

Received 6 May 2024, accepted 1 June 2024, date of publication 4 June 2024, date of current version 19 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3409413

RESEARCH ARTICLE

Analytical Validation and Integration of CIC-Bell-DNS-EXF-2021 Dataset on Security Information and Event Management

GYANA RANJANA PANIGRAHI¹, (Member, IEEE),
PRABIRA KUMAR SETHY^{1,2}, (Senior Member, IEEE),
SANTI KUMARI BEHERA³, (Senior Member, IEEE),
MANOJ GUPTA⁴, (Member, IEEE), **FARHAN A. ALENIZI⁵**,
PANNEE SUANPANG⁶, (Member, IEEE),
AND AZIZ NANTHAAMORNPHONG⁷, (Member, IEEE)

¹Department of Electronics, Sambalpur University, Sambalpur, Odisha 768019, India

²Department of Electronics and Communication Engineering, Guru Ghasidas Vishwavidyalaya, Bilaspur, Chhattisgarh 495009, India

³Department of Computer Science and Engineering, Veer Surendra Sai University of Technology (VSSUT), Burla 768018, India

⁴Department of Electrical Engineering, School of Studies of Engineering and Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur, Chhattisgarh 495009, India

⁵Department of Electrical Engineering, College of Engineering, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

⁶Department of Information Technology, Faculty of Science and Technology, Suan Dusit University, Bangkok 10700, Thailand

⁷College of Computing, Prince of Songkla University, Phuket 83120, Thailand

Corresponding authors: Aziz Nanthaamornphong (aziz.n@phuket.psu.ac.th) and Prabira Kumar Sethy (prabirsethy.05@gmail.com)

This study is supported via funding from Prince sattam bin Abdulaziz University project number (PSAU/2024/R/1445).

ABSTRACT Contemporary culture presents a substantial obstacle for cyber security experts in the shape of software vulnerabilities, which, if taken advantage of, can jeopardize the Confidentiality, Integrity, and Availability (CIA) of any system. Data-driven and modern threat intelligence tools can enhance cyber security, bolster resilience, and foster innovation across cloud, multi-cloud, and hybrid platforms. As a result, performance evaluation and accuracy verification have become essential for Security Information and Event Management (SIEM) to prevent cyber threats. The SIEM system offers threat intelligence, reporting, and security incident management through the collection and analysis of event logs and other data sources that are specific to events and their context. We propose a hybrid strategy to address threat intelligence, reporting, and security incident management consisting of two layers that utilize a predefined set of characteristics. Here, we use RStudio to assess how well a hybrid intrusion detection system (HIDS) handles the CIC-Bell-DNS-EXF-2021 dataset. Furthermore, we have incorporated our developed model into Multi-Criteria Decision Analysis Methods (MCDM) to enhance the methods' ability to identify complex DNS exfiltration attacks using machine learning algorithms: RF-AHP (RA), KNN-TOPSIS (KT), GBT-VIKOR (GV), and DT-Entropy-TOPSIS (DET). We consider several factors during the work, including accuracy, absolute error, weighted average recall, weighted average precision, kappa value, logistic loss, and root mean square deviation (RMSD). We use the Machine-Automated Model function to integrate and validate the models. According to the findings, GV has the highest level of accuracy, with a rate of 99.52%, while KT has the lowest level of authenticity, with a rate of 93.65%. Furthermore, these findings illustrate enhanced performance metrics for multiclass classification in comparison to previous approaches.

The associate editor coordinating the review of this manuscript and approving it for publication was Antonio J. R. Neves¹.

• **INDEX TERMS** Cyber security, SIEM, CIC-Bell-DNS-EXF-2021, HIDS, performance assessment, analytical validation, MCDM, machine learning.

I. INTRODUCTION

Cyberattacks are becoming longer and more intricate, and their harmful intentions are increasing over time. In today's society, any network connection must incorporate advanced security protocols to ensure dependable communication between multiple enterprises [1], [2], [3], [4]. An Intrusion Detection System (IDS) is a term used to describe any software or hardware that monitors network traffic for any unauthorized or malicious activity [5], [6], [7]. DNS is frequently utilized for the purpose of illicitly acquiring important corporate information and establishing a command-and-control pathway with a deceptive website. Due to the significant role played by DNS services, businesses often configure their firewalls to permit DNS traffic. This facilitates the transmission of encrypted data to a command server by malicious individuals following the compromise of certain items. This security measure can be implemented at various critical points within a network to mitigate the risk of potentially harmful actions [4]. Every instance of online traffic to and from connected devices is scrutinized and compared against recognized threats. DNS exfiltration is a method employed by hackers to stealthily extract confidential information from a network by concealing it within DNS queries or responses. To improve the identification of DNS exfiltration in a SIEM system, an important automated strategy is necessary. Here are some important areas need to develop in comparison to the current methodologies as threat intelligence integration, behavioral analysis, anomaly detection, and user and entity behavior analytics. Any infraction or detrimental behavior is reported utilizing a Security Information and Event Management (SIEM) component. This is the only reason where we planned to design a two-tier system with explicitly defined attributes to detect and differentiate between the removal of large amounts of illegal data and the rerouting of DNS traffic. The proposed approach is applicable to devices with abundant resources and the hybrid nature of the model, which combines both stateful and stateless characteristics. The proposed model will be integrated into stateful-based detection systems to identify sophisticated attacks on DNS-EXF-2021 systems, using a substantial dataset of DNS traffic exfiltration [3], [7], [8], [9]. This action would be undertaken with the aim of enhancing the security of the current systems. The naming system allocates IP addresses to domain names in order to enable browsers to access Internet resources. When an individual accesses "website1.eg1.net," their web browser or program transmits the hostname to a name server. The Domain Name System (DNS) is responsible for converting hostnames into their corresponding IP addresses. A DNS resolver verifies if the hostname is already stored in the host machine's cache. If it is not, the client-side browser proceeds to communicate with name servers until it retrieves the user's IP address. DNS communication is vulnerable to exploitation by hackers due to the lack of control over UDP port '53' in

organizations, unlike protocols such as HTTP, email, or FTP [1], [4]. Cybercriminals acquire a domain with the intention of initiating various types of attacks, such as DDoS attacks, protocol abnormalities, tunneling, escalation, and mirror traffic, among others [9]. The exfiltration of DNS data takes place in computer networks through the DNS protocol. The hacker possesses a server that is infected with malware, as well as a domain that directs to it. During the process of exfiltration, the attacker can manipulate the host to request the domain belonging to the attacker. When the DNS resolver directs the query, it establishes a tunnel between the hacker and the target, enabling the hacker to gain control of the host and carry out various malicious activities. The detection of data exfiltration attacks is challenging due to the congestion of DNS. DNS-based attacks present substantial risks to the security of organizations and can result in a range of malicious activities, such as unauthorized data extraction, disruptions caused by denial of service (DoS) or distributed denial of service (DDoS) attacks, fraudulent attempts to obtain sensitive information (phishing), dissemination of malware, and communication for the purpose of controlling compromised systems. It is crucial to identify and counteract these attacks in SIEM systems to safeguard network infrastructure, secure sensitive data, and maintain uninterrupted business operations.

This work utilizes both stateful and stateless DNS to rapidly identify malicious DNS traffic and evaluate the performance of SIEM systems by analyzing instances of DNS data exfiltration. One can install a hybrid threat detection system based on MCDM (Multi-Criteria Decision Making) on devices that have sufficient resources. This system can be extended to specifically identify detection systems that have stateless characteristics. We evaluate our proposed model against various Machine Learning (ML) models, including Random Forest (RF), K-Nearest Neighbor (KNN), Gradient Boosted Trees (GBT), and Decision Tree (DT), to showcase its efficacy. The GBT-VIKOR model achieves an accuracy rate of 99.52%, while the RF-AHP model achieves an accuracy rate of 98.70%.

Our proposed method enables the real-time detection of DNS data exfiltration attacks and the identification of resourceful traffic from an industry-standard network. This work proposes a two-tier framework as a possible solution to the aforementioned problems. Initially, a novel framework has been proposed to effectively rank the classification models into SIEM. A decision matrix, which has been approved, was created for this research by combining four distinct classification models with four additional multi-evaluation criteria. The MCDM (Multi-Criteria Decision Making) approach is employed to prioritize and select classification models. Initially, the assessment criteria, which are established with the aid of Analytic Hierarchy Process (AHP), constitute the initial segment of Multiple Criteria

Decision Making (MCDM) [10], [11]. Another technique, called TOPSIS and VIKOR, has been utilized to evaluate and rank different categorization models. The second layer selects the top two categorization models based on the findings of the first layer. The DT-Entropy-TOPSIS classification algorithm is utilized to construct the model for identifying instances of misuse incidents and threats. The model for identifying abnormal cyber threats is constructed using the K-nearest neighbors' algorithm. These models are derived from the initial layer.

