2025 3rd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) | 979-8-3315-2543-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICAECA63854.2025.11012512

AI Driven Exploit Mitigation for Zero Day Vulnerability Using SVM and Autoencoder

Padmalakshmi.R.S,
B. Tech Artificial Intelligence and Machine Learning,
Saveetha Institute of Medical and Technical Sciences,
padmalakshmi704@gmail.com

Shawn Basil.C.K,
B. Tech Information Technology, Saveetha
Institute of Medical and Technical Sciences,
shawnbasil1970@gmail.com

***Abstract*-This study is about the action taken to reduce the risk or impact of a "zero-day vulnerability" (flaw or weakness in a software that is unknown to the makers or the public but that can be exploited by attackers). This study helps to analyze the exploits in a software to prevent cyberattacks. This study evaluates the performance of SVM and Autoencoder to classify the "zero-day vulnerability" dataset. SVM classifier got an accuracy of 94.9%, precision and recall values of 20.83% and 30%, f1-score at 23% and specificity of 94%. Whereas, autoencoder got an accuracy of 81.6%, precision and recall values of 21% and 28%, f1-score at 21.05% and specificity of 81%. According to the achieved values, SVM provides better accuracy than autoencoder in analyzing the dataset.**

Keywords- Zero-day vulnerabilities, SVM (support vector machine), autoencoder, anomaly detection, intrusion detection systems (ids), cybersecurity, machine learning, deep learning, network security, attack detection, malware detection, security analytics, feature extraction, unsupervised learning, artificial intelligence in cybersecurity.

I. INTRODUCTION

The main goal of this study was to identify the risks that significantly affect "zero-day vulnerabilities," which are substantial threats to software that would offer the developer zero days to repair it before being discovered by researchers or attackers, making it a target for exploitation. By using algorithms to analyse datasets, the study aims to address these vulnerabilities, reducing or stopping attacks if addressed promptly. This study would allow us to address the most prevalent cause by employing algorithms to analyse the dataset of prior zero-day vulnerability attacks. If the developer regularly updates and analyses the system, these attacks can be reduced or stopped entirely. Exploitation would result if these problems aren't identified and addressed promptly. We used the autoencoder and SVM classifier algorithms to examine the data from "Google Zero Days in the Wild (Apr 2024)" and acquired accuracy percentages of 94.9% and 81.6%

II. ALGORITHM

SVM Classifier - Support Vector Machine (SVM) stands out as a widely recognized approach for obtaining optimal solutions in various learning tasks. Introduced by Vapnik, SVM is a kernel-based model extensively used for both classification and regression purposes. In recent years, there has been a surge of interest in SVM across the fields of data mining, pattern recognition, and machine learning, owing to its remarkable generalization performance, ability to find optimal solutions, and powerful discriminative properties. SVM has consistently proven its effectiveness in addressing complex real-world binary classification problems. It has been shown to deliver superior performance compared to many traditional supervised learning algorithms. Due to its solid theoretical foundation and excellent capacity to generalize from limited data, SVM has gained widespread popularity as a classification technique. Today, it remains one of the most trusted and effective tools in the domain of supervised learning.

Autoencoder - An autoencoder is a type of unsupervised learning model capable of autonomously extracting meaningful features from vast datasets, often utilized for dimensionality reduction tasks. With the rapid advancements in deep learning, autoencoders have garnered significant interest among researchers and practitioners alike. In response to various application domains, numerous enhanced versions of autoencoders have been introduced and explored. This paper outlines the fundamental concept behind the traditional autoencoder and examines the key stages involved in its evolution. Subsequently, we present a structured taxonomy of autoencoders based on their architectures and underlying principles. Various types of autoencoder models are thoroughly reviewed and critically evaluated. This study also emphasizes the recent advancements in the application of autoencoders across various domains, including image classification, natural language processing, and more.

III. RELATED WORKS

Support Vector Machines (SVMs) have attracted considerable research interest across numerous domains, especially in pattern recognition. Renowned for their effectiveness in both classification and regression tasks, SVMs used for

(979-8-3315-2543-9/25/$31.00 ©2025 IEEE)

