

A Supervised Machine Learning Framework for Intrusion Detection Using NSL-KDD Dataset and Ensemble Classification Models

Khtam Al-Meyah

Department of Architectural Engineering, College of Engineering, University of Basrah , Iraq.

khtam.gwam@uobasrah.edu.iq

ARTICLE INFO

Article history:

Received: 02 /04/2026

Rrevised form: 03 /00052026

Accepted : 05 /05/2026

Available online: 30/09/2026

Keywords:

Supervised Learning, Network Traffic Analysis, Anomaly Detection, Cyber Threat Detection, Classification Algorithms, Data Preprocessing, Feature Engineering, Model Evaluation, Cross-Validation, Ensemble Methods

ABSTRACT

We propose a supervised machine learning-based approach for intrusion detection using the NSL-KDD dataset. The primary aim is to develop and compare a reproducible offline Intrusion Detection System (IDS) that can classify normal network traffic against intrusions using experimental settings. The methodology encompasses the data cleaning, encoding of categorical features, scaling, balancing of class distribution, model building and evaluation through a common evaluation framework.

We tested five supervised learning models: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbors. The models are evaluated using accuracy, precision, recall, F1-score, ROC-AUC and false positive rate. The experimental results demonstrated that the best performance was achieved with the Random Forest classifier in terms of 97.2% accuracy, 97.6% precision, 96.9% recall, 97.2% F1-score, and 0.991 ROC-AUC, with the lowest false positive rate (2.1%). This suggests that ensemble classification techniques can achieve decent performance for intrusion detection experiments, with the aid of these census-based preprocessing and stratification procedures.

The research findings show Random Forest is the best choice among the tested algorithms for the proposed IDS framework, and similar preprocessing, cross-validation and security-related metrics enhance the reliability and repeatability of machine learning-based intrusion detection studies.

<https://doi.org/10.29304/jqcm.2026.18.32777>

1. Introduction

1.1 Background

Cyber- security threats have intensely increased in number and in sophistication and this has remained much owing to the increased reliance of organizations on cloud across the board, remote access technologies and dataStuff. Modern network pervasives generate huge data rates and velocity of traffic flows [1], and detecting the maliciousness in time has become a challenge in those situations when it is based on human observations or on fixed, rules-based systems [2]. This has resulted in the Intrusion Detection Systems (IDS) becoming an essential component of the cyber-security architecture, to detect the presence of unauthorized access, anomalies and possible attack vectors like the denial-of-service attacks, probing and increasing of the level of privilege and stealing data [3].

Traditionally, the IDS methods are divided into signature-basis and anomaly-basis. The signature-change analysis systems excel at detecting already known attack patterns [4], however, on novel or obfuscated threats will not work, and must be continuously modified. On the other hand, anomaly-based systems use anomalies with respect to normalized behaviour, however, are subject to noise [5], and hence sensitivity to noise may cause a system to large false alarms / false positive rates unless well calibrated. Knowing the above limitations, it has become an appealing field to supplement the effectiveness of IDS with machine learning (ML). Through learning the historic network traffic, ML based systems can be able to make sense of the tricky and unfamiliar attack signatures which have the capability to be observed with the enhanced acuity [6].

In most cases, MLLearn driven IDS investigation Intrusion Detection is tailored as a classification problem [7] which can either be binary, by attempting to distinguish between normal and malicious behaviour and multi-class by attempting to distinguish between various typologies of attack. Measures of statistical performance that have already been agreed upon, normally determine the adequacy of such models thereby ensuring the interoperability, reproducibility and meaningful comparisons across approaches [8].

1.2 Research Problem

As the machine learning-based intrusion detection system field has witnessed stupendous breakthrough, the success of many of these investigations that are still progressing is comforted with tiresomely complex implementations that involve deployment of networks in real time or bespoke monitoring ecclesiastical and state-of-the-art system integration or coding prerequisites. These complexities inevitably reduce the availability to interdisciplinary scholars and introduce hindrance to reproducibility in any of the situations in which there are divergences of evaluation conditions, unstandardized preprocessing regimes and conflicting validation strategies are of the order of the day. Besides this, several issues may be encountered during the training datasets of IDS, some of which are; class imbalance, presence of redundant features or features that are weakly correlated and also lack of uniformity in labeling the attacks under different independent categories. The risks of failing to systematize the tackling of these issues are that there will be bias in models and model performance reliability will be compromised [9].

The following is unavoidable: the adoption of a reproducible experimental paradigm, thus, becoming founded on publicly available standardized datasets, being harmonized in advance pipelines, and assessed through transparent procedures. Having the explicit assessment framework that brings in the relevant measures of performance, the effective validation measures and the focused visual tools to present the results, perhaps, we would be able to provide the compelling evidence of the effective practice of machine learning algorithms in the intrusion detection. This would allow a reasonable assessment without necessarily exposing the company to a wide-scale implementation of operations, or the need to involve the unidentified engineering structures [10].

1.3 Research Objectives

This work aims at designing and experimenting an intrusion detection system based on machine learning in a repeatable and offline experiment. The key objective is to work out the systematic and sound IDS assessment model relying on the machine learning classification approach and public benchmark data repositories. Besides, the study conducts a comparative study in relation to several supervised classification algorithms that are put under a common preprocessing procedures as well as common validation protocols that conducts to guarantee equity and uniformity of performance evaluation.

The results of the intrusion detection models are also taken into consideration by referring to a clear and sound statistical and security conditional assessment measures, such as precision, recall, F1 - score, ROC - AUC, and false/positive rate. These measures are full and detailed as they give a clear picture of what the model can perform and what it can detect.

1.4 Research Contributions

The contributions that ought to be anticipated as a result of this inquiry are the fabrication of an Intrusion Detection System (IDS) primitive investigation platform where it is planned to be an offline dataset investigation platform. This design not only obviates the requirement to deploy in vivo network subjects but is also not in any way meddled with the methodological integrity and it is an asset to provide researchers with a dependable and reproducing testing platform. At the same time, the paper contains a structured comparative analysis of multiple machine learning classifiers, all of which are conducted based on unified preprocessing conditions, a unified experimental design and transparent validation models, i.e., the findings of the resulting analysis is impartial and repeatable.

In addition to the empirical achievements of the work, one of the approaches to reporting in form of the work is based on an engaged visualization of the performance and on a brilliant analysis interpretation. This effort will provide ways of defensibly making comparative judgments between heterogeneous machine learning models and also will assist in making valid decisions in both educational and in real life in the intrusion detection domain.

