

Article

Not peer-reviewed version

---

# On Feature Selection Techniques for Detecting DoS Attacks with a Multi-class Classifier

---

Iuri A. Mundstock , Yuri Santo , Thiago L. T. da Silveira , [Roger Immich](#) , [André Riker](#) , [Bruno L. Dalmazo](#) \*

Posted Date: 27 March 2024

doi: 10.20944/preprints202403.1635.v1

Keywords: Network Security; Feature Selection; Denial of Service Attacks; Systematic Review; Smart City; Anomaly Detection









Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# On Feature Selection Techniques for Detecting DoS Attacks with a Multi-class Classifier

Iuri A. Mundstock<sup>1</sup> , Yuri Santo<sup>2</sup> , Thiago L. T. da Silveira<sup>3</sup> , Roger Immich<sup>4</sup> ,  
André Riker<sup>2</sup>  and Bruno L. Dalmazo<sup>1,\*</sup> 

<sup>1</sup> Computer Science Center (C3)— Federal University of Rio Grande, Brazil, Rio Grande, 96203-900, Brazil; iurimundstock@furg.br

<sup>2</sup> Institute of Exact and Natural Sciences (ICEN)— Federal University of Pará, Belém, 66075-110, Brazil; yuri.santo@icen.ufpa.br; ariker@ufpa.br

<sup>3</sup> Institute of Informatics — Federal University of Rio Grande do Sul, Porto Alegre, 91501-970, Brazil; tltsilveira@inf.ufrgs.br

<sup>4</sup> Digital Metropolis Institute — Federal University of Rio Grande do Norte, Natal, 59078-900, Brazil; roger@imd.ufrn.br

\* Correspondence: dalmazo@furg.br

**Abstract:** In a hyperconnected society, computer networks play a pivotal role in all activities that permeate our daily lives. In the context of smart cities, these networks play a pivotal role in enhancing urban efficiency and sustainability. The continuous flow of packets in these networks poses security challenges, from personal data leakage to service unavailability. Anomaly-based methods are commonly used to detect network attacks. In this context, Software-Defined Networking (SDN) emerges as a solution for anomaly detection in network traffic, facilitating comprehensive, real-time analyses and dynamic adaptation to changes. This study aims to present a systematic review of feature selection techniques and evaluate the effectiveness of attribute selection with a multi-class classifier in network anomaly detection. The objective of this paper is to inspire future research and identify trade-offs among these techniques for detecting Denial of Service (DoS) attacks. Based on the findings, the analysis of the entire dataset yields superior results in terms of precision (99%), but using the output of OneR, although resulting in a slight loss of precision compared to the complete dataset, presents the highest precision among other studied techniques, indicating a trade-off between precision and processing time efficiency.

**Keywords:** network security; feature selection; denial of service attacks; systematic review; smart city; anomaly detection

## 1. Introduction

Computer networks are ubiquitous in our daily lives, encompassing recreational and professional activities. Especially within the framework of smart cities, interconnected networks assume a central role by contributing to advancing urban efficiency and sustainability [1]. In this context, the web of connected systems and data flows is pivotal in optimizing various aspects of city life, fostering resource efficiency, and sustainable urban development [2]. In the vast majority of cases, this interaction is transparent to users. Still, we are constantly generating a flow of packets between destinations almost uninterruptedly. These packets that are transmitted if they are not secure can cause irreversible problems, ranging from personal data leakage, sharing of passwords, financial transactions, or even the unavailability of often essential services [3].

Two primary methods for detecting network attacks are Signature-based Detection (SD) and Anomaly-based Detection (AD). Signature detection involves scrutinizing network traffic to identify patterns that match predefined characteristics associated with malicious or unwanted activities. While the SD approach is highly effective at detecting known attacks, it can be ineffective against novel attack patterns [4]. In anomaly-based detection, there is no need to have prior knowledge of an attack pattern; instead, a detection system relies on well-defined rules for “normal” network behavior established through training. AD approach’s primary advantage is its ability to detect unknown attacks without prior knowledge, but it tends to generate a higher rate of false alarms as a notable drawback [5].

Numerous techniques are available for analyzing and detecting network anomalies, but they still present significant challenges, including dealing with the high dimensionality of data [6]. Currently,

Software-Defined Network (SDN) is employed for the detection of anomalies in network traffic [7]. Adopting an SDN-oriented approach makes it possible to conduct more comprehensive and real-time traffic analyses [8], as well as to provide data plane nodes that are able to establish secure channels between each other [9]. This enables a quicker response to potential anomalies. The flexibility provided by SDN also facilitates the implementation of advanced intrusion detection algorithms and dynamic adaptation to changes in the network environment, making it a strategic choice to enhance security and efficiency in identifying anomalous behaviors. In addition, SDN-based approaches enable effective management of the continuous generation of extensive data traffic on networks [10].

In this context, feature selection is one technique aimed at enhancing classifier quality, which involves carefully selecting the most relevant attributes available for detecting attacks. The feature selection process can substantially improve attack detection and significantly bolster the overall effectiveness of the security system by providing a rapid and efficient response [11].

The contributions of this work are twofold: (i) to present a systematic review covering the most prominent works in this field of research; and (ii) to assess the performance efficacy of feature selection techniques in anomaly detection within the network by using a dataset derived from real-world data. This investigation aims to gauge the effectiveness of various attribute selection methods concerning the classifier's objectives, specifically identifying techniques that provide time and performance advantages without significantly compromising the efficacy of attack detection. In essence, this study aspires to thoroughly compare attribute selection techniques, contributing to evaluating the most suitable technique for detecting Denial of Service (DoS) attacks.

The remainder of this paper is organized as follows. In Section 2, we present a brief overview of key aspects to understand this work better. In Section 3, we describe the related work, following a methodology used to select the most prominent studies on feature selection techniques for detecting DoS attacks. Section 4 presents an assessment methodology to compare attribute selection techniques. In Section 5, we detail the performance evaluation and discuss the obtained results. Finally, in Section 6, we conclude the paper with remarks.

## 2. Background

This section encompasses various types of existing attacks and Intrusion Detection Systems. Given that the primary objective of this work is review attribute selection techniques to enhance anomaly detection, we will narrow our focus to concepts relevant to attacks that generate anomalies in the network and those applicable to anomaly-based Intrusion Detection Systems.

### 2.1. Attacks

In the literature, some attacks generate traces as an anomaly in the network, each having a different way of affecting the user, whether from a simple denial of service that can cause immeasurable inconvenience or the theft of passwords and personal files. The primary attacks are Brute Force, DoS, Heartbleed, Web Attack, Infiltration, Botnet, and DDoS. In a DoS attack, the malicious node sends the message to the node and consumes the network bandwidth. The main objective of the malicious node is to occupy the network node. If a message from an unauthenticated node arrives, the receiver will not receive it because it is busy, and the initiator has to wait for the receiver's response [10].

In a DDoS attack, the attacker compromises a series of slave systems and installs flood servers. These compromised systems are then coordinated to launch a flood attack by contacting a set of servers and combining their transmission power. The utilization of a large number of slave systems amplifies the attack's potency and introduces complexity in defending against such attacks [12]. Brute Force is one of the most prevalent attacks utilized for password cracking and unveiling concealed pages and content within a web application. Essentially, it is a trial-and-error attack wherein the attacker may invest minutes to days or weeks before successfully penetrating the system [13].

Heartbleed arises from a bug in the OpenSSL cryptography library. When a Heartbeat request is sent, it always contains a token and a specified length. The server is expected to return the information

with a length equal to the requested size. However, a size check failure occurs, allowing the attacker to send more memory content information than intended [14]. The Infiltration attack involves a malicious file that can be delivered through various means, such as email attachments. When the user, i.e., the victim, downloads the file, it may contain different types of viruses. Unbeknownst to the user, malware can monitor the entire network in the background. This can lead to the theft of passwords, personal files, and other sensitive information [15].

Botnets are regarded as one of the most significant threats to network security, serving as a launching point for various illegal practices and attacks. Typically, bots undergo a well-defined life cycle comprising four stages: the exploration stage, rally stage, execution stage, and finally, the update and maintenance stage. The presence of botnets in a network can lead to anomalies, including a sudden surge in traffic volume or unusual activity targeting ports not commonly used [16].

In this study, our focus revolves around detecting of two specific types of DoS attacks:

- **DoS HTTP Unbearable Load King (HULK):** It serves as a web server DDoS tool that primarily intended for research purposes, aiding penetration testers in assessing server efficiency. Security specialists use HULK to identify vulnerabilities in their security measures against DDoS attacks and address them proactively, mitigating potential exploits by malicious actors. The tool generates multiple distinctive requests at irregular intervals from the same host, executing a DDoS attack and attempting to thwart the network's defense mechanisms by avoiding predictable attack patterns [17].
- **DoS GoldenEye:** GoldenEye, akin to an HTTP flood, constitutes a DDoS attack strategically crafted to inundate web servers' resources. It achieves this by incessantly soliciting single or multiple URLs from numerous source-attacking machines. GoldenEye introduces dynamic alterations to the generated requests, employing randomization of user agents, referrers, and nearly all pertinent parameters. To prolong the connection, GoldenEye incorporates a URL suffix, facilitating request bypass through several Content Delivery Network (CDN) systems, commonly known as *No Cache*. As the server reaches its limits of concurrent connections, legitimate requests from other users become unresponsive [18].

## 2.2. Intrusion Detection Systems

In network attacks, two approaches are employed for their detection: Signature Detection (SD) and Anomaly Detection (AD). There are specific characteristics and typical behaviors associated with attacks, and these features can be identified to prevent or respond to potential threats.

Intrusion Detection Systems (IDSs) using SD are usually pre-programmed to compare observed behavior with these predefined characteristics [19]. One of the main advantages of this approach is the low number of false positives since the attackers' behavior is already known, so detection is much faster. Still, it has the disadvantage of knowing in advance how all types of attackers attack the network; that is, it is not possible, for example, to detect new attacks that do not have known signatures.

