Cybersecurity Threats Detection in Intelligent Networks using Predictive Analytics Approaches

Shomili Duary
Kalinga Institute of Industrial Technology, India 2129104@kiit.ac.in

Pratyusha Choudhury
Kalinga Institute of Industrial Technology, India
2129084@kiit.ac.in

Sushruta Mishra
Kalinga Institute of Industrial Technology, India
sushruta.mishrafcs@kiit.ac.in

Vandana Sharma, SMIEEE
Department of Computational Sciences, CHRIST (Deemed to be University),
Delhi NCR, India
vandana.juyal@gmail.com

Deepak Dasaratha Rao
Independent Researcher
Consultant, Infosys, USA
Ddrb2011@gmail.com

Adedapo Paul Aderemi
School of Creative Technologies
University of Bolton, United Kingdom
apa1crt@bolton.ac.uk

***Abstract*— The modern scenario of network vulnerabilities necessitates the adoption of sophisticated detection and mitigation strategies. Predictive analytics is surfaced to be a powerful tool in the fight against cybercrime, offering unparalleled capabilities for automating tasks, analyzing vast amounts of data, and identifying complex patterns that might elude human analysts. This paper presents a comprehensive overview of how AI is transforming the field of cybersecurity. Machine intelligence can bring revolution to cybersecurity by providing advanced defense capabilities. Addressing ethical concerns, ensuring model explainability, and fostering collaboration between researchers and developers are crucial for maximizing the positive impact of AI in this critical domain.**

***Keywords*— AI (Artificial Intelligence), cyber security, cyber threats, threat detection**

I. INTRODUCTION

Artificial Intelligence or AI involves simulation or approximation of human intelligence in machines. The ability for artificial intelligence to reason and act in a way that maximizes the likelihood of reaching a given objective is the ideal feature. It is being employed in a variety of industries today, including healthcare and finance. These days, AI is widely applied in many other different contexts and to differing degrees of complexity. Popular AI applications include recommendation algorithms that make suggestions about what you would like to do next and chatbots that show up on websites or as smart speakers like Alexa or Siri. Machine learning with data analytics have become very relevant for prevention against cybersecurity threats, they can process huge quantity of information to detect risks like phishing and malware. To avoid detection, cybercriminals might alter virus code, nevertheless. Because machine learning can use information from previously identified malware to identify new varieties, it is perfect for anti-malware defense. This is effective even when malicious code is nested inside benign code. Tools for network monitoring with AI capabilities can monitor user activity, spot irregularities, and take appropriate action. The ever-evolving strategies of hackers make it difficult to identify every possible threat to a corporation. Because unknown risks can cause significant harm if left unnoticed, it is imperative that we employ cutting-edge solutions like artificial intelligence (AI) to successfully identify and prevent them. It has the potential to be resource-intensive and impractical. Cybercriminals may potentially employ AI to enhance their attacks. VPNs is a sector that gains from analytics since predictive intelligence enables them to defend users against AI-posed online dangers.

The practice of securing system nodes, systems, and data from damage, loss, illegal access, and other cyber threats is essentially known as cybersecurity. The increasing reliance on digital technologies in all aspects of our lives has led to the rise in importance and activity of the cybersecurity area. Cyber-threat environment, one of the variable factors of cybersecurity, has a threat landscape that is ever-changing as a result of different actors using sophisticated ways to exploit vulnerabilities for espionage, disruption, or financial gain. These actors include hackers, cyber criminals, nation-states, and hacktivists. Networking privacy, nodes authenticity, software privacy, information privacy, IAM etc are some of the key-elements of cybersecurity. Some typical cyber hazards such as malware which refers to malicious software, such as viruses are intended to damage or exploit computers, phishing which involves the act of assaulting or deceiving people into divulging private information, including passwords or bank account information, distributed denial of service (DDoS) that is the act of flooding a system or network with so much traffic that it becomes unusable for users, Man-in-the-Middle (MitM) etc diminishes the moral of cybersecurity. The advancement of cybersecurity depends upon artificial intelligence (AI), which provides cutting-edge capabilities to combat sophisticated threats, consisting of anomaly detection in which AI examines large datasets to determine typical behavior and identify anomalies. Behavioral analysis, AI-powered antivirus which uses machine learning to detect and prevent malicious files based on their behavior rather than just their known signatures, automation, user verification, biometric authentication, NLP etc helps in collaborating

with cybersecurity in order to enhance its functionalities. Although artificial intelligence (AI) can greatly improve cybersecurity, it is important to remember that AI is not a panacea. To effectively protect systems and data, a holistic cybersecurity strategy combines AI technologies, human knowledge, and other protection measures.

