

Artificial Intelligence for Web Application Firewall (WAF): A Comprehensive Review

^{1*}Aya A. Zaki, ²Saja J. Mohammed

^{1,2}Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

*Corresponding Author's E-mail: aya.23csp61@student.uomosul.edu.iq

Abstract - The increasing prevalence of cyberattacks that bypass traditional defenses necessitates prioritizing web application security. So, that create an urgent need to use "firewalls", especially with web applications. The paper submitted a summary of the search and analysis of the scientific literature on web applications, in addition to the studies that have been suggested model for a "web application firewall (WAF)" that employed features engineering and machine learning to identify frequent online threats. The existing research examined WAFs and test their effectiveness in identifying fraudulent requests using "machine learning algorithms" like "Naive Bayes", "k-Nearest Neighbors", "Support Vector Machines", and linear regression. The studies integration of AI algorithms with existing WAF has shown achieved accuracy rates ranging from 92% to 99% to be highly effective in mitigating attacks.

Keywords: Artificial intelligence, WAF, Machine learning, Deep learning, Network security.

I. Introduction

Internet-based data communication is susceptible to several possible "cyber-attacks". Hackers might transmit the data to unapproved parties and the integrity of network data can be compromised by breaches in infrastructure security [1][2][3].

To address security issues, numerous safety techniques have been employed at the various levels of defense, such as Intrusion Detection/Prevention Systems (IDS/IPS), Internet firewalls [4]. Encryption, with its various type, are also used to keep the confidentiality of transmitted data over the network it consider as other assist level for defense against attackers [5][6][7][8].

Servers and firewalls are crucial security tools that shield communication networks from external cyber threats. They act as network guardians, monitoring and controlling data traffic to prevent unauthorized access. By analyzing incoming and outgoing data packets, they filter out malicious content based on predefined rules. "Machine learning" techniques, "K-Nearest Neighbors (KNN)", "Decision Tree Method

(DTM)", "Majority Voting Method (MVM)", "Support Vector Machine (SVM)", "Artificial Neural Networks (ANN)", "Shallow Neural Network (SNN)", "Convolutional Neural Network (CNN)", can enhance cybersecurity by automatically classifying and identifying potential threats[9][2][10][11].

The selection of an appropriate approach depends on a variety of dataset-related and complexity-related parameters, including dataset size, quantity and kind of features, data structure, data labeling and clustering, data distribution, and dataset [1].

This paper is structured as follows: In the first section, we introduce artificial intelligence as the perfect tool for creating a model that can achieve high success rates, easily expand its knowledge base, and adapt to different types of malicious attacks. The second section provides a brief overview of WAFs. The third section summarizes machine learning concepts. The fourth section discusses related research, and the final section presents our conclusions.

II. Web Application Firewall

WAF are used to secure online applications, by filtering the Hypertext Transfer Protocol) HTTP (traffic that travels between the internet and web applications. Web applications are shielded against several forms of attacks, including cross-site scripting and injection, by using web application firewalls. There are several policies specified in WAFs. This guideline is intended to safeguard web applications against application vulnerabilities. The web application and the internet are separated by a WAF [12][13][14][15].

Requests are being received by WAF as soon as the training model and WAF are functioning. When an HTTP request to extract features reaches the WAF, the analysis module isolates it. The transaction-oriented HTTP protocol, which may be used in any client-server application that uses hypertext, is the foundation of the World Wide Web (WWW). Designed for distributed, hypermedia, and collaborative information systems, this protocol operates at the application level. There is frequently HTTP communication between a web server and a browser. To provide reliability in this case,

HTTP employs the Transmission Control Protocol (TCP) protocol, in which each transaction is processed independently. In the HTTP protocol, messages are used to communicate between clients and servers. Messages can be of two types: requests from clients to servers and responses from servers to clients [16][17][18].

