



Review Article

Review on Cyber Defence with Machine Learning: A System for Detecting and Monitoring Attacks

Amrapali Ambade^{1*} , Monika Ingole² and Prashant Tarone³

^{1*,2,3}Department of Computer Science and Engineering, Wainganga College of Engineering and Management, Nagpur, India

Article Information

Abstract

Article History

Received: 13 June 2026

Revised: 15 July 2026

Accepted: 10 August 2026

Published online: 16 September 2026

Keywords

Cybersecurity

Malware Detection

Machine Learning

Cyber Threat Intelligence

Cyber-attack

Correspondence*

amrapaliambade1693@gmail.com

ORCID

Amrapali Ambade
<https://orcid.org/0009-0004-0651-1466>

Monika Ingole
<https://orcid.org/0009-0004-3283-8665>

Prashant Tarone
<https://orcid.org/0009-0004-3870-000X>

Cybersecurity experts concentrate on identifying risk profiles and thinking up ways of handling them properly as the main focus of their work. One of the main objectives in this area is to set up strong and powerful ways to reinforce security measures. The use of machine learning has transformed the whole landscape of cyber defense systems in modern times to a great extent. It is the breakthroughs in data storage, processing power, and communication that have not only hastened the use of cloud services but also of advanced networks and programming languages that are gradually being replaced by more sophisticated ones. The ongoing digital transformation worldwide is building up the need for complex privacy and security issues to be handled, which in turn, puts pressure on the existing safeguards to further strengthen themselves against new and more efficient threats. The ever-growing computer system weaknesses are being looked upon as one of the root causes of the increase in the incidence of worldwide cyber terrorism. But on the other hand, the same technique is utilized to solve different global security issues through machine learning, which includes the detection of malicious software, the identification of ransomware, the detection of fraud, and the verification of spoofing attempts. This article looks into the dual role of online behavior modeling in attack and defense, revealing cyber risks through the application of machine learning tools and software. It also discusses the most common cyber threats and points out how machine learning helps in detection and prevention of attacks, vulnerability assessment, and open-source risk evaluation in the digital commerce sector.

© 2026 Centre for Research and Innovation (CRI). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

I. INTRODUCTION

The digitalization of the world at a faster rate has brought about a secure process on a worldwide scale which has never been seen before. The quests of the scientists for the breakthroughs and the technological advancements have become easier and more visible through the publication of their works in scientific journals with the help of online communication and the partial openness of societies [1–3]. The downside of this openness is that cybercriminals and state actors with malicious intent have equal rights to the access of cutting-edge research along with the scientists and researchers of the government who are legitimate. Among other things, the effect of machine learning on research has also been the uncovering of sophisticated algorithms that are capable of real-time detection of threats and can manage the situation effectively by responding [4]. “Internet barriers” is a broad term that includes the usage of technical

measures, protocols, and procedures that are specifically created to keep threats away and to secure the access to online services, computers, and sensitive data [4]. It is important to highlight that the advancements in artificial intelligence during the year 2016 were very prominent and they affected the whole spectrum of areas such as health care, voice support, and robot assistance to name a few. The digital forensic community has also taken advantage of AI technologies to gain insights from vast audit trails that can help in the detection of malicious actions [3].

Cyberattacks are not easy to stop. They can go through many rounds of attempts before the attackers finally gain access. The attacker's side always has an advantage in this digital battle whereas the defender's side has to maintain a very high level of security which is almost perfect in order to avoid getting hit. Research indicates that in 2017 only, a plethora of businesses, organizations, individuals, and

applications experienced data breaches that incidentally led to the exposure of private records, financial data, and even confidential information [5]. When such information is leaked to unauthorized persons or sold illegally, the negative impact can be disastrous. Cyber threats are substantial, and they affect individuals, corporations, and even institutions across the globe as reflected by numerous statistics:

The economic impact of security breaches has been shown by the fact that, in the past years, losses caused by cybercrime and the actions taken for theft prevention have exceeded \$3.9 billion. Besides that, the technology sector's rapid growth has been forecast to lead to the creation of more than 20 million IT jobs worldwide by 2022. The worldwide institutions will be spending at least \$20 million each year on data protection and cyber security measures to keep their vital information secure. Although the criminals are facing such challenges in their operations, stealing from the digital world is still extremely lucrative, and studies predict that the total annual hacker income from ransom and related cyberattacks might be as high as \$1 trillion [6][7].

II. PROBLEM IDENTIFICATION

The new AI technologies have improved the learning and detecting of cyberattacks systems based on AI learning considerably, with a lot of studies reporting spectacular results. Nevertheless, these improvements do not change the fact that the protection of IT infrastructures against the threats that are constantly changing, and the activities of sophisticated criminal networks, is still a very difficult task. Security measures and innovative solutions have to continuously evolve to match the emerging new style of attacks daily and the myriad of intrusions and malicious activity.