The main contribution of the paper are as follows:

- Highlights the immense potential of artificial intelligence to transform cybersecurity through pattern recognition, automation, and analyzing massive datasets.
- In this paper we proposed a model in the form of a flowchart that presents a thorough approach to vulnerability evaluation and cybersecurity threat identification that incorporates both conventional and historical security data sources.
- The field of AI cybersecurity is constantly evolving. While the potential benefits are significant, carefully considering and addressing these challenges is essential for responsible and effective AI deployment in this critical domain.
- AI holds great promise for advancing cybersecurity through its unparalleled data processing and pattern recognition capabilities. However, responsible development and pragmatic expectations are essential to maximize benefits and minimize risks when deploying AI for cyber defense.

II. RELATED WORKS

Authors proposed a model for AI driven Cybersecurity [1] whose primary goal is to act as a resource and set of recommendations for industry professionals and cyber-security researchers, particularly when it comes to AI-based technological aspects. The methods used for this purpose are K-Nearest Neighbor, Naive Bayes, Random Forests, Adaptive Boosting, RNN, LSTM, CNN, Hidden Markov Model which aids in fulfilling various purposes: Intrusion detection analysis, attack classification, DDos detection and analysis, preventing cyber terrorism etc. Henceforth, analytics driven structuring may be applied to several applications which ranges from risks assessment to abnormality detection which may result in a phishing risk as highlighted in the study. Various states of AI in cybersecurity have been examined in paper [2] L. Chan et al. who presented the analysis of machine intelligence in Cyber Security for Information Technology Management that leverages the present mode of AI in the cybersecurity field and describes several works based use cases of data analytics to assist the society identify various network risks and develop complex decisions. With the use of a database containing malware and threats, AI may be trained to classify files or specific behaviors as either supervised or unsupervised using supervised and unsupervised machine learning techniques. Neural Networking where each neuron in a neural network represents a point in a multidimensional space and is connected to its neighbors. When used in conjunction with clustering techniques, it aids in the detection of malicious IP traffic. By computing the behavior patterns that the system employs, Deep Learning—another crucial technique uses predictive powers to identify problems even before they arise. Consequently, AI technology will help to detect and diverge into various levels of networks risks as they develop which offers us excellent intrusion and detection capabilities, detects false positives, and uses predictive analytics to increase the security of information on the internet. Expert systems and a broad variety of neural networks, including both ANNs and DNNs, are still being released. This study [3] proposed by B. Thuraisingham explores the importance of cybersecurity and artificial intelligence in social media. The study examined several techniques of machine learning for social media platforms. Sentiment analysis techniques can detect user opinions and identify the spread of hazardous diseases or human trafficking activities. Machine learning also enables the detection of false information and malicious software on social platforms. The paper even looked at the security and privacy challenges, including access control methods and privacy-aware systems tailored for social media. Lastly, it discussed various ways in which AI and cybersecurity can be integrated to tackle threats to social platforms like adversarial attacks, inference risks, and data privacy issues. This paper [4] aims to highlight the dual usage of AI/ML in cybersecurity, for both defense and attacks, to propose a taxonomy classifying different types of AI/ML-enabled cyberattacks. Both AI/ML methods were used in the study to examine how AI/ML can be leveraged for good through cyber defense as well as weaponized to optimize attacks that can bypass traditional defenses. The combination of AI and ML techniques enhances cybersecurity through complementary capabilities, adaptive improvement, multi-pronged defense, accountable ML, scalable deployment, and efficient human-AI collaboration. The authors of the paper [5] proposed a Systematic Literature Review of Studies on Cybersecurity MOOCs which aims at the study of outcome of recent intelligent alternatives in network security. Failures in AI have the potential to affect society as a whole, thus it's critical to give citizens and cybersecurity professionals access to cutting-edge training as well as fundamental AI knowledge. In most cases, covered subjects rather than methods (such AI) are used by cybersecurity MOOCs to structure their educational content. [6] proposed which evaluates how AI might be used to enhance cybersecurity solutions by evaluating its advantages and disadvantages. It also highlights the potential benefits of AI-based cybersecurity solutions in diverting adversaries and reducing or eliminating data breaches. Various cybersecurity threats such as Denial of Service (DoS) attacks, Man-in-The-Middle (MiTM) attacks, Drive-by attacks, Eavesdropping attacks are being solved using cybersecurity solutions frequently carrying out traffic analysis, which categorizes Internet traffic as harmful or lawful. Few algorithms such as naïve Bayes, k-Nearest Neighbor (k-NN), decision tree which iteratively determines which feature best fits the samples of data. Until data samples with just one class are identified after a split, the repeated division generates a succession of rules for each side of the categories, producing a structure like a tree that is implemented in order to prevent the cyberthreats. Thus (AI)-driven cybersecurity solutions have predominantly concentrated on machine learning methodologies that employ intelligent agents to differentiate between malicious and authentic traffic. In the article, Applications of AI in Cybersecurity [7] proposed by authors signifies the key goal of applying AI and ML to help detect anomalous behavior that may indicate new cyber threats. Rather than a fully automated approach, the authors proposed a hybrid human-AI strategy where SMB users