When identifying intrusion caused by network anomalies, it is only necessary to understand the network's regular behavior. Thus, when abnormal behavior is detected, it is classified as an attack. One key advantage of this approach is its capability to detect new attacks. However, it comes with the drawback of a higher occurrence of false positives, often triggered by simple service unavailability, resulting in anomalies within the network [20].

Considering all the advantages and disadvantages presented regarding IDS types, this work will concentrate on the category of IDS based on AD. It has proven more efficient than SD, particularly in detecting attacks that are not yet known or have not been disclosed (*zero-day attacks*).

## 2.3. Feature Selection

The literature presents several techniques that can be used to improve AD. Among these, we have attribute selection techniques, whose primary purpose is to save time and reduce the effort to

detect attacks without compromising their effectiveness or, in some cases, obtaining improvements concerning the complete set.

Attribute selection is widely used in the current scenario due to the complexity of existing data sets, which is the major problem faced when detecting attacks.

During the literature review, it was noted that some models were more frequently used than others. The following list presents the main approaches to performing feature selection for IDS.

- **Information Gain:** It is one of the most used feature selection techniques for resizing datasets for anomaly detection. This technique consists of, through entropy calculation, ranking the attributes by assigning a “weight” to them, ranging from 0 to 1, depending on the degree of relevance of the attributes. This attribute selection technique is based on [6] filters. Entropy is used to infer the distribution of the characteristics of the features analyzed. Right after calculating the entropy, the gain formula is applied to obtain the weights of features.
- **OneR:** It is a simple attribute classification algorithm [21]. It consists of creating a rule for each attribute in the training data and selecting the attributes with the lowest error. It treats all numerically valued attributes as continuous and uses a direct method to divide the set’s range of values into several disjoint ranges. This is one of the most primitive techniques and produces simple rules based on just one resource [22].
- **PCA:** Principal component analysis (PCA) is a standard technique for most modern data analysis because it is a simple, non-parametric method for extracting relevant information from disorganized data sets. Simply and effectively, it can show how to resize the data set to improve processing. PCA generates new features using the existing features in a dataset, looking for the features that have underlays and are relevant at the same time [23].
- **Symmetrical Uncert:** Symmetric uncertainty is a commonly used technique in attribute selection to resize data sets in anomaly detection. This technique is based on the calculation of entropy and the ranking of attributes, assigning weights that vary from 0 to 1 according to their relevance. After calculating the entropy, we apply the gain formula to obtain the weights of the features, indicating their importance. [24].
- **Correlation Attribute:** The Attribute Correlation technique is used to measure the statistical relationship between pairs of attributes in a data set. It calculates the correlation between attributes to determine whether they are linearly related and to what degree. The correlation coefficient is a measure that varies from -1 to 1, indicating the direction sign (positive or negative) and strength (magnitude) of the relationship between attributes. This technique can be applied to identify highly correlated attributes, which may indicate redundancy or overlapping information. By removing or combining highly correlated attributes, we can reduce the dimensionality of the dataset and improve the efficiency of AD algorithms [25]. This technique has its objectives well defined during its execution, which are the extraction of the most relevant information in a dataset, the reduction of the size of the data set, preserving only the important data for detecting attacks, and finally carrying out a data simplification.

### 3. Literature Review

We carried out a systematic review of the literature to map existing work that explores the use of programmable networks in the context of feature selection applied to IDSs.

#### 3.1. Method of Selection

One of the goals of a systematic review is to be reproducible so the interested researcher can perform precisely the same steps and use the same criteria we employed to include papers in this section (Table 1). Our systematic review focused on covering research papers that addressed feature selection techniques in the context of programmable networks – we thus started with the following search terms:



((intrusion detection system OR ids) AND (software defined network OR sdn) AND feature selection)

**Table 1.** Paper Inclusion (IC) and Exclusion (EC) Criteria.

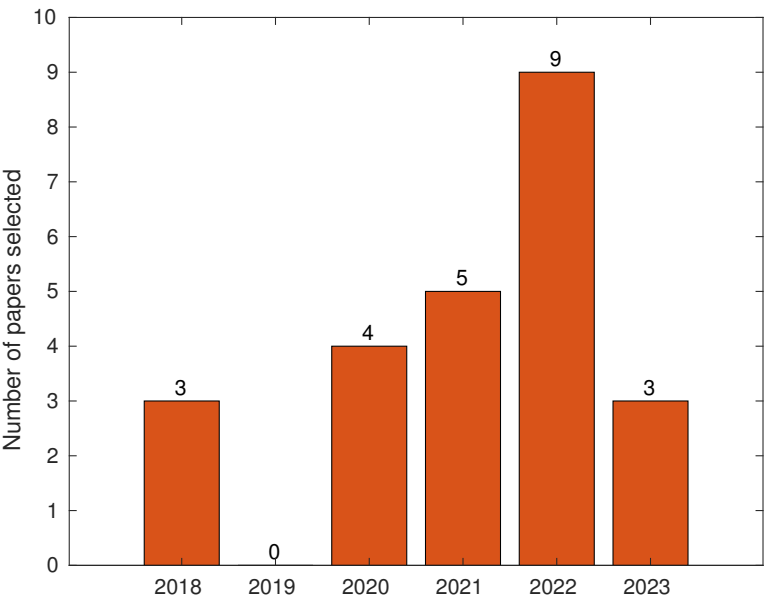
Inclusion Criteria	
IC	Published as a full paper in a conference, magazine or journal.
Exclusion Criteria	
EC	The paper was not published between 2018 and 2023.

We systematically queried the IEEE Xplore database systematically throughout 2023, with our last query performed on December 02, 2023, to retrieve the most recent works. In summary, we retrieved 24 distinct papers from this database, as can be observed in Table 2.

**Table 2.** Selection of studies from IEEE Xplore.

Results	Total
IEEE Xplore Search	31
Excluded by EC	5
Not satisfied IC	1
<b>Selected papers</b>	<b>24</b>

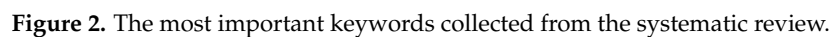
Figure 1 depicts the number of works selected per year of the interval. It is possible to notice 2022 as the year when most works were found.



**Figure 1.** Number of papers selected per year.

3.2. Discussion and Open Issues

The word cloud presented in Figure 2 summarizes the central theme in 24 review articles that explore specialized approaches to enhance network security through feature extraction. The prominence of keywords such as “software”, “learning”, “detection”, “intrusion”, and “network” highlights



AlMasri et al. [27] propose a combination of Machine Learning (ML) and network programmability to protect networks, specifically against DoS and Port Scanning (Probe) attacks. The proposed ML algorithm was constructed using Anova for feature selection. Four ML models were tested with the features selected by the Anova FS technique, utilizing the NSL-KDD training dataset, comprising 125,972 rows and 43 columns. The NSL-KDD test dataset was also employed, consisting of 22,543 rows and the same number of columns as the training dataset. Despite four models, the most prominent one was the Naive Bayes model, resulting in an 86.9% accuracy for DoS attack detection and 93.5% for Probe attacks. While the results are promising, as mentioned in the paper's conclusion, further tests, either in a real environment or virtual machines (VMs), should be conducted to validate their efficiency and accuracy. In summary, the proposed technique shows promise and has been effective in controlled environments and datasets. However, it lacks testing in a real environment, as noted in the paper's conclusion. Furthermore, performance measurements are needed to demonstrate its capability to handle high dataflow.

Sampath et al. [28] address the challenge of rapidly advancing attacks on SDN-based networks, proposing an automated solution for rule generation using the Genetic Algorithm (GA) in conjunction with IDSs. This approach aims to predict and prevent the formation of new attacks by automating the rule generation process. The distinctive feature of this proposal lies in its rule generation mechanism, leveraging the GA to create new rules based on data flow behavior. Additionally, a general ruleset serves as a foundation for generating rules that effectively block malicious activities. Inspired by biological concepts like natural selection and evolution, the GA is integrated with a feature selection technique to identify optimal rules and combinations. These enhanced flow rules can preemptively thwart attacks before they reach the data plane in the SDN architecture. The effectiveness of this approach was tested in a controlled environment comprising virtual machines with an SDN-based topology, including a controller, a switch, and two hosts. Results demonstrated the GA's innovation in generating rules to prevent the emergence of new attacks. While the proposal exhibited success in this controlled setting, further testing in a real-world SDN-based network environment is recommended to validate its performance and consider its potential as a disruptive technology for enhancing the security of SDN-based networks.

In accordance with Roy et al. [29], the Wireless Sensor Network (WSN) is a high-speed network designed to measure, monitor, and collect data from diverse contexts using a distributed, spatially sparse network of sensors. Addressing the challenge of intrusion detection in such networks, this study proposes a model employing feature selection to reduce the number of extracted features accurately. This minimizes the data volume traversing the network and alleviates the network traffic load, thereby expediting the intrusion detection process. Utilizing a Fully Convolutional Network (FCN) and the UNR-IDD dataset, the model demonstrated exceptional effectiveness and accuracy, achieving accuracy rates between 98% and 99%. Remarkably, the model attained 100% precision, correctly classifying all malicious traffic patterns. Despite the FCN's suitability for intrusion detection, its performance remains unresolved, rendering it unsuitable for real-time intrusion detection in WSNs. This limitation is particularly pertinent as WSNs typically operate with constrained computational resources and network bandwidth.

Jankowski et al. [30] introduce the Monitoring and Detection of Malicious Activities in SDN (MADMAS) systems, leveraging native SDN mechanisms and employing data exploration techniques to identify and process features for network traffic classification. The MADMAS system utilizes Independent Component Analysis (ICA) and PCA techniques to reduce the feature space, enhancing the efficiency of SDN traffic classification and notably increasing unauthorized activity detection. These techniques' significant benefits and impact underscore their pivotal role in successfully reducing the feature space for improved malicious traffic detection.

