International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V70I1P235
Volume 70 Issue 1, 299-312, January, 2022

DOS-DDOS Attacks Predicting : Performance Comparison of The Main Feature Selection Strategies

Kawtar Bouzoubaa[1], Youssef Taher[2], Benayad Nsiri[3]

[1]*M2CS, Research Center STIS, National Graduate School of Arts and Crafts of Rabat (ENSAM), Mohammed V University, Rabat, Morocco*

[2]*Center Of Guidance and Planning (COPE) Rabat, Morocco*

[3]*M2CS, Research Center STIS, National Graduate School of Arts and Crafts of Rabat (ENSAM), Mohammed V University, Rabat, Morocco*

[1]kawtar.bouzoubaa@um5r.ac.ma, [2]youssef.taher@laposte.net, [3]benayad.nsiri@um5.ac.ma

Abstract *— Today, cyberattacks are one of the largest threats to individuals and societies across the globe.*
Dealing with the complexity and variety of these threats by using classical solutions such as software/hardware firewalls and antivirus / antimalware becomes insufficient and presents many drawbacks.
To support and improve the efficiency of these traditional solutions, Machine Learning (ML) models can play an increasing role in detecting, preventing or disrupting cyberattacks at the earliest stage (near real-time).
In this context, the focus of the present paper is to analyze and assess how cyberattacks Feature Selection Strategies (FSS) can support the improvement of these ML models performances applied to cybersecurity, especially the case of DOS-DDOS attacks.
By reviewing more than one hundred and three references and using a hierarchical analysis model based on three levels of performance analysis, this paper has compared the performances of four main types of Feature Selection Methods (TFSM) (first analysis level), the main DOS-DDOS Features Selection Sub-Methods (FSSM) used in each TFSM (second analysis level) and DOS-DDOS datasets widely used in ML cybersecurity projects (third analysis level).

Keywords *— Cybersecurity; DOS-DDOS Attacks; Machine Learning; Feature Selection Strategies.*

I. INTRODUCTION

Due to the interconnectedness and digitalization, cyberattacks on systems and networks can have a devastating impact on individuals [1], businesses [2], the economy [3] and governments [1] everywhere. In table 1, we have summarized an example of the most significant and direct impacts of these threats.
Today, modern attacks against business applications, networks and IT infrastructure are becoming more complex and increasingly sophisticated.
Protecting organizations against these modern risks by using classical solutions such as software/hardware firewalls [4], antivirus/antimalware [5], etc. becomes insufficient and suffers from many drawbacks related to security performances and implementation costs (may not protect fully against internal threats, low detection accuracy, high running time, higher implementation costs, etc.).
To deal with these limitations, the integration of Artificial Intelligence (AI), especially Machine Learning (ML) technology, with these traditional security techniques can offer many advanced security benefits [6].

TABLE I. EXAMPLES OF CYBER ATTACKS TARGETS AND IMPACTS

Examples of cyber attacks targets	Examples of impacts
Individuals	▪ Stealing people's money and their identity. ▪ Social and psychological impacts ▪ ...etc.
Business & economy	▪ Higher costs from operational disruption and altered business practices. ▪ The substantial world economy and financial loss ▪ …etc.
Governments	▪ The real threat to democracy (by hacking, for example, political party computer system). ▪ Breaches of national security secrets ▪ …etc.

For example, by learning from network traffics experiences, ML algorithms can predict when and where many future advanced cyber attacks may occur at an earlier stage (near real-time) [7]. Consequently, organizations and governments can be more reactive and preventive.

However, the high-dimensionality of internet traffic data [8] is a significant challenge in these ML projects applied to cybersecurity (irrelevant and noisy attacks features, high computational time, the low performance of models..., etc.).

For example, the dimensionality of datasets covers much variety of modern attacks, as the case of the UNSW-NB15 dataset [9], which is based on more than forty-five cyberattacks features (forty-eight features for the UNSW-NB15 dataset).

To overcome the challenges mentioned above, cybersecurity predicting models arise from the output of the first ML key process: cyberattacks features selection.

Recently, to enhance and optimize this ML key process, researchers have been interested in conducting important improvements and innovations. This interest was demonstrated by more than one hundred recent references from research projects (≃one hundred and three references in DOS-DDOS feature selection between the years 2015 and 2021).

The remainder of this paper is structured as follows: Section II exposes the impacts of feature selection on cybersecurity datasets commonly used in Machine Learning. Related work is presented with results and discussion in section III. In section IV, we have made the conclusions.

II. IMPACT OF FEATURE SELECTION ON CYBERSECURITY DATASETS USED IN MACHINE LEARNING

A. Cyberattacks datasets challenges

Cyberattacks Machine Learning (ML) projects are based upon optimization techniques applied to cyberattacks datasets generated by internal and external network traffic ([10],[11]). This traffic often generates a high dimensional flow of data (high number of cyberattacks features) with noisy, redundant and irrelevant information. Table 2 summarizes an example of the higher dimensionality of benchmark datasets widely used in ML attacks projects.

The Knowledge Discovery in Databases (KDD_99) [10] was developed by the University of California in 1999. The dimensionality of this dataset is forty-one. This dataset has serious limitations, such as redundant and duplicate data samples unbalanced distribution of attacks between the training and testing phase.

The NSL_KDD dataset [12] was created by the University of California in 2009 with forty-one cyberattacks features. By eliminating redundant instances in the training and testing phase, NSL_KDD was an updated version of the KDD'99. However, it does not give a comprehensive representation of a modern low footprint of attack environment [9].

The UNSW_NB15 dataset [9] was created by the Australian Centre for Cyber Security (ACCS) with forty-fight

TABLE II. EXAMPLE OF CYBERATTACKS DATASETS DIMENSIONALITY

Dataset	Creation Year	Dimensionality	Number of attacks types
KDD'99	1999	41	4
NSL_KDD	2009	41	4
UNSW_NB15	2015	48	9
CIC_IDS 2017	2017	78	6
CIC_IDS 2018	2018	79	6

Cyberattacks features. UNSW_NB15 is composed of real modern behaviours and synthetically cyberattacks activities.

The CIC-IDS 2017 / CIC-IDS 2018 datasets [13] were created by the Canadian Institute of Cybersecurity (CIC) with seventy-eight / seventy-nine cyberattacks features. Compared to the earlier datasets, these datasets are composed of a new range of attacks generated from real network traffic. These datasets suffer from many drawbacks, such as a large number of data instances which complicates the cyberattacks data processing. The missing and redundant data records and high-class imbalance implies a low accuracy and high FPR of the used system.

The examples of cyberattacks datasets challenges discussed above impact negatively ML models applied to cybersecurity (degrading the ML models performance, making the process of learning very slow, the interpretability and generalization of ML cybersecurity models becoming very difficult [14]…etc.).

To deal with these challenges, ML cybersecurity models arise from the output of the first ML key process: cyberattacks Feature Selection Process (FSP).

In the paragraph below, we have summarized the principles of the four main strategies applied to FSP to select and optimize the most relevant cyberattacks features.

B. Feature selection on cybersecurity datasets

The existing ML projects applied to cybersecurity demonstrate that the available cybersecurity datasets are composed of an average of attack features between forty-one and seventy-nine (table 2). This great dimensionality is a real challenge of extracting relevant security information in terms of ML models performance, execution time and generalization. One of the key processes commonly used to mitigate these constraints is the cyberattacks Feature Selection Process (FSP).

Based on importance, similarity and performance, this pre-processing ML step identifies the most representative attacks features from the initial network traffic datasets. Consequently, it improves the performance of intrusions detection, reduces the overfitting and improves the generalization of the used ML ([15],[8],[16]).

Applied to ML cybersecurity models, cyberattacks feature selection strategies are principally grouped into four main/basic classes [17]: Filter, Wrapper, Hybrid and Embedded methods (figure 1). Many other feature selection improvements and innovations are derived from these main four classes [18].