contribute knowledge on normal vs abnormal activity to account for false positives/negatives. AI and ML algorithms have been utilized in this paper to detect abnormal actions. This human-in-the-loop approach with AI/ML improved performance and accuracy over fully automated systems. Authors proposed a paper [8] where the main objective of the study was to predict threats using Artificial Intelligence in the Cybersecurity Domain. This paper gives a clear view of the AI and ML techniques. An unsupervised learning method called clustering was used in the study which produces instances of equivalent cluster grouping. Thus, the process of clustering helps find patterns in data. In some circumstances, it is clustered to utilize the resulting classified data for supervised learning and unlabeled data. According to this study, the hunts that adhered to the structured threat model were more tightly centered with the project's goals. The phase of feedback is able to comprehend the significance of judgments made at the scale and stage through the official hunting purpose. Artificial Intelligence (AI) technologies have been incorporated into cyberspace using which report [9] signifies the impact of AI in cybersecurity and summarizes existing research in terms of benefits of AI in cybersecurity. Authors used various classification and regression algorithms, expert systems also known as knowledge-based systems along with decision tree, support vector machine, KNN, random forest in the cybersecurity domain. Several other algorithms feed forward, convolutional neural and recurrent neural networks etc come under DL algorithms. Threats history can be used in an AI-based system to know about the past threats and use this knowledge to predict similar attacks in the future applying various AI methodologies, such as bio inspired computing and ML/DL methods, or various learning approaches, including reinforcement learning and supervised learning, providing new insights. By strategically evaluating the dangers to virtual machines (VMs), authors in [10] suggested a unique multiple risks analysis-based VM threat prediction model (MR-TPM) to safeguard computational data and prevent adversary breaches. It takes into account several cybersecurity risk factors related to the setup and administration of virtual machines (VMs), in addition to analyzing user behavior. One of the algorithms in use, the random forest classifier (RFC), uses knowledge generated by extracted correlated patterns from users' previous data and the learning capacity of various base learners or decision trees to classify users based on their future behavior. The virtual risks level is recorded with scoring pattern like common vulnerability scoring system (CVSS) which determines the extreme degree of risks of an application. In the paper [11] author studied the role of Artificial Intelligence in the Cyber Security domain. Here, the main aim of AI was to create technologically based tasks that simulate human understanding in order to solve issues, to understand the various AI technologies and how important they are for cyber security and to measure how well AI instruments identify the various cyberattacks. To comprehend the function of artificial intelligence in cyber security so as to detect and stop cyberattacks and crimes, cybercrime-related data or reports Security incidents were gathered using secondary data sources which were then analyzed using two statistical tools , they were:- Measures of central tendency (simple and weighted average and Percentile analysis. The results showed that some of the cyber crimes are down, others are rising. Authors presented the paper [12] where their main aims were as to go about the various approaches and classifications of XAI, investigate the problems and obstacles that XAI is currently facing, ascertain which frameworks and datasets are available for XAI-based cyber defense mechanisms, investigate the applications of XAI in cyber security and the most recent, successful XAI-based solutions, ascertain the obstacles and unfulfilled research needs for XAI applications in cyber security and identify the most significant findings and futuristic innovative trends for XAI in cyber security. Thus, few methods were used to collect , analyze the data and some ML/DL techniques were used for prediction and study. After thorough analysis on XAI in cyber security applications the fundamental principles and taxonomies of the most recent XAI models were finalized with essential resources, including accessible datasets and a generic framework. Additionally, they looked into the XAI oriented network risks models from a variety of case studies scenarios, such as using XAI to defend against various cyberattacks categories in various industrial applications and identifying cyber threats that target XAI models and related defensive strategies.