To address the challenge of detecting intrusion in SDNs, Janabi et al. [31] propose a technique to enhance the performance of IDSs within SDNs, particularly catering to the demands of large enterprise networks. Their method minimizes the overhead during IDS operations through a two-stage process: optimized feature selection and extraction. In the first stage, a correlation-based feature selection (CFS) algorithm is employed to filter and select the most relevant features from the input data, thereby reducing the computational cost of subsequent analysis. In the second stage, PCA is applied to diminish the selected features' dimensionality further, resulting in fewer features used in the IDS. This approach accelerates the IDS process and mitigates overhead. However, further validation of the effectiveness of this technique in real-world enterprise network environments, especially in large-scale scenarios, would provide valuable insights into its practical applicability.

Friha et al. [32] worked on reducing feature dimensionality by selecting highly correlated features that can optimize resource usage and enhance accuracy. Additionally, employing distributed ML approaches, such as Federated Learning, can further improve the solution. The Pearson Correlation Coefficient technique and Chi-Square were chosen for feature selection in both the InSDN and EDGE-IIoTset datasets. Utilizing a convolutional neural network model and selecting the top 19 features, the achieved accuracy reached 96.06% for InSDN and 99.68% for EDGE-IIoTset, respectively, employing



PCC for feature selection. This outperformed the results without feature selection, which were 95.22% and 98.32%. It was also demonstrated a reduction in computation time from 21.05 seconds to 15.91 seconds for the first dataset and 9.70 to 7.10 seconds for the second dataset. While the Federated Learning model slightly decreased accuracy compared to a centralized model, the loss was minimal at most, 0.07%. However, it outperformed existing IDS models in accuracy and resource usage, reducing to 0.068 MFLOPs compared to 0.209 MFLOPs of the second-best model in the resource usage category.

While Industrial IoT networks are recognized as potential targets for attacks, it is essential to acknowledge that home IoT networks are also susceptible to such threats. In this context, the challenge lies in defining the relevant features within a dataset. While feature selection algorithms provide a solution, manual selection is also viable. This involves categorizing features and linking them to specific types of attacks. For instance, categories like Traffic features (time-based or connections to the same host or service) and Packet header features (IP and TCP headers) could be employed for DoS attacks. Manual selection may suffice for simple attacks, but fine-tuning feature selection becomes challenging to maintain accuracy in more complex attacks. The significance of feature selection is underscored, emphasizing that enhancements in Intrusion Detection Systems (IDS) can be achieved by improving feature selection rather than solely depending on the development of new ML models [11].

While feature selection brings the benefit of reducing computational complexity and has the potential to improve accuracy in various network types, it does not always guarantee an enhancement in accuracy. In a study utilizing a recurrent neural network with gated recurrent units and Long Short-Term Memory (LSTM), combined with ANOVA F-test and Recursive Feature Elimination (RFE), the false alarm rate was reduced to 0.76%. However, this approach resulted in an overall accuracy decrease to 87% compared to other models [33]. Another employed recurrent neural network (RNN) and LSTM approaches but differed in the feature selection algorithm and datasets. Utilizing the Information Gain (IG) filter method and Random Forest, this solution reduced the features from 48 to only 10. Achieving 98.76% accuracy for Random Forest and 99.5% for IG, this outperformed compared models in the literature mentioned in the paper, where the best accuracy was 96.5% for the CICIDS2017 dataset [7].

Feature selection can be executed by eliminating features from a dataset based on a correlation-defined rank. Another approach involves starting with zero features and gradually adding features until reaching the defined limit. Forward Feature Selection (FFS) adopts this iterative approach, evaluating each feature's correlation individually, pairing it with others, and incrementally adding features until the desired number is achieved. In a study employing FFS, out of a pool of 20, the 5 best features were selected to complement a hybrid model combining a CNN and a RNN. The resulting model achieved an impressive 98.09% accuracy with a minimal false positive rate of 0.02% [34].

One effective technique, known as the Gradient Boosting Feature Selection Module, assesses the importance of features in decision-making within individual decision trees, which are then compared across various trees. The findings from this approach [35] revealed that more than 75% of the features in the UNSWNB15 dataset lacked significant relevance for accurate classification. By integrating AdaBoost, a method that employs decision trees and stumps to converge towards a robust model, remarkable results were achieved. The study attained an accuracy of 97% within a training duration of less than 43 seconds, surpassing the performance of other methods evaluated.

Scaranti et al. [36] introduce an IDS based on Artificial Immune Systems (AIS) designed to detect anomalies in SDNs in near real-time. The AIS-IDS comprises three integrated modules within the SDN controller: Flow Collector, AIS Detection, and Mitigation. The Flow Collector acquires and preprocesses IP flows, while the AIS Detection module classifies network behavior as normal or abnormal using AIS. The Mitigation module then responds by creating forwarding rules to block malicious traffic. The proposal leverages key features of IP flows, such as source and destination IP addresses and ports, for precise anomaly detection. A sliding window technique adapts to dynamic SDN changes, ensuring rapid detector generation and improving detection capacity. The IDS identifies anomalies and proactively responds to block attacks, enhancing overall security. Experimental results in an

emulated environment demonstrate high efficacy, with an F-measure exceeding 99.9%. Evaluation using a public dataset of attacks further attests to the IDS's versatility and adaptability, achieving a performance exceeding 92% in detecting various attack types without prior information.

El Houda et al. [37] present the BoostIDS, a framework designed for the detection and mitigation of security threats in Smart Grid (SG) systems based on SDN. The framework uses ensemble learning to address common challenges in intrusion detection systems using ML and Deep Learning (DL). Given the critical role of SDN-based SG in electric power systems and the security challenges it faces, BoostIDS comprises two main modules: one for data monitoring and feature selection, utilizing an efficient boosting-based feature selection algorithm, and another for threat detection based on ensemble learning. Extensive experiments with real datasets (NSL-KDD and UNSW-NB15) showcase BoostIDS's effectiveness in efficiently detecting and mitigating threats in SDN-based SG systems. Performance metrics, including accuracy, detection rate, F1 score, and training time, demonstrate superior results to other ML/DL-based intrusion detection models. In conclusion, BoostIDS stands out as a prominent framework for enhancing cybersecurity in SDN-based SG systems, overcoming limitations through its ensemble learning approach, as validated by extensive experiments and optimization of training/test complexity.

Ganesan and Sarac [38] explore security threats in SDN environments, explicitly focusing on evasion-based intrusions. The research emphasizes the susceptibility of ML-based intrusion detection systems to evasion attacks, where adversaries manipulate packet features to avoid detection. The article proposes using multiple sets of reduced features to enhance intrusion detection capabilities in SDN environments instead of relying solely on complete datasets. This approach is grounded in Permutation Feature Importance (PFI), a method that evaluates the relevance of each feature in the effectiveness of ML models. The proposed strategy involves identifying important feature sets, training ML classifiers with these reduced feature sets, and using an ensemble of classifiers to improve NIDS system accuracy. Permutation Feature Importance (PFI) and Orthogonal Feature Ranking (OFR) are employed to identify crucial features in the dataset. Evaluations demonstrate that the hybrid multi-classifier system outperforms conventional classifiers when subjected to adversarial evasion attacks. Permutation Feature Importance involves thoroughly analyzing features to identify essential sets that enable ML classifiers to maintain robust performance in intrusion detection, even with reduced feature sizes. The primary objective is to enhance the resilience of ML classifiers against evasion attacks by diversifying and optimizing feature sets used for model training.

El Houda et al. [39] introduce a specialized multi-level machine learning framework tailored for advanced attack detection in SDN environments. The framework consists of three key modules: a Data Flow Collection (DFC) module utilizing the sFlow protocol, an Information Gain Feature Selection (IGF) module, and an unsupervised ML module employing Isolation Forest (ML-IF) for anomaly detection. While the exact features are not specified, the evaluation employs the UNSW-NB15 dataset, known for its diverse characteristics related to security threats. The IGF module streamlines the training and testing processes by selecting the most informative features. Based on isolation forests, the ML-IF module effectively identifies and classifies security threats in SDN environments. Experimental results in the OMNeT++ emulator with the UNSW-NB15 dataset showcase the framework's superiority over recent contributions, achieving a precision of 97% and a detection rate of 96%, while significantly reducing computational complexity. This contribution stands as a promising solution for addressing evolving security threats in SDN, contributing valuable insights to exploring feature selection techniques in SDN Intrusion Detection Systems.

Mbasuva and Zodi [40] address the vulnerability of SDN to DDoS attacks due to their centralized architecture. The proposed solution is an Ensemble Deep Learning-based IDSs tailored for detecting DDoS attacks in SDNs. Utilizing the CIC-IDS2017 dataset, the model combines Convolutional Neural Network (CNN), Deep Neural Network (DNN), and RNN architectures. Key features identified by literature, including Bwd Packet Length, Avg Packet Size, Flow Duration, and Flow IAT Std, are selected for accurate DDoS detection. The ensemble model outperforms individual and ensemble models in

the literature, demonstrating notable effectiveness in DDoS attack detection. Future work includes simulating the model on platforms like Mininet and OFNet, implementing mitigation measures for DDoS attacks, and extending detection to application and data layers, indicating a commitment to practical application and continuous improvement in real-world scenarios.

Firdaus et al. [41] propose addressing SDN vulnerabilities to DDoS by utilizing ML techniques, specifically employing an Ensemble Algorithm. The authors conducted experiments using the InSDN dataset, employing a two-stage methodology. The first stage involved feature selection, normalization, clustering, Ensemble algorithm classification. The second stage validated the detection in SDN using the Mininet emulator, utilizing Ensemble K-means++ and Random Forest algorithms. The approach, centered on Machine Learning and Ensemble Algorithms, aims to enhance SDN security by achieving more efficient detection of DDoS attacks. Using of the InSDN dataset and validation through the Mininet emulator contribute to the method's robustness and practical applicability.

Kanagaraj et al. [42] propose the application of deep learning to enhance IDS/IPS functionalities, aiming to reduce human effort in data preprocessing and feature selection. Multiple deep learning models are tested, with the selected model trained on the NSL dataset, a recognized benchmark. The integrated model offers intrusion detection, malware detection, and traffic analysis, contributing significantly to network security by providing a robust and effective defense against evolving threats. Incorporating deep learning strengthens the network's resilience and enhances its ability to remain secure and resilient against various attacks.