Cyber attacks have been not only increasing in number but also attacking in more sophisticated and complicated ways. Now the attacks got something more complex, it raises the critical question that has to be addressed at all times – Is cyber defense strategy getting enough reinforcements? By now, traditional techniques, such as intrusion detection systems (IDS) and deep packet inspection (DPI), are still in use as well as recommended but they usually do not cater for the rapidly evolving cyber threats.

The growing computing power and decreasing costs are making machine learning as a threat detection tool more popular as a complementary defense against spyware, botnets, and other malicious activities. This paper is taking a look at how effective machine learning is in classifying the harmful network traffic as an alternative approach. First, the NetFlow data is thoroughly scrutinized to identify the attributes. A feature selection phase follows where the features are compared to determine the importance of the features. In this study, the machine learning algorithms are tested on the NetFlow networks that have well-known botnets. The results show that the random forest classifier is capable of detecting more than 95% of the botnets in 8 out

of the 13 test cases and also getting over 55% detection rates in the most computation-intensive datasets, thereby proving its effectiveness.

III. LITERATURE REVIEW

A. Literature Survey

The discipline known as Machine Learning in computer science allows machines to process data and learn from it even if they are not directly trained on it. Instead, it uses computer simulations based on primary data and the latter is used to predict future training data [3]. Due to the fact that the machine learning technology is meant to be used in making decisions regarding the consumer's behavior, likes, and the healthcare needs, there are companies employing AI in a vast range of sectors such as online shopping. AI even used to predict outbreaks or the someone's condition whether they will recover from certain diseases like cancer, by looking at the patient's medical records [5]. Machine learning has been a main driver in improving the security measures during the whole system of detection and prevention of intrusions. The ML algorithms are categorized into two main classes: supervised and unsupervised. They are distinguished by the data that the algorithms are processing [4].

To clarify what distinguishes the markings, those who are familiar with labeled teaching could include simulators that are part of a regulated education methodology as one of the components. Unsupervised learning acts as a method of introducing the elusive learning formulations that are geared towards self-expansion of classes. The labeled data is often extremely sparse [5]. In the case of supervised training, the system's ability to predict the outcome with different learning algorithms is typically defined by a goal parameter. The application of machine learning techniques can be classified into two broad areas either being prediction or the recognition of patterns [4]. For instance, machine learning based model can be used to foresee the favorable online application usage and trend among users as well as whether a certain IP address has been used as the target center in a DDOS attack. Programming with various control systems utilizes a wide range of various diffusion techniques [1].

Moustafa & Slay [1], introduced the UNSW-NB15 dataset to address limitations of older benchmarks (KDD99), combining realistic normal traffic and diverse attack types to create a modern IDS benchmark. The paper includes information on the traffic generation, feature extraction (49 features), labeling methodology as well as statistical analysis to indicate the complexity and realism of this data set with respect to previous data sets. UNSW-NB15 use has increased around the world for testing ML based NIDS and raised problems of learnability, imbalance, overlapping feature extractors which need to be addressed by pre-processing and feature selection. Experiments can be conducted using the dataset with supervised and hybrid ML techniques on the detection of malware, DDoS, and

exploitation. The CTU-13 dataset (CTU-Prague), created by García *et al.* [2], consists of 13 real-world botnet capture scenarios which are combinations of botnet, normal and background traffic. These creators included further information about the specifics of the scenario such as type of the botnets, protocols, and to include labeled flow records for botnet detection research. The value of CTU-13 is the actual capture traces which allow researchers to analyze flow-based ML detectors under different conditions of a botnet. Follow-up analyses test clustering and supervised classification as well as behavior-based detection using CTU-13 but papers emphasize that the CTU scenarios are “dated” and should be used in conjunction with modern datasets to conduct “robust evaluations”. It is very widely available for download and preprocessed versions are available on GitHub and Kaggle.

The CICIDS2017 features flow-based labeled representations of benign and current attack traffic (e.g., DDoS, brute force, infiltration, botnets), that are generated from realistic traffic generators and CICFlowMeter [13]. It aids in realistic IDS experiments and is often employed to test the performance of ML models. Along with its advantages for multi-class classification as demonstrated by some of the studies presented at CICIDS2017, a number of inaccuracies in the labeling and flow-construction by independent research teams are also highlighted, and hence careful preprocessing and validation is recommended. CICIDS is still an important public tool to assess both classical ML and deep learning IDS methods in near-real settings.