WAF capabilities are offered by the most successful and efficient solutions [19][20]:

- Input protection filters out unwanted or harmful data, allowing only legitimate user input.
- By configuring the validation rules, HTTP validation finds HTTP vulnerabilities and stops assaults.
- Policies designed for commonly used applications are configured in accordance with particular needs and requirements. As a result, it offers real-time traffic data and safeguards apps against vulnerabilities.
- By locating, screening, and protecting the private data, data leakage prevention generates an alarm and stops any unexpected traffic or data leakage.
- Automated attack blocking offers technologies for automatically preventing harmful traffic from entering the network, hence thwarting assaults.

“Artificial Intelligence” is a good solution for WAF to create highly effective, scalable, and flexible models. These models can accurately identify malicious queries, leading to improved success rates. By expanding their knowledge base, AI-powered WAFs can scale to handle growing threats. Additionally, their ability to generalize patterns allows them to detect new types of attacks. Based on these assessments, the firewall can either allow, deny, or drop incoming traffic. All actions are recorded in log files for future analysis [21][22].

Using “machine learning” to automatically categorize and predict traffic that bypasses the firewall will enhance “cybersecurity” by providing deeper insights into firewall log data and helping to defend against threats. In fact, the discipline of cybersecurity and machine learning techniques have recently come together to provide reliable and effective security solutions for a range of systems and applications [23][24].

III. Machine Learning

Machine learning uses statistical methods to enable computers to learn and adapt without being explicitly programmed. “Machine learning” aims to learn from the data. These methods find use in data mining, image processing, predictive analytics, and other fields. The potential of machine learning to make algorithms capable of doing tasks automatically once they have mastered the handling of data is its main advantage. The topic of training computers to learn

independently without specific programming has been the focus of several academic studies [25][26] [27] [28]. Detecting online vulnerabilities in a variety of web apps is another use for machine learning. And also used to classify source code and identify potential online vulnerabilities [29][30][31]. Because cyberattacks are becoming more sophisticated, WAF rules are also becoming more complicated and challenging to manually test and manage. Consequently, in order to stop fraudulent requests from getting to web applications and services, automated testing methods for WAFs that leverage machine learning are essential [32][4][24]. In order to boost the detection rate and do away with the need for human construction and filter updates for input data, machine learning may be utilized to build the detection model. On the other hand, web logs that are automatically created by the web server for every website that is hosted might be utilized as the detection model's input [25].

IV. Scientific Literature Review

Researchers and information security specialists have made a concentrated effort in recent years to use artificial intelligence's capabilities to identify and thwart for WAF.

In 2021, there was an article [1] suggested a new firewall system that uses machine learning to automatically categorize network traffic. This system uses a special type of decision tree and a neural network to identify three different types of network traffic based on eleven specific characteristics. Finally, when compared to the state-of-the-art in the area, the acquired findings outperformed currently available automated categorization models for firewall activities advances this field of study.

In 2022, There was an article [18] suggested a WAF approach that employed feature engineering and machine learning techniques to identify frequent online threats. To extract generic and comprehensive features, feature engineering and prior knowledge in the software security industry were utilized. Basic features were extracted from (HTTP) requests: payload, headers, files, HTTP method, and absolute "Uniform Resource Locator (URL)". Four additional features were extracted: input length, attack weight, special character ratio, and alphanumeric character ratio. The proposed model was trained on "web server" data of WAF-protected web applications to enhance security. Naive Bayes, effective for binary classification, was primarily used, though other algorithms like Logistic Regression and DTM could be considered. 99.6% classification accuracy was attained by the suggested model using conventional datasets, and 98.8% accuracy using datasets that were really compromised web servers.

In 2022, there was an article [14] suggested that neural networks could effectively and precisely model firewall rules. The article also determined the optimal values for parameters like momentum, learning rate, number of epochs, and number of neurons in hidden layers for the simulated firewall rules. Furthermore, it was demonstrated that the "Random Forest," "J48," and multilayer perceptron algorithms produced the best results when used to classification problems involving the firewall dataset. The study found that the "k-means" technique outperformed the EM and DBSCAN algorithms in terms of accuracy and speed when it came to grouping firewall data.