Amarudin et al. [43] address the challenge of false positive detections in ML-based IDSs. False positives are attributed to harmful ML techniques, prompting the proposal of the S-SDN model. S-SDN, an Ensemble Learning (EL) model, is constructed through the stacking technique of three base learners (SVM, Decision Tree, and Naïve Bayes). It serves as a classifier in IDS for intrusion detection and is validated with the UNSW-NB15 dataset. Experimental results show that S-SDN outperforms the previous method based on a single classifier. S-SDN achieves an accuracy of 83.19%, surpassing the SVM with 75.89% accuracy and the ensemble classifier (Bagging-DT) with 80.09% accuracy. Despite promising results, the research emphasizes the ongoing need for improvements in EL-based IDS development, proposing an EL model with resource selection techniques and diverse base learners. Continuous advancements are deemed essential in this domain.

Abdulqadder et al. [44] focus on implementing key technologies like SDN and Network Function Virtualization (NFV) to support advanced 5G networks. Due to the challenges in security provisioning for the large number of users in 5G networks, an advanced attack-aware security provisioning scheme is proposed. The scheme involves the Initial Authentication Process, Packet Classification, and Switch Migration Process. Initial authentication employs a Secure Identity (SIA)-based scheme at the access point. Suspicious packets are identified and classified into Virtual Network Function (VNF) by the controller, using a Genetic Algorithm with Correlation (GAC) based feature selection algorithm leading to a Radial Basis Function with Extreme Learning Machine (RBF-ELM) classifier. Malicious packets are dropped in the VNF, and normal packets are redirected to the destination through the controller. To mitigate overload attacks on the flow table, an Enhanced Artificial Bee Colony (EABC) algorithm is introduced in the controller. Experimental results demonstrate superior performance regarding delay, redirected packets, detection accuracy, packet transmission rate, and packet loss rate for the proposed scheme.

Suresh et al. [45] present an IDS for software-defined IoT networks based on artificial intelligence, incorporating the self-adaptive energy-efficient BAT algorithm. The methodology involves three stages: preprocessing, feature extraction and selection using the information gain algorithm, and classification using SVM. The self-adaptive energy-efficient BAT algorithm is designed to optimize feature selection through a fitness-based parallel task processing strategy, improving scalability and energy efficiency. The study uses the NSL KDD CUP 1999 dataset for evaluation, considering parameters like accuracy, precision, recall, and additional time. The proposed algorithm enhances feature selection by dynamically adapting to changing environmental conditions, particularly in high-traffic environments. While

highlighting improvements over conventional BAT swarms, the study identifies limitations in energy efficiency and suggests areas for future enhancement, such as error-handling procedures and process revocation functionalities. Overall, the work contributes to advancing intrusion detection systems, artificial intelligence, and security in software-defined IoT networks.

Govindaraju et al. [46] address the increasing security concerns related to the widespread adoption of IoT services, particularly the threat of DDoS attacks targeting IoT devices. The study advocates using SDN as a secure management solution for these devices. The focus is on efficiently detecting DDoS attacks in SDN by employing optimized models based on deep learning. The proposal involves collecting normal and DDoS traffic characteristics from SDN datasets. The NSL-KDD dataset is recommended for feature selection to simplify models for readability and reduce training time. The research proposes a real-time DDoS attack detection system in SDN using a LSTM model. Applying an artificial gorilla troop optimizer for feature selection from the NSL-KDD dataset results in high classification accuracy. Their IDS achieves a notable detection accuracy of 97.59%, showcasing its effectiveness in reducing processing loads and execution times.

Ahn et al. [47] address the challenge of explicability in deep learning models used in traffic classification, particularly in network functions like SDN and network intrusion detection systems. While various methods have been proposed for classifying encrypted traffic without inspecting packet payloads, the lack of explainability in these models raises concerns, especially when dealing with malicious or incorrect data in the training set. To tackle this issue, the paper proposes an explainable artificial intelligence (XAI)-based method utilizing a genetic algorithm. The proposed method aims to elucidate how the deep learning-based traffic classifier functions, providing quantifiable importance measures for each feature. Additionally, a GA generates a feature selection mask that highlights the most significant features across the entire dataset. The practical implementation of this approach resulted in a deep learning-based traffic classifier with an accuracy of approximately 97.24%. These results indicate that the GA-based XAI method holds promise in offering valuable insights to enhance the understanding and reliability of traffic classification models in intricate network environments.

The mentioned studies highlight the significant role of feature selection as a valuable tool in enhancing efficiency and accuracy, emphasizing the need for careful consideration in evaluating attack detection. While leveraging established techniques from the literature is a prudent strategy, exploring novel feature selection methods remains crucial. In this context, the following section shows a comprehensive comparison of techniques for selecting attributes in datasets.

#### 4. Assessment Methodology

This work proposes a comparative and statistical study of attribute selection techniques. Also, this work aims to present the data obtained and serve as *baseline* for comparison between attribute selection techniques. The results presented here in this work make it possible to implement optimized IDSs.

##### 4.1. Assessment Model

Figure 3 depicts the basis of this evaluation model, highlighting its stages in detail, showing all steps of this process, its main components, and interactions.

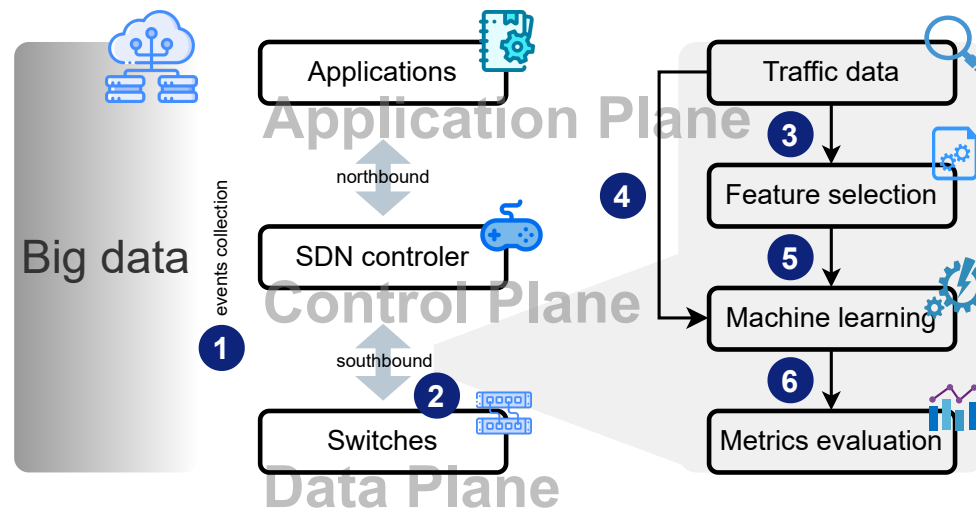


Figure 3. Evaluating Feature Selection Techniques Mechanism.

In step (1), events are gathered from the Big Data. In simple terms, the dataset for analysis is selected, encompassing both normal traces and traces with anomalies resulting from various attacks. The raw data collection is crucial to ensure that the study includes a diverse dataset, providing extensive training and enhancing detection accuracy.

Subsequently, in step (2), the data flowing through the SDN switches was meticulously examined to ensure the absence of redundant features. Any identified redundancies were eliminated to maintain the cleanliness of our dataset for training and analysis. This step involves a thorough scan, as datasets used for analysis may occasionally contain redundant columns that could impact the results.

In the subsequent step (3), the data is partitioned into training and testing sets to ensure the derivation of accurate and realistic results, facilitating the identification of various types of attacks. For this study, we adopted an approach in which 70% of the data was allocated for training and 30% for testing. This partitioning strategy is commonly employed in projects involving training and analysis of data from a dataset [6]. The division ensures a comprehensive training scope encompassing a diverse range of data.

The application of the subsequent collection of the outputs generated (4) and the feature selection techniques (5) represent crucial steps in our work. These outputs will be utilized in the subsequent layer. Shell scripts, Python 3.8.0, and the Weka software were employed to apply attribute selection techniques and gather the corresponding outputs. This toolkit offered a variety of methods for attribute selection, data processing, and classifiers, facilitating the preparation and comparison of data.

In step (6), following the acquisition of filtered data or the generation of new features from the preceding step, we apply these outputs to ML approach. In this work, we utilize a Multi-class Classifier. This step is deemed the most critical in the process, as it involves the application of the classifier to the raw or filtered data, providing us with metrics crucial for analyzing the efficiency of an IDS. These metrics include true positive rate, false positive rate, accuracy, and execution time. Then, a comprehensive analysis of all metrics obtained post applying the classifier is conducted. The strengths and weaknesses of attribute selection techniques for detecting these attacks are presented through tables and graphs, providing a detailed overview.

Ultimately, a comparison is undertaken to underscore the distinctions among attribute selection techniques. This allows us to conclude regarding the efficiency of these techniques, considering the trade-off between the potential loss of detection accuracy and the concurrent gain in processing time.

#### 4.2. Dataset and Sets of Features Analyzed

To validate the assessment model, we employed a widely recognized dataset containing all features extracted from the DUMP Dataset CICIDS2017 files, as detailed in the work by D. Stiawan



et al. [6]. This dataset encompasses various features crucial for detecting attacks, particularly those leading to anomalies. In the context of anomaly detection, the analysis of multiple behaviors within these features becomes imperative.

This dataset comprises eight monitoring sessions, spanning from 9:00 am on Monday to 5:00 pm on Friday. The file classifies traffic into two categories: “Benign” for normal traffic and “Attack” for malicious traffic. The dataset encompasses 14 types of attacks, as outlined in Table 3.

Table 3. Summary of the CICIDS-2017 dataset [6].

