]. In recent years, Machine Learning (ML) techniques have improved the performance of Intrusion Detection Systems by automatically learning attack patterns from stored data [3]. Because traditional intrusion detection systems were having difficulty detecting unknown and new attacks, researchers were prompted to explore solutions based on machine learning [1]. Several studies have compared the performance of various machine learning algorithms in the field of intrusion detection, and the results have varied considerably based on the dataset and evaluation metrics used [4]. Despite increasing research in this field, comprehensive comparisons of machine learning algorithms used in IDS across multiple datasets and different performance metrics remain limited. Accordingly, this paper presents a comparative review of ML techniques used in network intrusion detection through studies published between 2020 and 2025.

2. Methodology

This study used PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) to ensure the transparent and systematic selection of relevant scientific literature [5]. The search was conducted in April 2026 and used the following search string: ("machine learning" OR "deep learning" OR "neural network") AND ("intrusion detection" OR "network security" OR "anomaly detection") AND ("comparison" OR "performance" OR "evaluation"). The type of machine learning algorithm or network environment used was not restricted in this way to cover the widest possible range of relevant methods and approaches. The scientific research was searched using Google Scholar as the primary database, employing search terms that combined several concepts such as machine learning, intrusion detection, comparison, and performance.

The search results period for published studies was defined as 2020 to 2025. The studies were selected based on their application of at least one machine learning algorithm in the field of network intrusion detection, providing quantitative performance metrics such as accuracy or F1-Score, and being published in English between 2020 and 2025. Table 1 summarizes the inclusion and exclusion criteria applied during the screening process.

Table 1. Inclusion and exclusion criteria applied during the study selection process.

Criterion	Inclusion	Exclusion
Publication year	2020–2025	Before 2020
Language	English	Non-English
Topic	ML-based network intrusion detection	Unrelated to IDS or ML
Performance metrics	Reports accuracy or F1-score	No quantitative performance data
Publication type	Peer-reviewed journal or conference	Non-peer-reviewed (preprints, reports)

Studies lacking clear performance metrics, published before 2020, not focusing on areas related to network intrusion detection, or not undergoing peer review were excluded. Initially, 26 studies were collected from the research process, and 14 of them were selected for final analysis and comparison after meeting the required criteria. The selected studies had the required basic information extracted and placed in a comparison table containing the authors, year of publication, machine learning algorithm used in the study, standard dataset, and the accuracy to which it was assigned. Data extraction was performed by reviewing the full text of each selected study. For each study, the following information was recorded: (1) Authors' names and publication year; (2) Machine learning algorithms used; (3) Reference datasets used; (4) Marked accuracy rate; (5) F1-Score value, where available.

In studies that evaluated multiple algorithms, the highest marked accuracy was used as the representative value for that study. This approach was adopted to enable consistent cross-study comparison, acknowledging that it may favour algorithms tested under optimal conditions. Studies that did not report a single, summarized accuracy value were retained in the review but excluded from the calculation of the quantitative average accuracy. Figure 1 illustrates the PRISMA flowchart, which explains the stages of identifying, screening, and incorporating studies into the study selection process.

This review used Google Scholar as the only search database. While Google Scholar covers a wide range of academic literature, not including other databases such as Web of Science and Scopus limits the comprehensiveness of the search. This limitation is acknowledged and discussed further in Section 4.