In 2022, there was a paper [33] provided an overview of common web application attacks, their detection techniques, and a comparative analysis of these techniques. Recognizing the strengths and weaknesses of each method, the study used machine learning with a traditional method to improve WAF performance. Machine learning is commonly used in intrusion detection systems. Combining these methods makes intrusion detection systems smarter and better at stopping new attacks, as attackers can often find ways around traditional methods.

In 2023, there was an article [12] suggested several characteristics of the standard datasets ISCX, CISC, and CIC DDoS were examined. By offering application-level filtering that a standard network firewall system is unable to provide, the web WAF provided an additional layer of protection to online applications. And by considering various factors, attack and regular traffic were contrasted. Using data gathered from the simulation environment, to identify "DDoS", "XSS", and "SQL" injection attacks, a layered architectural approach was developed. The first layer, which was constructed with a 97.57% accuracy rate. The accuracy of the second layer's acquisition was 89.34% upon examination, the increased volume of HTTP traffic was sent to the second tier.

In 2023, there was a paper [13] investigated the effectiveness of several artificial intelligence models in

identifying malicious requests in WAF. The paper used a synthetic dataset with over 100,000 requests and examined models like "Naive Bayes", (KNN), (SVM) and linear regression. The research demonstrated that integrating AI algorithms and using them to filter requests could potentially mitigate the research successfully attacked existing systems with nearly flawless accuracy. This method enhanced security, even for modified data that would be difficult for humans to identify. Machine learning was applied to a vast dataset, which was processed using vectorization and obfuscation to optimize model performance. All models demonstrated exceptional results, correctly classifying fraudulent requests between 92% and 99% of the time.

In 2024, there was a paper [19] suggested described the creation and deployment of an advanced WAF powered by machine learning algorithms. The system continuously monitored incoming web traffic, extracted relevant information, and classified data as malicious or benign using a machine learning model. Trained on historical data, the model identified patterns and behaviors indicative of cyberthreats like cross-site scripting, SQL injection, and other common attack vectors. Through this learning process, the system gained proficiency in detecting harmful actions and adapting its security measures accordingly. Ultimately, the work successfully developed a KNN model with a 96.6% attack detection rate.

In 2024, there was an article [34] proposed creating an online (WAF) using modify security and the "Open Web Application Security Project" (OWASP) Core Rule Set. With WordPress and Damn Vulnerable Web Application (DVWA) as test objects, the process involved using UML to analyze the current system's functioning. The results demonstrated 100% detection of SQL Injection attacks and 99.8% detection of XSS attacks, while real-time attack logs were maintained. These findings emphasized the importance of integrating (WAFs) with built-in security features.

Table 1: A Review of WAF Research's

Ref.	year	Proposed System	Advantages	Gaps
[1]	(2021)	A reliable, automated machine-learning firewall was proposed to classify network traffic	"Machine learning" models were applied to a processed dataset to optimize malicious request detection. Vectorization and obfuscation techniques were used to enhance model performance.	SNNs might be too complex. Simpler models could be explored.
[18]	(2022)	"Machine learning" based web application firewall	Using genuine compromised web server datasets and standard datasets utilized in this field of study, good classification accuracy was achieved.	It was generalizable but had low classification accuracy compared to others and used a single dataset (CSIC 2010).
[14]	(2022)	Neural networks accurately modeled firewall rules. Optimal parameters were determined	Random Forest, J48, and multilayer perceptron were the most accurate algorithms for classifying firewall data. K-means clustering was faster and more accurate than EM or DBSCAN.	Highlighted the strengths of classification algorithms for firewall data but didn't discuss potential drawbacks. Naive Bayes might assume feature independence, while Random Forest can be computationally expensive and overfit.