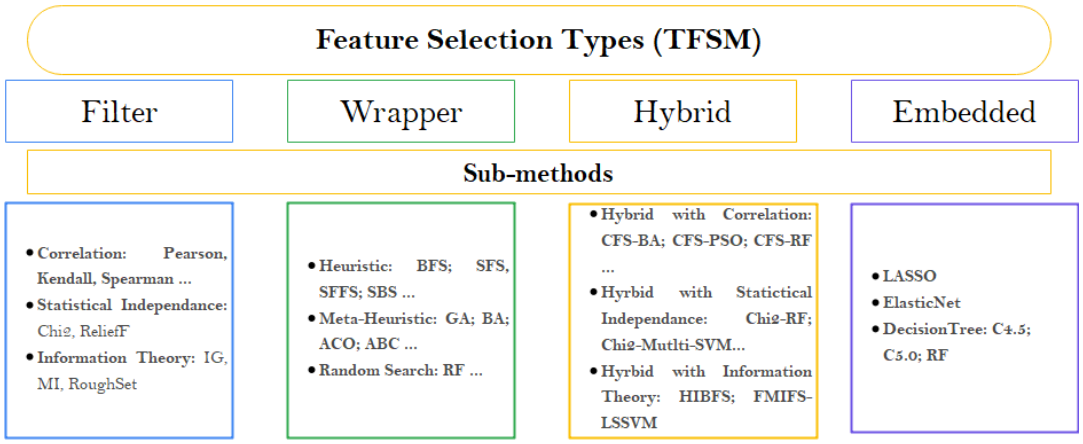

Fig. 1.Taxonomy of the main Feature Selection Types (TFSM) used in cybersecurity ML models.

Regardless of the used ML algorithm, Filter methods evaluate and select relevant attacks features by using different statistical measures such as consistency, correlation and information theory [18].

The Wrapper strategies choose the final subset by using a Learning Algorithm (LA). These strategies follow principally a schema based on two components: Search strategy and evaluation ([18], [19]).

The Hybrid methods combine two approaches: the filter method followed by a wrapper technique. In the first step, the filter method chooses the first subset of attacks features. In the second step, the wrapper process optimizes this first subset ([20], [21]).

The Embedded strategies use a regularization method to select the best attacks features [22]. For example, Lasso, Elastic Net and Random Forest algorithms have been used as embedded strategies by different researchers to select the most appropriate subset attacks features.

III. RELATED WORK

A. Research Objective

Employing Machine Learning (ML) models as part of cybersecurity strategy entails the optimization of attacks Feature Selection Process (FSP).

To select the most relevant attack features from the largest cyberattacks available datasets and provide accurate learning results, many attacks Feature Selection Strategies (FSS) have been proposed by many important and recent research projects.

Selecting and adopting the most appropriate solution among a considerable number of these research investigations are one of the key challenges of ML models applied to cybersecurity.

In this context, and based on the performance analysis model, the main objective of this paper is to build many simple performance dashboards that visually measure and showcases the key performance metrics of various main types of feature

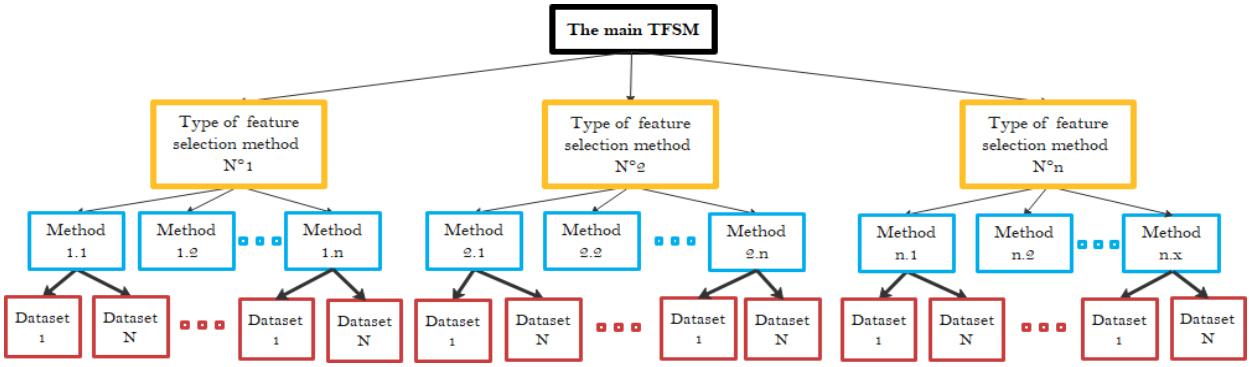

Fig. 2.Used Performance Analysis Model (PAM)

Selection methods (TFSM), the main DOS-DDOS Features Selection Sub-Methods (FSSM) used in each TFSM and DOS-DDOS datasets widely used in ML cybersecurity projects.

The dashboards are based on the assessment of more than one hundred DOS-DDOS feature selection references as a case study. Consequently, the findings presented by this study can customize and simplify the decision-making process in order to meet specific needs of DOS-DDOS cyberattacks Feature Selection Process (FSP) and ML predicting projects.

B. Used performance analysis model

To analyze and represent the relationships between TFSM (as the first level of performance analysis), FSSM (as the second level of performance analysis) and cyberattacks datasets (as the third level of performance analysis), we have used a Hierarchical Analysis Model (HAM) summarized in figure 2. This model allows us to take into account the influences of these three levels on FSS performances as well as the interaction between them.

To evaluate, compare and display the performance indicators of each level, we have used a range of different metrics summarized in table 3. These metrics are Accuracy, Detection Rate, ROC, Recall, FAR, FPR, Specificity, Precision and F-Measure.

TABLE III. PERFORMANCES METRICS USED

Used metrics	Formulas	Descriptions
FAR	FAR = (FPR+ FNR) / 2	The ratio of the misclassification of a case of no attack and classified as attack and vice versa [23]
FPR	FPR= FP / (FP +TN)	The probability when an alert occurs when there is no intrusion [24]
FNR	FNR = FN/ (FN+TP)	The probability that no alert occurs when there is an intrusion
Specificity= TNR	TNR= TN/ (TN+FP)	The percentage of false intrusions correctly classified to the total of negative intrusions existing on the dataset [7]
Recall= (Sensitivity, TPR)	TPR = TP/ (TP+ FN)	The percentage of true intrusions correctly classified to the total of positive intrusions existing on the dataset [7]
Precision	Precision= TP/ (TP+ FP)	The ratio of correctly true intrusions to all classified positive intrusions on the dataset.
Accuracy	Accuracy = (TP +TN) / (TN+FP+FN+TP)	Shows how many of the predictions are correct ([25], [16]).
F-measure	F-measure= 2.(Precision*Recall) / (Precision + Recall)	Computes the score between the precision and recall accuracies of the model for a given threshold [16].
ROC		(Represents the accuracy of the classifier) allows the visualization of the relation between detection rate and false-positive rate of a classifier [26].

C. Results and discussion

a) Performance analysis of the first level: main Types of Feature Selection Methods (TFSM)

As a first step of the experiment, we have started the performances analysis of DOS-DDOS Feature Selection Strategies (FSS) by assessing the impact of TFSM: Filter, Wrapper, Hybrid and Embedded.

By reviewing more than one hundred and three references and analyzing all used strategies by each Type of Feature Selection method (TFSM), we have determined the best value of each used metric for each TFSM.

We have compared the performances of TFSM by calculating the Performance Rate PR given by the equation below:

$$PR = \frac{(\text{Best Metric Value} - \text{Low Metric Value})}{\text{Best Metric Value}}$$

As shown in figure 3, the four main TFSM does not have a significant impact on the first seven used metrics (Accuracy, Recall, Precision, F-Msr, DR, Specificity and ROC).

The results are that the PR rate hasn't exceeded 0,08%, 0,21% for the recall, 0,28% for the precision, 0,25% for the F-Msr, 0,54% for the DR, 0,56% for the ROC and 1,35% for the specificity.

The maximum value of the accuracy (99.98%) was reached in two studies. The first one was based on Filter Correlation (CFS) [27] by selecting 17 DOS-DDOS features. The second one was based on Filter, Wrapper and Hybrid strategies by using CFS, Random Forest (RF) and selecting only the 9 best DOS-DDOS features [28].

The best recall value (100%) was reached in two important investigations. The first one was based on the Wrapper strategy Best First Search (BFS) by selecting 7 DOS-DDOS features [29]. The second one was based on the Hybrid

strategy HFSN by using Naive Bayes (NB) classifier and selecting 2 DOS-DDOS best features [30].
The maximum value of the precision (99.99%) was reached by selecting 12 DOS-DDOS features and combining many Filter strategies (MI and generalized entropy) [31].
Based on the MI-FS method, the best value of F-measure (99.99%) was obtained by the experiment of Ambusaidi et al. (2016) [32] by using 18 DOS-DDOS selected features.

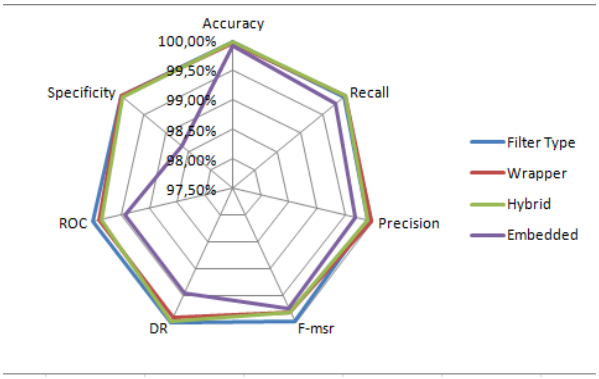

Fig. 3. Best performances of the Main Types of Feature Selection Methods (TFSM)

By selecting the 4 best DOS-DDOS features, the maximum values of DR and ROC metrics (100%) were reached in the study of Binbusayyis et Vaiyapuri (2019) [33] by combining many Filter strategies: CFS, Information Gain (IG), ReliefF, Consistency measure (CBF). This investigation has recorded the best FAR value (0%).