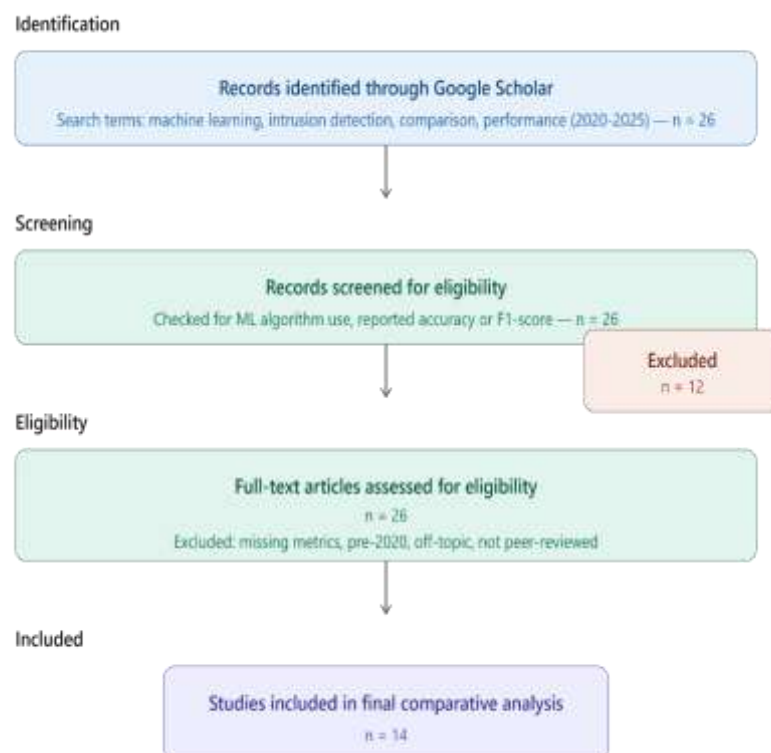


Figure 1. PRISMA flow diagram illustrating the study selection process.

3. Results

The selected and reviewed studies used machine learning algorithms that fall into three main categories: traditional models such as Random Forest, SVM, and Decision Tree; deep learning models such as CNN, DBN, and LSTM; and ensemble methods such as AdaBoost and XGBoost. The studies selected and reviewed spanned six years of publication, with 2021 being the most productive, contributing four studies, followed by 2020 and 2024 with three studies each, 2023 with two studies, and 2022 and 2025 with one study each. Random Forest was the most frequently tested algorithm, appearing in ten out of fourteen studies, followed closely by Decision Tree in nine, SVM in eight, the NSL-KDD dataset in six, UNSW-NB15 in five, and CSE-CIC-IDS2018, TON_IoT, and CICIDS2017 in one or two studies each. The reviewed studies used a range of benchmark datasets, from the older KDD Cup 99 to more recent IoT-specific datasets such as DS2OS and TON_IoT. Among the thirteen studies that reported a single summary accuracy value, the average accuracy was 96.42%. It should be noted that this average

is calculated across studies using different datasets, algorithms, and experimental conditions, and should be interpreted as a general indicator rather than a precise comparative benchmark. Ten of the thirteen studies reported accuracy exceeding 95%, while five studies exceeded 99%.

Table 2 presents a comparison of the algorithms, datasets, and reported accuracy rates in each of the fourteen reviewed studies.

Table 2. Comparison of Machine Learning Algorithms for Network Intrusion Detection (2020–2025).