The major contributions of this article are as follows. Using a hybrid detection model for security information and event management, this work suggests using a new adaptive multilayer framework to select the most effective DNS data exfiltration attacks to address the issues like threat intelligence integration, behavioral analysis, anomaly detection along with user and entity behavior analytics. The work employs machine learning to evaluate an MCDM-based hybrid threat detection system that uses the CIC-Bell-DNS-EXF-2021 for misuse and anomaly detection to improve SIEM performance. We assessed the proposed framework using common classification metrics as Accuracy (ACC), Absolute Error (AE), Weighted Average Recall (WAR), Weighted Average Precision (WAP), Kappa Value (KV), Logistic Loss (LL), and Root Mean Square Deviation. These tests show that resourceful devices can apply MCDM-based classification methods. Since most DNS traffic is benign in real life, the stateless and stateful methods would be useful. While earlier systems detected stateless behavior, the current method has the ability to detect stateful behavior. The method utilizes RStudio auto-pattern prediction for trend and behavior analysis in AI/ML, text analytics, and predictive modeling. This platform employs cloud-based ROI-centric IT software to prepare and preprocess data, as well as build and test models using block coding [28], [29]. Due to its enhanced classification accuracy, this method can be verified as a standalone module for integration into the SIEM application to effectively address the current issues and resolve the identified problems. High detection rates may be due to biases or the dataset's focus on simple attack vectors. We will expand our dataset pool, add more sophisticated attack simulations, and improve our model architectures to adapt to evolving cyber threats. We used machine learning algorithms to analyze the CIC-Bell-DNS-EXF-2021 dataset for SIEM. We acknowledge that simple attack scenarios in the dataset may limit us. We used rigorous analytical validation and invite further scrutiny and refinement of our methodologies to address this concern.

The subsequent article is structured in the following manner. Section II pertains to the examination of related studies. Section III provides an overview of the methodology. Section IV provides an overview of proposed approaches. Section V outlines the experimental analysis. Sections VI and VII present the experimental results and discussion. Sections VIII present the statistical findings, while the conclusions and future work are discussed last.

II. RELATED STUDIES

The issue of correctly assessing a SIEM's effectiveness has recently received much attention. It is currently regarded as amongst the most significant obstacles in cyber security [3], [10], [20], [21]. This suggests that SIEM requires specialized detection techniques, frameworks, and tools, including the capability to assist in quickly and accurately identifying security breaches. SIEM is an acronym that stands for Security Information and Event Management. It is a software program that offers immediate analysis of security alarms produced by apps and network devices. Security Information and Event Management (SIEM) systems gather log and event data from several sources inside an organization's IT infrastructure, including servers, firewalls, routers, and apps. Subsequently, this data is cross-referenced and scrutinized to detect any security risks and occurrences. SIEM systems commonly include functionalities such as log management, event correlation, alerting, and reporting, which aid companies in efficiently identifying, examining, and addressing cybersecurity risks. Machine learning algorithms have been integrated into various systems, including XGBoost, Python, Reccsys, Mahout, and Azure. This study investigates the usability of RStudio for HIDS machine learning [12]. A hybrid intrusion detection system (IDS) integrates multiple detection techniques to offer comprehensive security against diverse types of threats. Typically, these systems combine signature-based detection, anomaly-based detection, and occasionally rule-based detection methods. This approach will evaluate four different classifiers on the CIC-Bell-DNS-EXF-2021 dataset [3], [7], [13], [14]. Table 1 provides a comprehensive assessment of different datasets used for SIEM systems, utilizing various Machine Learning (ML) classification models. Although the main topic of the discussion has been DNS exfiltration, it is crucial to contemplate how the suggested techniques can be applied more broadly to identify various forms of attacks within SIEM systems. In this context, generalization refers to the ability of the detection mechanisms to identify different types of malicious activity beyond DNS exfiltration. The proposed hybrid techniques have several aspects that can be generalized, including behavioral analysis, threat intelligence integration, packet capture and analysis, and user and entity behavior analytics. While the process of analyzing and incorporating CIC datasets into SIEM systems is not new in the field of cybersecurity, its importance lies in the ability to apply these techniques to identify various types of attacks through a hybrid system which itself is beyond DNS exfiltration. SIEM systems are essential for protecting organizational assets from evolving cyber threats by utilizing various datasets, integrating threat intelligence, and consistently enhancing detection methodologies. Organizations can bolster their defenses against DNS-based attacks and seamlessly incorporate the CIC-Bell-DNS-EXF-2021 dataset into their SIEM environment by conducting a meticulous security analysis, creating a comprehensive threat model, and implementing specific mitigations. This methodology guarantees that the dataset is utilized

TABLE 1. Comparison between existing methods and our proposed method.

References	SIEM Integration	Dataset	Algorithms	MCDM Analysis Methods	Metrics	Detection System	Issues
[6]	No, Phase specific	UNSW-NB and Network TON_IoT	DT, GBT, MLP, LSTM, GRU, GIWRF	NO	F1-Score	IDS	Enhanced Dataset Diversity
[7]	No, Phase specific	NSL-KDD, and CSE-CIC-ID2018	DNN-DAE, AE, SAE,	Hyperparameter optimization	Accuracy	NIDS	Integration with Threat Intelligence Feeds
[13]	No, Phase specific	DNS EXF 2021	SVM, KNN, NB, DT, RG, IF	NO	Accuracy	DNS Channel Attacks	Real-Time Analysis and Response
[15]	Yes, Hybrid	CERT r4.2	KNN	Entropy-Vikor	Accuracy	Insider threat detection system	User-Centric Security Monitoring
[16]	No, Phase specific	Custom	ML	No	Accuracy	IDS	Robustness Evaluation
[19]	No, Phase specific	Custom	ML, DL, NLP	No	accuracy	Content validation	Adversarial Testing
Our	Yes, Hybrid	CIC-Bell-DNS-EXF-2021	ML	AHP, TOPSIS, VIKOR, and Entropy-Topsis	Accuracy	DNS Exfiltration	Regulatory Compliance

efficiently to enhance the identification of potential dangers, the handling of incidents, and the overall state of security. The comparison between existing methods and our proposed method can be referred in the following Table 1.

III. METHODOLOGY

This section will give a brief idea about the CIC dataset, ML techniques and classification algorithms, and multi-criteria decision analysis methods (MCDM).

A. DATASET

The classification methods were evaluated using the DNS-EXF-2021 dataset made accessible on UNB’s website [8], [9]. CIC-Bell-DNS2021 is designed to simulate real-world conditions by simulating overall benign traffic and various malicious domain types [17], [18], [19]. Even while the new dataset fixes some of the problems inherent in the older one, it still does not accurately portray the current actual networks. This is due to the dearth of publicly available data sets that may be used by network-specific SIEMs [20]. Alternatively, the researchers as an effective standard data set to categorize the various intrusion detection-based SIEM technologies. The primary reason for this is that both the training and the test datasets include records with suitable instances, providing research assessment outcomes that are similar and consistent [21]. In addition, since the training dataset does not contain any models identical to other examples, the tendency to favor often may be avoided. In contrast, these cases are not included, which prevents a bias towards the approaches that give improved detection rates on the examples that occur often. Some examples are picked across each difficulty level group, which is inversely related to the number of instances found in the preceding CIC-Bell-DNS-2021 described in Table 2. Because of this, the CIC-Bell-DNS-EXF-2021 is a

TABLE 2. Statistics of dataset.

Feature Class	Heavy Attack		Feature Class	
	Subclasses	Stateful Features	Subclasses	
Stateless Features	FQDN_count subdomain_length upper lower numeric entropy	rr_type rr_count rr_name_length rr_name_entropy rr_type_frequency rr_	loadmodule, rootkit, perl, xterm	
labels	Heavy Benign spy, multihop, phf, xlock, imap	distinct_domains reverse_dns	sendmail, xsnoop, httptunnel	
special labels	Light Attack nmap, portsweep, ipsweep, mscan	distinct_ns a_records unique_country unique_asn unique_ttl distinct_ip	saint, sendmail, snmpguess	
labels_max labels_average longest_word sld				
Light Benign longest_word len	snmpgetattack, xsnoop,	tfl_mean tfl_variance	named, ftp_write	

good way to measure how well different learning algorithms work. The dataset is highly regarded as a valuable resource due to its controlled environment, extensive coverage of the Domain Name System (DNS), large volume, diverse range of data, annotated labels, and the ability to utilize the dataset for various purposes such as development, training, testing, and empirical studies.