[33]	(2022)	This paper presents a comparative analysis of popular web application attacks and detection methods.	To improve accuracy, proposed combining machine learning methods, increasing attribute numbers, using regular expressions, and updating signature databases Future research will focus on cloud intrusion detection and firewall services	Analysis of information security incidents reveals that even advanced detection methods cannot guarantee 100% detection of web attacks.
[13]	(2023)	evaluates many artificial intelligence models, including support vector machines, Knn, linear regression, and naïve Bayes.	Machine learning models were trained on obfuscated and vectorized data to optimize malicious request detection. Results achieved 92%-99% accuracy.	The work use (SQL) injection attacks they can use another attacks like DDoS attacks
[12]	(2023)	New techniques are used to attack systems, stealing and destroying data. Defending against cyber threats like DDoS, SQL injection, and CSS is vital	Attacks using SQL injection, XSS, and DDoS were identified by a layered LSTM model. The accuracy of the first layer (DDoS) was 97.57%, while the second layer (XSS/SQL) was 89.34%. HTTP requests with a lot of traffic were first screened before going to the second layer.	Signature-based WAFs can detect some DDoS attacks, they are often ineffective against sophisticated and volumetric attacks.
[19]	(2024)	This paper developed a “machine learning” based WAF to enhance cybersecurity.	Proposed solution uses “machine learning” to dynamically detect and respond to cyber threats. By analyzing web traffic, it classifies requests as benign or malicious with high precision	The method for acquiring attack words, crucial for feature engineering, is not detailed. This model may struggle to adapt to real-world WAF changes.
[34]	(2024)	A WAF was developed utilizing Mod Security and the OWASP Core Rule Set to safeguard web applications from SQL Injection and XSS vulnerabilities.	By using UML, we assessed how the current system works. We tested DVWA and WordPress, and found that our system could detect almost all SQL Injection and XSS attacks. Logs showed attacks happening in real time	It focuses on SQL Injection and XSS attacks, neglecting other common web application vulnerabilities like CSRF, file inclusion, and remote code execution.

V. Conclusion

This paper highlights a comparative review of study between literatures of the critical Function of “artificial intelligence” especially “Machine Learning” in safeguarding web applications from evolving cyber threats. Machine learning-based web application firewall demonstrates significant potential in accurately identifying and mitigating common attacks. While existing studies have made valuable contributions, further research is needed to address limitations such as single-dataset reliance, computational complexity, and a narrow focus on specific attack types. Based on the provided researches, get the best results examined models like Naive Bayes, (KNN), (SVM) and linear regression. The integration of AI algorithms with existing WAF has shown to be highly effective in mitigating attacks, and showed that the work successfully developed a KNN model with a 96.6% attack detection rate. Future work should explore the integration of multiple datasets, investigate more efficient algorithms for resource-constrained environments, and broaden the attack detection's reach to include a greater variety of vulnerabilities. Addressing these shortcomings will pave the way for the development of more resilient and comprehensive WAFs that effectively protect against the ever-evolving landscape of online threats.

ACKNOWLEDGEMENTS

The authors are very grateful to the University of Mosul/College of Computer Science and Mathematics for their facilities, which helped improve the quality of this work.

REFERENCES

- [1] Al-Haijaa, Q. A., &Ishtaiwia, A. (2021). Machine learning based model to identify firewall decisions to improve cyber-defense. *International Journal on Advanced Science, Engineering and Information Technology*, 11(4), 1688-1695.
- [2] P. Kalariya and M. Jethva, “Progress Report: ML assisted Web application firewall,” Nov. 2023. Accessed: Nov. 20, 2023. [Online]. Available: https://brightspace.uwindsor.ca/d2l/lms/dropbox/user/f_older_user_vie_w_feedback.d2l?db=71493&grpId=0&isprv=0&bp=0&ou=146289.
- [3] Sharma, S., Zavarsky, P., &Butakov, S. (2020, May). Machine learning based intrusion detection system for web-based attacks. In *2020 IEEE 6th intl conference on big data security on cloud (BigDataSecurity), IEEE Intl conference on high performance and smart computing,(HPSC) and IEEE Intl conference on*