Ref.	Authors	Year	Algorithm(s)	Dataset	Accuracy	F1-Score	Notes
[3]	Sow & Mehdi	2025	RF, XGBoost, DNN	NSL-KDD	99.80%	0.9977	RF best performer
[9]	Hussain et al.	2024	LR, KNN, MLP, GRU, LSTM	UNSW-NB15	97.68%	0.9768	KNN=95.09%, MLP=96.38%, GRU=97.68% (best)
[4]	Alharthi et al.	2025	RF, XGBoost, DT, SVM, LSTM, CNN	SDN / IEC 60870-5-104	99.97%*	0.9957*	RF best on IEC; XGB best on SDN
[14]	Ajagbe et al.	2024	RF, XGBoost, DT	UNSW-NB15	~90.39% (RF)	N/A	Imbalanced dataset study
[6]	Zakariah et al.	2023	ANN, DNN, SVM, LSTM, KNN	NSL-KDD	97.50%	95.70%	
[11]	Hidayat et al.	2023	DT, MLP, LSTM, KNN, RF, AdaBoost	TON_IoT	99.80%	N/A	AdaBoost best (99.80%); DT=99.6%, MLP=99.2%
[13]	Disha & Waheed	2022	DT, MLP, LSTM, GRU, AdaBoost	UNSW-NB15 / TON_IoT	91.31%	0.9146	GIWRF used for feature selection; DT best performer
[15]	Shrestha et al.	2021	DT, KNN, LDA, LR, GNB, SGD	CSE-CIC-IDS2018	99.99%	1.00	Highest accuracy in review
[8]	Ahmad et al.	2021	RF, SVM, ANN	UNSW-NB15	98.67%	N/A	App & transport layer features
[7]	Alhajjar et al.	2021	MLP, SVM, RF, DT, NB, KNN, LR, LDA, QDA, BAG, GB	NSL-KDD / UNSW-NB15	83.27%	N/A	Lowest accuracy in review
[10]	Islam et al.	2021	RF, DT, SVM, LR, ANN, DBN, LSTM, Bi-LSTM	NSL-KDD / IoTDevNet / DS2OS / IoTID20 / IoT Botnet	99.40%	N/A	IoT-focused
[2]	Bagaa et al.	2020	J48, Bayes Net, RF, Hoeffding, SVM	NSL-KDD	98.86%	N/A	IoT security framework
[12]	Magán-Carrión et al.	2020	RF, SVM, DT, LR, DNN, RNN	UGR'16	Varies*	Reported	No single accuracy figure
[1]	Shaukat et al.	2020	SVM, DT, RF, NB, ANN, CNN	KDD Cup 99 / NSL-KDD	99.30%	N/A	

* Magán-Carrión et al. (2020) did not report a single summary accuracy figure; results vary by attack class.

* Alharthi et al. (2025): XGBoost achieved F1=0.9997 on SDN dataset; RF achieved F1=0.9357 on IEC dataset. Values marked * represent the highest reported result across both datasets.

Shrestha et al. achieved the highest accuracy rate of 99.99% using the KNN and Decision Tree algorithms on the CSE-CIC-IDS2018 dataset, followed by Sow and Mehdi with 99.80% using Random Forest on the NSL-KDD dataset. Eight of the fourteen studies tested algorithms from more than one class simultaneously by comparing traditional models with deep learning methods through an experimental setup, allowing for direct comparisons

within the same study. The lowest accuracy rate was 83.27%, achieved by Alhajjar et al. using a Multi-Layer Perceptron network between the NSL-KDD and UNSW-NB15 datasets. However, nine of the fourteen studies reported accuracy without assigning an F1-Score.

4. Discussion

The thirteen studies that reported a single summary accuracy value recorded an average of 96.42%. This figure varied across studies depending on the algorithm tested, the dataset used, and the experimental setup applied [6], [7].

Random Forest appeared in ten of the fourteen studies. It resists overfitting, handles unbalanced and high-dimensional data, and requires less parameter tuning than deep learning models [3], [1], [8]. These properties make it a practical choice for researchers working with datasets of varying sizes and class distributions. Deep learning models such as LSTM and DBN were tested in several studies. Bagaa et al. evaluated multiple algorithms on NSL-KDD and reported that Random Forest achieved 98.86% accuracy, outperforming other tested models [2]. Several other studies reported similar findings where traditional models matched or exceeded deep learning approaches [3], [9], [10].

NSL-KDD appeared in six of the fourteen studies. Despite its widespread use, it does not reflect modern network traffic patterns or contemporary attack vectors [7]. Studies using NSL-KDD reported accuracy above 99% in three cases. Studies using CICIDS2017 and TON_IoT reported accuracy between 94% and 97% [11]. IoT-based intrusions, encrypted traffic attacks, and adversarial examples were not captured in the NSL-KDD dataset [7]. The gap between accuracy scores reported on NSL-KDD and those reported on newer datasets was noted across multiple studies.