B. MACHINE LEARNING TECHNIQUES AND CLASSIFICATION ALGORITHMS

1) RF - AHP

Many decision trees capable of classifying independently are included in the random forest. The class that receives the

most votes is the one that is taken to represent the model's prediction. Because it draws on the capacity of several trees to make classifications, random forest is an effective method for increasing accuracy. The formation of lowly correlated decision trees inside the random forest is the most important thing that can be done in any case. In that case, the errors caused by the separate decision trees might build up, leading to erroneous classification. AHP calculates the relative weight of information security risk components. By sorting element weights to limit the number of indicators, the ideal indicators, which offer a strong basis for relevant metrics, can be chosen. Utilize randomization and bagging so that you may construct stochastic decision trees capable of providing a high level of accuracy, as exhibited in Table 3.

TABLE 3. RF parameters.

Variables	Trees Count	Criterion	Max. Depth
Values	160	Gini Impurity	16

2) KNN-TOPSIS

K-Nearest Neighbor is one more ML approach that is extensively utilized. This algorithm takes data from several classes and uses it to train a classifier of a new sample. The distance between the new test sample and all of the other points that already exist is computed by the classifier. Topsis was utilized to quantify data and prioritize criteria. High priority includes database and database concerns and challenges, inner application security, online billing security, and monitoring harmful software. It helps extract prioritized criteria from a decision-maker matrix and handles hierarchy, ambiguity, and many bars. TOPSIS combines process parameters depending on their near to the optimal solution. This strategy produced the optimal answer closest to the positive ideal and farthest from the negative. It uses linear normalization. It assigns a category to the newly sampled point according to the one corresponding to its immediate neighbors in the dataset, as presented in Table 4.

TABLE 4. KNN parameters.

Variables	K
Values	8

3) GBT-VIKOR

Gradient Boost tree uses series-connected decision trees. Each tree in the series strives to minimize the error caused by the three trees that came before it. The greedy machine learning technique, gradient boost tree, was developed to prevent prediction errors. The internal modules of the algorithm are subjected to the punishments imposed by the regularisation techniques. VIKOR uses vector normalisation. Although the sequential approach takes more time, it provides a high level of accuracy when applied to classification issues. The VIKOR

method calculates weight stability intervals for the experts' compromise option. The emphasis is on rating and choosing options. Table 5 outlines the GBT model's parameters as below.

TABLE 5. GBT parameters.

Variables	Tree Counts	Max. Depth	Min. Rows	@ Learning	@ Sample
Scores	80	16	16	0.16	1.6

4) DT- ENTROPY-TOPSIS

A decision tree is an example of an algorithm used in supervised machine learning and may be used to address issues of classification and regression. The objective weights of the indicators are calculated using entropy theory, and the subjective weights are utilized to fine-tune the results. TOPSIS is set up to deliver the ultimate assessment result by analyzing the evaluation data and indication weighting. This method improves upon TOPSIS by including indicators' objective and subjective importance. The data representation in this approach is a tree, and each leaf node holds a class label, whereas each inside node of the tree is where the characteristics are shown in Table 6.

TABLE 6. DT parameters.

Variables	Criterion	Max. Depth
Scores	Gini Impurity	16

C. MULTI-CRITERIA DECISION ANALYSIS METHODS (MCDM)

MCDM is a multi-objective variant of the theory of choice. MCDM evaluates options with conflicting criteria. MCDM can consolidate opposing standards into a special universal assessment [1], [5], [10], [15], [18]. A new paradigm for risk assessment in sociotechnical enterprises has lately evolved, namely SIEM. Because of this, SIEM is seen to be more suited to complicated socio-technical systems. SIEM is far more difficult to measure and model due to its multi-criteria structure and the inclusion of both descriptive and analytical hidden elements. Using the RA, KT, GV, and DET approaches, this project aims to build an ML hybrid MCDM model for measuring resilience [8], [22], [23] through various risk assessment comes based on MCDM, as summarized in Table 7.

D. MEASURES OF PERFORMANCE

Within the scope of this study, the following measures of performance have been utilized.

1) ACCURACY (ACC)

Accuracy refers to the proportion of correct predictions produced by a model after it has been put to the test. A learning

TABLE 7. A summary of the several weighted risk assessment approaches based on MCDM.

Techniques	Advantages	Disadvantages
AHP	User-friendly. Data Lite. The functional structure fits multiple-sized challenges platform agnostic. AHP is computationally more efficient than other methods.	Alternatives and metrics dynamic interactions. This may result in inconsistent assessment and ranking criteria. Rank Reversal. The weight assignment procedure is increasingly difficult with more evaluation.
TOPSIS	Straightforward pragmatism. Narrative coherence. Superior and computationally efficient. Ability to evaluate interim solution performance mathematically.	Desirable and undesirable numbers do not affect computations based on Euclidean distance. Strong divergence of one indication from the optimal answer affects outcomes. The strategy works when indicator values change much; the method works well when the signs of each option are very different.
VIKOR	VIKOR is used to tackle pertinent issues in distinct ways. It recognizes the best option immediately and reduces pairwise evaluations. Various choices and qualities are accessible.	Absence of weighing operationalization and reliability testing. It ignores semantic relatedness. As well as, performance restrictions may not regulate the process.
Entropy-TOPSIS	The larger the dispersion, the more information may be gleaned.	It cannot combine several behaviors; a single criterion can be assessed. If there are more criteria, there will be more pairwise comparisons.

algorithm’s confusion matrix is used to evaluate the model’s accuracy. When used, it allows for a general validation of the model, provided that the dataset is balanced. In machine learning algorithms for DNS exfiltration detection, accuracy means the model can correctly classify DNS queries as benign or malicious.

2) ABSOLUTE ERROR (AE)

It is the fraction of a classifier’s incorrect predictions, where the errors equal false and true positives. Various DNS exfiltration detection machine learning algorithms can benefit from absolute error as evaluation metrics, model tuning, anomaly detection, threshold selection, and feature importance.

3) WEIGHTED AVERAGE RECALL (WAR)

The recall gauges expected positives. A weighted mean is an average memory with probability weights. This is not simply means of recall. The algorithm’s ability to correctly identify all instances of a class out of all instances of that class is called recall in machine learning.

4) WEIGHTED AVERAGE PRECISION (WAP)

The level of trust in the efficiency of the adapted model may be determined by measuring the precision. Given the available information, it is the likelihood of properly forecasting a positive event. The weight that is equivalent to the

class probability is taken into consideration by the weighted mean precision method. It helps in precision, weighted average, evaluation of model performance and tuning model parameters.

5) KAPPA VALUE (KV)

Cohen’s kappa is a metric that determines how closely examples categorized by an artificial intelligence algorithm are similar to the underlying data [24], [25]. This statistic has a value that can vary from 0 to 1, with 0 denoting an entire disagreement and one denoting a perfect agreement. It is deemed more dependable than % agreement. It helps in finding classification task, ground truth, model prediction, comparison, evaluation, and iterative improvement.

6) LOGISTIC LOSS (LL)

The negative average of estimated probabilities indicates how near the projection may be to actuality; inverting the logarithm yields the desired result. It helps in finding probabilistic interpretation and its sensitivity, gradient descent optimization, and regularization.

7) ROOT MEAN SQUARE DEVIATION (RMSD)

RMSD is residuals’ standard deviation (prediction errors). Root-mean-square deviation (RMSD) indicates how distributed residuals are. The metric also measures how tightly the data fit the fitted line. RMSD helps in feature extraction, model training, prediction, evaluation, error calculation, and model tuning.

IV. PROPOSED APPROACHES

The question, response, authority, and extra are DNS message parts. This article implements ML algorithms in RStudio, as shown in Fig. 1. DNS record exfiltration and trafficking malware targets question-and-answer parts.