2. Literature Review

2.1 Conventional Intrusion Detection Systems

The Intrusion Detection Systems (IDS) as the term is typically used in the academic discussion is provided as security systems which are able to document the actions of either the host or network and the meaning of their programs is to detect the occurrence of phenomenon that can be quantified as the malicious action or breach of some security provisions. There are traditionally two types of IDS such as host-based IDS (HIDS) and network-based IDS (NIDS). Regarding the detection methodology of working force these systems traditionally therefore either apply signature-based or an anomaly-based paradigm. The classification function remains an essential framework within the framework of a study in the area of cybersecurity since it is the harmonious achievement of both the ability to detect known aggression accurately and, at the same time, unknown aggression incursion at high efficiency [11][12].

The fact that signature-based IDS has been using pre-defined patterns, known attack signatures has made it achieve high percentages of accuracy in the past. However, they work well inside a range of zero-day attacks or polymorphic malware, shooting in which threats do not have prior signatures. Moreover, keeping large signature repositories is an ongoing process of updates and resettlements hence an ongoing contribution to operational complexity and additional constraints upon operational agility in high-speed changing configuration across network medians [13]. Rule based detection systems are founded on explicitly held conditions and thresholds founded on expert acumen. Although these systems in general are both interpretable and computationally tractable, they have limitations to scalability when compared to modern and high volume network traffic. The opponent may also evade notice by simply changing the behavior in ways that the accumulation of their actions do not exceed certain fixed limits, which decreases the capability of the system to categorize the events as a covert attack or an attack that is occurring over a longer orientation. Besides this, increasing the use of encrypted transmissions, dynamic network conditions also serve as another condition to erode reliability of the fixed rule - based approaches [14].

Anomaly-based IDS were created to overcome these shortcomings by modeling the normal behavior of a system and raising the flag as possible intrusions of the typical behavior of a system. Even though this method would simplify the task of searching the never aids among more attacks, initial deployments of such have a high false-positivity rate of detection, particularly in cases where legitimate behavior changes are chronologically induced. The modern-day IDS assessment methods have therefore since moved towards utilizing composite evaluation measures rather than unified of accuracy to give more of a thorough evaluation of the capability to detect attack as well as the trustworthiness of the entire identification system [15].

2.2 Application of Machine Learning in Intrusion Detection Systems

The overall principles underlying machine learning-based intrusion detection system are models of learning salient pattern on the raw data instead of mere matching with existing signatures. In practice, the detection problem has now been posed as a binary classification problem in which the instances of the observed network traffic are the classification problems with the instances falling in one of the benign or malicious class. The importance of supervised learning methods in the research domain is significant in the ease of obtaining full-blown benchmarking data and the chlorine of added reproducibility through experimentation and the fact that it could readily compare the performance of algorithms in controlled conditions.

Some of the most common models that have been used amongst the family of the traditional machine learning methods are logistic regression, decision trees, random forests, support vectors machines, and k nearest neighbors. Such algorithms continue to do exceptionally well as a service of an advantageous trade-off among its understandability and computational expense. Specifically, ensemble techniques - whereby random forests have one - are more likely to provide more accurate prediction power because they encode highly non-linear interactions among features and at the same time make an algorithm less vulnerable to over-fitting. Nevertheless, it has always been emphasized throughout the literature that the efficacy of an algorithm is extremely conditioned by the nature of the data under analysis and the processing pipelines that are employed.

During the last several years, the intrusion detection community has been increasingly interested in using deep learning architectures. Convolutional and recurrent neural networks, and fully connected artificial neural networks have therefore the capability of automatically uncovering layered feature hierarchies and dynamical time models that are inherent in network traffic streams. Although these models may yield state-of-the-art results, they are bound to introduce certain computation overheads, reduce model transparency and become more sensitive to hyper-parameter selection - all of which creates a difficult proposition when it comes to deploying such models, unless already established usage frameworks exist.

Lastly, clustering, density-based anomaly detectors and reconstruction-based autoencoders, which are machine learning-based anomaly detectors, are applied to play a crucial role in instances where attack data is scarce and labeled. In this case, however, empirical studies show that these methods tend to incur a decline in performance in environments whose environment is characterized by a large amount of benign-traffic variability. This fact underscores the significance of the need to have well-designed validation procedures constructed in a manner that will ensure dependability in detection to operate reliably with varying operating conditions.

Data preprocessing is also another factor that has a significant role in the performance of the IDS. Big influence on the classification results is on feature normalization, categorical encoding, dimensionality reduction and feature selection. As now it is revealed that preprocessing choices can exhibit the same effect as that of algorithm choice itself and hence it is significant to have standard preprocessing to gain a just comparison [16][17].

Consequently, some of the recent assessments of the IDS systems have been founded on dataset based IDS evaluation frameworks where evaluations are founded on benchmark datasets that are publicly available as opposed to real-life deployment settings. This form of methodology contributes to the reproducibility of the experimentation and reduces the infrastructure required to conduct the research process of IDS to the advantage of accessibility in a manner that does not compromise on the aspect of scientific rigor [18][19].

Table 1. Summary of Related Works on Machine Learning-Based Intrusion Detection Systems

Study	Dataset / Scope	Methods / Techniques	Evaluation Focus	Main Findings	Relevance to Current Study
Buczak and Guven (2016)	Cybersecurity intrusion detection studies	Data mining and machine learning methods	Detection accuracy and applicability	Provided a broad survey of ML and data mining techniques for cybersecurity intrusion detection	Supports the use of ML as an effective approach for IDS development
Mishra et al. (2019)	IDS research and security datasets	Machine learning-based IDS techniques	IDS performance, challenges, and deployment issues	Showed that ML improves intrusion detection but requires careful preprocessing and validation	Justifies the need for a controlled and reproducible evaluation framework
Liu and Lang (2019)	IDS studies using ML and deep learning	ML and DL models	Detection capability and model limitations	Found that both ML and DL are useful for IDS, but model complexity and dataset quality affect performance	Supports comparing classical ML models under unified conditions
Kocher and Kumar (2021)	Recent IDS developments	Machine learning and deep learning models	Accuracy, scalability, and challenges	Highlighted recent progress and persistent challenges in IDS, especially in model selection and evaluation	Reinforces the importance of comparing multiple algorithms
Thakkar and Lohiya (2020)	Intrusion detection datasets	Dataset review	Dataset quality, imbalance, and redundancy	Reported that IDS datasets differ in structure, attack types, and reliability	Supports the selection of a benchmark dataset and standardized preprocessing
Moustafa and Slay (2015, 2016)	UNSW-NB15	Network anomaly detection evaluation	Dataset realism and attack classification	Introduced and evaluated UNSW-NB15 as a modern IDS dataset	Provides context for benchmark-based IDS evaluation