Nine of the fourteen studies reported accuracy without including an F1-score [6], [7]. Of the five that reported F1-score, values ranged from 0.91 to 0.99 [6], [13]. Normal traffic outnumbered attack traffic in most of the datasets used across the reviewed studies. In two studies, classifiers that correctly labeled most normal samples achieved high overall accuracy while performing poorly on attack classes [14], [6].

Shrestha et al. achieved 99.99% on CSE-CIC-IDS2018 using Decision Tree and KNN [15]. Alhajjar et al. reported 83.27% using MLP on NSL-KDD and UNSW-NB15 [7]. The remaining twelve studies fell between these two values, with ten of them exceeding 95%.

4.1. Limitations

Several limitations of this review should be acknowledged. First, the review is based on fourteen studies, which limits the statistical generalizability of the findings. Second, the search was conducted using Google Scholar as the sole database; the exclusion of Scopus and Web of Science may have resulted in relevant studies being missed. Third, only English-language publications were included. Fourth, the review covers studies published between 2020 and 2025 only, and findings from earlier periods are not considered. Fifth, the datasets, algorithms, and experimental setups used across the reviewed studies were not standardized, and direct quantitative comparisons between studies were therefore not always possible. These limitations should be considered when interpreting the findings of this review.

5. Conclusion

This paper reviewed fourteen studies published between 2020 and 2025 that applied machine learning techniques to network intrusion detection. Three questions were addressed: which algorithms appear most frequently, which datasets are most commonly used, and how accuracy varies across studies. Random Forest appeared in ten of the fourteen studies. Decision Tree appeared in nine and SVM in eight. These three algorithms recorded competitive accuracy across most of the studies that tested them. NSL-KDD appeared in six studies, more than any other dataset. UNSW-NB15 appeared in five. Both datasets were introduced before 2010. Accuracy ranged from 83.27% to 99.99%, with an average of 96.42% among the thirteen studies that reported a single value. Nine of the fourteen studies did not report F1-score. TON_IoT and CICIDS2017 were used in one and two studies respectively. Testing on these datasets alongside NSL-KDD, and reporting F1-score in addition to accuracy, are two steps not yet taken consistently across the field.

Declaration and Funding Statement

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper. The authors further confirm that no financial support, grants, or funding were received for this research.

List of Abbreviations

Abbreviation	Full Term
AdaBoost	Adaptive Boosting
ANN	Artificial Neural Network
Bi-LSTM	Bidirectional Long Short-Term Memory
CNN	Convolutional Neural Network
CSE-CIC-IDS2018	Canadian Institute for Cybersecurity Intrusion Detection System 2018
DBN	Deep Belief Network
DNN	Deep Neural Network
DT	Decision Tree
F1-Score	F1 Score
GB	Gradient Boosting
GNB	Gaussian Naive Bayes
GRU	Gated Recurrent Unit
IDS	Intrusion Detection System
IoT	Internet of Things
KDD	Knowledge Discovery in Databases
KNN	K-Nearest Neighbors
LDA	Linear Discriminant Analysis
LR	Logistic Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multi-Layer Perceptron
NB	Naive Bayes
NSL-KDD	Network Security Laboratory–Knowledge Discovery in Databases
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QDA	Quadratic Discriminant Analysis
RF	Random Forest
RNN	Recurrent Neural Network
SDN	Software-Defined Networking
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
UNSW-NB15	University of New South Wales Network-Based 2015
XGBoost	Extreme Gradient Boosting