File name	Traffic type	numbers of occurrences
Monday-WorkingHours.pcap_ISCX.csv	Benign	529,918
Tuesday-WorkingHours.pcap_ISCX.csv	Benign	432,074
	SSH-Patator	5,897
	FTP-Patator	7,938
Wednesday-WorkingHours.pcap_ISCX.csv	Benign	440,031
	DoS Hulk	231,073
	DoS GoldenEye	10,293
	DoS Slowloris	5,796
	DoS Slowhttptest	5,499
	Heartbleed	11
Thursday-WorkingHours-Morning-WebAtacks.pcap_ISCX.csv	Benign	168,186
	Web Attack-Brute Force	1,507
	Web Attack-Sql Injection	21
	Web Attack-XSS	652
Thursday-WorkingHours-Afternoon-Infiltration.pcap_ISCX.csv	Benign	288,566
	Infiltration	36
Friday-WorkingHours-Morning.pcap_ISCX.csv	Benign	189,067
	Bot	1,966
Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX.csv	Benign	127,537
	PortScan	158,930
Friday-WorkingHours-Afternoon-DDos.pcap_ISCX.csv	Benign	97,718
	DDos	128,07
Total de Registros		2,830,743

In this study, data from July 5, 2017, is utilized, as it represents the day with the highest concentration of DoS Hulk and DoS GoldenEye attacks – the specific focus of this investigation (highlighted in red in Table 3). This dataset encompasses normal (benign) traffic and attack traffic that induces anomalies in the network, including DoS, DDoS, and Heartbleed attacks.

Finally, the dataset comprises 77 features and a label, encompassing approximately 640 thousand instances on the analyzed day. Interested individuals can access this dataset through the Canadian Institute for Cybersecurity platform<sup>1</sup>.

<sup>1</sup> CIC Dataset download: <http://205.174.165.80/CICDataset/CIC-IDS-2017/>

5. Performance Assessment

This section first evaluates the classifier and the feature selection techniques studied. Secondly, all the results obtained are shown and discussed.

5.1. Evaluation of the Entire Dataset

The complete dataset, comprising 77 features and labels, was employed for the initial assessment. A Python script was executed to eliminate identified redundant features, and subsequently, the classifier was applied. Table 4 gives the results. It's noteworthy that utilizing the entire dataset, results in true positive rates exceeding 98% and false positive rates below 1%, indicating an exceptional performance in identifying network attacks. This approach underscores the robustness and effectiveness of the detection mechanism employed, based on a Multi-class Classifier. However, it's essential to acknowledge that this comprehensive detection process may incur higher computational costs due to the large number of features that need to be analyzed. When the classifier is applied to the entire dataset, the analysis of the confusion matrix displayed in Table 5 provides a detailed understanding of the precision in detecting attacks and classifying them.

Table 4. Results with the entire dataset.

Attack name	TP	FP
Benign	0.996	0.002
Dos Hulk	0.999	0.004
DoS GoldenEye	0.988	0

Table 5. Confusion matrix of the entire dataset.

Class	Benign	DoS Hulk	DoS GoldenEye
Benign	131626	501	25
DoS Hulk	47	69056	0
DoS GoldenEye	35	2	3103

Figure 4 depicts the accuracy of the entire dataset. It is possible to notice that when applying the classifier to the entire dataset, an accuracy of 99% was achieved for GoldenEye-type attacks, and a detection accuracy of 99.3% was achieved for HULK-type attacks.

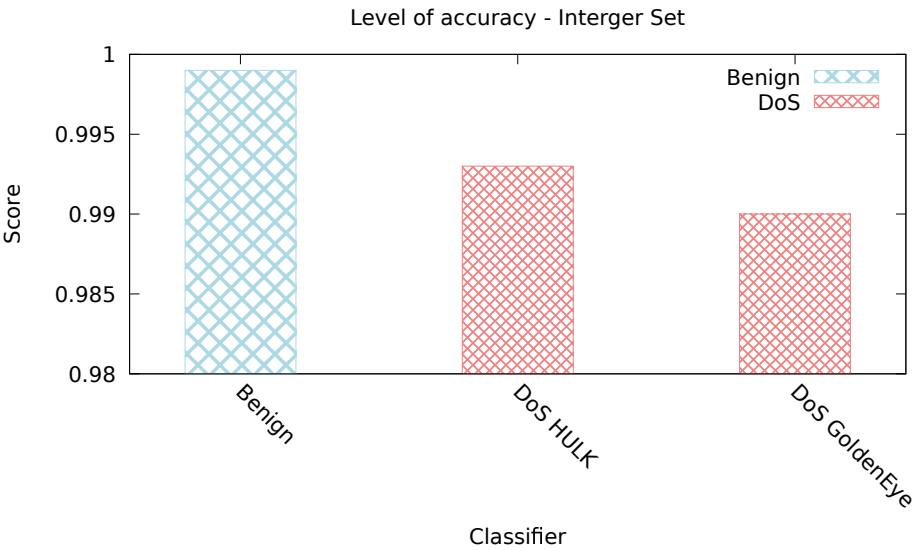


Figure 4. Accuracy of the entire dataset.

5.2. Evaluation Using Feature Selection

This subsection shows results when we applied Multi-class Classifier to the data with feature selection. A common observation among them is that the False Positive (FP) rate for GoldenEye is meager. However, all classifier methods faced challenges in analyzing this classification, with One R demonstrating the best results.

Next, we delve into a more detailed discussion of each analyzed result. We provide the confusion matrix and accuracy level for each method.

5.2.1. Correlation Attribute

Beginning with the Correlation Attribute (CA), the execution time for this method was 51.17 seconds, representing the shortest execution time among the tested classifiers. Notably, this method exhibited a commendable selection capacity, particularly in the case of the DoS Hulk method. Table 6 shows the confusion matrix, which provides an evaluation of the precision of attack detection and classification that is illustrated with more detail in Figure 5. Here, we can observe the scores achieved for both attacks and normal network traffic.

Table 6. Confusion matrix of the Correlation Attribute.

Class	Benign	DoS Hulk	DoS GoldenEye
Benign	306107	1783	408
DoS Hulk	53238	108217	31
DoS GoldenEye	1070	3273	2851

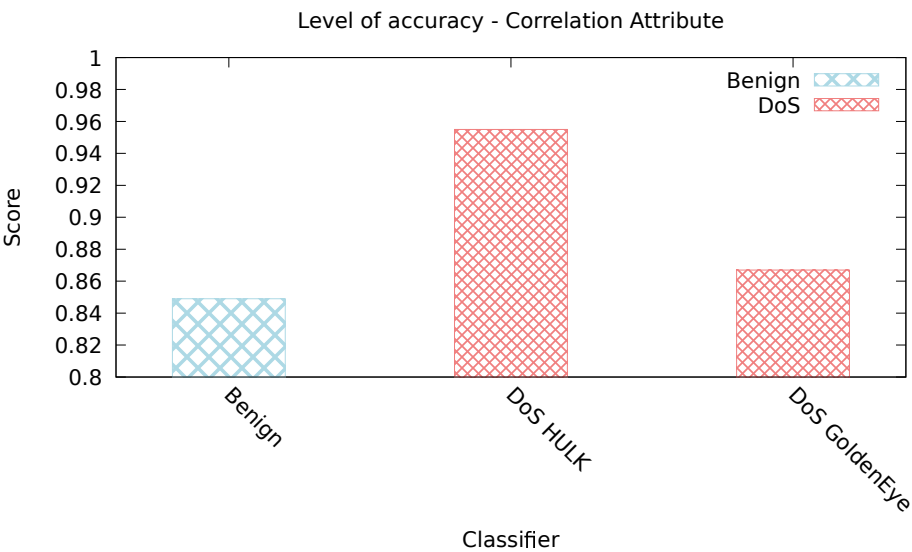


Figure 5. Correlation Attribute.

5.2.2. Symmetrical Uncert

In our second scenario, we employ the Symmetrical Uncert technique for attribute selection. This method assesses the level of uncertainty between attribute evaluations. Following the evaluation, the classifier was applied.

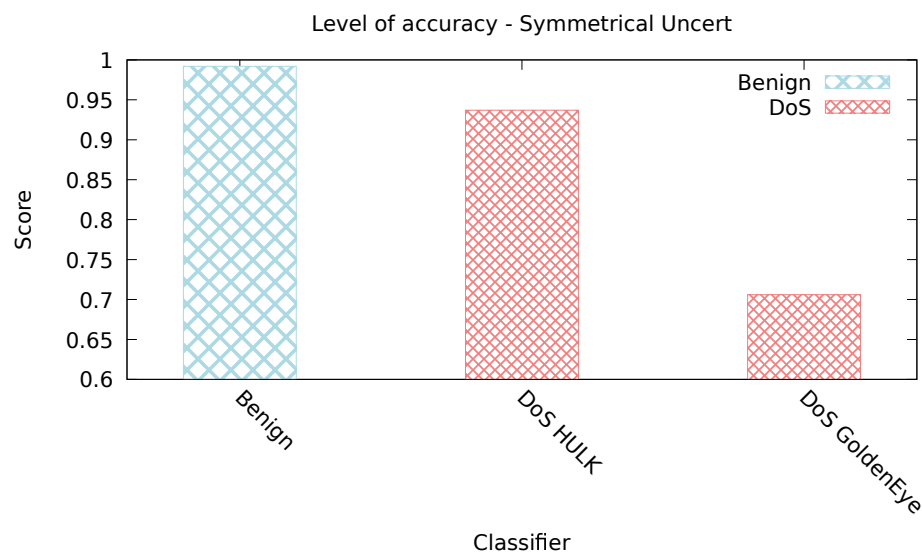
We can derive metrics such as accuracy by analyzing the confusion matrix concerning the application of Symmetrical Uncert. In this scenario, an accuracy of 99.4% was achieved for HULK attacks, while the detection accuracy for GoldenEye attacks was 17.8%. Upon applying the classifier with the Symmetrical Uncert technique, it is possible to find the confusion matrix (Table 7) and enables the extraction of metrics such as precision that means the proportion of examples correctly classified as

anomalies out of the total examples classified as anomalies, meaning it is the ratio of true positives to the sum of true positives and false positives.