III. PROPOSED MODEL

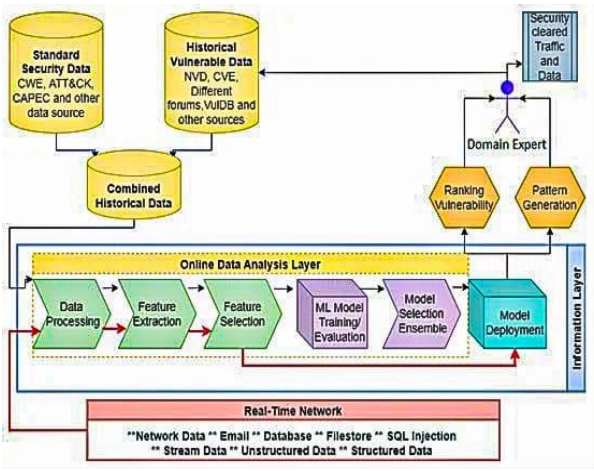

Fig.1 Flow chart of threat detection in cybersecurity using ML

The flowchart illustrates a comprehensive system for threat detection and vulnerability assessment which integrates a number of data sources and analytical techniques to find and rank potential security concerns. Each layer and component of the process has a distinct function within the larger system. Two main types of input data are displayed at the top of the flowchart: "Historical Vulnerable Data" and "Standard Security Data." Information from well-known cybersecurity databases and frameworks, such as CAPEC (Common Attack Pattern Enumeration and Classification), ATT&CK (Adversarial Tactics, Techniques, and Common Knowledge), CWE (Common Weakness Enumeration), and other comparable data sources, may be included in standard security data. These resources offer organized data regarding known security flaws and dangers. The National Vulnerability Database (NVD), CVE (Common Vulnerabilities and Exposures) entries, numerous forums, vulnerability databases (VulDB), and other sources are some of the sources from which the Historical Vulnerable Data is acquired. This data is a treasure trove of

historical details regarding exploits and vulnerabilities from the past that have been tracked down throughout time. After that, these two data streams are integrated to create "Combined Historical Data," which functions as a thorough store of security knowledge that includes both historical records of vulnerabilities and standardized frameworks. As we proceed to the next layer, we come across the "Online Data Analysis Layer," which is in charge of instantly processing and evaluating the merged historical data. This layer is divided into a number of steps that go sequentially, beginning with "Data Processing." In order to ensure that the raw data is in the right format and devoid of errors or unnecessary information, it is cleaned, normalized, and made ready for additional analysis [13-15].

