



Advanced Cyber Security Threat Intelligence with Hybrid Predictive Model Integration

SWARNA SUSEELAMMA¹, Department of CSE, Prakasam Engineering

College, Kandukur, India

Mr. K. SREEHARI², M. Tech, Assistant Professor, Department of CSE, Prakasam

Engineering College, Kandukur

Abstract: This extended study presents an enhanced cyber-attack detection framework using a hybrid stacked model that integrates Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), and Random Forest (RF) classifiers. The stacked ensemble combines the strengths of these base models to deliver high robustness and accuracy across heterogeneous benchmark datasets including NSL-KDD, CICIDS2017, CICDDOS2019, and X-IIOTID. The system is deployed through a Flask-based web interface, enabling real-time cyber-attack prediction and user interaction with uploaded test datasets. Incorporating Explainable AI (XAI) techniques provides interpretability by highlighting key feature contributions, supporting transparent and informed decision-making. Experimental results demonstrate superior performance, achieving up to 100% accuracy, proving the efficiency of the hybrid stacked framework as a scalable and intelligent solution for proactive cybersecurity defense.

Index terms —Cyber-attack prediction, machine learning, deep learning, hybrid stacked model, NSL-KDD, CICIDS2017, DNN, explainable AI

1. INTRODUCTION

Cybersecurity systems often utilize traditional machine learning and deep learning models like SVM, KNN, Random Forest, Naïve Bayes, and Multilayer Perceptron (MLP) to find bad behavior. These approaches look at past data to sort network traffic and discover problems. They can see threats they already know about, but zero-day attacks, high-dimensional data, and new cyber-attack patterns that come up quickly are hard to deal with. Because many AI models are black boxes, security experts have a hard time understanding how they work and responding to new threats.

To get over these problems, a full and flexible cyber-attack prediction framework is made by combining many machine learning and deep learning models. The system employs methods like Random Forest, SVM, KNN, MLP, Logistic Regression, Naïve Bayes, and DNN on standard datasets including NSL-KDD, CICIDS2017, CICDDOS2019, and X-IIOTID. Preprocessing, standardization, and data balance make the framework fair for learning and generalization. Explainable AI (XAI) solutions make it easier for analysts to understand and trust model findings by making feature importance clearer and more open.



This project builds on the idea of a hybrid stacking model that employs MLP, KNN, and Random Forest classifiers to make cyber-attack datasets more accurate and resistant to different types of attacks. An ensemble method uses predictions from several models to make it easier to find and classify difficult attack situations. The web interface built on Flask lets users interact with and make predictions about provided datasets in real time. Explainable AI makes things easier to understand by showing decision patterns and drawing attention to important details. This one-of-a-kind breakthrough makes it possible to detect cyberattacks in contemporary digital infrastructures in a way that is scalable, understandable, and high-performance.

2. LITERATURE SURVEY

a) Critical Role of Cyber Security in Global Economy

[Critical Role of Cyber Security in Global Economy](#)

The progress made in information technology has made national security more dangerous in many places. There has always been a big power difference between opposing armies in military history. This has always been the case. The unequal distribution of resources may be caused by changes in the economy, new technologies, or simply the number of parties involved. Information operations are becoming more important in modern battlefields. In this arena, all the fighters use the same weapons: computers.

Even though computers and network connections are everywhere in cyberspace, it is still a battlefield for conflicts that are still going on. International law is

determining the rules for this conflict, and the rules are still being made. This fight is about more than just obtaining or preserving sensitive information; it's also about protecting important infrastructures and institutions that are important for the economy and domestic security. The stakes couldn't be greater. The sessions today will put the significance of cyber security in protecting important infrastructure and its effect on economic security in context. This pertains to the presentations planned for today.

b) From information security to cyber security

[From information security to cyber security - ScienceDirect](#)

People commonly mix up cyber security with information security. This paper asserts that cyber security and information security are distinct, notwithstanding their intersection. The paper says that cyber security protects not only information resources but also other things, including people. Human aspects in information security mostly have to do with the responsibilities that people play in keeping information safe. The fact that people can be targets or unknowingly help with cyber attacks makes cyber security even more important. This new aspect has ethical implications for society since it is our duty to safeguard vulnerable groups, such as children.