Sharafaldin et al. (2018)	CICIDS2017	Dataset generation and traffic simulation	Modern attack scenarios	Proposed CICIDS2017 to represent realistic network traffic and contemporary attacks	Shows the importance of dataset selection in IDS studies
Widodo et al. (2024)	Multi-class imbalanced IDS dataset	Machine learning with SMOTE	Class imbalance and detection performance	Demonstrated that balancing techniques improve IDS performance on imbalanced data	Supports the use of class balancing in the current preprocessing pipeline
Sarhan et al. (2024)	ML-based IDS datasets	Feature extraction techniques	Feature quality and classification performance	Emphasized that feature extraction strongly influences IDS accuracy and robustness	Supports the importance of preprocessing and feature engineering
Shanmugam et al. (2025)	IDS datasets with class imbalance	Imbalance-handling methods	Recall, precision, and false alarms	Found that class imbalance affects detection reliability and must be addressed	Supports using recall, F1-score, and false positive rate in evaluation

2.3 Research Gap

Though abundant literature exists on the machine learning based intrusion detection systems, a lot of unanswered questions remain. First of them is the absence of a common approach to methodology; researchers tend to employ varying data collection sets, unequal assessment plans and a variety of other pre-processing tools. As a result, one cannot have fair and material comparisons of outcomes, thus obscuring the process of defining the truly superlative models.

The collections such as UNSW - NB15 and CICIDS2017 were meant to simulate the actual traffic of a network and the contemporary cases of cyber attack. Nevertheless, every dataset differs with respect to the contents of the features and how the attacks are defined and this can significantly affect the performance of the classifiers. Therefore application of common experimental frame is urgently required in order to compare different models in a fair and reliable manner [20][21].

Class imbalance is another crucial IDS study problem. The actual network traffic contains the nominal network instances typically by far many more than the instances of an attack, and certain forms of attacks are disproportionately underrepresented. Unless this type of imbalance is dealt with, there is a risk that models can be thought to have deceptive overall accuracy even though they are unable to recognize the minority classes of attacks. Based on this, recent evidence suggests that imbalance conscious metrics of evaluation and resampling, including SMOTE, are required in an attempt to augment the impartiality and dependability of detection [22].

Moreover, most of the available research studies merely compare and discuss a small number of algorithms, or fails to provide all the required performance statements. With different experimental settings and inexperienced evaluation criteria, it is then not necessarily straightforward to establish whether the monitored performance increments are a product of an inherent excellence of an algorithm or just a consequence of the peculiarities of a specific dataset. Therefore, the new research focuses on the need to have a series of reproducible evaluation frameworks that allow to the comparison of several algorithms with a consistent series of preprocessing and conditions of validation, including exhaustive performance measures.

These challenges collectively contribute to the motivation for this study, which proposes a reproducible, dataset orienting implementation of an IDS evaluation framework aimed at easing the task of making rigorous and equitable comparisons between machine learning models, without the need for elaborate system deployment and specialized infrastructure.

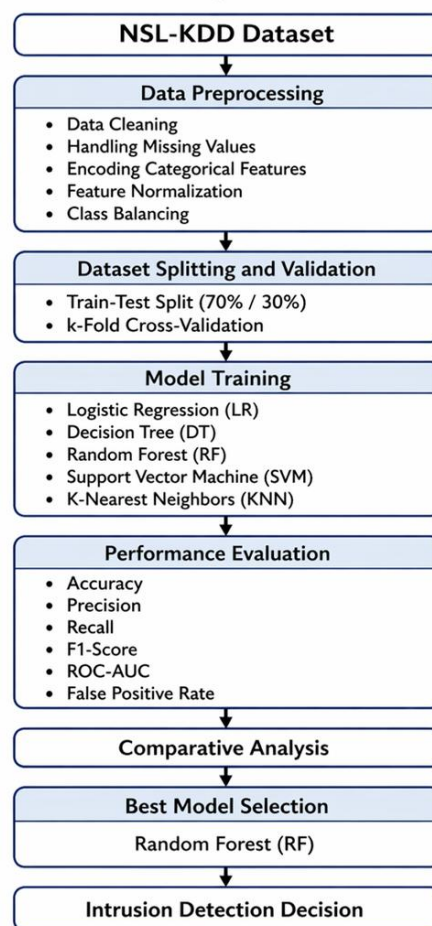
3. Materials and Methods

3.1 Research Design

In this paper we are using an experimental comparative approach to compare the performance of a machine learning based Intrusion Detection System (IDS) under controlled and reproducible conditions. The offline investigation is performed using a publicly available data set of benchmark problems and is pursued in accordance with the fixed experimental procedure to generate the produced findings that are replicable and the comparisons of the models are

fair. Instead of putting an IDS into a operational network, in this instance, the concept of implementation actually entails a systematic manner in which an entire pipeline of experimental IDS can be built, consisting of the following elements: (i) dataset preparation, (ii) preprocessing and features selection that is consistent, (iii) training of a set of machine learning classifiers, and (iv) evaluation using standardized cybersecurity metrics.

The rationale of the comparative design is premised on the fact that the findings of IDS can vary significantly with respect to the learning algorithm, the feature set selected, the type of strategy used to validate, and other variations in the methodological research. This means that each of the selected models is subject to the identical training procedures and testing plans, thereby guaranteeing that each is subjected to identical data partitions, preprocessing procedures and assessment indicators. This standardization is in conformity with the justice and validity generalizations in all the comparisons, etc., such that any performance differences that are determined can be ascribed to models and not to the differences of configuration in different experiments. Furthermore, the paper will not remain on the total accuracy and the false-positive rates and the capabilities of the system to identify attacks will also be considered, which is a highly significant concern in the environment of operating intrusion detection.



3.2 Dataset Selection

The use of this corpus was driven by the research desire to have labeled instances of traffic does not only capture normal traffic behavior but also span through the differently classified types of attacks, therefore doing away with the redundancy that plagued past KDD-type datasets. The data set is intrinsically ideal for both binary and multi-class intrusion detection problems. In the current study we limit our consideration to the binary case (Normal vs. Attack) in order to realize a uniform and meaningful comparison between different classical machine learning models, while still maintaining the diversity of attacks in the training and evaluation stage.