Sarhan, Layeghy *et al.* [3] analyze the surveyed NetFlow datasets to assess their appropriateness for use in tackling intrusion detection with ML. They summarize commonly used collections of flows (UNSW-NB15, CICIDS, CTU-13, MAWI), talk about features sets and pitfalls in preprocessing, and explain the issues with class imbalance and offer suggestions on the data selections and flow feature engineering that can be used to combat class imbalance. While highlighting the use of caution in the construction of flows, the importance of having consistent labelling requirements and feature drift is reinforced. Their expertise ensures the selection of appropriate datasets and the processing of these to optimize model generalizability and comparability between studies.

Sriram *et al.* [5], suggested a deep-learning-based approach to detect IoT botnets and suggested network flow features as indicators in their proposed DNN. They perform the flow to connection record conversion and analyze multiple deep architectures to detect the IoT devices compromised in smart cities. The empirical analyses over benchmark datasets demonstrate that DL models (CNNs/RNNs variants) can learn temporal and spatial flow patterns, and are able to often beat the classical classifiers for the IoT botnet detection, particularly given high-dimensional flow features. The study underscores the need for model

compression and optimization to enable deployment on resource-constrained environments.

Haddadi & Mortie [6] discuss the advantages and disadvantages of deep learning models over traditional ML for detecting botnets using NetFlow features. Based on the results of their experiments, they can conclude that CNNs and LSTMs have been found to be capable of learning sequential and temporal features of a botnet and enhance detection performance of classical methods with little feature engineering. They also talk about the computational compromises: Deep models need more resources and hyperparameter tuning for real-time applications. The paper proposes using a hybrid method by combining feature selection and lightweight deep models, to be able to maintain an accuracy that is acceptable when taking into account the feasibility of deployment.

This systematic review examines the techniques for botnets detection, including signature based, behavior based, and ML/DL based. In this systematic review, several techniques for detecting botnets: a signature-based, behavior-based, and ML/DL-based technique were discussed. Shinan *et al.* compare studies using supervised, unsupervised and hybrid models, discuss metrics for evaluating these models, and review data sets that have generally been used in order to benchmark these. It also outlines developing patterns, such as flow-based analytics, SDN-centric detection, and ensemble learning and recommends standardized evaluation procedures, cross-dataset validation, and focus on adversarial robustness. They recommend considering multiple datasets, monitoring and reporting of resources, and studying how to transfer learning to improve generalisation. Survey provides taxonomy for stop DDoS methods—supervised, unsupervised, hybrid and discusses catches, information sets, and DDoS mitigation strategies in the cloud. The review also notes high-performing solution techniques such as hybrid and ensemble approaches, but raises several challenges: Low and slow attacks, encrypted traffic and scalable, low latency detection pipelines. The paper is an actual guide to implementing cloudscale DDoS defense using ML algorithms, and promotes research efforts for developing adaptive models and studies on the deployment of ML models in realtime environments [8].

To prevent this, suggest an attack-specific feature analysis framework, which focuses on identifying important features for the different types of attacks, instead of analyzing the whole data set. It exploits two versions of Net Flow datasets - the enhanced NetFlow datasets for Internet of Things (NF-BoT-IoT-v2 and NF-ToN-IoT-v2) - and six filter based feature selection techniques such as Correlationbased Filter, RELIEFF, Relief, Chi Square, InfoGain and GiniImpurity with the unsupervised ML models Isolation Forest and Autoencoders. Experiments show that attack-specific feature selection can boost the detection accuracy for various attack classes, and that focusing on the prominent features per attack can also increase the efficiency of the detection system as well. This framework is especially

beneficial for resource-constrained IoT: where computation overhead is paramount [10].

M. Hossain *et al.* [10] proposes a Deep Q-Learning IDS (DQ-IDS) system that leverages the ability of Deep Q-Networks (DQN) to learn attack behaviors through replay of past experiences with a random adaptation exploration algorithm. The RL agent continuously improves the detection policies as it keeps getting rewarded for making correct classifications and gets penalized for the other otherwise. Unlike static ML models. The system has good capabilities on real network datasets, with detection accuracy ~97.18%, and false positives less than original classifiers. One of the most interesting features of DQ-IDS is its ability to learn and adapt over time to new threats, it also has no requirements for labeled data, and it is encouraging for the realtime environment of cyber security. In a systematic study, R Chinnasamy *et al.* surveyed the recent developments of deep learning (DL) techniques for network intrusion detection (NID), summarized, and identified the trends in DL architectures, data processing, and assessment metrics. It compares the different types of models including CNNs, RNNs, Autoencoders and Hybrid networks, which have proven effective in improving detection accuracy and reducing false alarms. Emerging datasets, and deployment challenges to realworlds, are also discussed. The key trends in question are flow of transfer learning, more emphasis on attention mechanisms and end-to-end DL pipeline. The paper suggests ways to improve the explainability and operationalization of DL-based IDS, given the scenario of poor explainability and scalability that exists in the current research [11].