Srivastava (2018) [34] reached the maximum value of ROC measure (100%) by using two strategies: Filter CFS with Wrapper Best First Search (BFS) by selecting 12 DOS-DDOS features.

Manjunatha et al. (2019) [35] and Ahmad and Aziz (2019) [36] have reached the best FPR value (0%). They have combined Mutual Information (MI) with Linear Correlation Coefficient (LCC) and selected only fourteen DOS-DDOS features.
In the study carried out by Ahmad et Aziz (2019) [36], they have selected 13 DOS-DDOS best features by adopting a Hybrid strategy and Particle Swarm Optimization (PSO) as Wrapper strategy.
However, this first level of performance analysis has shown that TFSM has a significant impact on training and test times.
The important differences between the lowest time and the best time values as a function of TFSM are shown in table 4.

TABLE IV. BEST PERFORMANCE VALUES OF TRAIN AND TEST TIME

	TRAIN_TIME (SECOND)	TEST_TIME (SECOND)
FILTER	0.0001	0.0006
WRAPPER	0.00041	0.21
HYBRID	0.03	0.09
EMBEDDED	0.6385	98.44

The lowest training time ($1E^{-4}$s) was reached in two important studies based on Combined Filter methods (CFS) [37] and Symmetric Uncertainty (SU) [38].
The best test time value (0.0006s) was reached in a study based on the Filter strategy Chi-Square (Chi2) by using the Core vector Machine (CVM) approach [39] and selecting only 10 best DOS-DDOS features.

b) Performance analysis of the second level: feature selection sub-methods (FSSM)

In this second level of performance analysis, we analyze and assess with more detail the effect of the main DOS-DDOS Features Selection Sub-Methods (FSSM) used in each Type of Feature Selection method (TFSM).
By grouping FSSM by TSFM (Filter sub-methods, Wrapper sub-methods, Hybrid sub-methods, Embedded sub-methods), the main objective in the second step is to classify the best sub-method that can be used to improve the process of DOS-DDOS attacks predicting.

1) Filter sub-methods

As shown in Table 5, we have selected in this subsection the five most commonly-used Filter Sub-Methods: Correlation, Statistical Independent (SI), Information Theory (IT) and the Combined Filter methods (CF).
We have evaluated, compared and displayed the best value of each performance metric used by these selected Sub-Methods.
According to the above findings (table 5), the best Filter strategy accuracy (99.98%) was recorded on the Correlation sub-method by Madbouly et al. (2016)[27]. In this investigation, the authors have used only 17 best DOS-DDOS features selected by the CFS Filter sub-method.
By selecting 25 DOS-DDOS features, the experiment carried out by Gupta and Kulariya (2016) [39], based on Filter sub-methods CFS and Chi2 has recorded the lowest accuracy value (35.4%).

The maximum values of precision (99.94%) and F-measure (99.99%) were reached by using Information Theory (MI) [31] by selecting 18 DOS-DDOS features.
The lowest precision value (83.8%) was recorded on the Combined Filter sub-method by the investigation carried out by Shahbaz et al. (2016) [36] and Bataghva et al. (2017)[37].
By selecting 28 DOS-DDOS attack features**,** the lowest F-measure value (82.3%) is based on the Information Theory (IG) sub-method [40].

Bhyan et al. (2016) [30] have reached the highest Recall value (99.99%) by selecting the 12 best DOS-DDOS and also using the Information Theory.
Gupta et Kulariya (2016) **[39]** have recorded the worst recall value (0.65%) by using the CFS Filter sub-method, which have selected the 25 best features subset.
By selecting 4 DOS-DDOS features, the investigation is based on combining 4 filter methods (CFS, IG, RelieF, CBF) and carried out by Binbusayyis et Vaiyapuri (2019)[33] has recorded the maximum and minimum values of detection rate metric (100%, 97.6%).

TABLE V. FILTER SUB-METHODS PERFORMANCE ANALYSIS

Sub-methods	Acc	Pr	Re	DR	F-msr	FPR	FAR	SPE	ROC	Train_ Time	Test_ Time
Correlation_ Min	75.2% [12]	92.27% [51]	92.12% [51]	n/a	92.25% [51]	0.030% [46]	0.02% [23]	99.9% [46]	79% [32]	0.04s [32]	0.35s [23]
Correlation_ Max	99.98% [27]	99.42% [51]	99.52% [53]	n/a	99.8% [46]	n/a	0.03% [23]	100% [23]	100% [32]	1356.73s [32]	632.400s [37]
Statistical_ Min	99.05% [36]	96.24% [44]	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.002s [36]	0.0006s [38]
Statistical_ Max	99.95% [38]	n/a	n/a	99.12% [36]	n/a	47.14% [36]	n/a	n/a	99.55% [38]	10.79s [38]	1.21s [36]
Information_ Min	73.79% [54]	85.1% [44]	86.7% [44]	98.76% [31]	82.3% [44]	0.1% [49]	0.2% [47]	n/a	88.4% [44]	0.05s [41]	4.02s [39]
Information_ Max	99.97% [55]	99.94% [31]	99.99% [30]	99.99% [30]	99.99% [31]	10.8% [44]	n/a	99.43% [39]	n/a	4102s [42]	270s [31]
Combined_ Min	35.4% [45]	83.8% ([35]; [43])	0.65% [40]	97.4% ([33];[45])	n/a	0% [33]	0% [45]	16.87% [40]	65% [45]	1E-4s ([35]; [43])	0.345s [40]
Combined_ Max	99.9% [40]	88.2% ([35]; [43])	99.41% [40]	100% [45]	98% [52]	15.9% ([43];[50])	0.0542% [48]	99.79% [40]	100% [45]	344.771s [40]	19.91s [40]

By selecting 14 DOS-DDOS best features, Manjunatha et Gogoi (2019) [34] have realized the minimum values of DR (97.6%) and FPR (0%) by combining the Correlation and Linear Correlation Coefficient Filter sub-method (CFS-LCC).
The worst FPR value (47.7%) was recorded by Divyasree et Sherly (2018) [38] on the Filter sub-method Chi2 by selecting 10 DOS-DDOS features.
The maximum FAR value (0.2%) was reached by using the Information Theoretical Filter method (IT) [42], which has selected a subset of 30 DOS-DDOS features.
However, the investigation of Binbusayyis et Vaiyapuri (2019)[33] has realized the minimum FAR value (0%) by combining four Filter sub-methods (CFS, IG, RelieF, and CBF).

The highest specificity value (100%) was reached by Idhammad et al. (2017) [23] by selecting two DOS-DDOS subsets with 6 and 5 features based on the Correlation sub-method.
The worst specificity value was recorded on the Combined filter methods by the investigation of Gupta and Kulariya (2016) **[39]** with 25 selected features.

The maximum ROC value (100%) was recorded on two filter sub-methods: Correlation (CFS) by selecting DOS-DDOS 11
features [33] and Combined Filter [32]. However, the investigation of Binbusayyis et Vaiyapuri (2019) [33] has realized the worst ROC value (65%).
The minimum value of training time ($1E^{-4}$s) was recorded on Combined Filter Sub-Methods by Shahbaz et al. (2016) [36] and Bataghva et al. (2017) [37].
The worst training time was reached on the Correlation sub-method [33].
Finally, the most interesting test time value (0.0006s) was recorded by using the statistical filter Chi2 and selecting only 17 DOS-DDOS features [42].
The highest test time value (632.400s) was recorded on the correlation sub-method by the experiment of Fitni et Ramli (2020)[43] by selecting the 23 best DOS-DDOS features.

2) Wrapper sub-methods

The main strategies commonly used in wrapper sub-methods [57] are classified into three main categories: Heuristic (HE), Meta-Heuristic (MH) and Random Search (RS).

In this subsection, we evaluate and display the best and lowest values of each performance metric used by these categories.

According to the above findings summarized in table 6, the best accuracy value (99.96%) is recorded by using the Meta-Heuristic category based on Simulated Annealing (SA), Support Vector Machine (SVM) and Decision Tree (DT) [58], by selecting 23 most representatives DOS-DDOS attacks features.