The corpus consists of both numerical and categorical attributes to describe connections of networks, such as protocol type, service, and flag status, and statistical attributes designed from traffic behavior. These features make the data set very suitable to train classifiers that are able to identify suspicious patterns without the need of real-time packet capture or active deployment of network.

Table 2. Characteristics of the NSL-KDD Dataset

Dataset	Records	Features	Attack Types	Normal Samples
NSL-KDD	125,973	41	Multiple categories (e.g., DoS, Probe, R2L, U2R)	67,343

3.3 Data Preprocessing

The preprocessing stage is used to ensure that the models are trained on inputs that are clean, consistent and similar. This step is of considerable significance since the datasets usually used for intrusion detection systems will often have different types of features, noisy attributes and unbalanced distribution of the classes, thus introducing bias into the learning process.

(1) Data cleaning

During the preprocessing stage, duplicate records, if any, are deleted and any inconsistent records are corrected. The dataset is then run through a rigorous process of checks and balances which removes logically invalid values (e.g. negative counts where the value does not make sense) and corrects any formatting anomalies.

(2) Handling missing values

Whilst the NSL-KDD dataset is mainly constituted by rigorously structured attributes, the computational pipeline depicts a systematic way of coping with missing data, thus ensuring both a level of robustness and a level of reproducibility. More specifically, any missing values are imputed with statistically justified defaults (the median for continuous values and the modal value for categorical values) so that problems in inadvertently excluding sizable data subsets are avoided.

(3) Encoding categorical variables

Network intrusion detection Transmission of such categorical information as the type of protocol, service, and flag into a quantitative form cannot even be ingested by analysis algorithms before this. This is commonly converted by a combination of many different encoding schemes starting as low as simple one hot coding to ordinal mapping techniques commonly involved in machine learning based classification piping.. By making use of such encoding methods, one would both preserve the semantic richness of the protocol and service descriptors on the one side, while still being able to meet the numerical input requirements of the subsequently employed learning algorithms.

(4) Feature normalization

In order to ensure greater stability of the model, numerical features are normalized in order to minimize scale variance, this is very beneficial especially in deploying distance-or margin-based classifiers like KNN/SVM. Normalization ensures that attributes with different scales add fairly equally to the learning process of the model.

(5) Dataset balancing (sampling techniques)

Intrusion detection datasets always span significant class imbalances, both in the way normal and malicious flow traffic are divided and in the range of different attack modes. To prevent a preference to the majority class and increase the system's sensitivity towards less instances of malicious patterns, such disproportions are usually handled during the model training phase using judicious sampling schemes. These techniques (whether under sampling the dominating class, oversampling the minority instances or using more complex and stratified resampling) play a decisive role in the improvement of model equity, recall rates for adversarial events and the maintenance of calibration of the false positive rate through rigorous tuning of the validation protocol..

3.4 Machine Learning Algorithms

To achieve accessibility as well as reproducibility with minimal implementation complexity, this study focuses on the usage of common supervised machine learning classifiers with easily accessible training and evaluation facilities of

standard analytical tools. Logistic Regression (LR) is a computationally efficient linear baseline method, especially adapted to high dimensional feature spaces, and is appreciated because of its interpretability through easy model coefficient interpretation.

The Decision Tree (DT) algorithm is also made to be part, as it gives high interpretability and the ability to model non-objective decision surfaces using a hierarchical, rule-based structure. This makes it particularly useful for elucidation of classification behavior. Moreover, the Random Forest (RF) algorithm, which is an ensemble method based on multiple decision trees, is applied to benefit from the ensemble learning to boost the generalization performance, make the model more robust and stable, i.e. to reduce the variance and lessen the overfitting problem.

Besides the Support Vector Machine (SVM), it is a highly competitive margin-based classifier that is also known to be very effective in handling the difficult class separation in high dimensional feature space. Lastly, one includes K - Nearest Neighbors (KNN) algorithm, a simple, distance-based algorithm, which may be used as benchmark indicating that similarity-based algorithms can be used in classification as well as in comparing their performance with more sophisticated learning algorithms.

Table 3. Comparative Overview of Machine Learning Models for Intrusion Detection

Algorithm	Learning Paradigm	Key Advantages	Functional Role in Study
Logistic Regression	Linear Classification	Computational efficiency and simplicity	Serves as a baseline model for performance comparison
Decision Tree	Rule-Based Learning	High interpretability and decision transparency	Utilized for extracting classification patterns
Random Forest	Ensemble Learning	High robustness and reduced overfitting	Selected as a primary predictive model
Support Vector Machine (SVM)	Margin-Based Classification	Strong performance in high-dimensional spaces	Applied for complex boundary separation
K-Nearest Neighbors (KNN)	Instance-Based Learning	Simplicity and ease of implementation	Used as a benchmark for distance-based classification

3.5 Performance Evaluation Metrics

The reason why comparative design was employed in the study is also paramount because comparative research on the outcome of IDS can be vulnerable to the distortions created by learning algorithms, feature selection and validation processes. Consequently, the same data divisions, preprocessing pipelines, and evaluation measures were used to train and test all of the selected models using the same data divisions.

These standardization make the comparative evaluation more reliable - and unbiased, to ensure that the apparent differences in the performance are not due to the models and to every heterogeneity of the experimental design. Moreover, neither is the investigation limited to simple accuracy, nor does it primarily deal with the false alarm rates and the model effectiveness to detect attacks, which validates the need to specify the parameters of the model as an effective intrusion detection system in real life.

Table 4. Definitions and Roles of Evaluation Metrics in Intrusion Detection Systems

Metric	Definition	Role in IDS Evaluation
--------	------------	------------------------

Accuracy	The proportion of correctly classified instances over the total number of observations	Measures the overall effectiveness of the model
Precision	The ratio of correctly predicted attack instances to all instances predicted as attacks	Reflects the reliability of attack predictions and reduces false positives
Recall (Detection Rate)	The ratio of correctly detected attack instances to all actual attack instances	Evaluates the model's ability to detect intrusions (minimize false negatives)
F1 Score	The harmonic mean of precision and recall	Provides a balanced measure of model performance under class imbalance

3.6 Validation Strategy

In order to increase the reliability of the model comparison and to reduce the danger of overfitting, this study uses two complementary validation strategies. The main method consists in a standard train-test split, in which 70% of the data is used to train the model and 30% of the data is used to test the model. This separation allows examination of model performance on completely unseen data - therefore providing a realistic indication of model generalization performance.