VZ Mohale *et al.*, [12] compare various ML models across IDS problems, and conclude that the use of machine learning techniques significantly enhances the IDS's ability to detect novel attacks, while reducing false alarms. This study verifies the superior performance of supervised and ensemble models (Random Forest, Gradient Boosting) at enhancing detection performance over rule-based models with respect to pattern recognition. It also highlights the need of having a proper preprocess and feature selection to have any reliable performance in different network environments. The paper proposes more using Elliptical and explainable mechanism with ML in next generation IDS. M Luay *et al.* [13], highlight the importance of temporal features, like inter-packet arrival time and flow duration, which were missing from most of the available NetFlow datasets, and seek to integrate them into IDS realm for ML-based studies. They point out that traditional NetFlow doesn't provide any time context, and this hampers the discovery of time-based attack patterns. The authors construct temporal NetFlow datasets and outperformed by doing a comprehensive temporal feature analysis and show the effectiveness of model discrimination between normal and malicious flows. The paper demonstrates that temporal quantification improves feature representations, thereby improving the accuracy of intrusion detection.

In this paper, J Li *et al* [14] suggest a Genetic Algorithm (GA) optimized ensemble deep transfer learning model named NFloT-GATE-DTL for intrusion detection of NetFlow traffic in IoT network. The method focuses on using GA to tune deep learning ensembles, which enhances their adaptability and accuracy in the context of IoT attack patterns. Experimental results demonstrate the effectiveness of the classification of various intrusions in lower training times and better generalization of the classification of different benchmark datasets. The proposed hybrid ensemble and transfer learning strategy can greatly improve the robustness of IoT IDS. In their paper G. Granadillo *et al.*, the authors study a wide range of supervised machine learning and deep learning models for anomaly and multi-class classification on a large dataset of NetFlow data containing more than 31 million connection records. They compare the accuracy of the binary model to that of the multi-class model and conclude with the idea that there is no single best model in all cases but provide a possibility to select an algorithm that delivers as high as 99.9% accuracy in certain attack categories. It highlights the significance of the diversity of the models used and the analysis of features to tackle multiple cyber threats such as DoS, brute force attacks, and port scanning. They conclude that ensemble methods and hybrid classifiers can be used for overall IDS performance [15].

B. Literature Summary

The most recent studies draw attention to the fact that machine learning and artificial intelligence have changed the ways traditional cybersecurity methods to the point where they now identify attack, threats, and perform monitoring in real-time. It has been proven that several learning models such as supervised, unsupervised and ensemble models are very efficient for detecting malware, the tracking of botnets, phishing and even detecting anomalies among the vast network traffic. However, the researchers do not agree that there are no solutions to some of the issues such as Fisher's rapidly changing attack patterns, the high volume of false positives, or lack of transparency within the system. There are several publications that state that now is the time for the use of an adaptive model, capable of handling heavy and dynamic data streams, with little human participation. The findings in this review of existing research confirm the need for applying modern learning techniques in conjunction with reliable, scalable and interpretable infrastructures to create products of next-generation cybersecurity to address both the complexity of the cyber threats and their emergence.

C. Research Gap

While a tremendous amount of progress has been made in machine learning and artificial intelligence for cyber threat detection and monitoring, there are many significant gaps left to fill. The existing systems mainly depend on unchanging datasets and as a result, they are not able to keep up with the latest and the sophisticated methods of

attack such as zero-day exploits or polymorphic malware. The majority of the studies concentrate on the accuracy of detection but at the same time do not consider the difficulties that come with practical implementation such as real-time processing, scalability for huge networks, and low false-positive rates. Furthermore, very little research has been conducted to investigate secure deployment of distributed learning approaches such as federated learning in order to guarantee safe handling of sensitive training data. Lack of explainable models behind deep learning technologies has also been a hurdle to secure the trust and acceptance of cybersecurity practitioners. Another barrier that has been stumbling the trust and acceptance of cybersecurity practitioners is their absence of explainable models behind deep learning technologies. The mentioned gaps bring the requirement of adaptive, interpretable and robust learning frameworks for the next-generation cybersecurity systems to the fore.

