Machine Learning and Deep Learning Approaches for CyberSecurity: A Review

1. N. Naveen Kumar, Professor of CS, School of Information Technology, JNTU Hyderabad, India.

2. Paladugu Shailaja, M. Tech(CNIS), School of Information Technology, JNTU Hyderabad, India.

ABSTRACT: The fast expansion and growth of the internet over the previous few decades has increased anxiety about cyber-attacks, which are constantly developing and changing. As a consequence, an effective intrusion detection system was necessary to secure data, and one of the most successful solutions to handle this challenge was the development of artificial intelligence's sub-fields, machine learning and deep learning. This study examined intrusion detection systems and explored the various learning techniques used by machine learning and deep learning to safeguard data from hostile activities. It highlights contemporary machine learning and deep learning work on developing an operational intrusion detection system using multiple network implementations, applications, algorithms, learning methodologies, and datasets.

Keywords *– Cybersecurity, machine learning, deep learning, intrusion detection system.*

1. INTRODUCTION

The internet is changing people's employment, learning, and lives, and it is now possible to integrate social life with the online, which raises security concerns in a variety of ways. What matters today is knowing how to recognise network dangers and intrusions, especially those that have already occurred. The process of creating cyber protective measures and policies to secure data, programmes, servers, and network infrastructures against unwanted access or alteration is referred to as cybersecurity. The bulk of our computer systems and network infrastructure are linked through the internet. As a consequence, cybersecurity has developed as the foundation for almost all sorts of enterprises, governments, and even individuals to safeguard data, expand their businesses, and retain privacy. People transmit and receive data via network infrastructure that may be hacked and controlled by outsiders, such as a router. Because of the rising usage of the internet, the volume and complexity of data has expanded, leading in the birth of big data. The continual growth of the internet and vast amounts of data demanded the development of a dependable intrusion detection system. Network security is a subcategory of cybersecurity that protects network-connected systems against harmful activities. The objective is to offer networked computers that provide data security, integrity, and accessibility. Current cybersecurity research is focused on developing an effective intrusion detection system capable of detecting both known and unknown assaults and threats with high accuracy and a low false alarm rate [1].

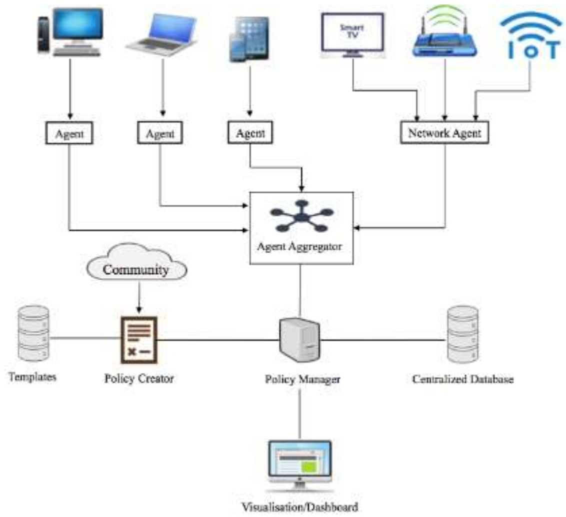


Fig.1: Example figure

Alan Turing said that general-purpose computers might learn and assess originality, paving the stage for the debate over whether computers should look at data to establish rules rather than rely on people to do it. Machine learning algorithms are data-driven algorithms that can learn and adapt. Machine learning algorithms are programmed to provide output depending on what they have learnt from data and examples. Such algorithms, for example, will enable a computer to pick and execute a specific job on innovative traffic detection without explicit knowledge [2]. Machine learning may be used to do efficient automatic assessments of assaults and security events such as spam mail, user identification, social media analytics, and attack detection [1].

2. LITERATURE REVIEW

Network intrusion detection system using deep learning technique:

The extensive usage of interconnection and interoperability of computer systems has become an unavoidable need for improving our everyday lives. Simultaneously, it opens the door to exploitable flaws that are well beyond the scope of human control. Because of the weaknesses, cyber-security procedures are required to assume communication exchange. Secure communication necessitates security measures to battle threats and advances in security measures to combat growing security risks. The use of deep learning architectures in the development of an adaptive and resilient network intrusion detection system (IDS) to identify and categorise network threats is proposed in this research. The focus is on how deep learning or deep neural networks (DNNs) might provide flexible IDS with learning power to identify known and novel or zero-day network behavioural traits, ejecting the system invader and lowering the risk of compromise. We utilised the UNSW-NB15 dataset, which reflects genuine current network communication behaviour with synthetically created attack activities, to illustrate the model's performance.