In addition, k - fold cross - validation is used as a secondary validation method. In this technique the data set is divided into k subsets (folds) which are exclusive and we train the model and check multiple times, every time we pick a different fold as the test set and rest of the data set is used for training. The final performance is calculated by taking an average of evaluation metrics of the all the folds which reduces the effect of any particular data split and increases the robustness of the result.

Furthermore, the hyperparameter tuning is performed systematically within predefined and reasonable ranges. The models are all tested in the same validation framework and therefore the comparison is fair, balanced, and uniform.

4. Results

This part outlines the results of a thoroughgoing comparative assessment of the models, using the metrics as they are stipulated above. The findings are lit up in regard to: detection capability(i), false alarms(ii), and overall discriminative proficiency(iii).

Table 5. Model Performance Results

Model	Accuracy	Precision	Recall	F1	ROC-AUC
LR	0.915	0.91	0.892	0.901	0.945
DT	0.938	0.934	0.921	0.927	0.958
RF	0.972	0.976	0.969	0.972	0.991
SVM	0.964	0.968	0.952	0.96	0.987
KNN	0.952	0.956	0.944	0.95	0.978

To examine in practice the proposed models, factors that went beyond the traditional performance measures were put into consideration. This must be mandatory since an effective system of intrusion detection should not just be capable of ensuring a high detection rate, but in practice the implementation has to be viable too. The obtained results indicate that the Random Forest model has offered the most optimal overall, along with a significantly low false positive rate which render it a very excellent candidate in the implementation of intrusion detection systems.

The Support Vector machine (SVM) was also performing well particularly in accuracy and ROC-AUC. However, it has relatively large computational specifications which may pose a challenge to its scalability to large systems. The results of the model were not very stable as in the case of Random Forest, but the Decision Tree model spanned up very much interpretable, easily understandable, and satisfactory results. Faster and smoother than before Toads lapsed by Logistic Regression, On intricate or non-lineage attack airs. At the same time, the K-Nearest Neighbors

(KNN) algorithm also had reasonable accuracy performance but was bounded by extremely large prediction time that can be explained by the extremely high cost of evaluating the distances, particularly when used on large data sets and is used in operation near real time.

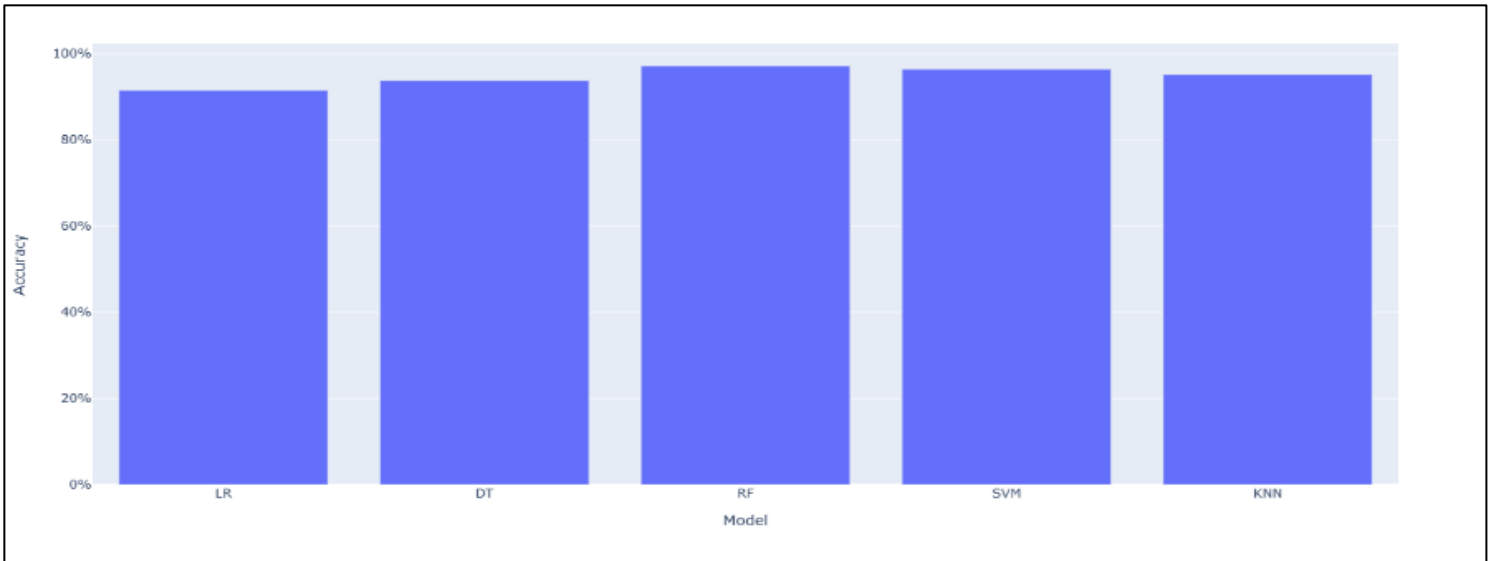


Figure 1. Comparative Analysis of Classification Accuracy across Models

A comparative test has been conducted as depicted in Figure 1 on machine learning models to evaluate the capability along with the capacity to correctly differentiate between the normal and malicious network traffic. The findings indicate contrasting glaring disparities on the properly predicted or cannot be predicted encroachments through algorithms, hence supporting the significance of model architecture on the effect of penetration on the efficiency of intrusion detectors.

The Random Forest denoted RF classifier was found to be the most accurate with 97.2% accuracy beating all other algorithms that were evaluated. This better performance can be explained by the assistance of the ensemble setup in which the accumulation of so many decision trees helps to decrease variance and generalization.

The Support Vector Machine (SVM) came in close behind with an accuracy of 96.4 per cent. This result is corroborative for the capacity of SVMs for complex decision boundary that is present in a high-dimensional feature space.

The other algorithms, K- Nearest Neighbors (KNN) was also able to provide competitive results with an accuracy of 95.2 per cent. This suggests that similarity-based classification is effective in its ability to capture traffic patterns, though it is an approach which relies on distance computations, which may result in some scalability constraints in a larger corpus.

The Decision Tree classifier (DT) achieved a moderate accuracy of 93.8 per cent depicting a moderate learning capacity but less robustness compared to ensemble techniques. Logistic Regression (LR) used as baseline linear model showed lowest accuracy of 91.5 per cent which depicts the ineffectiveness of linear boundary for representing complex non linear intrusion patterns present in a real-world scenario. Collectively, the insights revealed by Figure 1 reaffirm the fact that ensemble approaches and margin-based classifiers do trump simpler linear models in the domain of intrusion detection, and yield more accurate and reliable classification outcomes.