IV. METHODOLOGY

A. Criteria for Selecting this Study

1. *Emerging Threats*: As cybercriminals are becoming more ingenious, advanced detection systems are needed to monitor sophisticated, increasingly difficult-to-detect attacks like ransomware, phishing and advanced persistent threats (APTs). Immediate Response – Traditional system cannot respond to Zero-Day attacks, this necessitates the need for intelligent, adaptive detection systems.
2. *Utilization of AI/ML*: AI/ML to be used in anomaly detection speeds up detection of suspicious activities with minimal number of false positives in real time.
3. *Scalability & Adaptability*: The research is interested in strategies which can both cope with new threats, and apply to various network configurations and industries.
4. *Data-Driven Approach*: Decisions about the models trained and validated are based on a data-driven approach using APIs with huge amount of network load and behavior data.
5. *Industry Relevance*: By bringing this to the forefront of sectors needing data security, such as government, defense, healthcare and finance, it is a very relevant phenomenon in these industries, which are the primary focuses.
6. *Regulatory Compliance*: It aids in complying with data protection regulations such as GDPR, HIPAA, and NIST.
7. *Better accuracy*: They differentiate the solutions with regard to detection accuracy provided, and they have a nominal false alarm rate.
8. *Feasibility for Deployment*: The techniques used are assessed with respect to their implementation in a real-time security monitoring system.
9. *Contribution to Research*: The research aids to diminish gap between research and application in the following domain – Cyber-Attack Detection.

B. Method of Analysis

1. *Data Collection*: Generate real-time network traffic data and historical network traffic data in live environments and from trusted cybersecurity datasets.
2. *Feature Extraction*: Identify important features; such as packet size, flow duration, source/destination IPs and protocol type, which are beneficial for further analysis.
3. *Data Preprocessing*: Data inconsistency removal through cleaning, normalization, and encoding to enhance model accuracy.
4. *Algorithm Selection*: The application of machine learning models (like Random Forest, SVM, and Neural Networks) will be made to classify normal and malicious activities.
5. *Model Training & Testing*: Dataset division into training and testing sets for performance evaluation purposes.
6. *Performance Evaluation*: Metrics such as accuracy, precision, recall, F1-score, and ROC-AUC will be used for validation.
7. *Continuous Improvement*: Establish feedback loops for adaptive learning in response to changing cyber threats.

C. Evaluation of Methodologies Used in the Reviewed Studies

1. Machine learning (ML) and deep learning (DL) models such as SVM, Random Forest, CNN, and LSTM have been adopted in most reviewed studies for cyber-attack detection, which guarantees the improved accuracy and adaptability of the methods.
2. The dimensionality reduction methods such as PCA, correlation-based filtering, and mutual information have been applied extensively to enhance detection speed and to reduce the dimensionality of the data.
3. Hybrid approaches integrating ML/DL with statistical or rule-based methods show superior performance in the detection of zero-day and sophisticated attacks.
4. The datasets KDDCup99, NSL-KDD, CICIDS2017, and UNSW-NB15 are among the most widely used ones, but their use is limited because of outdated attack patterns and the absence of diversity in real-world traffic.
5. The accuracy, precision, recall, F1-score, and ROC-AUC have been used for model evaluation, and the false-positive rate (FPR) has been reported in some models to measure reliability.
6. Some studies have integrated real-time streaming analysis frameworks and thus are more appropriate for real-time network monitoring.
7. A moderate number of methods (but still very few or overfitting in training sets and inadequate testing cross validation to ensure that the methods are generalizable).
8. Future methodology will be robust through focus on adaptive learning, multi-source data fusion and large-scale, real world environment benchmark testing.

TABLE I COMPARISON AND ANALYSIS

Authors & Year	Dataset Used	Methodology / Algorithms	Key Findings / Performance
Moustafa & Slay (2015)	UNSW-NB15	Statistical analysis, ML baseline models	Introduced a realistic IDS dataset; highlighted class imbalance and feature overlap challenges.
Garcia <i>et al.</i> (2011)	CTU-13	Flow-based ML, clustering, supervised learning	Provided real botnet traffic scenarios; effective for botnet detection but limited by dated attack patterns.
CIC (2017)	CICIDS2017	Flow-based ML & DL benchmarking	Suitable for multi-class IDS; requires careful preprocessing due to labeling inconsistencies.
Sarhan <i>et al.</i> (2020)	UNSW-NB15, CTU-13, CICIDS, MAWI	Dataset survey, feature engineering analysis	Emphasized dataset selection, feature drift handling, and preprocessing best practices.
Sriram <i>et al.</i> (2020)	IoT Flow Datasets	CNN, RNN-based deep learning	Deep learning outperformed classical ML for IoT botnet detection; noted deployment challenges.
Haddadi & Mortier (2023)	NetFlow-based datasets	CNN, LSTM vs classical ML	DL models captured temporal flow patterns better; higher accuracy with increased computation cost.
Shinan <i>et al.</i> (2021)	Multiple public datasets	Survey of ML/DL botnet detection	Highlighted need for standardized evaluation, ensemble learning, and adversarial robustness.
Najafimehr (2023)	Cloud & Flow datasets	Supervised, unsupervised, hybrid ML	Hybrid and ensemble models effective for DDoS; challenges in encrypted and low-rate attacks.
Choi <i>et al.</i> (2025)	NF-BoT-IoT-v2, NF-ToN-IoT-v2	Feature selection + Isolation Forest, Autoencoders	Attack-specific feature selection improved detection accuracy and efficiency in IoT networks.
Hossain (2025)	Real network datasets	Deep Q-Learning (DQN)	Achieved ~97% accuracy; adaptive RL-based IDS reduced false positives and labeling dependency.