**Table 7.** Confusion matrix of the Symmetrical Uncert.

Class	Benign	DoS Hulk	DoS GoldenEye
Benign	301481	6804	13
DoS Hulk	457	160507	522
DoS GoldenEye	2002	3908	1284

Based on this scenario, the utilization of Symmetrical Uncert showed notably high true positive rates for HULK attacks (99.4%), suggesting accurate detection of these attacks, as depicted in Figure 6. However, for GoldenEye attacks, the true positive rate was only 17.8%, indicating lower effectiveness in detecting these specific types of attacks. Moreover, the false positive rates were low, with 1.5% for HULK attacks and 0.1% for GoldenEye attacks. These findings suggest that Symmetrical Uncert can effectively detect HULK attacks but is limited in detecting GoldenEye attacks.



**Figure 6.** Symmetrical Uncert.

### 5.2.3. InfoGain Classifier

The InfoGain attribute selection technique was also employed. This technique involves assigning weights to features to identify the most relevant ones. In this instance, the 15 features with the highest relevance were selected, yielding satisfactory results, as illustrated in Table 8.

**Table 8.** Confusion matrix of the Information Gain.

Class	Benign	DoS Hulk	DoS GoldenEye
Benign	129673	2215	207
DoS Hulk	13581	55307	4
DoS GoldenEye	619	286	2223

In this case, it was observed that applying this attribute selection technique did not yield satisfactory results compared to the entire set. True positive rates of 80% and 69.9% were obtained for the HULK and GoldenEye attacks, respectively.

In this scenario, Figure 7 shows that when applying Information Gain, an accuracy of 95.6% was achieved for HULK-type attacks, and a detection accuracy of 90.5% was obtained for GoldenEye attacks.

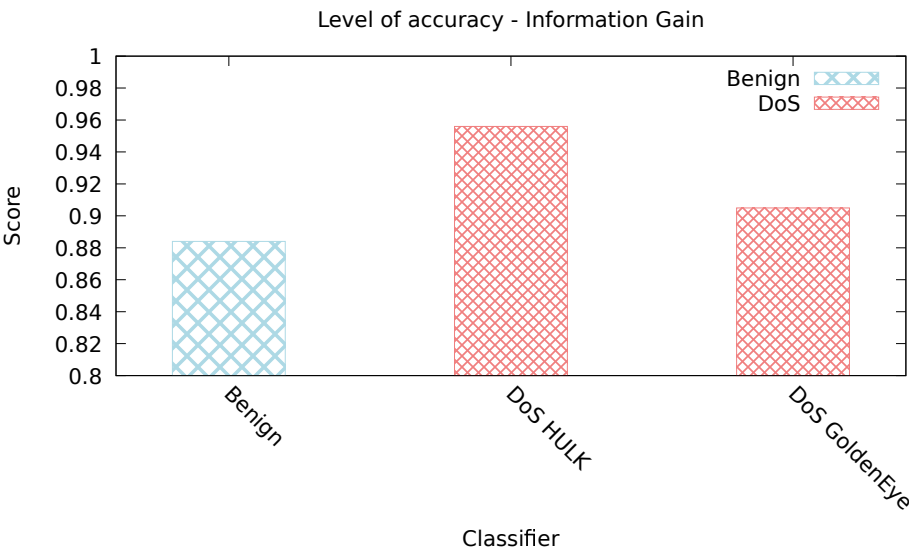


Figure 7. Information Gain.

5.2.4. OneR Classifier

In the fourth scenario, we applied OneR, one of the most primitive and efficient selection algorithms in the literature. By using the Multi-class Classifier to the selected set, we obtained the results presented in Table 9. In this scenario, it is observed that compared to the previous technique, we already have satisfactory results, with true positive rates of 87.3% and 83.9% for HULK and GoldenEye-type attacks, respectively. In contrast, their false positive rates were 1% and 0.1%, respectively.

Table 9. Confusion matrix of the One R.

Class	Benign	DoS Hulk	DoS GoldenEye
Benign	130842	1106	97
DoS Hulk	8703	60329	0
DoS GoldenEye	297	179	2635

When applying the OneR classifier, the confusion matrix above is observed, enabling the calculation of metrics such as accuracy, which allows for a comparative analysis of both attacks and normal network traffic. In this context, accuracy represents the model’s ability to correctly classify positive instances, i.e., anomalies, out of the total instances classified as positive by the model.

Figure 8 depicts that when applying OneR, an accuracy of 97.8% was achieved for HULK-type attacks, and a detection accuracy of 94.5% was obtained for GoldenEye-type attacks, demonstrating high precision and efficiency.



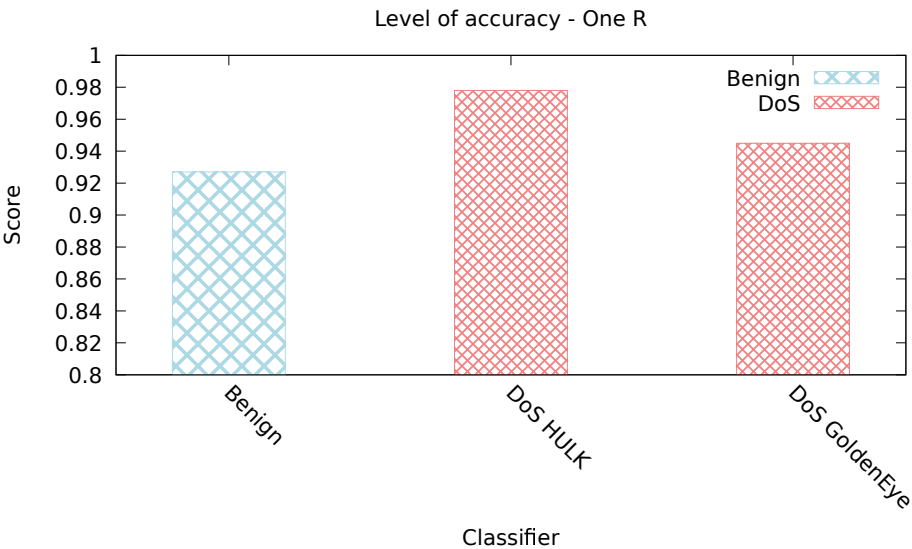


Figure 8. OneR.

5.2.5. PCA Classifier

In the final scenario, the PCA technique was employed for attribute selection, representing a modern and highly efficient data analysis approach, as presented in Table 10. Notably, significant variations in true positive rates were observed between the types of attacks when utilizing PCA. Specifically, a true positive rate of 95.1% was achieved for HULK-type attacks. In contrast, a substantially lower rate of 32.1% was observed for GoldenEye attacks, marking the lowest among all scenarios presented in this study.

Table 10. Confusion matrix of the PCA.

Class	Benign	DoS Hulk	DoS GoldenEye
Benign	12957	2654	69
DoS Hulk	3382	65713	0
DoS GoldenEye	558	1574	1008

When applying the PCA classifier, the displayed confusion matrix allows for deriving metrics such as precision. In this scenario, the application of PCA with the Multi-class Classifier reveals that the overall accuracy remains relatively unaffected despite not achieving a high true positive rate for GoldenEye attacks. Figure 9 that the obtained accuracies are 93.9% for HULK attacks and 93.6% for GoldenEye attacks, showcasing consistent performance.

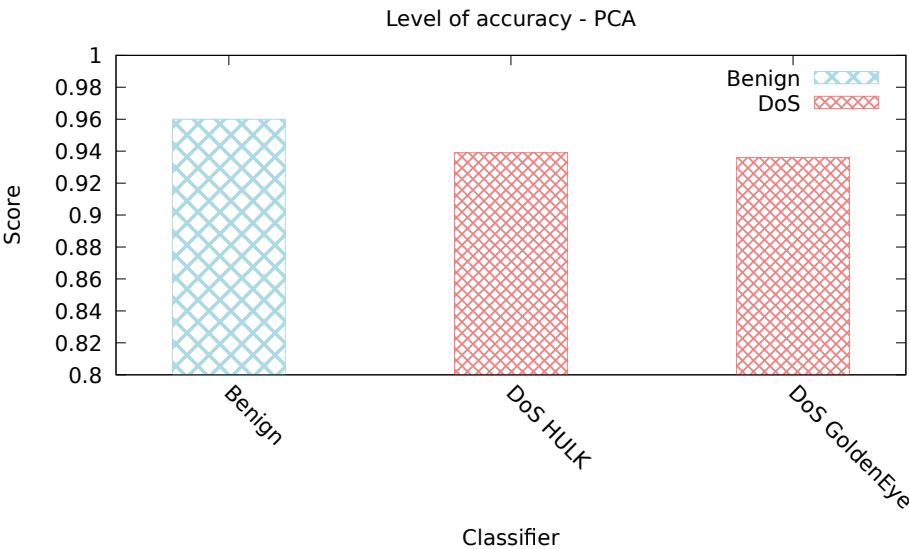


Figure 9. Principal Component Analysis.

5.3. Discussion of Results

Utilizing attribute selection techniques before data classification provides insights into several factors that can impact accuracy, offering crucial considerations for implementing a IDS. The results underscore that integrating attribute selection into the preprocessing stage allows for significant time savings in detecting attacks. Also, this efficiency gain is achieved without imposing a substantial compromise on the accuracy of the IDS, emphasizing the practicality and effectiveness of incorporating attribute selection methodologies in enhancing the overall performance of intrusion detection systems.

Among the attribute selection techniques scrutinized, OneR emerged as the most effective, showcasing the highest accuracy in the study. Notably, it achieved an accuracy of 97.8% for detecting DoS Hulk attacks and 94.5% for DoS GoldenEye attacks. Beyond its accuracy, implementing OneR yielded a substantial reduction in processing time, amounting to a noteworthy 98.25% reduction compared to the classification using the entire dataset. This underscores the efficiency gains achieved by employing OneR as an attribute selection technique in the context of intrusion detection systems. A summary is detailed in Table 11.

Table 11. General results of the methods.