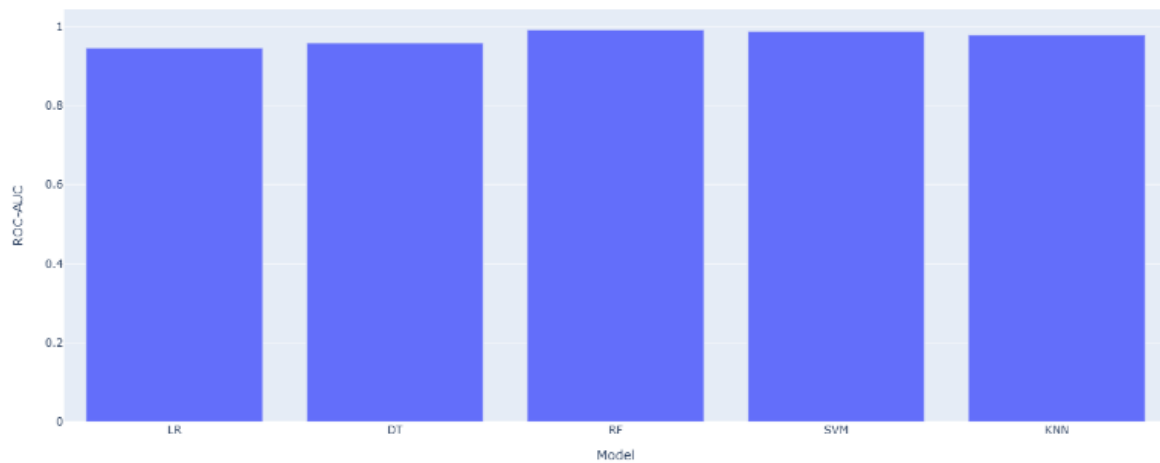


Figure 2. Comparison of ROC–AUC Performance Across Models

Figure 2 illustrates the comparison of classifiers in terms of Receiver Operating Characteristic- Area under the Curve (ROC-AUC) parameter. The ROC - AUC evaluates the discriminative power of a model on both normal and malicious traffic in all of the dynamics of any two classification thresholds; it is an accurate metric of a model discrimination.

The Random Forest algorithm possesses the highest ROC-AUC index of 0.991 which means impressive level of proficiency in differentiating between benign and malicious cases. This outcome is good ranking behavior and consistency in predictive behavior irrespective of the level of threshold employed hence depicts its robustness to varying ways of functioning. The Support Vector Machine (SVM) also exhibits a substantial degree of discriminative strength with a ROCAUC of 0.987, which therefore verifies its capability to attract an explicit decision divide between the two categories. The K-Nearest Neighbors (KNN) and Decision Tree (DT) classifiers have ROC-AUC values of 0.978 and 0.958 respectively signifying a satisfactory performance and comparing least stability when compared to the top performing models. Logistic Regression (LR) has registered the lowest ROC- AUC score at 0.945 which is consistent with LR having known shortcomings of modeling complex, non-linear relationships embedded in the data set..

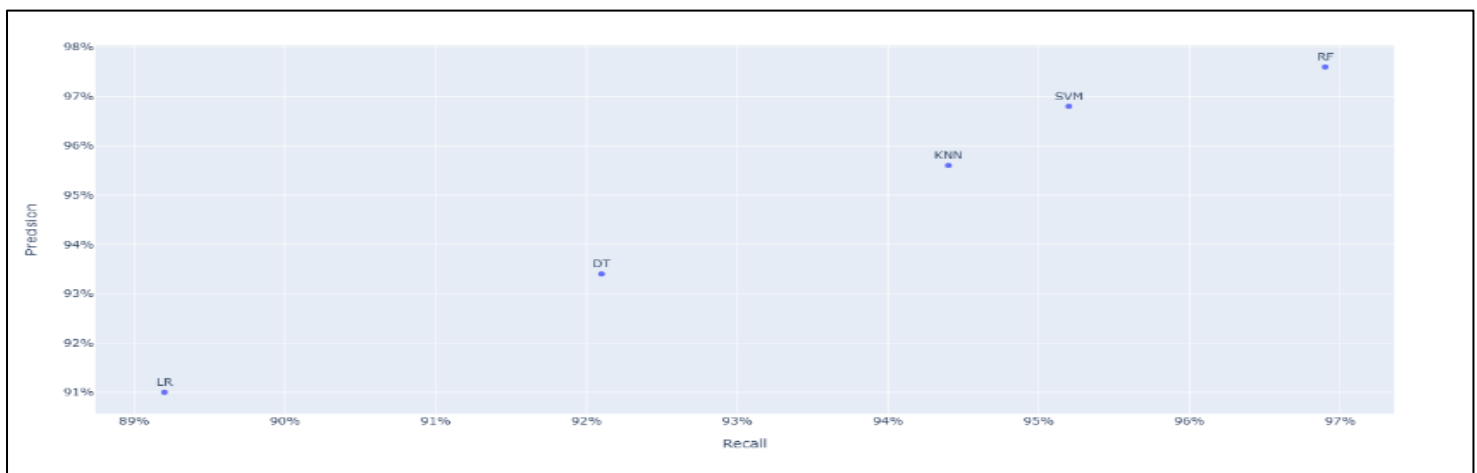


Figure 3. Precision–Recall Performance Comparison Among Algorithms

Figure represent what was shown in Figure 3, it can give some insight on the balance between effective attack detection and false alarm minimization: the precision- recall setting for all the evaluated models. Precision is related to the reliability of the predicted attack instances and recall (the detection rate) is the percentage of the actual attacks that are correctly found by the model.

Among the models that have been evaluated, the model Random Forest shows the most balanced performance with high precision and high recall (0.976 and 0.969 respectively). This means that the model is able to detect most attack instances while keeping a small proportion of false positives. The Support Vector Machine (SVM) also has a good performance, and a similar precision but a bit less recall, confirming that it is an effective algorithm to reduce classification uncertainty.

The K-Nearest Neighbors (KNN) model has a fairly balanced PR relationship, though it's performance is slightly below the other two, Random Forest and SVM; implying that there may be some sensitivity to the local distribution of the data. The Decision Tree model is balanced whereas the Logistic Regression has a lower recall rate which means some intrusion instances may go undetected.

Overall, the figure highlights one of the important principles of intrusion detection systems: high recall without serious loss of precision is the key to preventing the two extremes of loss of a missed attack and a large number of false alarms. In this context, ensemble based models seem to provide the best possible trade off..