The inquiry portion comprises the requested name or encrypted binary data sent to a malicious name

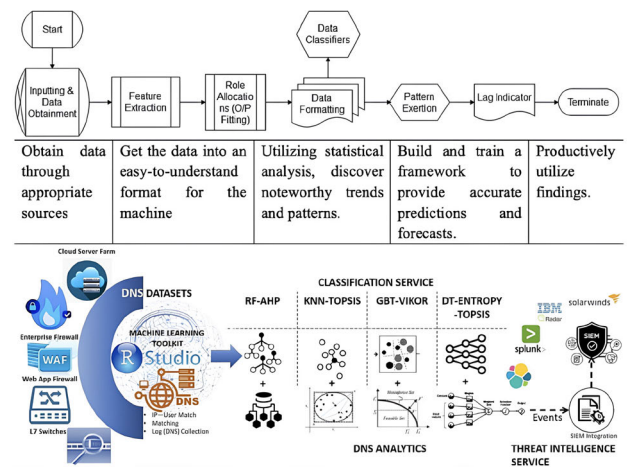


FIGURE 1. Proposed classification architecture.

server [13], [14]. The respond section contains resource records or arbitrary data from the attacker’s name server to the client. The server may abuse authority and different neighborhoods to send instructions and data to affected clients. Data exfiltration employing a DNS attack transfers data from a compromised workstation to an attacker’s external server. The data transmission might be manual by someone with access privileges to the system or automatic by malware across a network. The virus on a compromised system that uses DNS tunneling may include code that polls for orders to perform. Server-side malware rebuilds responses to record lookups to an attacker-controlled domain. Table 3 illustrates DNS data exfiltration properties. Stateless and stateful features are the most common. Stateless feature components are typically independent of domain or host time series and may be obtained from DNS query packets. This minimizes real-time computation overhead [13], [14], [19]. Stateful features examine many inquiries in a temporal span, increasing the detection system’s processing load. Stateful scanning of DNS logs handles delayed DNS attacks. Consider a hacked system infected with malware that sends information to “website2.eg2.com.” Attack scenarios are as follows: the attacker encodes data using RFC4648. Malware may create a DNS query with an encrypted subdomain, e.g., “website2.eg2.com.” First, local, and public name servers are verified. Since there is no subdomain, the website2.eg2.com name server resolves the IP. When the suspicious name system gets the query, the attacker extracts website2 and decodes the data. The name server might pick a benign reply for the question. Fig. 2 shows DNS requests exfiltrating a person’s name, SSN, and contact. Data is divided and given to “website2.eg2.com” to bypass the firewall in this example. If the query fails locally, it is transmitted to the attacker’s server. The server’s data almost satisfy the victim.

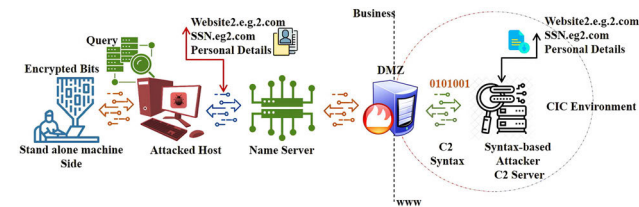


FIGURE 2. Shows DNS requests exfiltrating a user’s name, SSN, and details.

Stateless and stateful features are the most common sets wherein stateless components are typically independent of domain or time series and may be obtained from DNS query packets. This minimizes real-time computation overhead. Stateful components examine many inquiries in a temporal span, increasing the detection system’s processing load. Stateful scanning of DNS logs can manage slow attacks. This section explains If a DNS query is normal or malicious. We aim to develop hybrid threat detection on powerful devices using MCDM-based IDS for integration in SIEM [10], [24], [25].

As illustrated in Fig. 3, we use a two-tier proposed model to classify incoming DNS traffic exfiltration for SIEM integration. First, stateless characteristics are derived from the structured data. Each input sample is rated as benign, suspect, or malicious based on the classifier’s probability [7], [26]. If the number of anomalous samples is high enough, the whole traffic window will be reanalyzed with stateful features to help the trained classifier decide. Unless the incoming model is harmful, DNS transmission continues; otherwise, it stops. Stateful components may re-investigate DNS traffic if a large packet window appears suspicious. Classifier determines window severity using stateful properties, not each packet.

A. AUTO-MODEL IN RSTUDIO

In addition, RStudio makes it possible to speed up the process of generating and validating models by utilizing an extension of its own called the MCDM-based Machine Automated Model (MCDM-MAM) [8], [11], [22]. This feature can solve three of the most important issue classes: predictions (including regression and classification), segmentation, and identifying outliers, as given in Fig 1. After completing the preprocessing and data mapping for the CIC-Bell-DNS-EXF-2021, Machine Automated Model (MAM) offers a choice between RF, KNN, GBT, and DT [16], [18], [27]. In addition to using features engineering and MCDM optimization approaches, suitable parameters for each model were chosen [5], [7], [10].

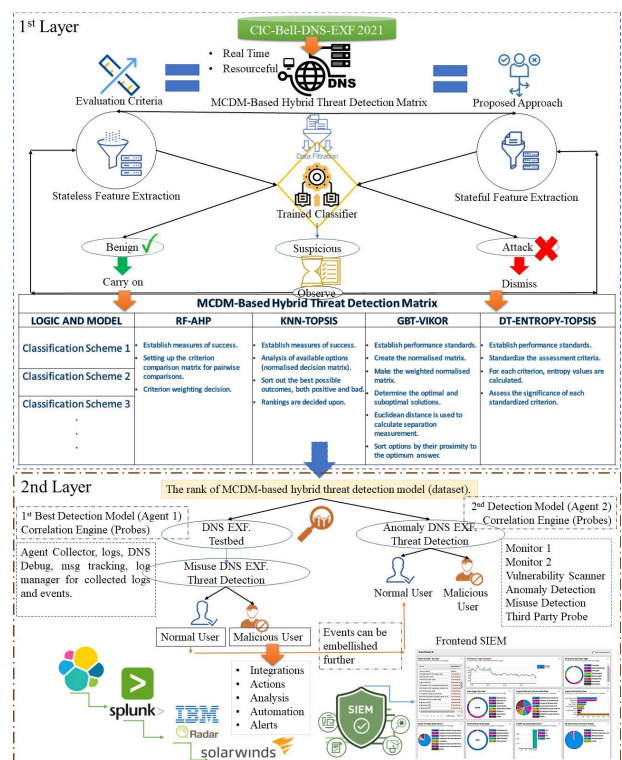


FIGURE 3. The proposed two tier-framework for MCDM-based hybrid SIEM integration.

TABLE 8. Summarizes the various forms of attacks and their duration.

Day	Benign	Audio	Compressed	.exe	Image	text	video
Day 1 20th July 2022 Benign	20.35 Hrs. (53,454 Domains)	–	–	–	–	–	–
Day 2 21st July 2022 Light Attack	05.13 Hrs. (14,934 Domains)	31 Mnts.	1.53 Hrs.	1.51 Hrs.	1.13 Hrs.	19 Mnts.	37 Mnts.
Day 3 22nd July 2022 Light Attack	5.25 Hrs. (14,934 Domains)	–	–	5.11 Hrs.	3.57 Hrs.	2.38 Hrs.	–
Day 4 23rd July 2022 Light Attack	7.58 Hrs. (21,549 Domains)	3.21 Hrs.	4.32 Hrs.	–	–	–	2.21 Hrs.
Day 5 24th July 2022 Heavy Attack	8.10 Hrs. (23,045 Domains)	1.56 Hrs.	–	–	4.32 Hrs.	–	5.41 Hrs.
Day 6 25th July 2022 Heavy Attack	13.69 Hrs. (38,947 Domains)	7.21 Hrs.	4.41 Hrs.	3.41 Hrs.	6.89 Hrs.	2.56 Hrs.	9.41 Hrs.
Day 7 26th July 2022 Heavy Attack	10.43 Hrs. (29,648 Domains)	–	3.32 Hrs.	7.53 Hrs.	–	4.33 Hrs.	–

V. EXPERIMENTAL ANALYSIS

A test bed is a controlled environment where we can test our methodologies, and solutions.

A. TEST-BED

Fig. 2 demonstrates a DNS-based DNS-EXF-2021 CIC testbed attack environment. The attacker’s name server receives DNS queries with encoded victim data. The attacker of the server side receives the most encoded file and acts rogue name server, and decryption occurs. Bitbuckets DNSExfiltrator lets us transfer a file over a DNS backdoor. The NS records led to the attacker’s server launching the script. DNS exfiltrates encoding technique and throttle time are RFC4648 and 400 ms. The maximum bytes and characters for each DNS request label are predefined.

B. DATASET

We utilized DNS log files to capture DNS data. One million Cisco Umbrella domains were mined for benign samples. Seven days of benign and heavy file attacks were employed to collect DNS traffic to detect audio, compressed, .exe, image, text, and video. The attack strategy is specified below in Table 8.