c) Solar Winds Hack: In-Depth Analysis and Countermeasures

[Solar Winds Hack: In-Depth Analysis and Countermeasures | IEEE Conference Publication | IEEE Xplore](#)



Organizations today rely heavily on technology, which makes them more vulnerable to cyberattacks. Cyberattacks have become more common and pose bigger hazards since technology became widely used. This story talks about SolarWinds, who were recently hit by one of the biggest cyberattacks. This paper also talks about supply chains in general, how hackers could target them, and what happens to the people who are affected. Lastly, the paper gives a list of suggestions for how to stop such assaults from happening again. We got the information by doing a lot of web investigation. We have noticed that hackers that go after supply chains frequently hunt for little weaknesses in the system or even create their own holes to get in and attack all the providers in the chain. So, all businesses need to have a highly robust security system to keep private or sensitive information safe.

d) Analyzing WannaCry Ransomware Considering the Weapons and Exploits

[Analyzing WannaCry Ransomware Considering the Weapons and Exploits | IEEE Conference Publication | IEEE Xplore](#)

Ransomware is becoming more popular, and its makers are taking advantage of our fears. The rapid rise in ransomware attacks shows that ransomware-as-a-service (RaaS) is becoming more common and that hacking tools are being combined. This article covers the investigation of the infamous WannaCry ransomware, which is one of the most disseminated and devastating viruses in 2017. The anatomy of ransomware attacks is explored to comprehend the multi-phased execution of WannaCry, including the deployment, installation, destruction, and command-and-control. The chain of WannaCry's execution

involves numerous hacking weapon components. WannaCry not only embeds the binary in the resource section for multi-phased execution, but also implements a strong encrypting algorithm and a key structure. A reverse engineering examination of each component, along with a network study of WannaCry's exploits, provides insight into the internal architecture of WannaCry. The findings of this research enhance contemporary security systems and inform future defense tactics.

e) Grand Challenge: Applying Artificial Intelligence and Machine Learning to Cybersecurity

[Grand Challenge: Applying Artificial Intelligence and Machine Learning to Cybersecurity | IEEE Journals & Magazine | IEEE Xplore](#)

Cybersecurity is a big issue for all businesses since the number of assaults is going higher. Cybersecurity threats are getting worse and are putting our lives at risk. AI (artificial intelligence) and ML (machine learning) can assist cyber analysts find dangers and provide them advice. To get more people to use AI and machine learning in cybersecurity, businesses, schools, and governments around the world need to work together.

3. METHODOLOGY

i) Proposed Work:

A hybrid stacked ensemble model that combines MLP, KNN, and RF classifiers makes systems that anticipate cyberattacks better. The models are trained independently on benchmark cybersecurity datasets such as NSL-KDD, CICIDS2017, CICDDOS2019, and X-IIOTID, which provide both



attack and traffic data. A meta-classifier may determine the most accurate conclusion for each case by stacking the predictions of various underlying models. This makes the results more accurate and applicable to a wider range of attack conditions.

Data consistency and bias removal during training are achieved via normalization, encoding, and shuffling. We employ Explainable AI (XAI) to assess model predictions and discover the most significant attack detection aspects. When automated cybersecurity decisions are easy to understand, people are more open and confident.

The system is built on a web app that uses Flask, which lets users submit test datasets and see the outcomes of their predictions right away. A web-based version of the framework illustrates how it may be used in real life and how it can grow. The hybrid layered model, explainability, and real-time deployment provide for a robust, flexible, and clear cyber-attack detection framework with almost flawless accuracy and proactive cybersecurity defense systems.

ii) System Architecture:

The suggested cyber-attack prediction methodology uses machine learning, deep learning, and explainable AI in a multi-stage pipeline. Benchmark cyber-attack datasets NSL-KDD, CICIDS2017, CICDDOS2019, and X-IIOTID provide the basis. Data preparation includes label encoding, normalization, and data shuffling to maintain consistency and eliminate bias before training. ML/DL techniques including Random Forest, KNN, SVM, MLP, Logistic Regression, Naïve Bayes, and DNN were used to identify and forecast probable attack types on processed data. A hybrid stacking model including

MLP, KNN, and Random Forest refines the predictions for greater accuracy and resilience. Cybersecurity specialists may input test data and see results on a Flask-based web interface. Integrating SHAP (SHapley Additive exPlanations) for explainability provides feature significance insights and supports transparent, data-driven security decisions.