The experiment carried out by [59] based on the Heuristic category and using the Best First search (BFS) and Decision Tree (DT) algorithms have recorded the lowest accuracy value (52.1%) by selecting 20 DOS-DDOS attacks features.

The maximum and minimum values of precision (99.97%, 36.09%) were recorded by the Meta-Heuristic category.

By selecting the 7 best DOS-DDOS features attacks and using the Heuristic category, the highest recall value (100%) was recorded in the study carried out by [29].

The experiment of Khammassi et al. (2017) [60] has reached the worst recall value (4.11%) by using the Meta-Heuristic category with selecting 20 best DOS-DDOS attacks features**.**

The maximum Detection Rate (DR) value (99.9%) was recorded by using the Heuristic category based on Sequential Forward (FS) and Random Forest (SFFS-RF)[61] and selecting the 10 best features.

By selecting 19 DOS-DDOS attacks features, the worst DR value (19.38%) was recorded on the Heuristic category based on the Decision Tree (DT) algorithm [62].

The lowest F-measure value (77.19%) was reached on the Heuristic category by using Ant Colony Optimization (ACO) and selecting 4 important DOS-DDOS features [63].

By using the Meta-Heuristic category based on forwarding Feature Selection and K-nearest Neighbour (FFS-KNN), Soodeh et al. (2019) [64] have reached the best F-measure value (99.8%). Their investigation has selected 11 DOS-DDOS attacks features.

TABLE VI. WRAPPER SUB-METHODS PERFORMANCE ANALYSIS

Sub-methods	Acc	Pr	Re	DR	F-msr	FPR	FAR	SPE	ROC	Train_ Time	Test_ Time
Heuristic_ Min	52.1% [58]	88.11% [77]	72% [67]	19.38% [78]	87.89% [77]	0.1% [60]	0.0006% [66]	82.6% [68]	88.7% [68]	0.05s [67]	0.21s [60]
Heuristic_ Max	99.9% ([66]; [67])	99.8% [63]	100% [67]	99.99% [60]	99.4% [68]	0.5% [75]	72.23% [58]	100% [67]	99.9% [68]	514.7s [68]	4.63s [72]
Meta-Heuristic_ Min	73.97% [64]	36.09% [59]	4.11% [59]	89.45% [79]	77.19% [62]	0.001% [65]	0.1% [71]	91.76% [74]	n/a	0.02s [73]	0.19s [71]
Meta-Heuristic_ Max	99.92% [80]	99.97% [59]	99.98% [59]	99.61% [65]	89.82% [62]	21.2% [64]	6.39% [59]	99.67% [74]	n/a	1795.94s [64]	13.84s [71]
Random_ Min	80.15% [69]	81.18% [69]	96.75% [70]	91.5% [69]	86.03% [69]	0.09% [76]	0.2% [69]	n/a	n/a	1.41*10-‘s [69]	n/a
Random Max	99.96% [57]	99.44% [70]	99.45% [70]	n/a	99.41% [70]	n/a	n/a	n/a	n/a	2590.6s [70]	1358.1s [69]

The highest FPR value (21.2%) was reached on the Meta-Heuristic category based on the Genetic Algorithm (GA)[65] by selecting 15 DOS-DDOS attacks features.

Mazini et Mahdavi (2019) [66] have recorded the best FPR value (0.001%) by using the Meta-Heuristic category based on Artificial Bee Colony (ABC) and selecting the 25 best DOS-DDOS features.

Al-Jarrah et al. (2014) [67] have recorded the minimum value of FAR (0.0006%) by selecting 15 DOS-DDOS features and using the Heuristic category based on Random Forest (RF) and Forward Feature Ranking (RF-FSR).

The higher FAR value (72.5%) was recorded by using the Heuristic category based on Decision Tree (DT) and selecting 20 DOS-DDOS attacks features [59].

The investigation carried out by Kavitha et al. (2010) [68] has realized the highest specificity value (100%) by selecting 7 DOS-DDOS features and using the Heuristic category based on Backward FS (BFS).

The lowest specificity value (86.9%) was recorded on the Heuristic category with the 6 best DOS-DDOS features selected by the CfsSubsetEval and BFS algorithms [69]. This experiment has realized two important ROC values. Depending on the used classifier, the highest value was (99.9%), and the lowest one was (88.7%).

By selecting the 9 best DOS-DDOS attacks features, the most interesting value of training time ($1.41*10^{-4}$s) was recorded on the Random category based on Differential Evolution (DE) [70].

By selecting 10 representative DOS-DDOS features with Random Forest (RF), Alrowaily et al. (2019) [71] have realized the highest training and test times (2590.6s, 1358.1s). However, the best test time (0.19s) was recorded on the Meta-Heuristic category based on Improved Clonal Search Algorithm (ICSA) by using a subset of 21 selected DOS-DDOS attacks features [72].

3) Hybrid sub-methods

In this subsection, we have analyzed the performances of Hybrid sub-methods by grouping the commonly used strategies into three main categories: Filter CFS, Statistical Filter (SF) and Information Theoretical (IT).

According to the above findings summarized in table 7, the maximum value of accuracy (99.98%) was reached on the CFS-RF algorithm by selecting 12 DOS-DDOS attacks features [82].

The lowest value of accuracy (80.07%) was recorded on the statistical sub-methods Chi2 and RF by selecting 7 DOS-DDOS attacks features [83].

By selecting 12 features, the minimum precision value (99.8%) was recorded on the CFS Bat algorithm [84].

The best precision value (99.9%) was recorded on CFS and Naïve Bayes (HFSN) [30] by selecting 2 features. This strategy has also realized the maximum Recall value (100%).

By selecting the 9 best DOS-DDOS features, Gu et al. (2019) [85] have recorded the lowest recall value (96.1%) by using Information Theoretical (IT) and Supervised K-means algorithm.

The minimum and maximum values of DR (1.76% [86**]**, 99.99% [28]) were reached on the Hybrid category based on the CFS sub-method. The lowest value was recorded by selecting a subset of the 13 best DOS-DDOS features and using CFS and K-means algorithms. The highest value was reached by selecting the 12 best features based on CFS-RF.

Song et al. (2019) **[**83] have realized the lowest F-measure value (80.27%) by selecting 7 DOS-DDOS features based on the statistical sub-method Chi2-RF. The highest F-measure value (99.8%) was recorded on CFS sub-method (CFS-BA) [84].

The highest FPR value (30.5%) was recorded on the Information Theoretical sub-method SKM-HFS [85]. The lowest value (0%) was recorded on the CFS sub-method based on CFS and Particle Swarm Optimization (PSO)[36].

The highest FAR value (2.46%) [87] was recorded by selecting 22 optimal DOS-DDOS features and using the statistical sub-methods Linear Discriminant Analysis (LDA), Chi2 and the modified Bayesian Net (LDA, Chi2, Modified BN). The best value (0.1%) was recorded on CFS sub-method CFS-BA [84].

The minimum value of training time (0.03s) was registered on the CFS sub-method [30]. The slowest training time (10235s) was recorded on the statistical sub-method Chi2 and Multiclass SVM algorithm by selecting 31 DOS-DDOS important features [88].

TABLE VII. HYBRID SUB-METHODS PERFORMANCE ANALYSIS

Sub-methods	Acc	Pr	Re	DR	F-msr	FPR	FAR	SPE	ROC	Train_ Time
Hybrid with CFS_ Min	95.03% FGLCC-LCA [90]	99.8% CFS-BAT [83]	99.9% CFS-PSO [34]	1.76% K-means-CFS [85]	95.46% FGLCC-LCA [90]	0% CFS-PSO [34]	0.1% CFS-BAT [83]	n/a	n/a	0.03 s HFSN [29]
Hybrid with CFS_ Max	99.98% CFS-RF [81]	99.9% HFSN [29]	100% HFSN [29]	99.99% CFS-RF [81]	99.8% CFS-BAT [83]	1.67% FGLCC-LCA [90]	7.03% K-means-CFS [85]	99.97% CFS-RF [81]	83.28% FGLCC-LCA [90]	43.50s FGLCC-LCA [90]
Hybrid with Statistical_ Min	80.07% Chi2-RF [82]	n/a	n/a	71.21% Chi2-RF [82]	n/a	n/a	0.13% Chi2_Multi-SVM [87]	n/a	n/a	0.11s Chi2-RF [82]
Hybrid with Statistical_ Max	98.87% Chi2_Multi-SVM [87]	n/a	n/a	97.78% LDA-Chi2-Modified BN [86]	80.27% Chi2-RF [82]	n/a	2.46% LDA-Chi2-Modified BN [86]	n/a	n/a	10235s Chi2_Multi-SVM [87]
Hybrid with Information_Min	88.36% MI-BGSA [20]	n/a	n/a	86.30% MI-BGSA [20]	n/a	0.10% HIBFS [89]	n/a	n/a	n/a	58.12s HIBFS [89]
Hybrid with Information_Max	99.70% HIBFS [89]	n/a	96.50% SKM-HFS [84]	99.46% FMIFS-LSSVM [88]	n/a	30.5% SKM-HFS [84]	n/a	n/a	n/a	603.6s FMIFS-LSSVM [88]

The best test time value (0.009s) was reached the CFS sub-method CFS-BA [84]. The highest value (276s) was recorded on the statistical sub-method Flexible Mutual Information (FMI) by using Least Square Support Vector Machine (FMIFS-LSSVM) [89].

i) Embedded sub-methods

In this subsection, we have analyzed the performance of Embedded sub-methods by grouping the commonly used strategies into three main categories: Least Absolute Shrinkage and Selection Operator (LASSO), Embedded Ensemble Optimal Feature Selection Algorithm (EEOFSA) and DT based Embedded.