D. Identifying trends, developments and issues

1. Trends:

- Growing adoption of AI and ML-driven advanced anomaly detection to detect complex attacks.
- The shift of learning in the direction of immediate threat surveillance and preemptive defense strategies.
- Combining the threat intelligence systems for co-battle security.
- Growth of behavioural analytics for insider threat and zero day exploit detection.

2. Advancements:

- Extremely high accuracy of deep learning models for malware and phishing detection.
- Collaborative training of the model without sharing information with others.
- Future network intrusion detection systems (NIDS) that can take automatic action in response to threats.
- Use of blockchain in cybersecurity to ensure data integrity and create tamper-proof logs.

3. Challenges:

- The technique that attackers use to try to avoid being detected.
- Blinded security by high false positives rates which result in fatigue with security alerts.
- The lack of balance between data in training sets that affects model performance.
- No detection in real time for extensive, fast-paced networks.
- Where AI-driven detection methods are being used, privacy and compliance ensured.

V. DISCUSSION

A. Report of Previously Published Work

The literature on cyber detection that has been reviewed points out a lot of progress in the use of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) for the purpose of detecting and stopping cyber threats. Across well documented research, hybrid models (supervised and unsupervised) have been shown to accurately detect anomalies. With Big Data analytics and real-time monitoring, it is possible to forewarn of threats. The cloud-based and distributed architectures used in the detection system have amplified the simplicity and flexibility of new attack vectors, which is arriving.

However, there are also some limitations for it: high False Positive rate, high computational complexity, limited diversity of datasets. So far, there has been a transition towards zero-day adaptative self-learning and automated cybersecurity systems. The findings highlight the importance of XAI to ensure greater transparency and trust. Sharing of threat intelligence between firms is proving to be a key strategy for cultivating cyber-resilience, in addition. The current research papers suggest that overall, the forthcoming progress will focus on the smart, adaptable, and resource-efficient detection methodologies that will prove capable to manage the growing level of complexity in the cyber threats.

B. Methodology for Future Research Directions

1. *Cyber Security Issues:* Machine Learning has been accepted in the last four areas of Cybersecurity as a major tool obfuscating Cybercrime detection systems, Malware analysis, Mobile malware detection, and Spam or Fraud detection. Intrusion detection becomes the most important

aspect, where unauthorized access by malicious programs or policy violations threatens the security of data. Different techniques are available for monitoring intrusion and among these two main categories are signature-based and anomaly-

based approaches [10]. In the case of the signature-based approach, packets are comparing with known attack patterns to detect insider actions that are suspicious.

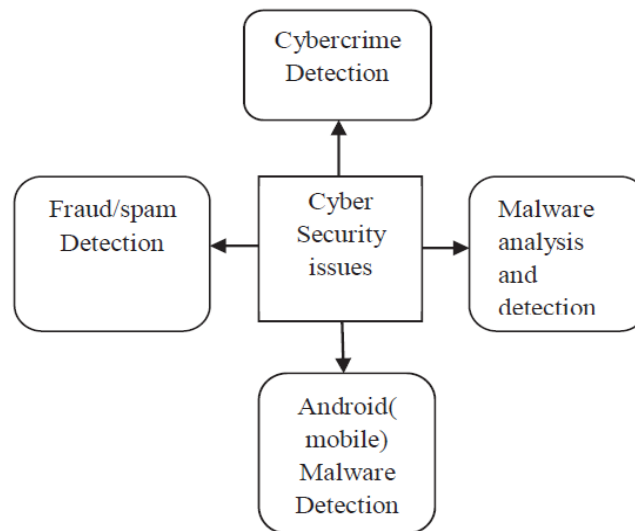


Fig.1 Cybers Security Issue

2. Cybercrimes Detection: Malware is an abbreviation for "malicious software," which refers to the set of programs that are created mainly for harmful purposes. Such applications are regularly error or crash damage systems, or steal, and even access data without permission. Among the hacking tools, ransomware is one of the most common and their types are worms, trojans, viruses, bugs, rootkits, and adware [11]. Besides, they can be classified into groups or families—for example, Charger, Jisut, Koler, Pletor, Svpeng, and Simplocker—that are tagged for ransom attacks [14].