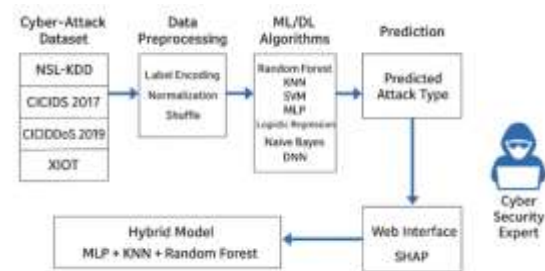


Fig 1: proposed architecture

iii) Modules:

a) Data Collection Module

- Collects benchmark datasets such as NSL-KDD, CICIDS2017, CICDDOS2019, and X-IIOTID containing various attack and normal traffic records.
- Ensures diverse data sources to improve model generalization and robustness.

b) Data Preprocessing Module

- Performs label encoding, normalization, and data shuffling to clean and standardize the input data.
- Removes inconsistencies and prepares datasets for efficient training and testing.

c) Model Training Module

- Trains multiple ML/DL algorithms including Random Forest, KNN, SVM, MLP, Logistic Regression, Naïve Bayes, and DNN on preprocessed data.



- Evaluates individual model performance to identify strong predictors for cyber-attack detection.

d) Hybrid Stacked Model Module

- Integrates MLP, KNN, and Random Forest classifiers using a stacking technique.
- Combines predictions from all base models to enhance accuracy and reduce false positives.

e) Explainable AI (XAI) Module

- Utilizes SHAP (SHapley Additive exPlanations) to interpret model predictions.
- Provides feature importance analysis for transparency and better understanding of decision-making.

f) Web Interface Module

- Implements a Flask-based interface allowing users or cybersecurity experts to upload test datasets and view predictions in real-time.
- Displays detected attack types and their explanations, supporting proactive security measures.

iv) Algorithms:

1. Random Forest (RF):

Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and combines their predictions to enhance accuracy and stability. It reduces overfitting by averaging the outcomes of individual trees, making it suitable for handling high-dimensional network data. In cyber-attack prediction, Random Forest effectively classifies complex patterns in traffic and identifies various attack types with high precision.

2. K-Nearest Neighbor (KNN):

KNN is a simple yet powerful classification algorithm that predicts the label of a sample based on

the majority class among its nearest neighbors. It uses distance measures such as Euclidean or Manhattan distance to determine similarity between data points. In cybersecurity, KNN helps detect anomalies and identify attacks by comparing new samples to known instances of network behavior.

3. Support Vector Machine (SVM):

SVM is a supervised learning algorithm that separates data points into classes using an optimal hyperplane. It works efficiently in both linear and non-linear classification through kernel functions. In cyber-attack detection, SVM excels at distinguishing between normal and malicious traffic, especially in cases where the data has clear separation boundaries.

4. Multilayer Perceptron (MLP):

MLP is a type of feedforward neural network consisting of multiple layers of interconnected nodes. Each layer transforms the input data using activation functions, allowing the model to learn complex, non-linear relationships. For cyber-attack detection, MLP is effective in identifying sophisticated attack patterns by capturing deep feature representations from network data.

5. Logistic Regression (LR):

Logistic Regression is a statistical model used for binary and multi-class classification. It predicts the probability of an event (e.g., attack or normal) using a sigmoid activation function. In this framework, LR serves as a baseline model for comparing advanced algorithms, offering interpretable and fast predictions for basic intrusion detection.

6. Naïve Bayes (NB):

Naïve Bayes is a probabilistic classifier based on Bayes' theorem that assumes independence among features. Despite its simplicity, it performs well on



large-scale datasets and is highly efficient for real-time detection. In cyber-attack prediction, it quickly classifies traffic by computing the likelihood of different attack types based on historical data.

7. Deep Neural Network (DNN):

DNN is a deep learning model composed of multiple hidden layers that can automatically extract high-level features from raw input data. It is capable of learning intricate patterns and relationships, making it suitable for detecting complex and evolving cyber threats. In this system, DNN contributes to achieving high accuracy and adaptability across heterogeneous datasets.