Deep learning approaches for network intrusion detection

Recently, computer networks have encountered a significant challenge in the form of increasing hostile assaults. One of the most important research topics in network and computer security is intrusion detection. Deep Learning (DL) strategies for enhancing the Intrusion Detection System are investigated and presented in this work (IDS). Furthermore, it offers a full comparison of assessing performance, deep learning algorithms for detecting assaults, feature learning, and datasets used to discover the benefits of utilising in strengthening network intrusion detection.

Machine learning and deep learning methods for cybersecurity

Cyber-attacks are evolving swiftly as the Internet evolves, and the cyber security picture is bleak. This survey report summarises significant literature reviews on machine learning (ML) and deep learning (DL) approaches for network analysis and intrusion detection, as well as providing a short instructional overview of each ML / DL method. Each method's papers were indexed, reviewed, and summarised based on their temporal or thermal correlations. Because data are so crucial in ML/DL approaches, we present some of the most widely used network datasets in ML/DL, highlight the obstacles of employing ML/DL for cybersecurity, and provide recommendations for future study.

Building effective network security frameworks using deep transfer learning techniques

Global network traffic is increasing at an alarming rate. According to the 2020 Cisco Annual Report, over two-thirds of the worldwide population will be connected to the internet by 2023. By the same year, the number of devices linked to IP networks will have tripled, as will the entire global population. The immensity of anticipated network infrastructure provides potential for new technologies and companies to emerge, but it also expands the surface of security risks. The global number of cyberattacks is increasing, and they are getting more diversified and sophisticated. Using different signature libraries, traditional network intrusion detection systems monitor a system to identify malicious activity and policy breaches in its information stream. Nonetheless, because of the high volume of network traffic in current network infrastructures, fraudsters may access systems unnoticed and effectively steal or destroy information assets. The speed and efficiency of traditional network intrusion detection designs also fall short in a real-time processing environment. Given the aforementioned constraints, the goal of this thesis is to provide unique methodology for designing and architecting network intrusion detection systems utilising applied deep learning techniques. Neural networks can deduce patterns and signatures from a raw dataset and then utilise the learnt signatures to anticipate the nature and categorise incoming data at breakneck speed. The resilience of neural network architecture may be enhanced to provide a real-time and efficient network security framework. We will look at several machine learning and deep learning principles and approaches in this article. We will create a hybrid network intrusion detection system utilising the CNN-LSTM architecture by combining the capabilities of the current models in terms of latent feature extraction, memory retention, and classification skills. Furthermore, we will compare our findings to contemporary studies in this area of study.

Anomaly-based network intrusion detection using machine learning

Network infiltration is now the most serious risk in network communications. The growing threat of network assaults is a crippling challenge for network services. Several studies have previously been undertaken in order to identify an effective and efficient method to prevent network intrusion and protect network security and privacy. Machine learning is a powerful analytical method for detecting suspicious occurrences in network data flows. In this study, we created a network intrusion detection classifier model based on SVM and Random Forest techniques. The NSL-KDD dataset, a much upgraded version of the original KDDCUP'99 dataset, was utilised to assess our algorithm's performance. Our detection algorithm's primary goal was to determine if incoming network traffic was regular or malicious, using 41 attributes that described each pattern of network traffic. SVM and Random algorithms were used to obtain detection accuracy of higher than 95%. The outcomes of two algorithms were compared, and it was discovered that the Random Forest method outperformed the Support Vector Machine technique.

3. METHODOLOGY

The internet is changing people's employment, learning, and lives, and it is now possible to integrate social life with the online, which raises security concerns in a variety of ways. What matters today is knowing how to recognise network dangers and intrusions, especially those that have already occurred. The process of creating cyber protective measures and policies to secure data, programmes, servers, and network infrastructures against unwanted access or alteration is referred to as cybersecurity. The bulk of our computer systems and network infrastructure are linked through the internet. As a consequence, cybersecurity has developed as the foundation for almost all sorts of enterprises, governments, and even individuals to safeguard data, expand their businesses, and retain privacy. People transmit and receive data via network infrastructure that may be hacked and controlled by outsiders, such as a router. Because of the rising usage of the internet, the volume and complexity of data has expanded, leading in the birth of big data. The continual growth of the internet and vast amounts of data demanded the development of a dependable intrusion detection system. Network security is a subcategory of cybersecurity that protects network-connected systems against harmful activities.