Figure 4. Analysis of False Positive Rate versus Detection Rate

Figure 4 is a close expression of the relation between the False Positive Rate (FPR) and Detection Rate (Recall) and brings into the perspective an important aspect of the work of the modern intrusion detection systems (IDS). The effectiveness of an IDS is featured by how many malicious phenomena are detected and alleviated at the same time, thus the former can and in fact has negative consequences to the cost of the security analysts and the latter can negatively affect its effectiveness.

Random Forest model is depicted to be optimal equilibrium in comparative paradigm with exemplary detection rate of 96.9 per cent and minimal FPR value of 2.1. This is performance that testifies to the sound knack that this model has in not only differentiating between malicious and benign traffic but also intent to parameter against legitimate activity. Support Vector Machine (SVM) has also demonstrated good results with high detection rate and significant FPR of 2.8% and low rates of detection but high specificity of 2.8%.

Instead, the algorithms K-Nearest Neighbors (KNN) and Decision Tree (DT) have moderate false positive and, consequently, show a bigger propensity to misdecide the fair traffic as a malicious one. The FPR of Logistic Regression

(LR) is highest with the highest FPR of 6.6 percent to 6.8 percent and the least detection efficacy when combined with the most complex intrusion pattern of the models we tested resulting in a weak capacity to deal with complex intrusion patterns.

Combined with the figure, the point is that it is not possible to measure the performance of IDS accurately in a vacuum. Most efficient models can be found in whelmed single points under which the models achieving high level detects beget and the same time get reduced false alarms to gain greater operation effectiveness and capturing into the term world.

5. Discussion

This study supports the hypothesis that the selection of machine learning classifier can impact the effectiveness of intrusion detection systems. Random Forest showed the highest accuracy with 97.2%, precision of 97.6%, recall of 96.9%, F1-score of 97.2%, ROC-AUC of 0.991 and the lowest false positive rate (2.1%). This study suggests that ensemble classifiers are the best choice for intrusion detection to achieve higher efficiency because they use a collection of base learners and they overcome the instability that may arise from a single classifier like Decision Tree or Logistic Regression.

This observation is in line with that of Ajagbe et al. [23], who experimented with various machine learning algorithms, such as Linear SVC, Logistic Regression, Random Forest, Decision Tree, and XGBoost, for UNSW-NB15 with an unbalanced data distribution. They found that Random Forest and XGBoost performed better than the other models, confirming that ensemble and tree-based classifiers are more suitable to model nonlinear distributions of intrusion patterns than linear models. But while their work has mainly dealt with UNSW-NB15 and an unbalanced dataset setting, this study has used NSL-KDD with a standardised data pre-processing and validation approach. So, the present findings can be employed to exhibit the capabilities of traditional supervised learning algorithms, under a controlled and consistent process.

The results of present study, showing the best performance achieved by Random Forest, are also in line with Hamidou and Mehdi [24], who benchmarked Random Forest, XGBoost and Deep Neural Network using NSL-KDD. They found that Random Forest achieved very good performance with 99.80% accuracy, 0.9988 AUC after using SMOTE and hyperparameter tuning. In the present study, the accuracy is lower than that reported in the literature; however, this is because their method included further optimisation (i.e. SMOTE and Optuna for the hyperparameter optimisation). Thus, the present study offers a simpler and more comprehensive comparison of the systematic supervised classifiers, whereas their study added to the demonstration of the use of advanced optimization on Random Forest.

The same can be assumed from the study of Waghmode et al. [25], who developed a model for intrusion detection using Least Square Support Vector Machine with exhaustive feature selection. Their model achieved 99.3% accuracy on NSL-KDD, 99.5% on CIC-IDS2017, and 93.3% on UNSW-NB15. This shows that feature selection and margin-based optimized learning can result in excellent IDSs. But our research is different from theirs, in that it has compared several common classifiers rather than just an optimized model. Thus, the current study is significant in that it offers a more comparative perspective and that Random Forest performs well without an intricate feature selection method.

The findings are also consistent with Maodah et al. [26], who assumed the use of voting and stacking ensemble classifiers on cloud-based intrusion detection using NSL-KDD data. Their results demonstrate that ensemble methods are superior to their constituent classifiers, and achieve close-to-perfect recognition rates in terms of accuracy, precision, recall and F1-score. This reinforces the view that taking multiple learning decisions could lead to more robust IDS models. Their work and ours differ in the sense that this current study does not stack or vote ensembles; it compares the performance of Random Forest as a typical ensemble model to Logistic Regression, Decision Tree, SVM, and KNN. In this sense, the present results can be considered as baseline results for future work that may propose the extension of the framework to hybrid or stacked ensemble models.

A second insight from the comparison with other recent studies is that more variables besides accuracy should be considered in models for intrusion detection. In intrusion detection systems, high accuracy does not necessarily provide the right model when it is not accompanied by a low false positive rate and high recall. For this reason, we used recall, F1-score, ROC-AUC, and false positive rate supplementing accuracy in the current study. Random Forest's high recall and low false positive rate indicate that it can detect attacks with minimal false positives. This is in line with recent IDS research, which moved towards more balanced metrics in addition to relying solely on accuracy.

In conclusion, comparing with recent research demonstrates that the results of this study are in line with the trend of recent intrusion detection research. Indeed, there is evidence that ensemble-based classifiers, modified SVM variants, feature selection, class balancing, and hybrid ensemble-based classifiers can improve IDS accuracy. In this regard, the present research sheds some light in this direction by providing a clear, repeatable, and controlled comparison of

standard supervised learning algorithms for the NSL-KDD dataset. The findings show that Random Forest offers the best overall combination of accuracy, recall, ROC-AUC and false positive rate, and is thus the best choice among the tested algorithms for the proposed IDS.

6. Conclusion

The current study proposed an approach for intrusion detection using supervised machine learning algorithms with the NSL-KDD dataset. The study had the objective of developing and testing a consistent machine learning-based IDS system to classify network traffic as benign or malicious using a controlled experimental setup. In order to realise this aim, this research adopted a uniform preprocessing process, which consisted of data cleaning, dealing with missing values, encoding dichotomous features, normalizing features, and balancing class. Five supervised machine learning models - Logistic Regression, Decision Tree, Random Forest, Support Vector Machine and K-Nearest Neighbors - were applied and evaluated by using the pre-processed dataset.