8. Hybrid Stacked Model (MLP + KNN + RF):

The Hybrid Stacked Model combines the strengths of MLP, KNN, and Random Forest through a stacking mechanism. Each base model generates its predictions, which are then integrated by a meta-classifier to produce the final output. This ensemble approach enhances robustness, minimizes false detections, and delivers near-perfect accuracy, making it the most reliable solution for real-time cyber-attack prediction.

4. EXPERIMENTAL RESULTS

We used four benchmark datasets—NSL-KDD, CICIDS2017, CICDDOS2019, and X-IIOTID—to test the effectiveness of several machine learning and deep learning algorithms for predicting cyber attacks. To make sure that the training was fair and balanced, each dataset went through preprocessing stages including label encoding, normalization, and shuffling. We trained and evaluated the individual models, such as SVM, KNN, Random Forest, MLP, Logistic Regression, Naïve Bayes, and DNN, to see how accurate, precise, and reliable they were. The

hybrid stacked model (MLP + KNN + Random Forest) worked the best out of them. It got 99.90% accuracy on the NSL-KDD dataset and 100% accuracy on the CICIDS2017, CICDDOS2019, and X-IIOTID datasets. Adding Explainable AI (SHAP) made the model much easier to understand by showing which factors had the most effect on predictions. These results show that the suggested hybrid model is better at delivering a strong, accurate, and clear cyber-attack detection system that can function in different and changing network contexts.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$\text{Accuracy} = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

$$\text{Precision} = \frac{\text{True positives}}{(\text{True positives} + \text{False positives})} = \frac{TP}{(TP + FP)}$$

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.



$$Recall = \frac{TP}{(FN + TP)}$$

mAP:One ranking quality statistic is Mean Average Precision (MAP). It takes into account the quantity of pertinent suggestions and where they are on the list. The arithmetic mean of the Average Precision (AP) at K for each user or query is used to compute MAP at K.

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$

$AP_k = \text{the AP of class } k$
 $n = \text{the number of classes}$

F1-Score:A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot Precision)}{(Recall + Precision)}$$

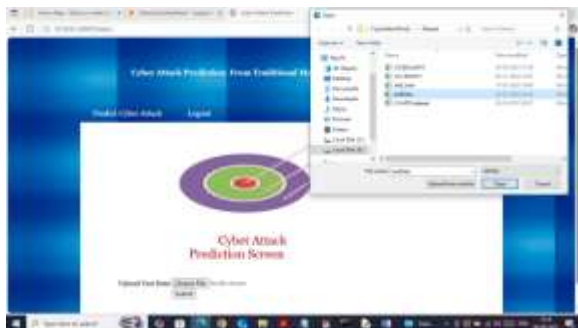


Fig 2 uploading test data file



Fig 3 results

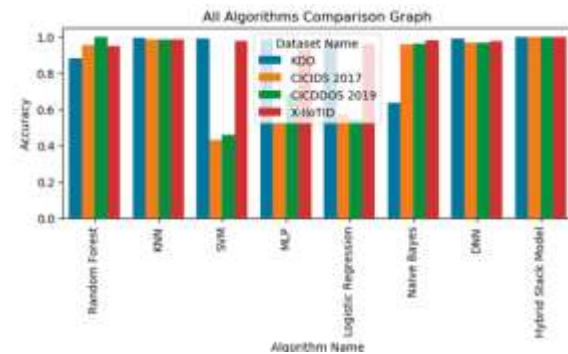


Fig 4 Accuracy graph

	Dataset Name	Algorithm Name	Accuracy
0	KDD	Random Forest	0.880250
1	CICIDS 2017	Random Forest	0.954596
2	CICDDOS 2019	Random Forest	1.000000
3	X-IloTID	Random Forest	0.950000
4	KDD	KNN	0.993250
5	CICIDS 2017	KNN	0.983389
6	CICDDOS 2019	KNN	0.984981
7	X-IloTID	KNN	0.984000
8	KDD	SVM	0.988250
9	CICIDS 2017	SVM	0.429679
10	CICDDOS 2019	SVM	0.458073
11	X-IloTID	SVM	0.973500
12	KDD	MLP	0.986250
13	CICIDS 2017	MLP	0.521595