We used a Python script to make HTTP requests to the web servers of the domains we identified and stored the packets that returned an OK answer as a traffic record. Each day, we chose a new region to collect real-world data. On the victim’s side, Networkminer 2.7.3 (open source) was used to record all benign and attack traffic, and the obtained data were tagged according to the timestamps of the traffic. We successfully captured 295MB, 181MB, 54MB, and 125MB of DNS packets, respectively, representing light, light benign, heavy, and heavy benign traffic that was not malicious, as given in Table 9 for the detailed dataset’s statistical information. DNS information was gathered from DNS log files using active data gathering. Table 10 and Fig. 4 present the calculations for the accuracy, F1-score, and precision of the suggested ML models for both light and heavy attacks. Cisco Umbrella’s top 1 million domains were mined for samples. Seven days of mild and heavy file attacks were employed to gather DNS traffic to identify audio, compressed,.exe, image, text, and video file types. 7 KB to 1.1 MB are “light” files, while 9.7 MB to 51 MB are “heavy.” We used a Python script to perform HTTP queries to the web servers of the sites we gathered and stored the successful (ok) packets as a traffic

TABLE 9. The dataset’s statistical information.

Attack Category	DNS Packets	Stateful Features	Stateless Features
Light	295 MB	144056	503340
Light Benign	181 MB	312028	805534
Heavy	54 MB	22590	85366
Heavy Benign	125 MB	219532	562328

TABLE 10. Accuracy, F1-score, and Precision ML models estimations for light and heavy attacks.

ML-MODELS	ATTACK TYPES					
	LIGHT (in %)			HEAVY (in %)		
	Prec.	F1-S	Acc.	Prec.	F1-S	Acc.
RF-AHP	93.21	76.94	71.45	85.43	73.54	72.32
KNN-TOPSIS	97.34	96.67	95.43	96.56	95.23	94.84
GBT-VIKOR	99.52	99.52	99.52	99.63	99.63	99.63
DT-ENTROPY-TOPSIS	94.11	95.12	96.89	81.45	85.45	78.45

log. Each day we chose a new place to create a real-world dataset.

VI. EXPERIMENTAL RESULTS

We will use the section six model to assess our MCDM-based hybrid SIEM strategy. This model’s first layer recognizes stateless DNS traffic based on suspicious sample ratios. Following that, inbound DNS traffic will be statefully filtered. Heavy and light traffic is used to test domain name data exfiltration threats. Preprocessing features without times- tamps prevents ML overfitting. Replace nan values with zero to clean the data.

We are encoding stateful and state- less categorical characteristics. Stateful absolute elements include rr, rr_type_frequency, rr_name_entropy, rr_name_length, rr_count, rr_type, distinct do- mains, and reverse DNS.

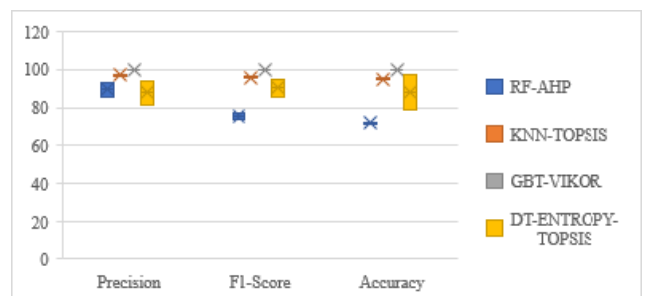


FIGURE 4. Graphical plot for attack estimation.

Stateless features include entropy, subdomain length, and labels max. We also average the unique asn lists. Pre-processing of feature timestamps prevents ML overfitting. We remove nan from the data. Stateful and stateless characteristics are encoded. The performance of the proposed architecture is measured in terms of accuracy, absolute error, weighted average recall, weighted average precision, kappa value, logistic loss, and root mean square deviation (RMSD). The mathematical details of these measures are in [7] and [15]. The packet size of the window is 500 bytes, and the number of 's' in a sliding window is 150. For false positives, we set it low. If the stateless classifier finds one rogue packet, we toss out everything. Placing the window size to a big amount may block out harmless DNS activity, which is not desired in real life. Using Python's mlr3 package, we evaluated the RF-AHP (RA) shown in Tables 11 and 12, the KNN-TOPSIS (KT) shown in Tables 13 and 14, the

TABLE 11. Performance assessment of RA.

Performance Metrics in %	Values of RA
ACC	98.70%
AE	1.30%
WAR	82.11%
WAP	92.08%
KV	0.973
LL	0.298
RMSD	0.113

TABLE 12. Performance matrices and confusion matrix for RA.

(PA-RA)	True Normal	True Heavy Attack	True Heavy Benign	True Light Attack	True Light Benign	True Normal	Total=119816	Class Precision in %
Pred. Heavy Attack	61154	90	128	164	60		61596	99.27
Pred. Heavy Benign	118	42454	4	376	2		42954	98.83
Pred. Light Attack	130	4	3942	19	10		4105	96.02
Pred. Light Benign	234	159	27	10700	6		11126	96.17
Pred. Normal	8	3	3	3	18		35	51.42
Total=119816	61644	42710	4104	11262	96		118268	
Class Recall in %	99.20	99.40	96.05	95	18.75			

TABLE 13. Performance assessment of PA-KT.

Performance Metrics in %	Values of KT
ACC	93.65%
AE	6.35%
WAR	57.27%
WAP	59.49%
KV	.890
LL	0.341
RMSD	0.247

TABLE 14. Performance matrices and confusion matrix for KNN-TOPSIS.

(PA-KT)	True Normal	True Heavy Attack	True Heavy Benign	True Light Attack	True Light Benign	Total=118796	Class Precision in %
Pred. Heavy Attack	61400	2524	3026	1344	96	68390	89.78
Pred. Heavy Benign	16	40180	2	226	0	40424	99.40
Pred. Light Attack	0	0	0	0	0	0	0.00
Pred. Light Benign	228	6	76	9672	0	9982	96.89
Pred. Normal	0	0	0	0	0	0	0.00
Total=118796	61644	42710	3104	11242	96	111252	
Class Recall in %	99.61	94.41	50	88	50.00		

TABLE 15. Performance assessment of GV.

Performance Metrics	Values of GV in %
ACC	99.52
AE	0.48
WAR	91.25
WAP	94.21
KV	0.988
LL	.393
RMSD	.423

TABLE 16. Performance matrices and confusion matrix for GBT-VIKOR.

(PA-GV)	True Normal	True Heavy Attack	True Heavy Benign	True Light Attack	True Light Benign	Total=110439	Class Precision in %
Pred. Heavy Attack	69008	46	60	24	16	69154	99.64
Pred. Heavy Benign	52	25001	4	54	0	25111	99.56
Pred. Light Attack	94	0	3380	20	32	3526	95.85
Pred. Light Benign	44	22	34	12468	4	12572	98.00
Pred. Normal	0	2	14	4	56	76	73.68
Total=110439	69198	25071	3492	12570	108	109913	
Class Recall in %	99.50	99.72	97	98	51.85		

GBT-VIKOR (GV) shown in Tables 15 and 16, and the DT-ENTROPY-TOPSIS (DET) shown in Tables 17 and 18 to determine their cumulative performances cum percentages. We scrambled the full dataset before dividing it (80%-20%). Table 19 shows how well five algorithms classify heavy and light attacks. Table 20 shows that RF-AHP operates effectively for heavy and benign strikes. GBT-VIKOR excels in powerful strikes [8], [22], [23]. The total classification results show that our technique can detect heavy and light DNS traffic successfully. Fig. 4 exhibits the class accuracy

TABLE 17. Performance assessment of DET.

PA-DET	
Performance Metrics in %	Values of DET
ACC	97.47%
AE	2.53%
WAR	97.40%
WAP	97.70%
KV	.958
LL	.320
RMSD	0.156

TABLE 18. Performance matrices and confusion matrix for DT-Entropy-TOPSIS.

(PA-DET)	True Nor mal	True Hea vy Atta ck	True Hea vy Beni gn	True Ligh t Atta ck	True Ligh t Beni gn	Total = 1185 60	Class Precis ion in %
Pred. Heavy Attack	5999 4	44	88	12	0	6013 8	99.76 %
Pred. Heavy Benign	208	421 02	0	256	0	4256 6	98.90 %
Pred. Light Attack	672	0	2928	250	48	3898	75.11 %
Pred. Light Benign	770	564	78	104 88	4	1190 4	88.10 %
Pred. Normal	0	0	10	0	44	54	81.48 %
Total= 118560	6164 4	427 10	3104	110 06	96	1155 56	
Class Recall in %	97.3 2	98.5 7	94	95	45.8 3		

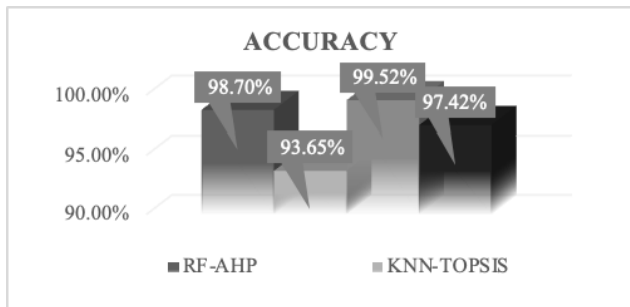


FIGURE 5. Overall class accuracy for the best-proposed algorithms.

for the heavy attack, heavy benign, light attack, light benign, and normal assessed on RA, KT, GV, and DET, respectively.