Method	Benign		Dos Hulk		DoS GoldenEye	
	TP	FP	TP	FP	TP	FP
Correlation Attribute	0.993	0.322	0.670	0.016	0.396	0.001
Symmetrical Uncert	0.978	0.015	0.994	0.034	0.178	0.001
Info Gain	0.981	0.225	0.8	0.018	0.699	0.001
One R	0.99	0.137	0.873	0.01	0.839	0.001
PCA	0.978	0.071	0.951	0.951	0.321	0

Figure 10 provides a comparative illustration of the time spent in seconds through attribute selection techniques. A logarithmic scale has been applied to the y-axis to enhance clarity in visualizing the values. The comprehensive results affirm that attribute selection techniques play a pivotal role in implementing intrusion detection systems. The pronounced time and resource savings achieved underscore their significance, especially in real-time attack detection, where efficiency is paramount. It's worth noting that all experiments in this study were conducted on a computer equipped with an Intel Core i7 processor featuring six processing cores, 16 GB RAM, and an NVidia Quadro T2000 4GB GDDR5 video card for comparative purposes.

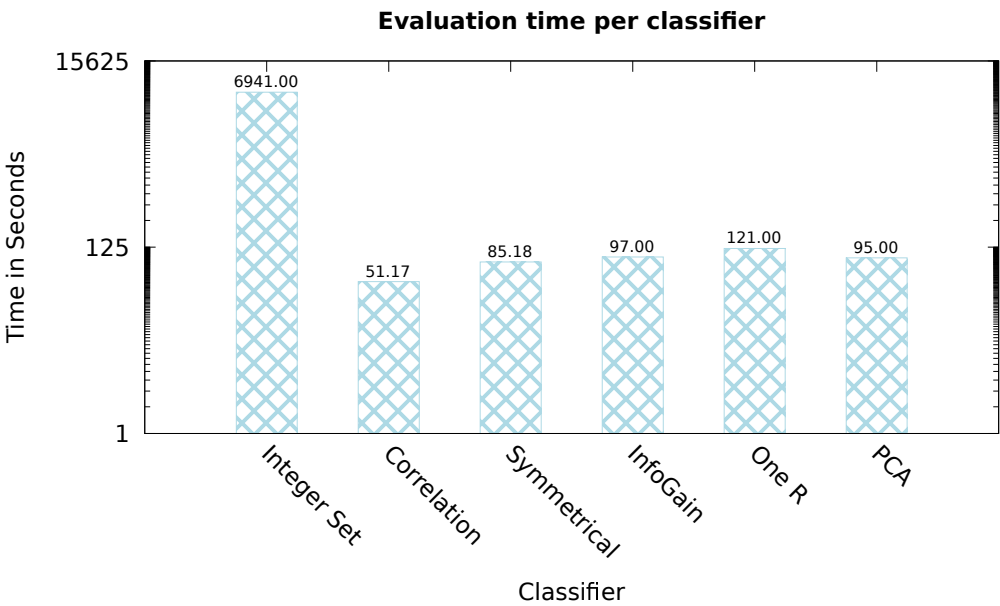


Figure 10. Training time comparison.

6. Conclusions and Future Works

This work conducted a comprehensive literature review that delved into the utilization of programmable networks concerning feature selection to detect anomalies generated by attacks. Also, the proposed investigation compared attribute selection techniques: Correlation Attribute, Symmetrical Uncert, Information Gain, Principal Component Analysis, and OneR. Based on the findings generated by this work, it can be observed that regarding precision and processing cost, the analysis of the entire set provides better results, reaching 99% of precision, but also concerning these metrics, it was observed that classifying using the output of OneR resulted in a slight loss of precision compared to the precision of the entire dataset, being the technique with the highest precision among those studied. Furthermore, it is observed that even if there is a loss in detection efficiency, we have a gain in processing time that must be considered when defining the method to be implemented.

Finally, in future work, we intend to use the knowledge from this work to support a real Intrusion Detection System in the context of Smart Cities. It is expected that using this knowledge, it will be possible to choose an attribute selection technique with a favorable trade-off, that is, one that has efficient results while not using all available resources, saving memory and time processing.

**Author Contributions:** Conceptualization, B.L.D., I.A.M.; methodology, B.L.D., I.A.M., Y.S., A.R; software, I.A.M.; validation, I.A.M.; investigation, B.L.D., I.A.M., Y.S., R.I, A.R, T.L.T.S; writing–original draft preparation, B.L.D., I.A.M.; writing–review and editing, B.L.D., I.A.M., Y.S., R.I, A.R, T.L.T.S; supervision, B.L.D., A.R; All authors read and approved the final manuscript.

**Funding:** The authors would like to thank the FAPERGS agency for the project (23/2551-0000773-8), which partially supported this research. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

- do Prado, P.F; Peixoto, M.L.M.; Araújo, M.C.; Gama, E.S.; Gonçalves, D.M.; Silva, M.V.S.; Immich, R.; Madeira, E.R.M.; Bittencourt, L.F, Mobile Edge Computing for Content Distribution and Mobility Support in Smart Cities. In *Mobile Edge Computing*; Mukherjee, A.; De, D.; Ghosh, S.K.; Buyya, R., Eds.; Springer International Publishing: Cham, 2021; pp. 473–500. doi:10.1007/978-3-030-69893-5\_19.
- Akabane, A.T.; Immich, R.; Pazzi, R.W.; Madeira, E.R.M.; Villas, L.A. Exploiting Vehicular Social Networks and Dynamic Clustering to Enhance Urban Mobility Management. *Sensors* **2019**, *19*, 3558. doi:10.3390/s19163558.

3. dos Santos, R.L.; Wickboldt, J.A.; Lunardi, R.C.; Dalmazo, B.L.; Granville, L.Z.; Gaspary, L.P.; Bartolini, C.; Hickey, M. A solution for identifying the root cause of problems in IT change management. 12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops, 2011, pp. 586–593. doi:10.1109/INM.2011.5990563.
4. Shenfield, A.; Day, D.; Ayesh, A. Intelligent intrusion detection systems using artificial neural networks. *ICT Express* **2018**, *4*, 95–99.
5. Jose, S.; Malathi, D.; Reddy, B.; Jayaseeli, D. A survey on anomaly based host intrusion detection system. *Journal of Physics: Conference Series* **2018**, *1000*, 012049.
6. Stiawan, D.; Idris, M.Y.B.; Bamhdi, A.M.; Budiarto, R.; others. CICIDS-2017 dataset feature analysis with information gain for anomaly detection. *IEEE Access* **2020**, *8*, 132911–132921.
7. Elsayed, M.; Le-Khac, N.A.; Azer, M.; Jurcut, A. A Flow Based Anomaly Detection Approach with Feature Selection Method Against DDoS Attacks in SDNs. *IEEE Transactions on Cognitive Communications and Networking* **2022**, *PP*, 1–1. doi:10.1109/TCCN.2022.3186331.
8. Neto, E.P.; Silva, F.S.D.; Schneider, L.M.; Neto, A.V.; Immich, R. Seamless MANO of multi-vendor SDN controllers across federated multi-domains. *Computer Networks* **2021**, *186*, 107752. doi:https://doi.org/10.1016/j.comnet.2020.107752.
9. Oliveira, I.; Neto, E.; Immich, R.; Fontes, R.; Neto, A.; Rodriguez, F.; Rothenberg, C.E. dh-aes-p4: On-premise encryption and in-band key-exchange in P4 fully programmable data planes. 2021 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), 2021, pp. 148–153. doi:10.1109/NFV-SDN53031.2021.9665012.
10. Dalmazo, B.L.; Marques, J.A.; Costa, L.R.; Bonfim, M.S.; Carvalho, R.N.; da Silva, A.S.; Fernandes, S.; Bordim, J.L.; Alchieri, E.; Schaeffer-Filho, A.; others. A systematic review on distributed denial of service attack defense mechanisms in programmable networks. *International Journal of Network Management* **2021**, *31*, e2163.
11. Illy, P.; Kaddoum, G.; Kaur, K.; Garg, S. ML-based IDPS Enhancement With Complementary Features For Home IoT networks. *IEEE Transactions on Network and Service Management* **2022**, *PP*, 1–1. doi:10.1109/TNSM.2022.3141942.
12. Paxson, V. An analysis of using reflectors for distributed denial-of-service attacks. *ACM SIGCOMM Computer Communication Review* **2001**, *31*, 38–47.
13. Sharafaldin, I.; Lashkari, A.H.; Ghorbani, A.A. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp* **2018**, *1*, 108–116.
14. Torres, G.; Liu, C. Can data-only exploits be detected at runtime using hardware events? A case study of the Heartbleed vulnerability. In *Proceedings of the Hardware and Architectural Support for Security and Privacy 2016*; ACM, New York, NY, United States, 2016; pp. 1–7.
15. Basnet, R.B.; Shash, R.; Johnson, C.; Walgren, L.; Doleck, T. Towards Detecting and Classifying Network Intrusion Traffic Using Deep Learning Frameworks. *J. Internet Serv. Inf. Secur.* **2019**, *9*, 1–17.
16. Alieyan, K.; Almomani, A.; Manasrah, A.; Kadhum, M.M. A survey of botnet detection based on DNS. *Neural Computing and Applications* **2017**, *28*, 1541–1558.
17. Mahjabin, S. Implementation of DoS and DDoS attacks on cloud servers. *Periodicals of Engineering and Natural Sciences (PEN)* **2018**, *6*, 148–158.
18. Charlier, J.; Singh, A.; Ormazabal, G.; State, R.; Schulzrinne, H. SynGAN: Towards generating synthetic network attacks using GANs. *arXiv preprint arXiv:1908.09899* **2019**.
19. Kumar, V.; Sangwan, O.P. Signature based intrusion detection system using SNORT. *International Journal of Computer Applications & Information Technology* **2012**, *1*, 35–41.
20. Dalmazo, B.L.; Vilela, J.P.; Curado, M. Triple-Similarity Mechanism for alarm management in the cloud. *Computers & Security* **2018**, *78*, 33–42. doi:https://doi.org/10.1016/j.cose.2018.05.016.
21. Singh, J.; Singh, G.; Singh, R. Optimization of sentiment analysis using machine learning classifiers. *Human-centric Computing and information Sciences* **2017**, *7*, 1–12.
22. NOVAKOVIĆ, J.; STRBAC, P.; BULATOVIĆ, D. TOWARD OPTIMAL FEATURE SELECTION USING RANKING METHODS AND CLASSIFICATION ALGORITHMS. *Yugoslav Journal of Operations Research* **2011**, *21*, 119–135.
23. Gupta, V.; Mittal, M. KNN and PCA classifier with autoregressive modelling during different ECG signal interpretation. *Procedia Computer Science* **2018**, *125*, 18–24.