According to the above findings summarized in table 8, the maximum accuracy and precision value (99.88%) was recorded by the experiment of Serkani et al. (2019)[92]. By selecting the 9 best DOS-DDOS features, these important metric values are obtained by using the DT sub-method and Least Square SVM algorithm (DT-LSSVM).

The lowest accuracy value of (87.3%) was recorded by [22] by selecting 20 features and combining Lasso-ElasticNet and CFS filter methods.

By selecting 17 DOS-DDOS features, the best values of precision (97%), F-measure (97.37%) and specificity (98.65%) were registered on the Lasso sub-method based on SVM and L1-regularization algorithms[11].

TABLE VIII. EMBEDDED SUB-METHODS PERFORMANCE ANALYSIS

Sub-methods	Acc	Pr	Re	DR	F-msr	FPR	FAR	SPE	ROC	Train_ Time	Test_ Time
LASSO_ Min	87.3% [22]	73.45% [22]	n/a	92.81% [92]	86.55% [22]	2.50% [92]	n/a	n/a	n/a	0.6385s [94]	n/a
LASSO_ Max	97.08% [11]	97% [11]	97.7% [11]	99.21% [94]	97.37% [11]	2.85% [92]	n/a	98.65% [11]	99.44% [11]	8296.92s [92]	98.44s [92]
EEOFSA	98.67% [90]	n/a	n/a	99.46% [89]	n/a	n/a	0.32% [89]	n/a	n/a	0.68s [89]	n/a
DT_ Max	89.65% DT-LSSVM [91]	n/a	90.23% DT-LSSVM [91]	n/a	n/a	0.3% DSSVM [93]	0.099% DT-LSSVM [91]	n/a	94.70% DT-LSSVM [91]	0.099s DT-LSSVM [91]	0.1447s DT-LSSVM [91]
DT_Min	99.88% DT-LSSVM [91]	n/a	99.88% DT-LSSVM [91]	97.2% DSSVM [93]	n/a	n/a	10.86% DT-LSSVM [91]	n/a	99.85% DT-LSSVM [91]	209.92s DSSVM [93]	483.39s DSSVM [93]

The lowest values of precision (73.45%) and F-measure (86.55%) by using the Lasso sub-method [22].

The maximum and minimum values of the Recall measure (99.88%, 90.23%) were recorded on the DT sub-method [92].

By using EEOFSA and selecting 13 DOS-DDOS features, Vekatarathinam et al. (2018) [90] have recorded the highest value of DR (99.46%). The lowest value of this metric (92.81%) was recorded on Lasso embedded sub-method based on Linear Nearest Neighbour Lasso Step (LNNLS-KH) [93]. This strategy has selected a subset of 14 DDOS features.

By using the distance sum SVM method (DSSVM) and selecting 5 DOS-DDOS important features, the minimum value of FPR (0.3%) was recorded on the DT embedded sub-method [94].

The highest value of this metric (2.85%) was recorded by the model carried out by [93] by selecting a subset of 10 DOS-DDOS features. This model has recorded the slowest training time (8296.92s).

However, Serkani et al. (2019) [92] have recorded the best value of train time (0.099s) by using the DT embedded sub-method and selecting 9 DDOS features.

By selecting only 5 DDOS features and using the DT embedded sub-method, the worst value of test time (483.39s) was recorded on the experiment carried out by [94]. The best value of this time (0.1447s) was recorded on the DT embedded sub-method [92] by selecting the 9 best DOS-DDOS features.

This section has shown that the embedded sub-method based DT has realized the best performances on most metrics.

c) Performance analysis of the third level: used cyberattacks datasets

In this section (the third level of performance analysis), we have paid particular attention to the impact of DOS-DDOS datasets on feature selection strategies discussed above.

We have selected the main five used datasets widely used in ML cybersecurity projects: KDD_99, NSL_KDD, UNSW_NB15, CIC-IDS2017 and CIC-IDS2018.

Indeed, the maximum accuracy (99.98%) was recorded on the KDD'99 dataset [27]. However, the lowest accuracy (73.79%) was recorded on the CIC-IDS2017 dataset [55].

The interesting precision value (99.9%) was attempted on the KDD [60]. The lowest value of this metric (36.09%) was recorded on the UNSW_NB15 dataset [30].

The best recall value (100%) was recorded on two datasets: the KDD [29] CIC-IDS2017 [30**]** datasets. The lowest value was reached by [60] on the UNSW_NB15 dataset.

The maximum value of DR (100%) was found by Binbusayyis and Vaiyapuri (2019) [33] on the KDD dataset. The lowest value (42.1%) was found by Bagui et al. (2019)[86] on the UNSW_NB15 dataset.

The best F-measure value (99.99%) was recorded on the KDD dataset [32]. However, the minimum value (77.19%) was found on the NSL_KDD dataset by Housseizadeh Aghdam and Kabiri (2016) [63].

The lowest rate of FPR (0%) was recorded on the KDD [36] and NSL_KDD [35] datasets. However, the worst value of this metric (47.14%) was recorded on the KDD dataset [39].

The best FAR value (0%) was recorded on the KDD dataset [33]**.** And the highest value (10.86%) was found on the UNSW_NB15 dataset [92].

The highest value of DOS-DDOS specificity (100%) was detected on three datasets: KDD[29]**, (**NSL_KDD and UNSW_NB15) by Idhammad et al. (2017) [23]. The lowest value of specificity (97.33%) was recorded on the NSL_KDD [40].

Binbusayyis et Vaiyapuri (2019) [33] has recorded the best ROC value (100%) on three different datasets: KDD, NSL_KDD and the CIC_IDS2017. But the same method has shown a lower value of ROC measure (65%) on the UNSW_NB dataset.

The best train time ($1E^{-4}$ s) was recorded on the NSL_KDD by Shahbaz (2016) [37] and Bataghva (2017)[38]**.** Nevertheless, the same dataset has recorded the highest training model time (8296.92s)[93].

TABLE IX. DATASETS PERFORMANCE ANALYSIS

Datasets	Acc	Pr	Re	DR	F-msr	FPR	FAR	SPE	ROC	Train_Time	Test_Time
KDD_ Min	92.13% [40]	81.66 % [63]	93.8% [50]	95.23% [90]	89.82 % [62]	0.0% [34]	0.0% [45]	99.79 % [40]	92% [32]	0.0006s [36]	0.09s [83]
KDD_ Max	99.98% ([27]; [81])	99.97 % [59]	100% [67]	100% [45]	99.99 % [31]	47.14 % [36]	1.40% [95]	100% [67]	100% [45]	603.6s [88]	483.30s [93]
NSL_KDD_ Min	79.09% [64]	79.22 % [62]	69.28 % [40]	71.21% [82]	77.19 % [62]	0% [33]	0.001% [45]	97.33 % [40]	88.4% [44]	$1E^{-4}$s ([35] ; [43])	0.09s [83]
NSL_KDD_ Max	99.90% [96]	99.69 % [96]	99.9% [47]	99.99% ([30]; [31])	99.8% [63]	17.2% [64]	0.22% [58]	100% [23]	100% [45]	8296.92s [92]	98.44s [92]
UNSW_NB15_ Min	81.42% [59]	36.09 % [59]	4.11% [59]	42.1% [85]	n/a	0.03% [46]	0.0069 % [55]	n/a	65% [45]	0.010s [45]	0.1627s [91]
UNSW_NB15_ Max	99.88% [58]	n/a	97% [23]	99.31% [61]	n/a	14.8% [34]	10.86% [91]	100% [23]	94.70 % [91]	0.7715s [91]	0.46s [23]
CIC_IDS2017_ Min	73.79% [54]	97.5% [83]	96.50 % [84]	92.81% [92]	98.1% [83]	2.50% [92]	0.002% [45]	n/a	n/a	0.009s [45]	0.26s [83]
CIC_IDS2017_ Max	99.9% [42]	99.9% [29]	100% [29]	99.9% [45]	99.8% [46]	30.5% [84]	2.4% [83]	99.9% [46]	100% [45]	2590.6s [70]	1358.1s [70]
CIC_IDS2018	98.8% [37]	n/a	n/a	n/a	97.9% [37]	n/a	n/a	n/a	n/a	n/a	632.4s [37]

The lowest test time (0.09) was recorded by Zhou et al . (2019) [84] on two datasets: KDD and NSL_KDD. The highest test consuming time (1358.1s) was found on the CIC-IDS2017 dataset [71].