The "Feature Extraction" step happens after data processing. In this stage, particular traits or qualities that are important for locating vulnerabilities are taken out of the processed data [16-17]. The analytical models will use these properties to identify trends and forecast outcomes. The next step is "Feature Selection," in which the extracted set's most important features are selected. This is an important stage since it helps to focus on the most informative qualities and lower the dimensionality of the data, which improves the effectiveness of the models. It is followed by "ML Model Training/Evaluation." The chosen features are used to train machine learning (ML) models, which help them identify patterns and anticipate vulnerabilities. The efficacy and accuracy of these models in spotting possible dangers are also assessed. The term "Model Selection Ensemble" implies that various machine learning models can be employed in tandem, with the top-performing models or a blend of models (ensemble) being chosen to yield the most precise outcomes. "Model Deployment" is the last stage in the Online Data Analysis Layer, where the selected model or group of models is put into use for threat detection and real-time analysis. A feedback loop from the "Real-Time Network" layer, which symbolizes the operational setting in which the deployed models are actively collecting and evaluating data, is also depicted in the flowchart. This layer contains a variety of data kinds, including emails, databases, file storage, stream data, SQL injections, network traffic, and both organized and unstructured data. There is an interface with a "Domain Expert" on the right side of the flowchart, who probably offers expert knowledge and insights to improve the system. Additionally, the Domain Expert might contribute to "Security cleared Traffic Data," which could be utilized as extra training data for the ML models or as a means of validating the results of the system. Additionally, "Ranking Vulnerability," which ranks vulnerabilities according to severity, impact, or other factors, is a result of the Domain Expert's input. In order to concentrate efforts on the most important topics, this ranking is crucial. Finally, as a result of the Domain Expert's contribution, "Pattern Generation" is mentioned, implying that new patterns of attacks or vulnerabilities may be found and introduced to the system for improved detection capabilities. To summarize, the flowchart presents a multifaceted and intricate strategy for cybersecurity that combines machine learning, expert knowledge, and historical data to identify, evaluate, and handle vulnerabilities instantly. The system is meant to be dynamic, constantly improving its threat identification and prioritizing skills by absorbing new information and expert insights.

IV. RESULT ANALYSIS

Table.1 contrasts the accuracy of various predictive analytics models like XGBoost, bagging, boosting, decision tree, random forest and stacking on a specific dataset. XGBoost can very well tackle uneven data samples and complicated data, reduce overfitting, optimized for speed and scale, ensembling improves stability and achieves high accuracy in various classification and regression problems since it shows the highest accuracy among the listed algorithms hence the best algorithm for performing the task assigned.

Table.1 Ensemble Machine Learning Classification Techniques.

Algorithms	**Accuracy**
XGBoost	98%
Bagging	97.62%
Boosting	94%
Decision Tree	93.49%
Random Forest	95.77%
Stacking	92.32%

Table.2 Types of Cyber Threats with training and testing accuracy using XGBoost

Cyber-Threats	**Training Accuracy**	**Testing Accuracy**
DDoS	97.8%	92.76%
DoS	96.2%	94.34%
MiTM	95.47%	92.33%
Malware	96.72%	93.19%
Phishing	92.21%	93.58%
Spyware	92.44%	94.29%
Password Attack	99.18%	98.26%
Cryptojacking	92.82%	93.46%

Table.2 signifies that most of the cyber threats have a high detection accuracy rate, both during training and testing. Nonetheless, there are some differences in accuracy throughout various categories of threats. For instance, the accuracy rate of password attacks is above 98%, although the accuracy rate of phishing attempts is just about 93%. With the majority of threats having a testing accuracy of more than 90%, the security system was able to identify a significant portion of cyber attacks.

V. CONCLUSION

The rapid advancement of cyberattacks creates an urgent need for innovative defenses. Artificial intelligence presents immense potential to radically enhance cybersecurity through its unmatched abilities to process enormous datasets, detect subtle patterns, and automate security workflows. By harnessing AI, we can equip our systems with intelligent protection that evolves along with emerging threats. However, for AI to fulfill its promise in cybersecurity, we must thoughtfully address ethical risks, prioritize model interpretability, and promote synergy between academia and industry. If we collaborate to steer AI's development responsibly, it can provide indispensable reinforcements in the endless battle against cyber crime. But we must remain vigilant - neither underestimating the ingenuity of hackers nor overestimating the capabilities of machines. AI is not a panacea, but rather a powerful asset whose limitations we must recognize. With care and wisdom, AI can become a crucial ally in the fight for our digital future.