3. Mobile Malware Detection: As the global leading mobile operating system, Android is the most common target for malware creators. The rapid growth of Android applications every day complicates the task of recognizing and categorizing bad apps more and more. Different solutions exist, for example, DroidMat, which applied machine learning techniques like k-Nearest Neighbors (K-NN) and k-means clustering on static app features to identify malicious apps.

4. Fraud/Spam Detection: Another significant problem in data management today is fraud detection. One of the most common forms of spam, which is usually described as unsolicited messages, is made up of irrelevant emails, social media posts, or misleading ads. To seduce users, a lot of

spam messages act as if they are coming from trustworthy sources such as banks—causing users to lose their time and possibly money [11]. Researchers have used advancement of technology towards the use of machine learning techniques as the main way to detect and filter the communications that have been declared fraudulent.

5. Machine Learning for Cyber Security: Figure 2 shows examples of machine learning models that can help solve particular cyber security problems. Advanced threat detection often involves using computer vision algorithms and a robust Classifier like a recurrent neural network (RNN). Artificial neural networks (ANNs) and convolutional neural networks (CNNs) have been found to perform better than the traditional detection methods. Many times, the malware code is first transformed into something resembling images to be processed by CNN. The current machine learning strategies and fusion framework, however, come with limitations which pose problems for comprehensive botnet detection.

The process of threat detection starts with data collection and preprocessing that are very detailed, as depicted in Figure 3. The final evaluation relies on the experimental results, which confirm the efficiency of the proposed detection framework.

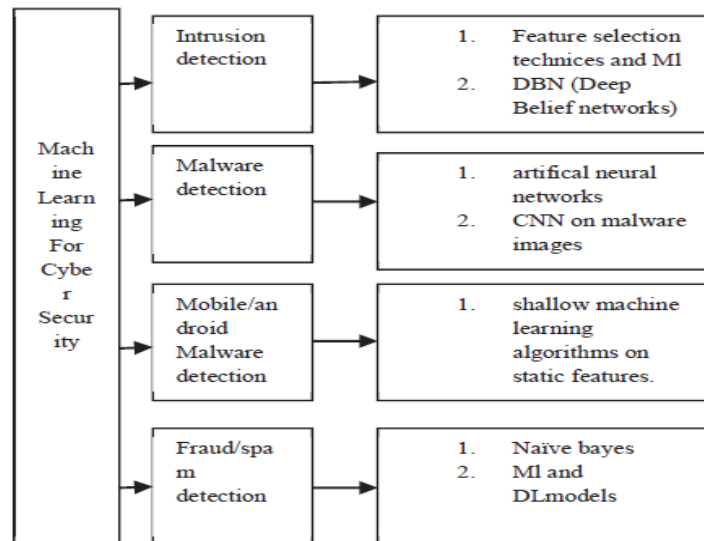


Fig.2 Machine Learning in Cyber Security

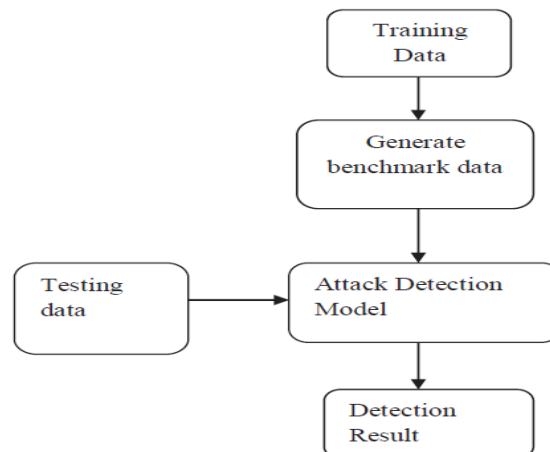


Fig.3 Flow Chart for Cyber Attack Detection

VI. CONCLUSION

The application of machine learning methods to cybersecurity has become a trend, giving rise to new ways of thinking due to the combination of AI and advanced analytical techniques. Nevertheless, choosing the right method for every single task is necessary in order to guarantee productivity. To keep the defense models against malware very strong, micro-level operations are required to be done, which in turn will lead to the development of global and very accurate systems. The choice of the proper algorithm framework is of utmost importance especially when it comes to resolving complicated cryptographic issues. Our method started with arranging security functions according to their importance, followed by the crafting of an authentication system that was both simple and effective. The system was based on decision trees that were made out of the most significant features pinpointed during the selection process. By the intentional use of lightweight strategies during the model building process, we were able to cut down on computation costs, and at the same time, increase the accuracy of the threat defense, even in the case

of uncertain testing conditions. And this led to the final tree structure being optimized.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation

The author(s) utilised generative AI technologies solely to improve language clarity and refine grammar. No images were generated, manipulated, or altered using AI-assisted technologies.