In conclusion, we have observed that the KDD dataset has realized the best performances on most metrics. But we have to notice that the KDD dataset has known an over-representation of the DOS attack class. And the standard machine learning algorithms were designed to be applied to balanced data. Then on unbalanced data, the results will bias towards the majority class.

IV. CONCLUSION

Nowadays, the integration of Artificial Intelligence (AI), especially Machine Learning (ML) technologies, with traditional security techniques is transforming the roles of cybersecurity. Indeed, ML models applied to cybersecurity allow private and public sectors to innovate and efficiently transform the security of their systems strategies.

To improve these ML models, the innovation of cyberattacks Feature Selection Process (FSP) can play an increasing role.

In this context, and based on a Hierarchical Analysis Model (HAM) by using the main FSP performance indicators, we highlight in the present paper how the four main types of Feature Selection Methods (TFSM) (HAM first analysis level), the main DOS-DDOS Features Selection Sub-Methods (FSSM) used in each TFSM (HAM second analysis level) and DOS-DDOS datasets widely used in ML cybersecurity projects (HAM third analysis level) can support the improvement of these ML models performances applied to the case of DOS-DDOS attacks.

To meet the specific needs of DOS-DDOS cyberattacks Feature Selection Process (FSP) and ML predicting projects, we have built seven simple performance dashboards that visually measure and showcase the key performance metrics of various main strategies used in FSP.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the research supervisor Pr. Youssef TAHER to suggest the main idea of this project, Pr. Benayad NSIRI and my family for the continuous support of my study and related research, for their patience, motivation, and immense knowledge.

REFERENCES

[1] M. Bada et J. R. C. Nurse, Chapter 4 - The social and psychological impact of cyberattacks, in Emerging Cyber Threats and Cognitive Vulnerabilities, V. Benson et J. Mcalaney, Éd. Academic Press, (2020) 73-92. doi: 10.1016/B978-0-12-816203-3.00004-6.

[2] SBIR.gov, Tutorial 1: The Impact of Cybercrime on Small Business |, SBIR.gov. https://www.sbir.gov/tutorials/cyber-security/tutorial-1 (consulté le oct. 05, 2021).

[3] B. Cashell, W. D. Jackson, M. Jickling, et B. Webel, The Economic Impact of Cyber-Attacks, (2004).

[4] Y. V. Srinivasa Murthy, K. Harish, V. Varma, K. Sriram, et B. Revanth, Hybrid Intelligent Intrusion Detection System using Bayesian and Genetic Algorithm (BAGA): Comparative Study, International Journal of Computer Applications, 99 (2014) 1-8, doi: 10.5120/17342-7808.

[5] J. Sen et S. Mehtab, Machine Learning Applications in Misuse and Anomaly Detection, in Machine Learning Applications in Misuse and Anomaly Detection, IntechOpen, (2020). doi: 10.5772/intechopen.92653.

[6] N. Alqudah et Q. Yaseen, Machine Learning for Trafic Analysis: A Review, Warsaw Poland, 170 (2020) 911-916.

[7] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, et M. Xu, A Survey on Machine Learning Techniques for Cyber Security in the Last Decade, IEEE Access, 8 (2020) 222310-222354, doi: 10.1109/ACCESS.2020.3041951.

[8] R. Abdulhammed, H. Musafer, A. Alessa, M. Faezipour, et A. Abuzneid, Features Dimensionality Reduction Approaches for Machine Learning Based Network Intrusion Detection, Electronics, 8(3) (2019), doi: 10.3390/electronics8030322.

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

[10] S. K. Sahu, S. Sarangi, et S. K. Jena, A detailed analysis on intrusion detection datasets, in 2014 IEEE International Advance Computing Conference (IACC), févr. (2014) 1348-1353. doi: 10.1109/IAdCC.2014.6779523.

[11] D. H. Hagos, A. Yazidi, Ø. Kure, et P. E. Engelstad, Enhancing Security Attacks Analysis Using Regularized Machine Learning Techniques, in IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), mars (2017) 909-918. doi: 10.1109/AINA.2017.19.

[12] L. Dhanabal et S. P. Shantharajah, A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms, International Journal of Advanced Research in Computer and Communication Engineering, 4(6) (2015) 446, doi: 10.17148/IJARCCE.2015.4696y.

[13] A. Thakkar et R. Lohiya, A Review of the Advancement in Intrusion Detection Datasets, Procedia Computer Science, 167 (2020) 636-645, doi: 10.1016/j.procs.2020.03.330.

[14] J. Miao et L. Niu, A Survey on Feature Selection, Procedia Computer Science, 91 (2016) 919-926, doi: 10.1016/j.procs.2016.07.111.

[15] Y. Feng, H. Akiyama, L. Lu, et K. Sakurai, Feature Selection for Machine Learning-Based Early Detection of Distributed Cyber Attacks, in 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), août (2018) 173-180. doi: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00040.

[16] W. Mostert, K. M. Malan, et A. P. Engelbrecht, A Feature Selection Algorithm Performance Metric for Comparative Analysis, Algorithms, 14(3) (2021), doi: 10.3390/a14030100.

[17] M. Torabi, N. I. Udzir, M. T. Abdullah, et R. Yaakob, A Review on Feature Selection and Ensemble Techniques for Intrusion Detection System, International Journal of Advanced Computer Science and Applications, 12(5) (2021)16.

[18] V. R. Balasaraswathi, M. Sugumaran, et Y. Hamid, Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms, J. Commun. Inf. Netw., 2(4) (2017) 107-119, doi: 10.1007/s41650-017-0033-7.

[19] S. Wang, J. Tang, et H. Liu, Feature Selection, (2016) 1-9. doi: 10.1007/978-1-4899-7502-7_101-1.

[20] H. Bostani et M. Sheikhan, Hybrid of binary gravitational search algorithm and mutual information for feature selection in intrusion detection systems, Soft Comput, 21(9) (2017) 2307-2324 doi: 10.1007/s00500-015-1942-8.

[21] S. Solorio-Fernández, J. A. Carrasco-Ochoa, et J. Fco. Martínez-Trinidad, A review of unsupervised feature selection methods , Artif Intell Rev, 53(2) (2017) 907-948, X doi: 10.1007/s10462-019-09682-y.

[22] S. Murugesan, Application of Machine Learning Models for Network Intrusion Detection Systems Based on Feature Selection Approach »,

masters, Dublin, National College of Ireland, (2019) (2021). [En ligne]. Disponible sur: http://norma.ncirl.ie/4302/

[23] M. Idhammad, K. Afdel, et M. Belouch, DoS Detection Method based on Artificial Neural Networks, International Journal of Advanced Computer Science and Applications (IJACSA), 8(4) (2017), doi: 10.14569/IJACSA.2017.080461.

[24] M. Labonne, Anomaly-based network intrusion detection using machine learning, Institut Polytechnique de Paris, (2020).

[25] R. Magán-Carrión, D. Urda, I. Diaz-Cano, et B. Dorronsoro, Towards a Reliable Comparison and Evaluation of Network Intrusion Detection Systems Based on Machine Learning Approaches, Applied Sciences, 10 (2020) 1775, doi: 10.3390/app10051775.

[26] A. Khraisat, I. Gondal, P. Vamplew, et J. Kamruzzaman, Survey of intrusion detection systems: techniques, datasets and challenges, Cybersecur, 2(1) (2019) 20, doi: 10.1186/s42400-019-0038-7.

[27] A. I. Madbouly et T. M. Barakat, Enhanced relevant feature selection model for intrusion detection systems, Int. J. Intell. Eng. Inform., 4(1) (2016) 21-45, doi: 10.1504/IJIEI.2016.074499.

[28] M. H. Kamarudin, C. Maple, et T. Watson, Hybrid feature selection technique for intrusion detection system, International Journal of High-Performance Computing and Networking, 13 (2019) 232, doi: 10.1504/IJHPCN.2019.097503.

[29] B. Kavitha, Dr. S. Karthikeyan, et P. Sheeba Maybell, An Ensemble Design of Intrusion Detection System for Handling Uncertainty Using Neutrosophic Logic Classifier, Know.-Based Syst., 28 (2012) 88-96, doi: 10.1016/j.knosys.2011.12.004.

