International Conference on Science, Engineering & Management Trends

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

An Analysis of Requirements of Deep learning Based Classification of Attacks over Big Data Security

Assistant Professor Dr. Banita, Ms. Jyoti Ahlawat
Department of Computer Science & Engineering, BMU, Rohtak

Abstract- The rapid growth of big data across diverse digital ecosystems has made cyber security a critical concern, especially as traditional intrusion detection systems struggle to scale and adapt. This study explores the necessity and advantages of deploying deep learning-based techniques for the classification of cyber-attacks in big data environments. We will analyze the limitations of conventional machine learning models in handling high-volume, high-velocity, and high-variety data streams, emphasizing the unique challenges posed by modern attack vectors. The paper evaluates various deep learning architectures—including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models—on their ability to detect complex and evolving threats. Additionally, we will address the infrastructural and computational requirements, such as distributed processing frameworks like Apache Spark and the role of GPU acceleration, to support deep learning at scale.

Keywords- Encryption, Attacks, Big data security, Deep learning.

I. INTRODUCTION

Deep learning for big data security requires access to diverse and labeled datasets, which can be scarce and challenging to obtain for specific attack types. Additionally, the resource-intensive nature of deep learning models may pose scalability and operational challenges for organizations with extensive big data infrastructure. Moreover, false positives and false negatives can be a concern, demanding continuous fine-tuning and adjustment to minimize misclassifications. Despite these challenges, the potential benefits of using deep learning in attack classification within big data environments are substantial, as they enable real-time threat detection and proactive security measures, ultimately reinforcing the overall security posture of these critical data systems.DL can be a valuable tool in enhancing the security of big data systems due to its ability to process and analyze vast amounts of data, adapt to evolving threats, and detect complex patterns. Here are some key ways to use deep learning for securing big data:Securing big data using deep learning is an ongoing process that requires continuous monitoring, model training, and adaptation to evolving threats.

To protect big data systems against these types of attacks, several security measures can be implemented:

- **Encryption:** Encrypting data in transit and at rest can prevent MitM attacks by ensuring data confidentiality and integrity.
- **Access Control:** Implement strict access controls and authentication mechanisms to protect against brute force attacks.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

- **IDS:** Utilize IDPS to detect and mitigate DoS attacks by identifying abnormal traffic patterns and responding in real-time.
- **Network Segmentation:** Isolate critical components of the big data system from the public network to reduce the attack surface and limit the impact of attacks.
- **Monitoring and Logging:** Keep a close eye on the system for any unusual or suspicious activity and record all important events for analysis after an occurrence.
- **Patch and Update Management:** Apply security fixes to software and systems on a regular basis to fix known vulnerabilities that hackers may use.
- **Incident Response Plan:** Create a solid incident response strategy that specifies what to do in the case of an attack, from preventing it to investigating it and recovering from it.
- Protecting big data systems against MitM, DoS, and brute force attacks is essential for ensuring data integrity, availability, and confidentiality, which are fundamental for maintaining the reliability and security of these complex and valuable environments.
- Requirement of deep learning based classification of Attacks over big data
- The need for classifying MitM, DoS, and brute force attacks on big data using a deep learning approach is driven by several important considerations:
- **Scalable and Resource-Efficient:** Deep learning models can be deployed at scale without significant resource overhead. This is important for big data environments that handle vast amounts of data and require efficient attack detection methods.
- **Scale and Complication of Big Data:** Big data environments process and store massive volumes of data, making them attractive targets for attackers. Traditional methods of attack detection and prevention may not scale to handle the complexity and size of big data systems.
- **Real-Time Detection:** Big data systems often require real-time processing, and attacks can have immediate and severe consequences. Deep learning models can analyze data streams in real-time, allowing for faster detection and response to attacks.
- **Advanced Attack Techniques:** Attackers are continually evolving their tactics. Deep learning models can adapt and learn from data, making them more capable of identifying both known and emerging attack patterns.
- **Zero-Day Attacks:** Traditional rule-based systems are less effective at detecting new, previously unseen attack patterns. Deep learning models can identify anomalies and unknown attack vectors without relying on predefined rules.
- **Anomaly Detection:** Big data attacks often involve subtle anomalies in data patterns. Deep learning models are well-suited for anomaly detection because they can learn the expected data distribution and identify deviations from it.
- **Mitigation of False Positives:** Deep learning models can reduce false positives by learning to distinguish between normal fluctuations in data and actual attacks, improving the efficiency of security operations.
- **Automation and Efficiency:** Automating the process of attack detection and classification using deep learning allows security teams to focus on responding to confirmed attacks, rather than sifting through large volumes of data and alerts.
- **Multi-Vector Attacks:** Attackers often use a combination of techniques in a coordinated manner. Deep learning models can handle complex, multi-vector attacks by learning the relationships between various attack indicators.
- **Continuous Learning:** Deep learning models can adapt and improve over time by continuously learning from new data. This adaptability is crucial for staying ahead of evolving attack strategies.
- **Reduced Manual Rule Creation:** Deep learning reduces the reliance on manually creating rules for attack detection. This can save time and resources and enable security teams to respond more effectively to new and complex threats.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