REFERENCES

- [1] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in Proc. Military Communications and Information Systems Conf. (MilCIS), Canberra, Australia, 2015, pp. 1–6.
- [2] S. García, A. Zunino, and M. Campo, "The CTU-13 dataset: A benchmark for botnet detection," Stratosphere IPS Project, Czech Technical University, Prague, Czech Republic, 2011.

- [3] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. Int. Conf. Inf. Syst. Security Privacy (ICISSP), Funchal, Portugal, 2018, pp. 108–116.
- [4] M. Sarhan, A. Layeghy, N. Moustafa, and M. Portmann, "NetFlow datasets for machine learning-based network intrusion detection systems," in Big Data Technologies and Applications, Springer, 2020, pp. 117–135.
- [5] S. Sriram, R. Vinayakumar, M. Alazab, and K. P. Soman, "Network-flow based IoT botnet attack detection using deep learning," in Proc. IEEE INFOCOM Workshops, Paris, France, 2019, pp. 189–194.
- [6] H. Haddadi and R. Mortier, "NetFlow-based botnet detection using deep learning," Security and Communication Networks, vol. 2023, Art. no. 2856912, 2023.
- [7] M. Shinan, A. Abusitta, and M. Guizani, "Machine learning-based botnet detection: A survey," Symmetry, vol. 13, no. 4, pp. 1–29, 2021.
- [8] A. Najafimehr, "DDoS attacks and machine-learning-based detection: A taxonomy and survey," Engineering Reports, vol. 5, no. 3, pp. 1–23, 2023.
- [9] D. Choi, J. Kim, and H. Kim, "Attack-specific feature analysis for NetFlow IoT intrusion detection," Journal of Information Security and Applications, vol. 79, pp. 103–118, 2025.
- [10] M. Hossain, "Deep Q-learning intrusion detection system (DQ-IDS)," in Proc. Int. Conf. Information and Communication Technology (ICT), Dhaka, Bangladesh, 2025, pp. 1–7.
- [11] R. Chinnasamy, "Deep learning-driven methods for network-based intrusion detection systems: A systematic review," ICT Express, vol. 11, no. 2, pp. 1–12, 2025.
- [12] V. Z. Mohale, S. B. Jagtap, and A. R. Patil, "Evaluating machine learning-based intrusion detection systems," Frontiers in Computer Science, vol. 7, Art. no. 1520741, 2025.
- [13] M. Luay, A. Alshamrani, and M. Alenezi, "Temporal analysis of NetFlow datasets for network intrusion detection systems," arXiv preprint arXiv:2503.04404, 2025.
- [14] J. Li, Y. Zhang, and X. Wang, "NFloT-GATE-DTL: GA-tuned ensemble deep transfer learning for IoT intrusion detection," Expert Systems with Applications, vol. 242, Art. no. 122987, 2025.
- [15] G. G. Granadillo, R. Pries, and M. Bagnulo, "AI-based anomaly detection and classification of traffic using NetFlow," in Proc. Int. Conf. Security and Cryptography (SECRYPT), Rome, Italy, 2025, pp. 321–331.
- [16] D. Wagner and R. Soto, "Mimicry Attacks on Host-Based Intrusion Detection Systems," in Proc. 9th ACM Conf. Computer and Communications Security (CCS'02), pp. 255–264, 2002.
- [17] C. Kruegel, D. Mutz, W. Robertson, and F. Valeur, "Bayesian Event Classification for Intrusion Detection," in Proc. 19th Annual Computer Security Applications Conf. (ACSAC), pp. 14–23, 2003.
- [18] S. Benferhat and K. Tabia, "A Bayesian Approach for Intrusion Detection in Large-Scale Networks," in Proc. 5th Int. Conf. Intelligent Data Engineering and Automated Learning (IDEAL 2004), pp. 41–49, 2004.
- [19] L. Koc, T. A. Mazzuchi, and S. Sarkani, "A Network Intrusion Detection System Based on a Hidden Naive Bayes Multiclass Classifier," Expert Syst. Appl., vol. 39, no. 18, pp. 13492–13500, 2012.