Disadvantages:

1. Because of the rising usage of the internet, the volume and complexity of data has expanded, leading in the creation of big data.

2. The steady growth of the internet and vast amounts of data demanded the development of an effective intrusion detection system.

This study examined intrusion detection systems and explored the various learning techniques used by machine learning and deep learning to safeguard data from hostile activities. It highlights contemporary machine learning and deep learning work on developing an operational intrusion detection system using multiple network implementations, applications, algorithms, learning methodologies, and datasets.

Advantages:

1. Deep learning is better than machine learning in many research and trials because it can handle more complex issues with higher accuracy and lower false alarm rates.
2. They used different datasets, architectures, learning approaches, and learning algorithms each time to protect data from assaults and threats.

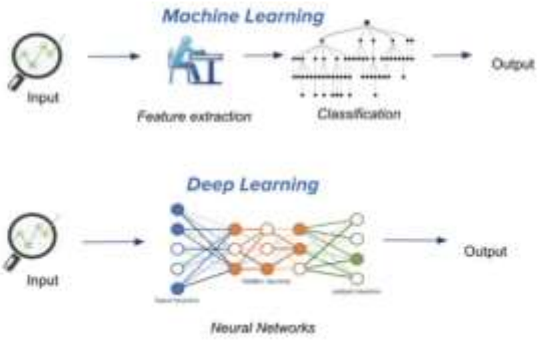


Fig.2: System architecture

MODULES:

To carry out the aforementioned project, we created the modules listed below.

- Data exploration: we will put data into the system using this module.
- Processing: we will read data for processing using this module.
- Using this module, data will be separated into train and test groups.
- Model generation: the model will be built. DT - Decision Tree - KNN - Naive Bayes - LDA - LR - Logistic Regression - K-Means - CNN - CNN + LSTM - AutoEncoder.
- User signup and login: Using this module will result in registration and login.
- User input: Using this module will result in predicted input.
- Prediction: final predicted shown

4. IMPLEMENTATION

ALGORITHMS:

SVM: Support Vector Machine (SVM) is a supervised machine learning technique that may be used for classification and regression. Though we call them regression issues, they are best suited for categorization. The SVM algorithm's goal is to identify a hyperplane in an N-dimensional space that clearly classifies the input points.

Random forest: Random Forest is a well-known machine learning algorithm that belongs to the supervised learning approach. It may be used to both classification and regression issues in machine learning. No assumptions are made about the data or its distribution by random forest. As a result, data transformations are usually modest. Because the random forest approach employs random selections of features, it can perform admirably with high-dimensional datasets (a dataset with a large number of features)

Voting Classifier: Voting Classifier is a machine-learning method that Kagglers often employ to improve the performance of their model and move up the rank ladder. Voting Classifier may also be used to increase performance on real-world datasets, although it has significant limitations.

DT: A decision tree is a non-parametric supervised learning technique that may be used for classification and regression applications. It has a tree structure that is hierarchical and consists of a root node, branches, internal nodes, and leaf nodes.

KNN: The k-nearest neighbours method, often known as KNN or k-NN, is a non-parametric, supervised learning classifier that employs proximity to create classifications or predictions about an individual data point's grouping.

Naïve Bayes: Naive Bayes classifiers are a group of classification algorithms based on Bayes' Theorem. It is a family of algorithms that all share a similar premise, namely that every pair of characteristics being categorised is independent of each other.

LDA: Linear discriminant analysis (LDA) is employed in this case to limit the number of features to a more manageable quantity prior to the classification phase. Each of the newly created dimensions is a linear combination of pixel values that constitute a template.

LR - Logistic Regression (LR): Logistic regression is a statistical approach for developing machine learning models using dichotomous dependent variables, i.e. binary. Logistic regression is a statistical technique used to explain data and the connection between one dependent variable and one or more independent variables.

K-Means:

The K-means clustering technique computes centroids and then repeats the process until the best centroid is discovered. It is assumed that the number of clusters is known. The flat clustering method is another name for it. The letter 'K' in K-means denotes the number of clusters discovered from data by the approach.

CNN: A CNN is a kind of network architecture for deep learning algorithms that is primarily utilised for image recognition and pixel data processing jobs. There are different forms of neural networks in deep learning, but CNNs are the network design of choice for identifying and recognising things.