Table 1 all algorithms accuracy on each dataset

5. CONCLUSION

It is still very hard to find cyberattacks since they are getting more complicated and varied, and they are targeting digital infrastructures. The assessment of several machine learning and deep learning



algorithms on benchmark datasets—NSL-KDD, CICIDS2017, CICDDOS2019, and X-IIOTID—revealed considerable performance fluctuation, underscoring the necessity for resilient prediction models. The hybrid stacking model, which incorporates MLP, KNN, and Random Forest classifiers, was the most accurate of the methods studied. It got 99.90% on NSL-KDD and 100% on CICIDS2017, CICDDOS2019, and X-IIOTID datasets. This shows that using ensemble techniques to find different patterns in datasets that are not uniform works. Also, using Explainable AI approaches helped us understand how different features contributed, which made things clearer and easier to understand. This is important for making smart decisions in cybersecurity. The findings demonstrate that hybrid models, which amalgamate many learning paradigms, surpass individual algorithms, providing nearly flawless detection capabilities across diverse cyber-attack situations. These results offer a dependable standard for choosing high-performance detection algorithms and stress the need to use machine learning, deep learning, and generative AI together to make predictions more accurate and systems more resistant to new cyber threats. The research highlights a trajectory towards more flexible and comprehensible cyber-attack prediction systems.

6. FUTURE SCOPE

In the future, this framework for predicting cyberattacks might be expanded to include real-time streaming data from live networks, which would allow for immediate detection and reaction to threats that are happening right now. Integrating Generative AI (GenAI) can make adaptability even better by recreating attack patterns that have never been

observed before or that happen on the same day as the attack for model retraining. Federated learning will let several businesses train their models in a safe, dispersed way without putting data privacy at risk. Also, using automated incident response systems and deploying on the cloud will make the system more scalable and efficient. The device may also become an intelligent cyber defense helper that gives security experts clear explanations and suggestions on how to stop threats before they happen.

REFERENCES

- [1] G. S. Emile and M. Kala, "Critical role of cyber security in global economy," *Open J. Saf. Sci. Technol.*, vol. 13, no. 4, pp. 231–248, 2023, doi: 10.4236/ojsst.2023.134012.
- [2] R. von Solms and J. van Niekerk, "From information security to cyber security," *Comput. Secur.*, vol. 38, pp. 97–102, Oct. 2013.
- [3] J.W. Goodell and S. Corbet, "Commodity market exposure to energy firm distress: Evidence from the colonial pipeline ransomware attack," *Finance Res. Lett.*, vol. 51, Jan. 2023, Art. no. 103329.
- [4] R. Alkhadra, J. Abuzaid, M. AlShammari, and N. Mohammad, "Solar winds hack: In-depth analysis and countermeasures," in *Proc. 12th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Jul. 2021, pp. 1–7.
- [5] Cobalt. (Jun. 20, 2024). 11 Biggest Cybersecurity Attacks in History. [Online]. Available: <https://www.cobalt.io/blog/biggest-cybersecurity-attacks-in-history>
- [6] D.-Y. Kao, S.-C. Hsiao, and R. Tso, "Analyzing WannaCry ransomware considering the weapons and