The need for classifying MitM, DoS, and brute force attacks on big data using a deep learning approach is driven by the complexity, scale, and evolving nature of big data systems and the attacks targeting them. Deep learning offers an adaptive and effective solution for identifying and responding to various types of attacks, enhancing the security and reliability of big data environments. Classifying MitM, DoS, and Brute Force attacks on big data using a deep learning approach presents several notable challenges and issues. First and foremost, acquiring labeled datasets for deep learning models can be a formidable obstacle, as obtaining diverse and representative data for specific attack types is often a complex and resource-intensive task. The resource demands of deep learning models pose another challenge, particularly in big data environments, where scalability, processing power, and memory requirements can strain existing infrastructure. Additionally, the interpretability of deep learning models can be a concern, making it difficult to explain the rationale behind the model's classifications and leading to a potential lack of trust in automated security systems. Furthermore, false positives and false negatives in attack classification are common challenges, necessitating continuous fine-tuning and adjustments to reduce misclassifications and enhance accuracy. The evolving nature of cyber threats means that deep learning models must be regularly updated to adapt to new attack strategies and vulnerabilities, requiring a commitment to ongoing maintenance and improvement. Despite these hurdles, the potential advantages of employing deep learning in attack classification within big data ecosystems, such as real-time threat detection and proactive security measures, make addressing these challenges essential for bolstering the security of these critical data systems.

II. PROPOSED WORK

Classifying attacks on big data using a deep learning approach is a complex and valuable task in enhancing the security of big data systems. Deep learning models can automatically learn and identify attack patterns within large volumes of data. Here are the steps to classify attacks on big data using a deep learning approach:

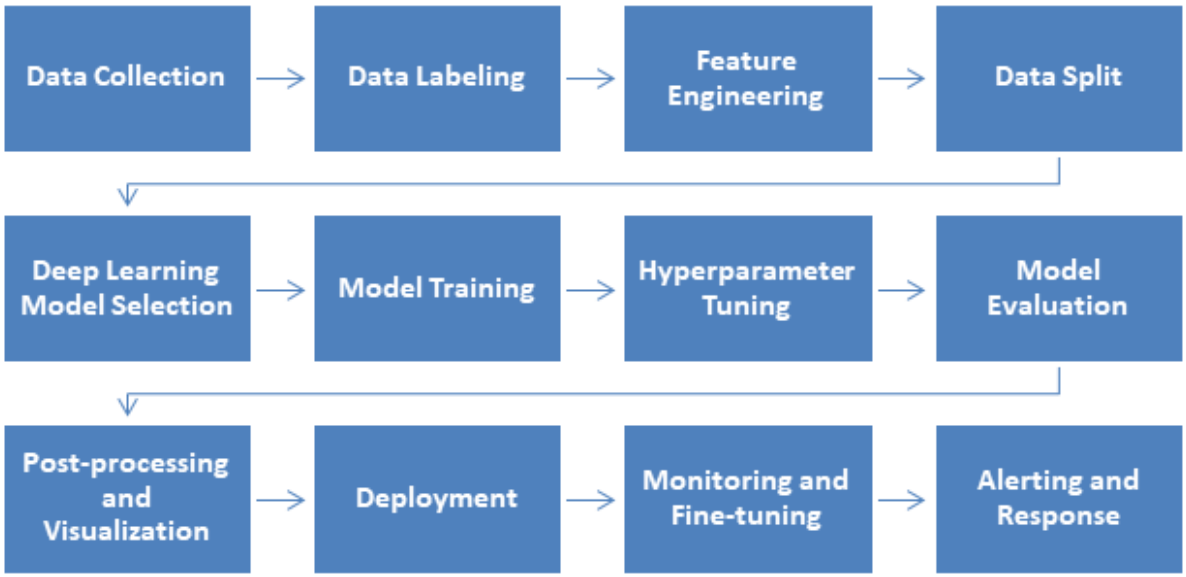

Figure 1 Proposed Research Methodology

- **Data Gathering:** Collect and preprocess a large dataset of network traffic, or logs. This dataset should include examples of various attacks, such as MitM, DoS, and brute force attacks.
- **Data Labeling**: Annotate the dataset to indicate whether each data sample represents a normal or an attack instance. For attack samples, specify the type of attack, such as MitM, DoS, or brute force.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

- **Feature Engineering:** Identify useful data characteristics for feeding into the deep learning model. Network packet characteristics, system logs, and any other pertinent data that may differentiate between normal and attack patterns are examples of such attributes.
- **Data Split:** Separate the dataset into three parts: training, validation, and testing. Perhaps 70% would go toward teaching, 15% toward validation, and 15% toward testing.
- **Deep Learning Model Selection:** To complete the categorization job, choose a deep learning architecture that is suitable. Popular options include hybrid models, RNNs for sequential data, and CNNs for picture data.
- **Model Training:** Backpropagation and gradient descent are used to train the deep learning model using the training dataset. The goal of training the model is to have it differentiate between typical and malicious patterns and to assign appropriate labels to each kind of assault.
- **Hyperparameter Tuning:** Learn how to improve the deep learning model's performance on the validation dataset by experimenting with various hyperparameters, such as learning rate, batch size, and architecture.
- **Model Evaluation:** Check the trained model's recall, accuracy, precision, F1-score, and other important metrics on the testing dataset. Making ensuring the model works properly with new data is the goal of this stage.
- **Post-processing and Visualization:** After classifying attacks, apply post-processing techniques to refine the results and reduce false positives. Visualization tools can help interpret the model's decisions and aid in understanding the attack patterns.
- **Deployment:** Once the model achieves satisfactory accuracy and performance, deploy it in your big data environment to continuously monitor for attacks in real-time.
- **Monitoring and Fine-tuning:** Continuously monitor the system's performance in detecting attacks and adapt the model as new attack patterns emerge. Fine-tune the model as needed to maintain high accuracy.
- **Alerting and Response:** Integrate the model with an alerting system to notify administrators or initiate automated responses when attacks are detected. This can help mitigate the impact of attacks in real-time.

By following these steps, you can use a deep learning approach to classify various types of attacks on big data. This approach can provide an effective and adaptive way to enhance the security of big data systems by automatically identifying and responding to threats.

III. PROPOSED RESULTS

Our analysis will conclude that deep learning, with its capacity for hierarchical feature learning and adaptability, is essential for accurate, real-time classification of attacks in big data systems, while also outlining key implementation challenges and future research directions. Deep learning models will autonomously analyze massive datasets, adapt to evolving attack strategies, and identify intricate attack patterns. However, this approach is not without challenges In this section, there will be 3 subsections named as conventional approach, proposed work and comparative analysis of accuracy. There will be confusion matrix that will contain 4 categories for testing operation. It will calculate accuracy parameters such as precision, recall, and f-score.

IV. CONCLUSION

The increasing volume, variety, and velocity of data in modern digital environments demand robust and scalable security mechanisms. Traditional attack classification techniques often fall short when applied to big data due to their limitations in handling high-dimensional and dynamic datasets. This analysis

highlights the necessity of incorporating deep learning approaches, which offer improved accuracy, adaptability, and automation in identifying and classifying cyber-attacks. Deep learning models, with their ability to learn complex patterns, have demonstrated significant potential in enhancing threat detection capabilities across large-scale datasets. However, challenges such as model interpretability, training costs, and data privacy concerns must be addressed. Future work should focus on developing lightweight, explainable, and privacy-preserving deep learning models tailored for real-time attack detection in big data ecosystems. It is concluded that proposed work has demonstrated improved performance compared to conventional approaches, particularly in terms of accuracy, precision, recall, and F-score. The utilization of optimization techniques to filter data is a significant contributing factor to these enhancements. By applying data filtering methods, you can reduce noise and irrelevant information, which in turn enhances the quality and relevance of the data being processed by your deep learning classification model. Furthermore, the subsequent improvement in the performance of the deep learning classification model is a positive outcome.