VII. DISCUSSION

Each model is evaluated using a variety of measures, including weighted average recall, accuracy, absolute error, kappa values, weighted average precision, root mean squared deviation (RMSD), and logistic loss. The correlation of these values facilitates the evaluation of the effectiveness of the learning algorithm. Fig. 5 compares the machine learning algorithms in terms of their class-wise accuracy. A classification model generates predictions for each class, and the confusion matrix summarizes those predictions. DNS traffic is generated during the exfiltration of any file formats, from the smallest to the largest, including .exe, audio,

```

1  {"MHS": {
2    "Version": "EXF_21",
3    "reference": "Cisco_Umbrella",
4    "attacks": [
5      {"attack_type": "Anm_Det_SIEM",
6       "confidence": "0.98546785"}]},
7  "MHS": {
8    "Version": "EXF_21",
9    "reference": "Cisco_Umbrella",
10   "attacks": [
11     {"attack_type": "rootkit",
12      "confidence": "0.68"},
13     {"attack_type": "xterm",
14      "confidence": "0.9"}]},
15  "MHS": {
16    "Version": "EXF_21",
17    "reference": "Cisco_Umbrella",
18    "attacks": [
19     {"attack_type": "Benign",
20      "confidence": "0.58"}]}

```

FIGURE 6. An example of SIEM security detection of anomaly object.

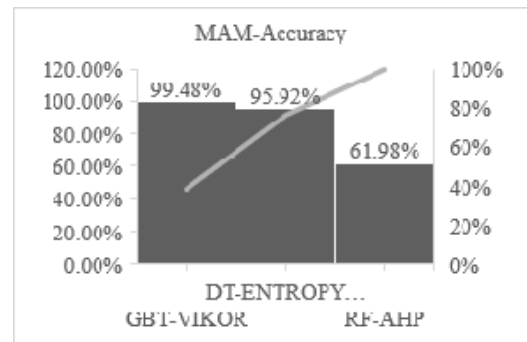


FIGURE 7. Overall accuracy performances in MAM.

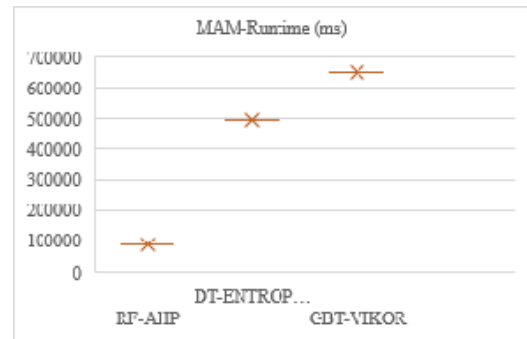


FIGURE 8. Overall runtime performances in MAM.

picture, compressed, video, and text files. The dataset and testbed are rather sizable, totaling 655 MB. Because of our built-in feature extractor, thirty features could be derived from DNS packets. This means 107956 heavy-attack samples were included in the final structured dataset, and samples of 647396 for light-attack, 781860 heavy distinct benign samples, and 1117562 distinct light samples for benign. Our experimental evaluation utilizing multiple ML methods on our dataset confirmed the continued success of our hybrid detection system in the presence of existing DNS traffic. It has been shown in the study. Fig. 6 depicts such a case in point.




Accuracy	Mdel	Accuracy	Standard Deviations	Gains	Total Time	Training Time
	RF-AHP 	61.98%	±0.2%	10031	1 min 04 sec	17 ms
	GBT-VIKOR 	99.48%	±0.1%	43216	5 min 31 sec	56 ms
	DT-ENTROPY-TOPSIS 	95.92%	±0.1%	57347	6 min 27 sec	1 s

FIGURE 9. Exposition of the classifier’s efficiency in MAM.

TABLE 19. Confusion matrix for RF-AHP in MAM.

(MAM-RA)	True Normal	True Heavy Attack	True Heavy Benign	True Light Attack	True Light Benign	Total=83874	Class Precision in %
Pred. Heavy Attack	4196	222	132	377	60	69342	60.51
Pred. Heavy Benign	1834	821	48	222	2	12316	66.66
Pred. Light Attack	74	32	818	90	4	1018	80.35
Pred. Light Benign	78	54	68	996	2	1198	83.14
Pred. Normal	0	0	0	0	0	0	0.00
Total=83874	43948	30518	2260	7080	68	51986	
Class Recall in %	95.48	26.90	36	14	0.00		
Accuracy	61.98	±0.2%					
Classification Error	38.02	±0.2%					

TABLE 20. Confusion matrix for GBT-VIKOR in MAM.

(MAM-GV)	True Normal	True Heavy Attack	True Heavy Benign	True Light Attack	True Light Benign	Total=69748	Class Precision in %
Pred. Heavy Attack	3671	32	46	32	12	36836	99.67
Pred. Heavy Benign	4	256	34	4	22	25696	99.76
Pred. Light Attack	32	0	155	8	34	1642	94.88
Pred. Light Benign	50	8	24	545	6	5540	98.48
Pred. Normal	0	2	4	0	28	34	82.35
Total=69748	36830	25676	1636	5544	62	69390	
Class Recall in %	99.69	99.84	95	98	45.16		
Accuracy	99.48	±0.1%					
Classification Error	0.52	±0.1%					

In this case (lines 5-6), the attack type includes instantiated JSON with a particular item. This term is used to describe a machine-automated intrusion with a 98.54% degree of confidence in identifying an abnormal traffic pattern (labeled ANOMALY) to the rr_name_entropy that was employed. Since the protocol only supports expressing individual events, it cannot simulate relationships between several occurrences that may all be linked with the same attack.

VIII. STATISTICAL FINDINGS

DT has the greatest weighted mean recall (97.40%), while KNN has the lowest (57.27%). DT’s weighted mean accuracy is 97.70%, while KNN’s is 59.49%. GBT has the highest Kappa (0.988), and KNN has the lowest (59.54%). GBT has the biggest logistic loss (0.393), and RF is the lowest (0.298). RF’s RMSD (0.113) is the lowest, whereas GBT’s is the greatest (0.423). It has been noted that all of the classifiers have successfully detected the heavy attack, which was carried out using DNS attacks. The KNN classifier, on the other hand, was unable to recognize the phf or xsnoop attacks. In addition, every classifier has a poor precision and accuracy score for the heavy benign class and all of the subclasses that fall under it. This may be because light benign, light attack, and heavy benign cases only make up 12% of the dataset. In a comparison of the accuracies provided by the various machine learning algorithms, it has been found

that the GBT-VIKOR model offers the highest accuracy (99.52%). The KNN-TOPSIS algorithm provides the lowest accuracy (93.65%). The highest accuracy was achieved through the use of the GBT-VIKOR model.

A. STATISTICS OF MACHINE AUTOMATED MODEL (MAM)

The results of the models constructed with the machine automated model are comparable to those with an accuracy of 99.48% and a classification error of 0.52%; GBT-VIKOR has excelled over both DT-Entropy-TOPSIS and RF-AHP in terms of performance. The accuracy provided by DT-Entropy-TOPSIS is 95.92%, while the error rate for classification is 4.08%. On the other hand, RF-AHP achieves an accuracy of 61.98% while having a classification error of 38.02%. According to the model, AM-GVT possesses the best-class recall for heavy attacks (99.84%) and the highest-class precision for serious attacks (99.84%). For normal, AM-GVT had the best-class accuracy (82.35%) and the highest-class recall (99.69%). The AM-GBT possesses the best level of accuracy in the class for heavy benign (99.76%) and the highest level for normal (95%). The AM-GBT has the best accuracy in its class for light benign (98.48%) and the highest recall in its category for normal (45.16%). Figs. 7 and 8 provide a more detailed comparison of accuracy performance vs. runtime for the best GBT-VIKOR specified machine automated model.