[30] M. Babiker, E. Karaarslan, et Y. Hoşcan, A hybrid feature-selection approach for finding the digital evidence of web application attacks, Turkish Journal of Electrical Engineering and Computer Sciences, 27 (2019) 4102-4117, doi: 10.3906/elk-1812-18.

[31] M. H. Bhuyan, D. K. Bhattacharyya, et J. K. Kalita, A multi-step outlier-based anomaly detection approach to network-wide traffic, Information Sciences, 348 (2016) 243-271, doi: 10.1016/j.ins.2016.02.023.

[32] M. A. Ambusaidi, X. He, P. Nanda, et Z. Tan, Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm, IEEE Trans. Comput., 65(10) (2016) 2986-2998, doi: 10.1109/TC.2016.2519914.

[33] A. Binbusayyis et T. Vaiyapuri, Identifying and Benchmarking Key Features for Cyber Intrusion Detection: An Ensemble Approach, IEEE Access, 7 (2019) 106495-106513, doi: 10.1109/ACCESS.2019.2929487.

[34] D. Srivastava, Feature Classification and Outlier Detection to Increased Accuracy in Intrusion Detection System, International Journal of Applied Engineering Research, vol. 13, p. 7249-7255, nov. 2018.

[35] B. A. Manjunatha et M. T. Gogoi, Data Mining based Framework for Effective Intrusion Detection using Hybrid Feature Selection Approach, IJCNIS, 11(8) (2019) 1-12, doi: 10.5815/ijcnis.2019.08.01.

[36] T. Ahmad et M. N. Aziz, Data preprocessing and feature selection for machine learning intrusion detection systems, ICIC Express Letters, 13 (2019) 93-101, doi: 10.24507/icicel.13.02.93.

[37] M. B. Shahbaz, X. Wang, A. Behnad, et J. Samarabandu, On efficiency enhancement of the correlation-based feature selection for intrusion detection systems, in IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), (2016) 1-7. doi: 10.1109/IEMCON.2016.7746286.

[38] M. Bataghva, Efficiency and Accuracy Enhancement of Intrusion Detection System Using Feature Selection and Cross-layer Mechanism, Electronic Thesis and Dissertation Repository, University of Western Ontario, Ontario, (2017). [En ligne]. Disponible sur: https://ir.lib.uwo.ca/etd/5160

[39] T. H. Divyasree et K. K. Sherly, A Network Intrusion Detection System Based On Ensemble CVM Using Efficient Feature Selection Approach, Procedia Computer Science, 143 (2018) 442-449, doi: 10.1016/j.procs.2018.10.416.

[40] G. P. Gupta et M. Kulariya, A Framework for Fast and Efficient Cyber Security Network Intrusion Detection Using Apache Spark, Procedia Computer Science, 93 (2017) 824-831, doi: 10.1016/j.procs.2016.07.238.

[41] M. Abdullah, A. Balamash, A. Al-Shannaq, et S. Almabdy, Enhanced Intrusion Detection System using Feature Selection Method and Ensemble Learning Algorithms, International Journal of Computer Science and Information Security, 16 (2018) 48-55.

[42] Z. Foroushani et Y. Li, Intrusion Detection System by Using Hybrid Algorithm of Data Mining Technique, in ICSCA 2018: Proceedings of the 2018 7th International Conference on Software and Computer Applications, Kuantan, Malaysia, févr. (2018) 119-123. doi: 10.1145/3185089.3185114.

[43] S. M. Othman, F. M. Ba-Alwi, N. T. Alsohybe, et A. Y. Al-Hashida, Intrusion detection model using machine learning algorithm on Big Data environment, Journal of Big Data, 5(1) (2018) 34, doi: 10.1186/s40537-018-0145-4.

[44] Q. R. S. Fitni et K. Ramli, Implementation of Ensemble Learning and Feature Selection for Performance Improvements in Anomaly-Based Intrusion Detection Systems, in IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), (2020) 118-124. doi: 10.1109/IAICT50021.2020.9172014.

[45] B. Setiawan, S. Djanali, et T. Ahmad, Increasing Accuracy and Completeness of Intrusion Detection Model Using Fusion of Normalization, Feature Selection Method and Support Vector Machine, International Journal of Intelligent Engineering and Systems, 12 (2019) 378-389, doi: 10.22266/ijies2019.0831.35.

[46] K. K. Myint et N. S. M. Kham, Feature Selection in Hybrid Intrusion Detection System, févr. (2016) (2021). [En ligne]. Disponible sur: https://onlineresource.ucsy.edu.mm/handle/123456789/2364

[47] Kurniabudi, D. Stiawan, Darmawijoyo, M. Y. Bin Idris, A. M. Bamhdi, et R. Budiarto, CICIDS-2017 Dataset Feature Analysis With Information Gain for Anomaly Detection, IEEE Access, 8 (2020) 132911-132921, doi: 10.1109/ACCESS.2020.3009843.

[48] M. Hammad, N. Hewahi, et W. Elmedany, T-SNERF: A novel high accuracy machine learning approach for Intrusion Detection Systems », IET Information Security, 15(2) (2021) 178-190, doi: https://doi.org/10.1049/ise2.12020.

[49] D. Kshirsagar et S. Kumar, An efficient feature reduction method for the detection of DoS attack, ICT Express, (2021), doi: 10.1016/j.icte.2020.12.006.

[50] J. Yogendra Kumar et Upendra, Intrusion Detection using Supervised Learning with Feature Set Reduction, International Journal of Computer Applications, 33(6) (2011) 22-31.

[51] Akashdeep, I. Manzoor, et N. Kumar, A feature reduced intrusion detection system using ANN classifier, Expert Systems with Applications, 88, (2017) 249-257, doi: 10.1016/j.eswa.2017.07.005.

[52] G. Farahani, Feature Selection Based on Cross-Correlation for the Intrusion Detection System, Security and Communication Networks, vol. 2020, p. e8875404, sept. 2020, doi: 10.1155/2020/8875404.

[53] Y. Wahba, E. ElSalamouny, et G. ElTaweel, Improving the Performance of Multi-class Intrusion Detection Systems using Feature Reduction, International Journal of Computer Science Issues, 12(3) (2015) 255-262.

[54] M. A. Siddiqi et W. Pak, Optimizing Filter-Based Feature Selection Method Flow for Intrusion Detection System, Electronics, 9(12) (2020), doi: 10.3390/electronics9122114.

[55] S. Sambangi et L. Gondi, A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression, Proceedings, 63(1) (2020), doi: 10.3390/proceedings2020063051.

[56] A. Binbusayyis et T. Vaiyapuri, Comprehensive analysis and recommendation of feature evaluation measures for intrusion detection, Heliyon, 6(7) (2020), doi: 10.1016/j.heliyon.2020.e04262.

[57] K. Bouzoubaa, Y. Taher, et B. Nsiri, Predicting DOS-DDOS Attacks: Review and Evaluation Study of Feature Selection Methods based on Wrapper Process, International Journal of Advanced Computer Science and Applications, (2021).

[58] S.-W. Lin, K. Ying, C. Lee, et Z.-J. Lee, An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion

detection, Appl. Soft Comput., 12(10) (2012) 3285-3290, doi: 10.1016/j.asoc.2012.05.004.
[59] M. A. Umar et Z. Chen, Effects of Feature Selection and Normalization on Network Intrusion Detection. (2020). doi: 10.36227/techrxiv.12480425.
[60] C. Khammassi et S. Krichen, A GA-LR wrapper approach for feature selection in network intrusion detection, Computers & Security, vol. 70 (2017) 255-277, doi: 10.1016/j.cose.2017.06.005.
[61] J. Lee, D. Park, et C. Lee, Feature Selection Algorithm for Intrusions Detection System using Sequential Forward Search and Random Forest Classifier, KSII Transactions on Internet and Information Systems, 11(10) (2017) 5132-5148.
[62] M. A. Umar et C. Zhanfang, Effects of Feature Selection and Normalization on Network Intrusion Detection, (2020), doi: 10.36227/techrxiv.12480425.v2.
[63] M. Hosseinzadeh Aghdam et P. Kabiri, Feature Selection for Intrusion Detection System Using Ant Colony Optimization, International Journal of Network Security, 18 (2016) 420-432.
[64] H. Soodeh et A. Mehrdad, The hybrid technique for DDoS detection with supervised learning algorithms, Computer Networks, 158 (2019) 35-45, doi: 10.1016/j.comnet.2019.04.027.
[65] D. P. Gaikwad et R. C. Thool, Intrusion Detection System Using Bagging with Partial Decision TreeBase Classifier, Procedia Computer Science, 49 (2015) 92-98, doi: 10.1016/j.procs.2015.04.231.
[66] M. Mazini, B. Shirazi, et I. Mahdavi, Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms, Journal of King Saud University - Computer and Information Sciences, 31(4) (2019) 541-553, doi: 10.1016/j.jksuci.2018.03.011.
[67] O. Y. Al-Jarrah, A. Siddiqui, M. Elsalamouny, P. D. Yoo, S. Muhaidat, et K. Kim, Machine-Learning-Based Feature Selection Techniques for Large-Scale Network Intrusion Detection, in IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW), (2014) 177-181. doi: 10.1109/ICDCSW.2014.14.
[68] B. Kavitha, S. Karthikeyan, et B. Chitra, Efficient Intrusion Detection with Reduced Dimension Using Data Mining Classification Methods and Their Performance Comparison, in Information Processing and Management, V. V. Das, R. Vijayakumar, N. C. Debnath, J. Stephen, N. Meghanathan, S. Sankaranarayanan, P. M. Thankachan, F. L. Gaol, et N. Thankachan, Éd. Berlin, Heidelberg: Springer Berlin Heidelberg, 70 (2010) 96-101. doi: 10.1007/978-3-642-12214-9_17.
[69] S. Alabdulwahab et B. Moon, Feature Selection Methods Simultaneously Improve the Detection Accuracy and Model Building Time of Machine Learning Classifiers, Symmetry, 12(9) (2020), doi: 10.3390/sym12091424.
[70] F. Almasoudy, W. Al-Yaseen, et A. Idrees, Differential Evolution Wrapper Feature Selection for Intrusion Detection System, Procedia Computer Science, 167 (2019) 1230-1239, doi: 10.1016/j.procs.2020.03.438.
[71] M. Alrowaily, F. Alenezi, et Z. Lu, Effectiveness of Machine Learning Based Intrusion Detection Systems, in Security, Privacy, and Anonymity in Computation, Communication, and Storage, Cham, (2019) 277-288. doi: 10.1007/978-3-030-24907-6_21.
[72] C. Yin, L. Ma, et L. Feng, Towards accurate intrusion detection based on improved clonal selection algorithm, Multimed Tools Appl, 76(19) (2017) 19397-19410, , doi: 10.1007/s11042-015-3117-0.
[73] L. Yinhui, J. Xia, S. Zhang, J. Yan, X. Ai, et K. Dai, An efficient intrusion detection system based on support vector machines and gradually feature removal method, Expert Systems with Applications, 39 (2012) 424-430, doi: 10.1016/j.eswa.2011.07.032.
[74] W. Xing-zhu et H. Changde, ACO and SVM Selection Feature Weighting of Network Intrusion Detection Method, 9 (2015) 129-270. doi: 10.14257/ijsia.2015.9.4.24.
[75] M. Samadi Bonab, A. Ghaffari, F. Soleimanian Gharehchopogh, et P. Alemi, A wrapper-based feature selection for improving performance of intrusion detection systems, International Journal of Communication Systems, 33 (2020), doi: 10.1002/dac.4434.
[76] F. Zhang et D. Wang, An Effective Feature Selection Approach for Network Intrusion Detection, in 2013 IEEE Eighth International Conference on Networking, Architecture and Storage, (2013) 307-311. doi: 10.1109/NAS.2013.49.
[77] M. N. Chowdhury, K. Ferens, et M. Ferens, Network Intrusion Detection Using Machine Learning, in Computer Science, (2016) 30-35. [En ligne]. Disponible sur: /paper/Network-Intrusion-Detection-Using-Machine-Learning-Chowdhury-Ferens/a30c16f5598ba18ffd7d9c533515cf671d54b382
[78] H. Polat, O. Polat, et A. Cetin, Detecting DDoS Attacks in Software-Defined Networks Through Feature Selection Methods and Machine Learning Models, Sustainability, 12(3) (2020) 1-16.
[79] M. A. Umar, C. Zhanfang, et Y. Liu, A Hybrid Intrusion Detection with Decision Tree for Feature Selection, arXiv:2009.13067 [cs], Consulté le, (2021). [En ligne]. Disponible sur: http://arxiv.org/abs/2009.13067
[80] A. Enache, V. Sgârciu, et M. Togan, Comparative Study on Feature Selection Methods Rooted in Swarm Intelligence for Intrusion Detection, in 2017 21st International Conference on Control Systems and Computer Science (CSCS), (2017) 239-244. doi: 10.1109/CSCS.2017.40.
[81] W. Jun, L. Taihang, et R. Rongrong, A real time IDSs based on artificial Bee Colony-support vector machine algorithm, Suzhou, Jiangsu, China, (2010) 91-96. doi: 10.1109/IWACI.2010.5585107.
[82] M. H. Kamarudin, C. Maple, T. Watson, et N. S. Safa, A LogitBoost-Based Algorithm for Detecting Known and Unknown Web Attacks, IEEE Access, 5 (2017) 26190-26200, doi: 10.1109/ACCESS.2017.2766844.
[83] J. Song, W. Zhao, Q. Liu, et X. Wang, Hybrid feature selection for supporting lightweight intrusion detection systems, J. Phys.: Conf. Ser., 887 (2017) 012031, doi: 10.1088/1742-6596/887/1/012031.
[84] Y. Zhou, G. Cheng, S. Jiang, et M. Dai, An Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier. (2019).
[85] Y. Gu, K. Li, Z. Guo, et Y. Wang, Semi-Supervised K-Means DDoS Detection Method Using Hybrid Feature Selection Algorithm, IEEE Access, 7 (2019) 64351-64365, 2019, doi: 10.1109/ACCESS.2019.2917532.
[86] S. Bagui, E. Kalaimannan, S. Bagui, D. Nandi, et A. Pinto, Using machine learning techniques to identify rare cyber-attacks on the UNSW-NB15 dataset, Security and Privacy, 2 (2019), doi: 10.1002/spy2.91.
[87] I. S. Thaseen et Ch. A. Kumar, Intrusion Detection Model Using Chi Square Feature Selection and Modified Naïve Bayes Classifier, in Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC – 16'), Cham, (2016) 81-91. doi: 10.1007/978-3-319-30348-2_7.
[88] I. Sumaiya Thaseen et C. Aswani Kumar, Intrusion detection model using fusion of chi-square feature selection and multi class SVM, Journal of King Saud University - Computer and Information Sciences, vol. 29(4) (2017) 462-472, doi: 10.1016/j.jksuci.2015.12.004.
[89] A. Kumar K S, A. K., L. M. N., et S. M., A Novel Approach for Intrusion Detection System Using feature Selection algorithm, 13 (2017) 1963-1976.
[90] R. Venkatarathinam, Dr. V. Cyril Raj, et Dr. V. Victo Sudha George, A Novel Hybrid Iterative Backward Feature Selection Framework for Intrusion Detection System, International Journal of Applied Engineering Research, 13(5) (2018) 2780-2785.
[91] S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, et H. Karimipour, Cyber intrusion detection by combined feature selection algorithm, Journal of Information Security and Applications, 44 (2019) 80-88, doi: 10.1016/j.jisa.2018.11.007.
[92] E. Serkani, H. Gharaee Garakani, et N. Mohammadzadeh, Anomaly Detection Using SVM as Classifier and Decision Tree for Optimizing Feature Vectors, The ISC International Journal of Information Security, 11(2) (2019) 159-171., doi: 10.22042/isecure.2019.164980.448.
[93] X. Li, P. Yi, W. Wei, Y. Jiang, et L. Tian, LNNLS-KH: A Feature Selection Method for Network Intrusion Detection, Security and

Communication Networks, (2021) e8830431, doi: 10.1155/2021/8830431.

[94] C. Guo, Y. Zhou, Y. Ping, Z. Zhang, G. Liu, et Y. Yang, A distance sum-based hybrid method for intrusion detection, Applied Intelligence, 40 (2014), doi: 10.1007/s10489-013-0452-6.

[95] A. S. Alzahrani, R. A. Shah, Y. Qian, et M. Ali, A novel method for feature learning and network intrusion classification, Alexandria Engineering Journal, 59(3) (2020) 1159-1169, doi: 10.1016/j.aej.2020.01.021.

[96] N. Acharya et S. Singh, An IWD-based feature selection method for intrusion detection system, Soft Comput, vol. 22, nº 13, p. 4407-4416, juill. 2018, doi: 10.1007/s00500-017-2635-2.

[97] H. Nkiama, S. Z. M. Said, et M. Saidu, A Subset Feature Elimination Mechanism for Intrusion Detection System, International Journal of Advanced Computer Science and Applications (IJACSA), 7(4) (2016), doi: 10.14569/IJACSA.2016.070419.