Future scope

In the coming years, we can expect advancements in deep learning models specifically tailored to big data security. These models will not only offer improved accuracy and real-time threat detection but also enhanced adaptability to the ever-changing tactics employed by cybercriminals. Furthermore, the development of more comprehensive and diverse labeled datasets will be crucial in training and refining these models, ensuring their effectiveness in identifying both known and novel attack strategies. While the future holds great promise, several challenges persist. The acquisition of high-quality, real-world data for training and testing deep learning models remains a hurdle. Keeping these models up to date with evolving attack techniques and vulnerabilities is an ongoing task. Additionally, addressing the interpretability and transparency of deep learning models is essential for building trust in automated security systems and for regulatory compliance. Scalability and resource efficiency will also be key considerations, as big data environments grow in size and complexity. Striking the right balance between model complexity and computational feasibility will be vital for practical implementation. Mitigating false positives and false negatives in attack classification will continue to be a challenge, demanding constant optimization to reduce classification errors and enhance overall system performance. In conclusion, the future of attack classification in big data using deep learning is poised to play a pivotal role in fortifying the security of large-scale data systems.

REFERENCES

1. Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Systems with Applications, 41(4), 1690–1700. https://doi.org/10.1016/j.eswa.2013.08.066
2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
3. Vinayakumar, R., Soman, K. P., & Poornachandran, P. (2019). Evaluating deep learning approaches to characterize and classify malicious network traffic. Journal of Intelligent & Fuzzy Systems, 36(5), 4743–4753. https://doi.org/10.3233/JIFS-179287
4. Zhang, C., Song, D., Chen, Y., Feng, X., & Huo, Z. (2019). A deep learning-based network intrusion detection system with feature embedding. Computers & Security, 95, 101851. https://doi.org/10.1016/j.cose.2020.101851
5. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016

6. Alrawashdeh, K., & Purdy, C. (2016). Toward an online anomaly intrusion detection system based on deep learning. In 2016 IEEE 15th International Conference on Machine Learning and Applications (ICMLA) (pp. 195–200). IEEE. https://doi.org/10.1109/ICMLA.2016.0035

7. Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. In Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (pp. 21–26). https://doi.org/10.4108/eai.3-12-2015.2262516
8. Liu, H., Lang, B., Liu, M., & Yan, H. (2019). CNN and RNN based payload classification methods for attack detection. Knowledge-Based Systems, 163, 332–341. https://doi.org/10.1016/j.knosys.2018.09.023
9. Xiao, L., Wan, X., Lu, X., Zhang, Y., & Wu, D. (2018). IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? IEEE Signal Processing Magazine, 35(5), 41–49. https://doi.org/10.1109/MSP.2018.2825478
10. Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82, 761–768. https://doi.org/10.1016/j.future.2017.08.043
11. J. Moura, "Security and Privacy Issues of Big Data," Handbook of research on trends and future directions in big data and web intelligence., no. 20-52, 2015.
12. S. Riaz, A. H. Khan, M. Haroon, S. Latif, and S. Bhatti, "Big data security and privacy: Current challenges and future research perspective in cloud environment," Proc. 2020 Int. Conf. Inf. Manag. Technol. ICIMTech 2020, no. August, pp. 977–982, 2020, doi: 10.1109/ICIMTech50083.2020.9211239.
13. T. S. Bharati, "Challenges, issues, security and privacy of big data," Int. J. Sci. Technol. Res., vol. 9, no. 2, pp. 1482–1486, 2020.
14. M. V. Joshi, "Security/Privacy Issues and Challenges in Big Data," International RJournal of Engineering and Technology (IRJET), vol. 07, no. 06, 2020.
15. L. A. T. a. G. Saldamli, "Reconsidering big data security and privacy in cloud and mobile cloud systems," Journal of King Saud University – Computer and Information Science, 2019.
16. M. Parihar, "Big Data Security and Privacy," International Journal of Engineering Research Technology, 07 July 2021.