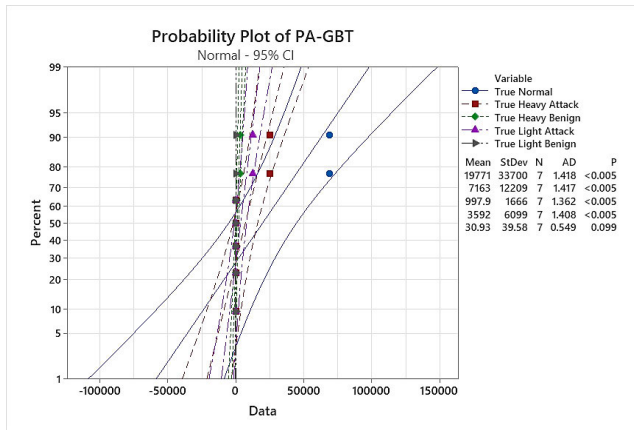


FIGURE 10. Probability plot of PA-GBT (proposed).

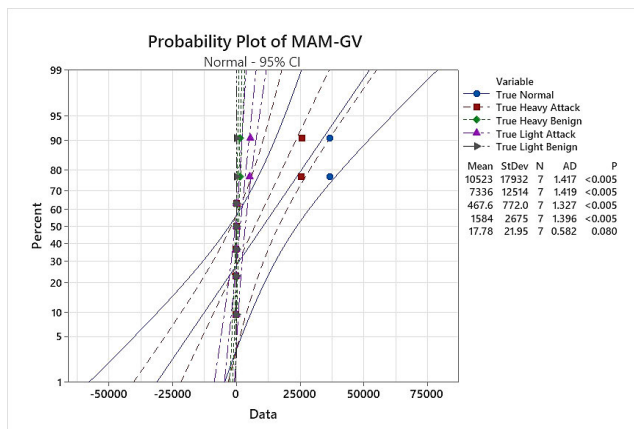


FIGURE 11. Probability plot for MAM-GV (actual).

Fig. 9 depicts the comprehensive exposition of the classifier’s efficiency. Figs. 10 and 11 show the real exhibition outcome of the suggested vs. actual. It can be seen that the Anderson-Darling (AD) Goodness of Fit efficiency test for the basic is highly satisfactory.

IX. CONCLUSION

With the rising interconnectedness between computers, intrusion detection software has become a necessary component of all secure network infrastructures. Researchers have developed models such as classification and clustering through machine learning techniques such as RF, KNN, GBT, and DT. The CIC-Bell dataset comprises 19% DoS attacks and 39% regular packets, whereas the combined subclasses account for 12% of the whole data set. This study compares four machine learning classification techniques using an MCDM- based hybrid detection system that includes RA, KT, GV, and DET on the CIC-Bell-DNS-EXF-2021 dataset. According to this study, the GBT classification algorithm outperformed all others in generated designs and the Machine Automated Model feature. Correctness attained by the KNN-TOPSIS algorithms was the lowest (93.65%), whereas the GBT- VIKOR algorithms achieved maximum accuracy of 99.52%. In addition, it has been discovered that implementing machine

learning models is made much simpler by using the Machine Automated Model (MAM) feature in RStudio. This is particularly the case, which is not only strategy, and space-and-time-efficient, but it also lessens the strain of integrating SIEM models through complicated syntax. These machine learning classifiers all exhibit accuracy on the CIC-Bell-DNS-EXF-2021 to the degree that it is considered acceptable. When assessing upcoming work that aim to improve upon the latest advancements in development, three key elements can be integrated: increased diversity in datasets, advanced machine or deep learning techniques, and real-time analysis and responses for seamless integration with SIEM systems.

ACKNOWLEDGMENT

This study is supported via funding from Prince sattam bin Abdulaziz University project number (PSAU/2024/R/1445).

REFERENCES

- [1] H. Ho, R. Ko, and L. Mazerolle, “Situational crime prevention (SCP) techniques to prevent and control cybercrimes: A focused systematic review,” *Comput. Secur.*, vol. 115, Apr. 2022, Art. no. 102611.
- [2] N. M. Chayal, A. Saxena, and R. Khan, “A review on spreading and forensics analysis of windows-based ransomware,” *Ann. Data Sci.*, vol. 1, pp. 1–22, Jun. 2022.
- [3] M. D. Iannacone and R. A. Bridges, “Quantifiable & comparable evaluations of cyber defensive capabilities: A survey & novel, unified approach,” *Comput. Secur.*, vol. 96, Sep. 2020, Art. no. 101907.
- [4] L. Zhang and V. L. L. Thing, “Three decades of deception techniques in active cyber defense—retrospect and outlook,” *Comput. Secur.*, vol. 106, Jul. 2021, Art. no. 102288.
- [5] A. Al-Qarafi, F. Alrowais, S. S. Alotaibi, N. Nemri, F. N. Al-Wesabi, M. Al Duhayyim, R. Marzouk, M. Othman, and M. Al-Shabi, “Optimal machine learning based privacy preserving blockchain assisted Internet of Things with smart cities environment,” *Appl. Sci.*, vol. 12, no. 12, p. 5893, Jun. 2022.
- [6] R. A. Disha and S. Waheed, “Performance analysis of machine learning models for intrusion detection system using Gini impurity-based weighted random forest (GIWRF) feature selection technique,” *Cybersecurity*, vol. 5, no. 1, p. 1, Dec. 2022.
- [7] Y. N. Kunang, S. Nurmaini, D. Stiawan, and B. Y. Suprpto, “Attack classification of an intrusion detection system using deep learning and hyperparameter optimization,” *J. Inf. Secur. Appl.*, vol. 58, May 2021, Art. no. 102804.
- [8] E. Zarei, B. Ramavandi, A. H. Darabi, and M. Omidvar, “A framework for resilience assessment in process systems using a fuzzy hybrid MCDM model,” *J. Loss Prevention Process Industries*, vol. 69, Mar. 2021, Art. no. 104375.
- [9] E. M. Bärli, A. Yazidi, E. H. Viedma, and H. Haugerud, “DoS and DDoS mitigation using variational autoencoders,” *Comput. Netw.*, vol. 199, Nov. 2021, Art. no. 108399.
- [10] L. Rosa, T. Cruz, M. B. D. Freitas, P. Quitério, J. Henriques, F. Caldeira, E. Monteiro, and P. Simões, “Intrusion and anomaly detection for the next-generation of industrial automation and control systems,” *Future Gener. Comput. Syst.*, vol. 119, pp. 50–67, Jun. 2021.
- [11] R. Verma and S. Chandra, “Interval-valued intuitionistic fuzzy-analytic hierarchy process for evaluating the impact of security attributes in fog based Internet of Things paradigm,” *Comput. Commun.*, vol. 175, pp. 35–46, Jul. 2021.
- [12] B. A. Tama and S. Lim, “Ensemble learning for intrusion detection systems: A systematic mapping study and cross-benchmark evaluation,” *Comput. Sci. Rev.*, vol. 39, Feb. 2021, Art. no. 100357.
- [13] S. Wang, L. Sun, S. Qin, W. Li, and W. Liu, “KRTunnel: DNS channel detector for mobile devices,” *Comput. Secur.*, vol. 120, Sep. 2022, Art. no. 102818.
- [14] S. MahdaviFar, A. Hanafy Salem, P. Victor, A. H. Razavi, M. Garzon, N. Hellberg, and A. H. Lashkari, “Lightweight hybrid detection of data exfiltration using DNS based on machine learning,” in *Proc. 11th Int. Conf. Commun. Netw. Secur.*, Dec. 2021, pp. 80–86.