- exploits,” in Proc. 21st Int. Conf. Adv. Commun. Technol. (ICACT), Feb. 2019, pp. 1098–1107.
- [7] K. Bresniker, A. Gavrilovska, J. Holt, D. Milojicic, and T. Tran, “Grand challenge: Applying artificial intelligence and machine learning to cybersecurity,” *Computer*, vol. 52, no. 12, pp. 45–52, Dec. 2019, doi: 10.1109/MC.2019.2942584.
- [8] M. Husák, J. Komárková, E. Bou-Harb, and P. Celeda, “Survey of attack projection, prediction, and forecasting in cyber security,” *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 640–660, 1st Quart., 2019.
- [9] N. Mohamed, “Current trends in AI and ML for cybersecurity: A state of-the-art survey,” *Cogent Eng.*, vol. 10, no. 2, pp. 1–11, Dec. 2023, doi: 10.1080/23311916.2023.2272358.
- [10] L. Chan, I. Morgan, H. Simon, F. Alshabanat, D. Ober, J. Gentry, D. Min, and R. Cao, “Survey of AI in cybersecurity for information technology management,” in Proc. IEEE Technol. Eng. Manage. Conf. (TEMSCON), Jun. 2019, pp. 1–8.
- [11] G. Disterer, “ISO/IEC 27000, 27001 and 27002 for information security management,” *J. Inf. Secur.*, vol. 4, no. 2, pp. 92–100, 2013.
- [12] J. H. Li, “Cyber security meets artificial intelligence: A survey,” *Frontiers Inf. Technol. Electron. Eng.*, vol. 19, no. 12, pp. 1462–1474, 2018, doi: 10.1631/FITEE.1800573.
- [13] I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, “Cybersecurity data science: An overview from machine learning perspective,” *J. Big Data.*, vol. 7, pp. 1–29, May 2020, doi: 10.1186/s40537-020-00318-5.
- [14] D. Aggarwal, D. Sharma, and A. B. Saxena, “Role of AI in cybersecurity through anomaly detection and predictive analysis,” *J. Informat. Educ. Res.*, vol. 3, no. 2, pp. 1–12, 2023.
- [15] G. Srivastava, R. H. Jhaveri, S. Bhattacharya, S. Pandya, Rajeswari, P. K. R. Maddikunta, G. Yenduri, J. G. Hall, M. Alazab, and T. R. Gadekallu, “XAI for cybersecurity: State of the art, challenges, open issues and future directions,” 2022, arXiv:2206.03585.
- [16] I. H. Sarker, M. H. Furhad, and R. Nowrozy, “AI-driven cybersecurity: An overview, security intelligence modeling and research directions,” *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–18, May 2021.
- [17] D. Ucci, L. Aniello, and R. Baldoni, “Survey of machine learning techniques for malware analysis,” *Comput. Secur.*, vol. 81, pp. 123–147, Mar. 2019.
- [18] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, “A survey of deep learning-based network anomaly detection,” *Cluster Comput.*, vol. 22, no. S1, pp. 949–961, Jan. 2019.
- [19] R. A. Nafea and M. A. Almaiah, “Cyber security threats in cloud: Literature review,” in Proc. Int. Conf. Inf. Technol. (ICIT), Jul. 2021, pp. 779–786.
- [20] A. Kuppa and N. A. Le-Khac, “Blackbox attack on explainable artificial intelligence (XAI) methods in cyber security,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2020, pp. 1–8.
- [21] K. D. Ahmed and S. Askar, “Deep learning models for cyber security in IoT networks: A review,” *Int. J. Sci. Bus.*, vol. 5, no. 3, pp. 61–70, Jan. 2021.
- [22] J. Gerlings, A. Shollo, and I. Constantiou, “Reviewing the need for explainable artificial intelligence (xAI),” 2020, arXiv:2012.01007.



- [23] G. Jaswal, V. Kanhangad, and R. Ramachandra, AI and Deep Learning in Biometric Security: Trends Potential and Challenges. Boca Raton, FL, USA: CRCPress, 2021.
- [24] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” 2018, arXiv:1811.10154.
- [25] 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,” IEEE Access, vol. 10, pp. 93104–93139, 2022, doi: 10.1109/ACCESS.2022.3204051.
- [26] A. Bandi, P. V. S. R. Adapa, and Y. E. V. P. K. Kuchi, “The power of generative AI: A review of requirements, models, input–output formats, evaluation metrics, and challenges,” Future Internet, vol. 15, no. 8, Jul. 2023. [Online]. Available: <https://www.mdpi.com/1999-5903/15/8/260>
- [27] J. Babcock and R. Bali, Generative AI With Python and Tensorflow 2: Create Images, Text, and Music With VAEs, GANs, LSTMs, Transformer Models. Birmingham, U.K.: Packt, 2021.
- [28] ChatGPT. Accessed: Jun. 22, 2023.[Online]. Available: <https://chat.openai.com/>
- [29] Dall·eNowAvailableWithoutWaitlist. Accessed:Jun. 22,2023.[Online]. Available: <https://openai.com/blog/dall-e-now-available-without-waitlist>
- [30] B. Hitaj, P. Gasti, G. Ateniese, and F. Perez-Cruz, “PassGAN: A deep learning approach for password guessing,” 2017, arXiv:1709.00440.