Machine Learning and Deep Learning Methods for Cybersecurity

Nihal[1], Nidhi[2], Niranjan Hiremath[3], P Keerthi Reddy [4]

[1] *Student, Computer Science and Engineering (IOT & Cyber-security Including Blockchain Technology), Alvas Institute of Engineering and Technology, Karnataka, India*
[2] *Student, Computer Science and Engineering (IOT & Cyber-security Including Blockchain Technology), Alvas Institute of Engineering and Technology, Karnataka, India*
[3] *Student, Computer Science and Engineering (IOT & Cyber-security Including Blockchain Technology), Alvas Institute of Engineering and Technology, Karnataka, India*
[4] *Student, Computer Science and Engineering (IOT & Cyber-security Including Blockchain Technology), Alvas Institute of Engineering and Technology, Karnataka, India*

ABSTRACT

Internet usage has surged, resulting in a fluctuating cyber-threat landscape that complicates cyber security. This study examines machine learning (ML) and deep learning (DL) techniques applied to intrusion detection for network analysis. Each one of these methods is adequately discussed by the authors based on an extensive literary review. The survey represents the progress of these approaches over time in terms of major themes. Additionally, we have considered some commonplace datasets used in ML/DL-based network studies because data is key. Nevertheless, ML/DL has limitations and challenges that should be taken note of in this section.

Keyword: - *Cybersecurity, Machine learning, Deep learning*

1. Introduction

In daily life, the internet's widespread incorporation has affected major transformations on how people learn and work. Nevertheless, it has also exposed us to more security threats. There is a pressing need to identify different network attacks including those that were previously unknown. Cybersecurity refers to technologies and processes developed to protect computers, networks, programs and data from unauthorized access, alteration or destruction. Some of the important elements in a network security system are firewalls used for scanning packets in order to determine if they are authorized or not; antivirus software which guards against malicious software; and Intrusion Detection Systems (IDSs) which can detect possible threats coming from within the system.

The reason why security may be compromised is due to external and internal intrusions, which include the misuse-based (or signature-based), anomaly-based and hybrid detection. Misuse based techniques depend on known attack signatures to detect known threats, but they may have challenges in identifying new (zero-day) attacks. On the other hand, anomaly-based techniques examine normal network and system behavior for deviations which has a possibility of detecting zero-day attacks but may result in high false alarm rates. Hybrid detection is a combinationof both misuse and anomaly detection that increases detection rates for known intrusions while reducing false positives for unknown attacks; many machine learning (ML) and deep learning (DL) methods fall into this category.

This research paper presents a review of the literature on cybersecurity applications of ML and DL in network intrusion detection. It presents in detail various techniques in ML and DL using examples of their applications to cybersecurity. It focuses on ML and DL techniques for anomaly detection but also highlights other approaches such as signature based and hybrid methods. Furthermore, this paper looks into how these ML and DL techniques are applicable when it comes to intrusion detection in wired and wireless networks – considering that the latter has been found to be more prone to malevolent intrusions than the former. The survey is organized into similarities and differences between ML and DL, datasets used in ML and DL for cyber security, methods including related papers for ML and DL within cybersecurity, current studies being conducted concerning this topic area, as well asanticipated future directions which will hopefully offer a more complete comprehension of ML/DL based network intrusion detection.

2. Similarities and Differences in Ml and Dl

The two DL and ML are both based on data for predictions but they vary in complexity and methodology. Simple models with manual feature engineering work well with the machine learning while automated feature extraction is done through deep neural networks by the deep learning ones. DL more computational power for its complex architecture, both are important but in different field of high depth of representation and computation.

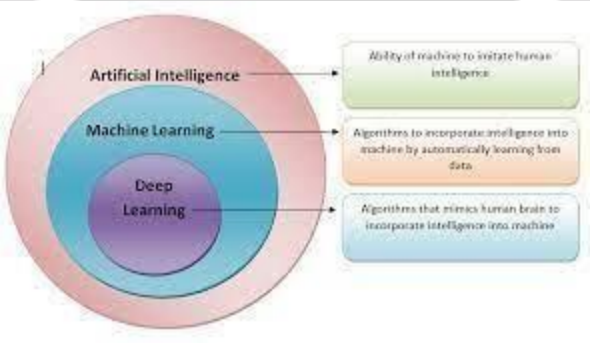

Fig-1: Similarities and Differences in Ml and Dl

3. Network Security Data Set

A network security dataset usually contains several logs, events and traffic data that are gathered from diverse sources in a specific network infrastructure. These datasets may be used to develop machine learning models for improved detection of network intrusions over normal traffic.

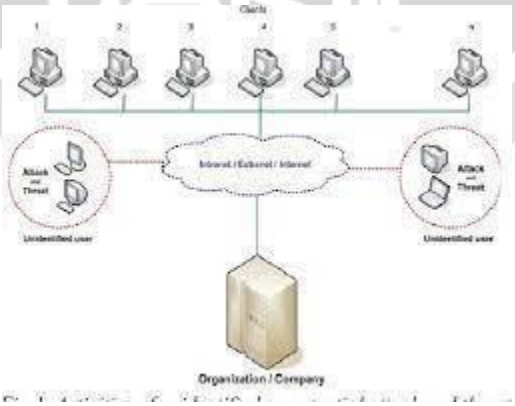

Fig-2: Network security data set

Some of the main parts of network security dataset can be like below.
1. Network traffic logs: Information about communications occurring between devices on a network. For example, IP addresses, port and protocol information associated with incoming and outgoing communications along with timestamps.
2. Intrusion detection alerts (created by an intrusion detection system IDS or IPS indicating possible malicious activity/ security breaches).
3. System logs (networking devices -> servers and services) => Provides the log information regarding the networking devices, and systems of one or more networks. System actors log events such as administrative changes made in network device configuration, login/logout attempts on system service/processes maintenance functions accessed during specific time ranges etc.; many important security relevant sensors are capable pf acquiring this data
4. Not more than Packet Captures are the raw network traffic data captured at packet level (means they do have information what individual packets together having + context.
5. The NSL-KDD dataset, UNSW-NB15 dataset and CICIDS2017 dataset: These are the datasets used for the research work related to security tool development. These datasets consist of normal activities as well as intrusion activities, so that researchers can perform evaluation of their algorithms on different cyber threats.