large datasets, multi-class problems, and imbalanced data scenarios. To enhance classification performance and parameter tuning, SVM had also been combined with advanced techniques such as evolutionary algorithms. This paper provides an overview of SVMs, exploring their applications, current challenges, emerging trends, limitations, and prospects, with a focus on specific areas of interest. [1] This paper explores the evolution of the autoencoder, an unsupervised learning model designed to extract data features from large datasets. It presents a taxonomy of autoencoders, examines their architectures and underlying principles, and discusses their applications in areas such as image classification and natural language processing. Additionally, it highlights the limitations of current autoencoder algorithms and outlines potential directions for future development. [2] Zero-day attacks exploit unknown vulnerabilities to evade detection by traditional cybersecurity mechanisms. These attacks pose a significant threat, as signature-based detection techniques often fail to recognize new or evolving exploits. In response, machine learning-based detection methods have emerged as promising solutions due to their ability to learn patterns and anomalies in data. This survey paper provides a comprehensive review of ML-based techniques for zero-day attack detection. It evaluates and compares various models, the datasets used for training, and the corresponding evaluation metrics and results. Despite their potential, current ML-based approaches still face challenges in achieving high accuracy, recall, and consistency when applied to diverse types of zero-day attacks. The paper identifies major limitations in existing methods, including generalization issues, lack of robust datasets, and difficulties in real-time detection. Furthermore, it discusses the need for more adaptive and intelligent systems capable of evolving with emerging threats. Finally, it outlines key challenges and offers recommendations for future research to enhance the effectiveness and reliability of ML-based zero-day attack detection systems.[3] This paper introduces a novel approach called the Transferred Deep-Convolutional Generative Adversarial Network (tDCGAN) for detecting zero-day malware, which typically evades traditional antivirus systems. The proposed method generates synthetic malware samples and learns to differentiate them from real malware using altered feature representations. The detection model utilizes both real and modified data derived from a random distribution, while a deep autoencoder (DAE) is employed to stabilize the training process of the GAN.[4] Cybersecurity seeks to safeguard systems against digital threats in the Information Age. Adversarial machine learning, a growing area of research, concentrates on the creation and detection of adversarial examples. While it has been widely studied in the context of image recognition, its application in intrusion and malware detection remains an emerging field. Survey findings indicate that although several adversarial attacks have proven effective in malware and intrusion detection, their real-world practicality in intrusion scenarios has yet to be fully evaluated. Additionally, there is a limited availability of diverse datasets for intrusion detection tasks.[5] Computer systems are susceptible to both known and zero-day attacks, which pose significant challenges in ensuring their security. Addressing these threats properly is complex due to their unpredictable and evolving nature. The k-zero-day safety metric is commonly used to evaluate the potential risks associated with such attacks. However, most existing algorithms designed to compute this metric are not scalable and often require extensive computational resources. To overcome these limitations, this paper introduces a set of polynomial-time algorithms that estimate k-zero-day safety more efficiently, eliminating the need to precompute the entire attack graph. This approach significantly enhances scalability while maintaining accuracy in risk assessment. The proposed method is not only computationally efficient but also practical for large-scale network environments. Experimental results demonstrate that the algorithm provides reliable estimations, making it a valuable tool for real-time cybersecurity analysis. Additionally, the study discusses the potential applications of this approach in proactive defence strategies and future research in attack graph modelling.[6] This study explores zero-day vulnerabilities within the digital ecosystem, emphasizing their impact on network security and the challenges they pose to businesses. It discusses preventive strategies such as intrusion detection systems and the use of artificial intelligence while also underscoring the importance of responsible vulnerability disclosure, timely patching, and continuous research to combat these elusive and hard-to-detect threats.[7] Zero-day attacks exploit unknown vulnerabilities in networks, leading to potential damage or disruption of programs. The most effective defence against such attacks involves timely detection and response. This study proposes a novel detection and prevention mechanism for zero-day attacks in software-defined networks (SDN) using Cuckoo, which effectively mitigates zero-day malware by isolating the compromised client.[8] A novel framework for detecting zero-day attacks, employing a hybrid learning approach. This method integrates supervised learning to identify

(979-8-3315-2543-9/25/$31.00 ©2025 IEEE)

known attacks with unsupervised learning for detecting previously unseen attack classes. The framework encompasses various intrusion detection phases, processes real-time network flow data, and incorporates cluster segmentation. Notably, it achieves a significant reduction in the false detection rate (FDR) without relying exclusively on the optimization of machine learning or deep learning algorithms. The effectiveness of the framework has been validated using two separate datasets.[9] Zero-day attacks take advantage of unknown vulnerabilities to evade detection by conventional cybersecurity tools. Traditional signature-based approaches are largely ineffective, whereas machine learning (ML)--based detection methods show considerable promise. This paper reviews various ML-based approaches for detecting zero-day attacks, comparing different models, training datasets, and evaluation outcomes. However, existing methods face challenges in terms of accuracy, recall, and consistency across diverse types of zero-day attacks. Key challenges are identified, and recommendations for future research directions are provided.[10]

IV. METHODOLOGY

Data collection:

The dataset used for analysis of Zero-day vulnerability in this research paper is "Google Zero Days in the Wild (Apr 2024)" from Kaggle, for analysing various reasons and cause for the exploitation. The various aspects like vendor, date of discovery, date patched, cause, etc. were analysed for the research work from the dataset.

Data pre-processing:

The data is separated as training data and testing data for better analysis of the dataset when algorithms are implemented to get accuracy values.