References

- [1] Shaukat, K.; Luo, S.; Varadharajan, V.; Hameed, I.A.; Xu, M. Performance Comparison and Current Challenges of Using Machine Learning Techniques in Cybersecurity. *Energies* 2020, 13, 2509. <https://doi.org/10.3390/en13102509>
- [2] Bagaa, M.; Taleb, T.; Bernabe, J.B.; Skarmeta, A. A Machine Learning Security Framework for IoT Systems. *IEEE Access* 2020, 8, 114066–114077. <https://doi.org/10.1109/ACCESS.2020.2996214>
- [3] Sow, M.; Mehdi, S. Enhancing IDS Performance Through a Comparative Analysis of Random Forest, XGBoost, and Deep Neural Networks. *Machine Learning with Applications* 2025, 19, 100738. <https://doi.org/10.1016/j.mlwa.2025.100738>
- [4] Alharthi, A.; Alaryani, M.; Kaddoura, S. A Comparative Study of Machine Learning and Deep Learning Models in Binary and Multiclass Classification for Intrusion Detection Systems. *Array* 2025, 26, 100406. <https://doi.org/10.1016/j.array.2025.100406>
- [5] Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* 2021, 372, n71. <https://doi.org/10.1136/bmj.n71>
- [6] Zakariah, M.; AlQahtani, S.A.; Al-Rakhami, M.S. Intrusion Detection System with Customized Machine Learning Techniques for NSL-KDD Dataset. *Computers, Materials & Continua* 2023, 76, 3025–3054. <https://doi.org/10.32604/cmc.2023.043752>
- [7] Alhajjar, E.; Maxwell, P.; Bastian, N. Adversarial Machine Learning in Network Intrusion Detection Systems. *Expert Systems with Applications* 2021, 186, 115782. <https://doi.org/10.1016/j.eswa.2021.115782>

- [8] Ahmad, I.; Ur Rehman, M.; Ali, A.; Khan, M.I.; Khan, F.A. Intrusion Detection in Internet of Things Using Supervised Machine Learning Based on Application and Transport Layer Features. *EURASIP Journal on Wireless Communications and Networking* 2021, 2021, 1–16. <https://doi.org/10.1186/s13638-021-01893-8>
- [9] Hussain, A.; Khatoon, A.; Aslam, A.; Tariq; Khosa, M.A. A Comparative Performance Analysis of Machine Learning Models for Intrusion Detection Classification. *Journal of Cybersecurity* 2024, 8, 1–18. <https://doi.org/10.32604/jcs.2023.046915>
- [10] Islam, R.; Nur Alam, A.S.M.; Khatun, T.; Uddin, M.A.; Alenezi, F.; Masud, M. Towards Machine Learning Based Intrusion Detection in IoT Networks. *Computers, Materials & Continua* 2021, 69, 1801–1821. <https://doi.org/10.32604/cmc.2021.018466>
- [11] Hidayat, I.; Ali, M.Z.; Arshad, A. Machine Learning-Based Intrusion Detection System: An Experimental Comparison. *Journal of Computational and Cognitive Engineering* 2023, 2, 97–104. <https://doi.org/10.47852/bonviewJCCE2202270>
- [12] Magán-Carrión, R.; Urda, D.; Díaz-Cano, I.; Dorronsoro, B. Towards a Reliable Comparison and Evaluation of Network Intrusion Detection Systems Based on Machine Learning Approaches. *Applied Sciences* 2020, 10, 1775. <https://doi.org/10.3390/app10051775>
- [13] Disha, R.A.; Waheed, S. Performance Analysis of Machine Learning Models for Intrusion Detection System Using Gini Impurity-Based Weighted Random Forest. *Cybersecurity* 2022, 5, 1–22. <https://doi.org/10.1186/s42400-021-00103-8>
- [14] Ajagbe, S.A.; Awotunde, J.B.; Oladele, T.O.; Abiodun, M.K.; Imoize, A.L. Intrusion Detection: A Comparison Study of Machine Learning Models Using Unbalanced Dataset. *SN Computer Science* 2024, 5, 1–18. <https://doi.org/10.1007/s42979-024-03369-0>
- [15] Shrestha, R.; Kim, S. Machine-Learning-Enabled Intrusion Detection System for Cellular Connected UAV Networks. *Electronics* 2021, 10, 1549. <https://doi.org/10.3390/electronics10131549>

Disclaimer/Publisher’s Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **JIBAS** and/or the editor(s). **JIBAS** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.