REFERENCES

[1] Sarker, I.H., Furhad, M.H., & Nowrozy, R. (2021). AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. SN Computer Science, 2.

[2] Chan, L., Morgan, I., Simon, H., Alshabanat, F., Ober, D., Gentry, J., Min, D., & Cao, R. (2019). Survey of AI in Cybersecurity for Information Technology Management. 2019 IEEE Technology & Engineering Management Conference (TEMSCON), 1-8.

[3] Thuraisingham, B.M. (2020). The Role of Artificial Intelligence and Cyber Security for Social Media. 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 1-3.

[4] Kamoun, F., Iqbal, F., Esseghir, M.A., & Baker, T. (2020). AI and machine learning: A mixed blessing for cybersecurity. 2020 International Symposium on Networks, Computers and Communications (ISNCC), 1-7.

[5] Laato, S., Farooq, A., Tenhunen, H., Pitkamaki, T., Hakkala, A., & Airola, A. (2020). AI in Cybersecurity Education- A Systematic Literature Review of Studies on Cybersecurity MOOCs. 2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT), 6-10.

[6] Zeadally, S., Adi, E., Baig, Z.A., & Khan, I.A. (2020). Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity. IEEE Access, 8, 23817-23837.

[7] Hofstetter, M., Riedl, R., Gees, T., Koumpis, A., & Schaberreiter, T. (2020). Applications of AI in cybersecurity. 2020 Second International Conference on Transdisciplinary AI (TransAI), 138-141.

[8] Sree, V.S., Koganti, C.S., Kalyana, S.K., & Anudeep, P. (2021). Artificial Intelligence Based Predictive Threat Hunting In The Field of Cyber Security. 2021 2nd Global Conference for Advancement in Technology (GCAT), 1-6.

[9] Morovat, K., & Panda, B. (2020). A Survey of Artificial Intelligence in Cybersecurity. 2020 International Conference on Computational Science and Computational Intelligence (CSCI), 109-115.

[10] D. Saxena, I. Gupta, R. Gupta, A. K. Singh and X. Wen, "An AI-Driven VM Threat Prediction Model for Multi-Risks Analysis-Based Cloud Cybersecurity," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 53, no. 11, pp. 6815-6827, Nov. 2023, doi: 10.1109/TSMC.2023.3288081.

[11] Shamiulla, Arab Mohammed. "Role of artificial intelligence in cyber security." International Journal of Innovative Technology and Exploring Engineering 9.1 (2019): 4628-4630.

[12] Z. Zhang, H. A. Hamadi, E. Damiani, C. Y. Yeun and F. Taher, "Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research," in IEEE Access, vol. 10, pp. 93104-93139, 2022, doi: 10.1109/ACCESS.2022.3204051.

[13] Biswal, A. K., Avtaran, D., Sharma, V., Grover, V., Mishra, S., & Alkhayyat, A. (2024). Transformative Metamorphosis in Context to IoT in Education 4.0. EAI Endorsed Transactions on Internet of Things, 10.

[14] Verma, S., Mishra, S., Sharma, V., Nandal, M., Garai, S., & Alkhayyat, A. (2024). Distinctive Assessment of Neural Network Models in Stock Price Estimation. EAI Endorsed Transactions on Scalable Information Systems.

[15] Das, U., Sharma, V., Das, M., Mishra, S., Iwendi, C., & Osamor, J. (2023, December). Vehicular propagation velocity forecasting using open CV. In Proceedings of ICCAKM 2023: 4th International Conference on Computation, Automation and Knowledge Management. IEEE.

[16] Sharma, S., Pandey, A., Sharma, V., Mishra, S., & Alkhayyat, A. (2023, November). Federated Learning and Blockchain: A Cross-Domain Convergence. In 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS) (pp. 1121-1127). IEEE.

[17] Ajmani, P., Sharma, V., Sharma, S., Alkhayyat, A., Seetharaman, T., & Boulouard, Z. (2023, September). Impact of AI in Financial Technology-A Comprehensive Study and Analysis. In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 6, pp. 985-991). IEEE.