- [15] M. N. Al-Mhiqani, R. Ahmad, Z. Z. Abidin, K. H. Abdulkareem, M. A. Mohammed, D. Gupta, and K. Shankar, "A new intelligent multi-layer framework for insider threat detection," *Comput. Electr. Eng.*, vol. 97, Jan. 2022, Art. no. 107597.
- [16] A. Seyfollahi and A. Ghaffari, "A review of intrusion detection systems in RPL routing protocol based on machine learning for Internet of Things applications," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–32, Aug. 2021.
- [17] S. Rameem Zahra, M. Ahsan Chishti, A. Iqbal Baba, and F. Wu, "Detecting COVID-19 chaos driven phishing/malicious URL attacks by a fuzzy logic and data mining based intelligence system," *Egyptian Informat. J.*, vol. 23, no. 2, pp. 197–214, Jul. 2022.
- [18] S.-S. Lin, S.-L. Shen, A. Zhou, and Y.-S. Xu, "Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods," *Autom. Construct.*, vol. 122, Feb. 2021, Art. no. 103490.
- [19] S. Dadkhah, F. Shoeleh, M. M. Yadollahi, X. Zhang, and A. A. Ghorbani, "A real-time hostile activities analyses and detection system," *Appl. Soft Comput.*, vol. 104, Jun. 2021, Art. no. 107175.
- [20] P. Radoglou-Grammatikis, P. Sarigiannidis, E. Iturbe, E. Rios, S. Martinez, A. Sarigiannidis, G. Eftathopoulos, Y. Spyridis, A. Sesis, N. Vakakis, D. Tzovaras, E. Kafetzakis, I. Giannoulakis, M. Tzifas, A. Giannakoulis, M. Angelopoulos, and F. Ramos, "SPEAR SIEM: A security information and event management system for the smart grid," *Comput. Netw.*, vol. 193, Jul. 2021, Art. no. 108008.
- [21] M. Di Mauro, G. Galatro, G. Fortino, and A. Liotta, "Supervised feature selection techniques in network intrusion detection: A critical review," *Eng. Appl. Artif. Intell.*, vol. 101, May 2021, Art. no. 104216.
- [22] M. Nazim, C. Wali Mohammad, and M. Sadiq, "A comparison between fuzzy AHP and fuzzy TOPSIS methods to software requirements selection," *Alexandria Eng. J.*, vol. 61, no. 12, pp. 10851–10870, Dec. 2022.
- [23] U. Hacıoglu, D. Chlyeh, M. K. Yilmaz, E. Tatoglu, and D. Delen, "Crafting performance-based cryptocurrency mining strategies using a hybrid analytics approach," *Decis. Support Syst.*, vol. 142, Mar. 2021, Art. no. 113473.
- [24] C. H. L. Resende, C. A. S. Geraldes, and F. R. Lima, "Decision models for supplier selection in Industry 4.0 era: A systematic literature review," *Proc. Manuf.*, vol. 55, pp. 492–499, May 2021.
- [25] R. Deb and S. Roy, "A software defined network information security risk assessment based on Pythagorean fuzzy sets," *Expert Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115383.
- [26] R. Leszczyna, "Review of cybersecurity assessment methods: Applicability perspective," *Comput. Secur.*, vol. 108, Sep. 2021, Art. no. 102376.
- [27] M. Khan, M. T. Mehran, Z. U. Haq, Z. Ullah, S. R. Naqvi, M. Ihsan, and H. Abbass, "Applications of artificial intelligence in COVID-19 pandemic: A comprehensive review," *Expert Syst. Appl.*, vol. 185, Dec. 2021, Art. no. 115695.
- [28] M. Ahmad Khan, S. Mahmood Khan, and S. K. Subramaniam, "A systematic literature review on security issues in cloud computing using edge computing and blockchain: Threat, mitigation, and future trends," *Malaysian J. Comput. Sci.*, vol. 36, no. 4, pp. 347–367, Dec. 2023.
- [29] M. A. Khan, S. M. Khan, and S. K. Subramaniam, "Secured dynamic request scheduling and optimal CSP selection for analyzing cloud service performance using intelligent approaches," *IEEE Access*, vol. 11, pp. 140914–140933, 2023.



PRABIRA KUMAR SETHY (Senior Member, IEEE) received the M.Tech. degree from IIT (ISM) Dhanbad and the Ph.D. degree from Sambalpur University. From 2009 to 2013, he was an Engineer in Doordarshan, Prashar Bharati. He is currently an Associate Professor with the Department of Electronics and Communication Engineering, Guru Ghasidas Vishwavidyalaya, Bilaspur, Chhattisgarh, India. He has nine years of teaching, research, and administrative experience, and four years of industry experience. He has published 80 research papers in different reputed journals and conferences. In addition, he has two patents. His research interests include image processing, machine learning, and deep learning. He is a frequent reviewer of many journals and the session chair of international conferences.



SANTI KUMARI BEHERA (Senior Member, IEEE) has been an Assistant Professor with the Department of Computer Science and Engineering, Veer Surendra Sai University of Technology, since 2015. She has published 70 research papers in different reputed journals and conferences. In addition, she has one patent and one copyright. Her research interests include image processing, machine learning, and deep learning.



MANOJ GUPTA (Member, IEEE) received the B.Tech. degree from the Institute of Engineering and Technology, M. J. P. Rohilkhand University, Bareilly, Uttar Pradesh, India, the M.Tech. degree from HNB Garhwal Central University, Uttarakhand, India, and the Ph.D. degree from the University of Rajasthan, Jaipur, India. His overall experience is more than 18 years. He has published more than 60 research papers in international journals and national/international conferences and book chapters. He has two patent grants and published five patents in his credit and currently having two Indian research copyrights on his credit. His research interests include wireless and multimedia communication, biomedical image and signal processing, soft computing, intelligent computing and computational intelligence, the Internet of Things (IoT), artificial intelligence and machine learning, healthcare informatics, and embedded systems. He is a member of many professional bodies, such as IACSIT, ISTE, IAENG, and many more. He has served as a Technical Programme Committee (TPC) Member and a Reviewer in various international conferences, such as ICCIA 2020, ICCIA 2019 ICSIP 2018, ICSIP 2017, ICSIP 2016, AIPR 2017, AIPR 2016, ICCIA 2018, ICCIA 2017, ICCIA 2016, and ICNIT 2018. He is the Editor-in-Chief (EiC) of the Book Series *Advances in Antenna, Microwave and Communication Engineering* (Scrivener, John Wiley, USA), *Advances in Antenna Design, Wireless Communication and Mobile Network Technology* (CRC Press, Taylor and Francis, USA), and *Advances in Digital Signal Processing and Image Processing for Industrial Applications* (CRC Press, Taylor and Francis, USA), and *AAP Book Series on Digital Signal Processing, Computer Vision and Image Processing* (Apple Academic Press, USA). He was invited as a Keynote Speaker/Invited Speaker at the 2017 2nd IEEE International Conference on Signal and Image Processing (ICSIP 2017), Nanyang Executive Centre, Singapore, on August 4–6, 2017, and Invited as a Keynote Speaker at the 2017 International Conferences on Public Health and Medical Sciences (ICPHMS 2017), Xi'an, China, May 23–24, 2017. He is an editor, associate editor, and reviewer of many international journals. His name has been listed in Marquis Who's Who in Science and Engineering®USA and Marquis Who's Who in the World®USA.



GYANA RANJANA PANIGRAHI (Member, IEEE) received the Ph.D. degree in electronics from the Department of Electronics, Sambalpur University, Sambalpur, Odisha, India. He is currently a member of Microsoft and the EC-Council. His research interests include cyber security, digital forensics, physical cyber systems, AI/ML, communication, wireless communication, data communication and networking, the IoT, and storage area networks (SAN).



FARHAN A. ALENIZI received the B.Sc. and M.Sc. degrees in electrical engineering from King Saud University, Saudi Arabia, in 1999 and 2006, respectively, and the Ph.D. degree in electrical engineering and computer science from the University of California at Irvine, Irvine, CA, USA, in 2017. Since 2010, he has been a Faculty Member with the Department of Electrical Engineering, Prince Sattam Bin Abdulaziz University, Saudi Arabia, where he is currently an Assistant

Professor. Besides his academic experience, he worked for ten years with Saudi Telecom Company (STC), a leading telecommunication company in the Middle East as a Designer and a Consultant in the satellite and optical fiber transmission networks. Moreover, he is also involved in drone spoofing and jamming projects. His research interests include images and video processing, signal processing, discrete signal processing (DSP), images and video watermarking, 3D-mesh objects watermarking, and secure multimedia exchanges.



AZIZ NANTHAMORNPHONG (Member, IEEE) received the Ph.D. degree from The University of Alabama, USA. He is currently an Associate Professor and the Dean of the College of Computing, Prince of Songkla University, Phuket Campus, Thailand. With an extensive academic background, he specializes in empirical software engineering and data science, and among other areas. His research significantly contributes to the development of scientific software and leverages

data science in the field of tourism. In addition to his core focus, he is also deeply engaged in the study of human–computer interaction, pioneering innovative approaches to foster a beneficial interplay between humans and technology.

• • •



PANNEE SUANPANG (Member, IEEE) received the B.I.T. and M.I.S. degrees from Griffith University, Australia, in 1997 and 2001, respectively, and the D.Tech. degree in science from the University of Technology, Sydney, Australia, in 2005. She is currently an Associate Professor with the Department of Information Technology, Suan Dusit University. Until now, she has published eight Scopus/WoS-indexed documents. Her research interests include advanced information

technology in agricultural, big data, the IoT, and smart tourism. She is an editor and a reviewer of many famous journals in the world.