The results of this study show that machine learning approaches can enhance the ability to identify intrusions compared with conventional fixed-signature-based methods. Random Forest exhibited the best performance across all the models, with an accuracy of 97.2%, precision of 97.6%, recall of 96.9%, F1-score of 97.2%, ROC-AUC of 0.991 and the lowest false positive rate (2.1%). The findings suggest that Random Forest strikes a good balance between detection and low false positive rate, which is critical to the design and development of real-world intrusion detection systems.

Our findings also demonstrated that lightweight models (such as Logistic Regression and Decision Tree) were fast to train and easily interpretable, but they were not as successful in capturing complex intrusion patterns. Support Vector Machine and the K-Nearest Neighbors method also performed well, but their applicability may prove to be costly and complex in large and/or near-real-time detection systems. Thus, Random Forest is a favourable choice, among all tested algorithms, in the proposed framework.

The novelty of this paper is the controlled and reproducible nature of the experiments. The use of identical preprocessing steps, validation techniques and evaluation methods across all models allows for a valid comparison of supervised machine learning classifiers for intrusion detection. Furthermore, consideration of different performance evaluation metrics, such as accuracy, precision, recall, F1-score, ROC-AUC, and false positive rate, adds to the robustness of the evaluation, rather than being solely based on accuracy.

But the study is constrained by the use of the NSL-KDD dataset, which while popular, may not cover all aspects of current network traffic and attack techniques. This framework can also be extended by testing more recent data sets like CICIDS2017, UNSW-NB15 or real network traffic data. Future work might also explore deep learning, hybrid ensemble models, different feature selection strategies and real-time applications.

References:

- [1] World Economic Forum. (2025). Global Cybersecurity Outlook 2025 (in collaboration with Accenture). World Economic Forum.
- [2] European Union Agency for Cybersecurity (ENISA). (2024). ENISA Threat Landscape 2024. ENISA.
- [3] Verizon. (2024). 2024 Data Breach Investigations Report (DBIR). Verizon.
- [4] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), 1153–1176. doi:10.1109/COMST.2015.2494502
- [5] Mishra, P., Varadharajan, V., Tupakula, U. K., & Pilli, E. S. (2019). A detailed investigation and analysis of using machine learning techniques for intrusion detection. *IEEE Communications Surveys & Tutorials*, 21(1), 686–728. doi:10.1109/COMST.2018.2847722
- [6] Liu, H., & Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences*, 9(20), 4396. doi:10.3390/app9204396
- [7] Kocher, G., & Kumar, G. (2021). Machine learning and deep learning methods for intrusion detection systems: Recent developments and challenges. *Soft Computing*, 25, 9731–9763. doi:10.1007/s00500-021-05893-0
- [8] Pinto, A., Herrera, L.-C., Donoso, Y., & Gutierrez, J. A. (2023). Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure. *Sensors*, 23(5), 2415. doi:10.3390/s23052415
- [9] Thakkar, A., & Lohiya, R. (2020). A review of the advancement in intrusion detection datasets. *Procedia Computer Science*, 167, 636–645. doi:10.1016/j.procs.2020.03.330
- [10] Widodo, A. O., Setiawan, V., & Indraswari, R. (2024). Machine learning-based intrusion detection on multi-class imbalanced dataset using SMOTE. *Procedia Computer Science*, 234, 578–583. doi:10.1016/j.procs.2024.03.042
- [11] Liu, H., & Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences*, 9(20), 4396. <https://doi.org/10.3390/app9204396>
- [12] Mishra, P., Varadharajan, V., Tupakula, U. K., & Pilli, E. S. (2019). A detailed investigation and analysis of using machine learning techniques for intrusion detection. *IEEE Communications Surveys & Tutorials*, 21(1), 686–728. <https://doi.org/10.1109/COMST.2018.2847722>
- [13] Pinto, A., Gonçalves, G., & Silva, S. (2023). Survey on intrusion detection systems based on machine learning techniques. *Sensors*, 23(5), 2415. <https://doi.org/10.3390/s23052415>
- [14] Kocher, G., & Kumar, G. (2021). Machine learning and deep learning methods for intrusion detection systems. *Soft Computing*, 25, 9731–9763. <https://doi.org/10.1007/s00500-021-05893-0>

-
- [15] Hozouri, A., Mirzaei, A., & Effatparvar, M. (2025). Advances in intrusion detection systems using machine learning. *Discover Artificial Intelligence*, 5, 314.
 - [16] Sarhan, M., Layeghy, S., Moustafa, N., Gallagher, M., & Portmann, M. (2024). Feature extraction for ML-based intrusion detection. *Digital Communications and Networks*, 10(1), 205–216.
 - [17] Umar, A., Mahmoud, M. A., et al. (2025). Preprocessing techniques for ML-based IDS. *arXiv preprint arXiv:2507.13314*.
 - [18] Thakkar, A., & Lohiya, R. (2020). Advancement in intrusion detection datasets. *Procedia Computer Science*, 167, 636–645.
 - [19] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset (CICIDS2017). *ICISSP Proceedings*.
 - [20] Moustafa, N., & Slay, J. (2015). UNSW-NB15 dataset for network intrusion detection. *MilCIS Conference Proceedings*.
 - [21] Moustafa, N., & Slay, J. (2016). Evaluation of network anomaly detection systems using UNSW-NB15 dataset. *Information Security Journal*, 25(1–3), 18–31.
 - [22] Shanmugam, V., Razavi-Far, R., & Hallaji, E. (2025). Addressing class imbalance in intrusion detection. *Electronics*, 14(1), 69.
 - [23] Ajagbe, S. A., Awotunde, J. B., & Florez, H. (2024). Intrusion Detection: A Comparison Study of Machine Learning Models Using Unbalanced Dataset. *SN Computer Science*, 5, 1028. <https://doi.org/10.1007/s42979-024-03369-0>
 - [24] Hamidou, S. T., & Mehdi, A. (2025). Enhancing IDS performance through a comparative analysis of Random Forest, XGBoost, and Deep Neural Networks. *Machine Learning with Applications*, Article 100738. <https://doi.org/10.1016/j.mlwa.2025.100738>
 - [25] Waghmode, P., Kanumuri, M., El-Ocla, H., & Boyle, T. (2025). Intrusion detection system based on machine learning using least square support vector machine. *Scientific Reports*, 15, 12066. <https://doi.org/10.1038/s41598-025-95621-7>
 - [26] Maodah, K. A., Alhomdy, S., & Thabit, F. (2025). Detecting intrusions in cloud-based ensembles: evaluating voting and stacking methods with machine learning classifiers. *Frontiers in Computer Science*, 7, 1623375. <https://doi.org/10.3389/fcomp.2025.1623375>