4. ML and DL Algorithm for Cybersecurity

ML and DL Algorithm for Cybersecurity: The traditional Machine Learning algorithms and Deep Learning techniques are used respectively in the cybersecurity domain.

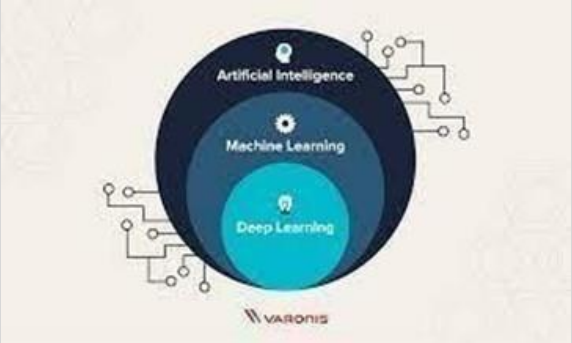

Fig-3: ML and DL Algorithm for Cybersecurity

Examples of each are the following: Machine Learning Algorithm:
1. Random Forest: Random Forest comes as one of the most used Machine Learning algorithms in cybersecurity tasks like intrusion detection. It trains many decision trees to output the mode of the classes in case of classification problems or the mean prediction in case of regression problems of individual swayed trees.
2. SVM: They are the best algorithms that classify the data into different categories. Hence this aids in malware detection and network intrusion detection for finding out the hyperplane which separates the different classes of data in the best possible manner.
Deep Learning (DL) Algorithm:
1. Convolutional Neural Networks (CNNs): They are massively applied in cybersecurity, specifically in areas where tasks—like malware detection or image-based intrusion detection—can profit from this kind of models. They have been effective in learning hierarchical representations of data; thus, they are very suitable, particularly in complex pattern analyses such as network traffic or binaries of malware.
2. Recurrent Neural Networks: RNN and its variants, including Long Short-Term Memory networks, are trained for analysis of sequential data applied to such tasks as analysis of the network traffic or pinpointing anomalies in time-series datasets.
ML & DL: Both algorithms of ML and DLL have their own benefits, hence applied based on relative requirement in cybersecurity. Tasks suited to ML algorithms, that are more suited for tasks in the DL realm needing a lot of data and complex patterns to be learned. Discussion and Future Direction The merging of machine learning with deep learning in cybersecurity allows all this to be affected more efficiently in detection, response, and prevention.

5. Discussion and Future Direction

The convergence of Machine Learning (ML) and Deep Learning (DL) in cybersecurity has the capability to significantly improve threat detection, response, and prevention. These two cyber patterns help to detect possible threats [11] For analyzing large data, ML techniques were used a lot and with new DL approaches due for handling complex structures of the data pattern detection accuracy is improved.

Chart-4: Cybersecurity Paradigms

Going forward, some of the key areas that advancements in ML and DL will cater to are:
1. Artificial Intelligence (AI) is expected to additionally develop and its utilization in industry for threat detection will increase as well; we should expect Machine Learning (ML) and Deep learning models which can really detect zero-day attacks or even advanced persistent threats (APTs).
2. Behavioral analysis: ML and DL algorithms will be fine-tune to perform behavioral based risk analytics of users/entities for detection in case a user/entity behaves oddly (means insider threats or the account is compromised).
3. Automated Response Systems - ML driven automation will be key in responding to real-time cyber threats, where machines are competent enough to fight back while the attack is taking place by autonomously mitigating attacks or isolating compromised systems.
4. Explainable AI (XAI): Focus on providing more interpretability and transparency into ML/ DL models, giving cyber security staff the ability to understand why an AI made a specific decision in relation of their Security System.
5. Adversarial Machine Learning: This will involve hardening machine learning and deep learning models against adversarial attacks, ensuring robustness to attempts of tampering with these models or bypassing their detection mechanisms.
6. ML and DL models are tailored to run on edge devices for advanced, real-time threat detection in distributed IoT environments Edge Computing & IoT Security.
7. Privacy-preserving techniques will involve an addition of encrypted features in ML and DL algorithms to protect the source data privacy while retaining effective threat detection & analysis.
To conclude, the future of ML and DL within cybersecurity purely adds in its natural path developing defenses that are much more tailored to beat new age cyber threats across newer but constrained pathways.

6. CONCLUSIONS

In this literature review, we have presented the state-of-the-art methodologies based on machine learning (ML) and deep learning methods for intrusion detection in network security throughout recent years. Extensive research has yet to declare any single approach a winner - each comes with its own set of strengths and weaknesses. Yet, a key challenge remains the availability of adequate high-quality datasets: due to issues such as data imbalance and obsolescence. In support of this, future research will also need to consider incremental and lifelong learning models as the rapid evolution of network data means that retraining must be continuous.

7. REFERENCES

[1]. Machine Learning and Deep Learning Methods for Cybersecurity: Journal Article