Model development:

Machine learning algorithms are used to analyse the "zero-day vulnerability" dataset, (SVM and autoencoder algorithm). SVM effectively separates data classes and performs well in generalization, particularly when applied to high-dimensional or unbalanced datasets. The ability of autoencoders to develop effective data representations and spot uncommon patterns makes them excellent in unsupervised anomaly identification. When used in tandem, they can employ autoencoders to pre-process data, identify anomalies, and then use SVM to precisely classify these anomalies. This combination makes it easier to find intricate patterns in big datasets, such as zero-day vulnerabilities.

Evaluation and comparison:

Various values like accuracy, precision, recall, etc. are analysed and compared to find the algorithm which performs better than the other. So that for further analysis better performing algorithms can be used.

V. RESULT

This study evaluates the performance of SVM and Autoencoder to classify the "zero-day vulnerability" dataset. The SVM classifier achieved an accuracy of 94/9% with precision and recall values of 20.83% and 30%, f1-score at 23% and specificity of 94% whereas, the autoencoder achieved an accuracy of 81.6% with precision and recall values of 21% and 28%, f1-score at 21.05% and specificity of 81%. According to the achieved values, SVM provides better accuracy than autoencoder in analysing the dataset.

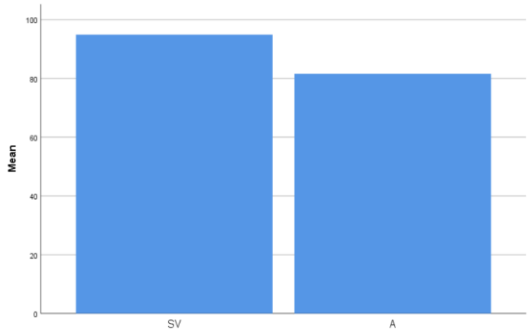

Fig. 1. In the present work, SVM is compared with Autoencoder it depicts that SVM algorithm gives more accuracy when compared with the other.

Group Statistics

	VAR1	N	Mean	Std. Deviation	Std. Error Mean
VAR2	SV	10	94.9000	1.44914	.45826
	A	10	81.6000	.69921	.22111

Fig. 2. Mean, standard deviation and standard error mean with an accuracy rate comparison of SVM over Autoencoder algorithm.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
VAR2	Equal variances assumed	4.224	.055	26.139	18	.000	13.30000	.50881	12.23103	14.36897
	Equal variances not assumed			26.139	12.975	.000	13.30000	.50881	12.20057	14.39943

Fig. 3. Significant threshold value of an accuracy rate comparison of SVM and Autoencoder algorithm.

VI. CONCLUSION

Evaluate the effectiveness of Support Vector Machine (SVM) and autoencoder models in detecting zero-day vulnerabilities. SVM outperforms autoencoder in terms of accuracy and performance metrics. SVM classifier achieved an accuracy of 94/9% with precision and recall values of 20.83% and 30%, f1-score at 23% and specificity of 94%. Whereas, the autoencoder achieved an

(979-8-3315-2543-9/25/$31.00 ©2025 IEEE)

accuracy of 81.6% with precision and recall values of 21% and 28%, f1-score at 21.05% and specificity of 81%. Despite all, the autoencoder's overall vulnerability detection performance was weaker compared to SVM.

VII. REFERENCE

[1] Jair Cervantes, et al. “A comprehensive survey on support vector machine classification: Applications,challenges and trends.” vol. 408, 30 September 2020, pp. 189-215.

[2] Penzhi Li, et al. “A comprehensive survey on design and application of autoencoder in deep learning.” vol. 138, May 2023.

[3] Yang Guo. “A review of machine learning based zero-day attack detection: challenges and future directions.” vol. 198, 15 January 2023, pp. 175-185.

[4] Jin-Young Kim, et al. “Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders.” vol. 460-461, September 2018, pp. 83-102.

[5] Nuno Martins, et al. “Adversarial machine learning applied to intrusive and malware scenarios: a systematic review.” vol. 8, February 2020, pp. 35403-35419.

[6] Massimiliano Albanese, et al. “An efficient approach to assessing the risk of zero-day vulnerabilities.”

[7] Raheela Zaib, and Kai-Quing Zhou. “Zero-day vulnerabilities: Unveiling the threat landscape in network security.” 2020, pp. 57-64.

[8] Huthifh Al-Rushdan, et al. “Zero-day attack detection and prevention in software defined networks.” December 2019.

[9] Almamy Touré, et al. “A framework for detecting zero-day exploits in network flows.” vol. 248, June 2024.

[10] Yang Guo. “A survey of machine learning based zero-day attack detection: challenges and future directions.”

(979-8-3315-2543-9/25/$31.00 ©2025 IEEE)