CNN + LSTM: The CNN-LSTM technique creates a shallow CNN to extract the molten pool image's major characteristics. The CNN's extracted feature tensor is then translated into the feature matrix. Finally, the feature matrix's rows are input into the LSTM network for feature fusion.

AutoEncoder:

An autoencoder is an unsupervised learning strategy for neural networks that trains the network to disregard signal "noise" in order to develop efficient data representations (encoding). Autoencoders may be used for picture denoising, compression, and, in certain situations, image data creation.

5. EXPERIMENTAL RESULTS

CyberSecurity
Machine Learning and Deep Learning Approaches for CyberSecurity: A Review
Get started

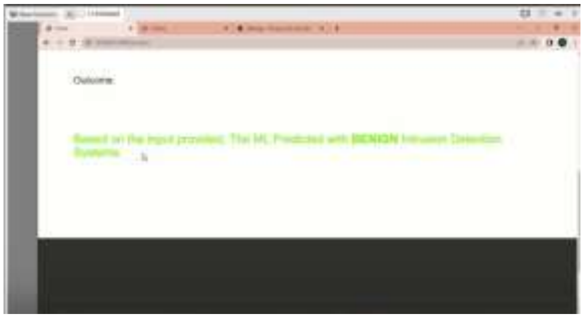
Fig.3: Home screen

Fig.4: User registration

Fig.5: user login

Fig.6: Main screen

Fig.7: User input


Outcome:

Based on the input provided, The ML Predicted with BENIGN Intrusion Detection Systems

Fig.8: Prediction result

6. CONCLUSION

Intrusion detection systems were a critical topic in the cybersecurity field. Many academics are working on a solution to protect data against harmful behaviour. Other uses of learning algorithms, such as creating a new dataset or merging algorithms, are also being researched. As a consequence, in this work, we describe the notion of an intrusion detection system, different sorts of assaults, and how to decide whether or not we have an effective system. It was obvious that datasets have an influence on research in this industry, since some judge it out of current or includes redundant information. As a consequence, the study compares the most frequently used datasets for detecting threats over the previous decade. The last phase in this endeavour was to investigate how other individuals saved their data. According to recent study, there are several data protection implementations. Initially, they used machine learning for a variety of objectives, and several research were undertaken to establish which algorithm would give more accuracy or which datasets would yield a lower false alarm rate. After significant research and testing, they arrived at deep learning. Many research and tests have shown that deep learning outperforms machine learning in handling more complex issues with better accuracy and lower false alarm rates. Previous work has been applied in a number of ways. They used different datasets, architectures, learning approaches, and learning algorithms each time to protect data from assaults and threats.

REFERENCES

[1] D. I. Edeh, ‘‘Network intrusion detection system using deep learning technique,’’ M.S. thesis, Dept. Comput., Univ. Turku, Turku, Finland, 2021.

[2] G. C. Fernandez, ‘‘Deep learning approaches for network intrusion detection,’’ M.S. thesis, Dept. Comput. Sci., Univ. Texas at San Antonio, San Antonio, TX, USA, 2019.

[3] H. Benmeziane, ‘‘Comparison of deep learning frameworks and compilers,’’ M.S. thesis Comput. Sci., Inst. Nat. Formation Informatique, École nationale Supérieure d’Informatique, Oued Smar, Algeria, 2020.

[4] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, and M. Gao, ‘‘Machine learning and deep learning methods for cybersecurity,’’ IEEE Access, vol. 6, pp. 35365–35381, 2018, doi: 10.1109/ACCESS.2018.2836950.

[5] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[6] H. Dhillon, ‘‘Building effective network security frameworks using deep transfer learning techniques,’’ M.S. thesis, Dept. Comput. Sci., Western Univ., London, ON, Canada, 2021.

[7] M. Labonne, ‘‘Anomaly-based network intrusion detection using machine learning,’’ Ph.D. dissertation, Inst. Polytechnique de Paris, Palaiseau, France, 2020.

[8] A. Kim, M. Park, and D. H. Lee, ‘‘AI-IDS: Application of deep learning to real-time web intrusion detection,’’ IEEE Access, vol. 8, pp. 70245–70261, 2020.

[9] P. Wu, ‘‘Deep learning for network intrusion detection: Attack recognition with computational intelligence,’’ M.S. thesis, School Comput. Sci. Eng., Univ. New South Wales, Sydney NSW, Australia, 2020.

[10] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, ‘‘A survey of network-based intrusion detection data sets,’’ Comput. Secur., vol. 86, pp. 147–167, Sep. 2019.