24. al kaaf, H.; Ali, A.; Shamsuddin, S.; Hassan, S. Feature selection for malicious android applications using Symmetrical Uncert Attribute Eval method. *IOP Conference Series: Materials Science and Engineering* **2020**, 884, 012060. doi:10.1088/1757-899X/884/1/012060.
25. Gnanambal, S.; Thangaraj, M.; Meenatchi, V.; Gayathri, V. Classification algorithms with attribute selection: an evaluation study using WEKA. *International Journal of Advanced Networking and Applications* **2018**, 9, 3640–3644.
26. Zainudin, A.; Akter, R.; Kim, D.S.; Lee, J.M. Towards Lightweight Intrusion Identification in SDN-based Industrial Cyber-Physical Systems. 2022 27th Asia Pacific Conference on Communications (APCC), 2022, pp. 610–614. doi:10.1109/APCC55198.2022.9943641.
27. AlMasri, T.; Snober, M.A.; Al-Haija, Q.A. IDPS-SDN-ML: An Intrusion Detection and Prevention System Using Software-Defined Networks and Machine Learning. 2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS), 2022, pp. 133–137. doi:10.1109/APICS56469.2022.9918804.
28. Sampath, N.; Jerlin, M.; Krithika, L.; Anitha, A. Intrusion Detection in Software Defined Networking using Genetic Algorithm. 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1–5. doi:10.1109/ic-ETITE47903.2020.464.
29. Roy, B.; Acharya, I.; Papalkar, D.; Joseph, M. Top-Performing Unifying Architecture for Network Intrusion Detection in SDN Using Fully Convolutional Network. 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), 2023, pp. 1340–1344. doi:10.1109/ICIRCA57980.2023.10220608.
30. Jankowski, D.; Amanowicz, M. A study on flow features selection for malicious activities detection in software defined networks. 2018 International Conference on Military Communications and Information Systems (ICMCIS), 2018, pp. 1–9. doi:10.1109/ICMCIS.2018.8398697.
31. Janabi, A.H.; Kanakis, T.; Johnson, M. Overhead Reduction Technique for Software-Defined Network Based Intrusion Detection Systems. *IEEE Access* **2022**, 10, 66481–66491. doi:10.1109/ACCESS.2022.3184722.
32. Friha, O.; Ferrag, M.A.; Shu, L.; Maglaras, L.; Choo, K.K.; Nafaa, M. FELIDS: Federated Learning-based Intrusion Detection System for Agricultural Internet of Things. *Journal of Parallel and Distributed Computing* **2022**, 165. doi:10.1016/j.jpdc.2022.03.003.
33. Dey, S.; Rahman, M.M. Flow Based Anomaly Detection in Software Defined Networking: A Deep Learning Approach With Feature Selection Method. 2018, pp. 630–635. doi:10.1109/CEEICT.2018.8628069.
34. Matsa, L.S.; Zodi-Lusilao, P.G.A.; Bhunu-Shava, P.F. Forward Feature Selection for DDoS Detection on Cross-Plane of Software Defined Network Using Hybrid Deep Learning. 2021 3rd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), 2021, pp. 1–7. doi:10.1109/IMITEC52926.2021.9714561.
35. Abou El Houda, Z.; Khoukhi, L. A Hierarchical Fog Computing Framework for Network Attack Detection in SDN. ICC 2022 - IEEE International Conference on Communications, 2022, pp. 4366–4371. doi:10.1109/ICC45855.2022.9838560.
36. Scaranti, G.F.; Carvalho, L.F.; Barbon, S.; Proença, M.L. Artificial Immune Systems and Fuzzy Logic to Detect Flooding Attacks in Software-Defined Networks. *IEEE Access* **2020**, 8, 100172–100184. doi:10.1109/ACCESS.2020.2997939.
37. El Houda, Z.A.; Brik, B.; Khoukhi, L. Ensemble Learning for Intrusion Detection in SDN-Based Zero Touch Smart Grid Systems. 2022 IEEE 47th Conference on Local Computer Networks (LCN), 2022, pp. 149–156. doi:10.1109/LCN53696.2022.9843645.
38. Ganesan, A.; Sarac, K. Mitigating Evasion Attacks on Machine Learning based NIDS Systems in SDN. 2021 IEEE 7th International Conference on Network Softwarization (NetSoft), 2021, pp. 268–272. doi:10.1109/NetSoft51509.2021.9492526.
39. El Houda, Z.A.; Hafid, A.S.; Khoukhi, L. A Novel Machine Learning Framework for Advanced Attack Detection using SDN. 2021 IEEE Global Communications Conference (GLOBECOM), 2021, pp. 1–6. doi:10.1109/GLOBECOM46510.2021.9685643.
40. Mbasuva, U.; Zodi, G.A.L. Designing Ensemble Deep Learning Intrusion Detection System for DDoS attacks in Software Defined Networks. 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM), 2022, pp. 1–8. doi:10.1109/IMCOM53663.2022.9721785.
41. Firdaus, D.; Munadi, R.; Purwanto, Y. DDoS Attack Detection in Software Defined Network using Ensemble K-means++ and Random Forest. 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2020, pp. 164–169. doi:10.1109/ISRITI51436.2020.9315521.



42. Kanagaraj, G.; Primya, T.; Subashini, G.; Senthilkumar, V.; Gomathi, S. Hybrid Intrusion Detector using Deep Learning Technique. 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1–5. doi:10.1109/ICAECA52838.2021.9675648.
43. Amarudin.; Ferdiana, R.; Widyawan. New Approach of Ensemble Method to Improve Performance of IDS using S-SDN Classifier. 2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), 2022, pp. 463–468. doi:10.1109/COMNETSAT56033.2022.9994302.
44. Abdulqadder, I.H.; Zou, D.; Aziz, I.T.; Yuan, B. Enhanced Attack Aware Security Provisioning Scheme in SDN/NFV Enabled over 5G Network. 2018 27th International Conference on Computer Communication and Networks (ICCCN), 2018, pp. 1–9. doi:10.1109/ICCCN.2018.8487339.
45. Suresh, G.M.; Madhavu, M.L. AI Based Intrusion Detection System Using Self-Adaptive Energy Efficient BAT Algorithm for Software Defined IoT Networks. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1–6. doi:10.1109/ICCCNT49239.2020.9225415.
46. Govindaraju, S.; Metia, R.; Girija, P.; Baranitharan, K.; Indirani, M.; R, M. Detection of DDoS Attacks using Artificial Gorilla Troops Optimizer based Deep Learning Model. 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2023, pp. 385–391. doi:10.1109/ICAIS56108.2023.10073935.
47. Ahn, S.; Kim, J.; Park, S.Y.; Cho, S. Explaining Deep Learning-Based Traffic Classification Using a Genetic Algorithm. *IEEE Access* **2021**, 9, 4738–4751. doi:10.1109/ACCESS.2020.3048348.

## Short Biography of Authors



**Iuri A. Mundstock** is a master's student in the Graduate Program in Computing (PPGComp) at the Federal University of Rio Grande (FURG). He received a bachelor's degree in Computer Engineering in 2024 from the Federal University of Rio Grande, Brazil. His primary interests are focused on security and feature extraction methods.



**Yuri Santo** is a Ph.D. student in Graduate Program in Computer Science (PPGCC) at the Federal University of Para, Brazil. Yuri received a Master's degree in Computer Science in 2023 from the Federal University of Para, Brazil. His main research interests include Internet of Things, Federated Learning models, Open Radio Access Network, and data security and privacy.



**Thiago L. T. Silveira** holds a D.Sc. degree in Computer Science (2019) from the Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil, and a M.Sc. degree in Computer Science (2016), and B.Sc. degrees in Information Systems (2015) and Computer Science (2013) from the Federal University of Santa Maria (UFSM), Santa Maria, Brazil. Thiago is currently an Assistant Professor at UFRGS. His interests are computer vision, pattern recognition, and signal processing.



**Roger Immich** is a Professor at the Digital Metropolis Institute (IMD) of the Federal University of Rio Grande do Norte (UFRN). He earned his Ph.D. in Informatics Engineering from the University of Coimbra, Portugal (2017). He was a visiting researcher at the University of California at Los Angeles, United States (UCLA) in 2016/2017, a postdoctoral researcher at the Institute of Computing of the University of Campinas (UNICAMP) in 2018/2019, and a Visiting Professor of University of Málaga (UMA), Spain, in 2021. His research interests encompass various areas, including Smart Cities, IoT, 5G, Quality of Experience, as well as Cloud and Fog computing.



**Andre Riker** is a computer scientist who holds a bachelor's and master's degree in computer science. He received his Ph.D. degree in information science and technology from the University of Coimbra, Coimbra, Portugal, in 2019. He is currently Assistant Professor with the Computer Science Faculty, Federal University of Pará, Brazil. His main research interests include Internet of Things, optimization models, federated learning models, and data security and privacy.



**Bruno L. Dalmazo** received his Ph.D. degree in Information Science and Technology from the University of Coimbra in 2018. He completed his Master's degree in Computer Science in 2011 at the Federal University of Rio Grande do Sul, Brazil. Bruno also received a Bachelor's degree in Computer Science in 2008 from the Federal University of Santa Maria, Brazil. Currently, he holds the position of Associate Professor at the Center for Computational Sciences of the Federal University of Rio Grande - FURG, working both in undergraduate and graduate programs. His primary research interests involve network traffic prediction, as well as security and privacy in software-defined network environments.

**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 MDPI and/or the editor(s). MDPI 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.