- intelligent data and security (IDS) (pp. 227-230). IEEE.
- [4] E. Ucar, E. Ozhan, "The Analysis of Firewall Policy Through Machine Learning and Data Mining", *Wireless Personal Communication*, Springer, vol. 96, p.p. 2891–2909, 2017.
- [5] Mohammed, S.J., Taha, D.B. Paillier cryptosystem enhancement for Homomorphic Encryption technique. *Multimed Tools Appl* **83**, 22567–22579 (2024). <https://doi.org/10.1007/s11042-023-16301-0>.
- [6] Mohammed SJ, Taha DB (2021) Privacy Preserving Algorithm using Chao-Scattering of Partial Homomorphic Encryption. *J Phys: Conf Ser*. <https://doi.org/10.1088/1742-6596/1963/1/012154>.
- [7] Mohammed SJ, Taha DB (2021) From Cloud Computing Security towards Homomorphic Encryption: A Comprehensive Review. *Telkomnika (Telecommunication Computing Electronics and Control)* **9**(4). <https://doi.org/10.12928/telkomnika.v19i4.16875>.
- [8] Mohammed, S.J. (2024). Developing a Hybrid Pseudo-Random Numbers Generator. In: Rasheed, J., Abu-Mahfouz, A.M., Fahim, M. (eds) *Forthcoming Networks and Sustainability in the AIoT Era. FoNeS-AIoT 2024. Lecture Notes in Networks and Systems*, vol 1036. Springer, Cham. https://doi.org/10.1007/978-3-031-62881-8_23.
- [9] Brain. G. Caspi, "Introducing Deep Learning: Boosting Cybersecurity with an Artificial Informa Tech" *Dark Reading, Analytics* <http://www.darkreading.com/analytics>, 2016.
- [10] Hammadi, Dhafar S., Ansam N. Younis, Fawziya M. Ramo. (2021) Hybridization and modification of the pso algorithm and its use in personal recognition by opg x-ray. *Journal of Engineering Science and Technology* **16.1** pp: 325-338.
- [11] Q. A. Al-Haija, S. Zein-Sabatto, "An Efficient Deep-Learning-Based Detection and Classification System for Cyber-Attacks in IoT Communication Networks" *Electronics*, MDPI, vol. 9, no. 12: paper no. 2152., 2020.
- [12] Dawadi, B. R., Adhikari, B., & Srivastava, D. K. (2023). Deep learning technique-enabled web application firewall for the detection of web attacks. *Sensors*, **23**(4), 2073.
- [13] Román-Gallego, J. Á., Pérez-Delgado, M. L., Viñuela, M. L., & Vega-Hernández, M. C. Artificial Intelligence Web Application Firewall for advanced detection of web injection attacks. *Expert Systems*, e13505. (2023).
- [14] Čisar, P., Popović, B., Kuk, K., Čisar, S. M., & Vuković, I. (2022). Machine Learning Aspects of Internet Firewall Data. In *Security-Related Advanced Technologies in Critical Infrastructure Protection: Theoretical and Practical Approach* (pp. 43-59). Dordrecht: Springer Netherlands.
- [15] Q. Niu and X. Li, "A high-performance web attack detection method based on CNN-GRU model," in *Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 804–808, IEEE, Chongqing, China, June 2020.
- [16] R. Kumari and S. K. Srivastava, "Machine learning: a review on binary classification," *International Journal of Computer Application*, vol. 160, p. 7, 2017.
- [17] H. Fadhil and A. R. Hakim, "Classification Model of Web Application Attacks," *2021 6th International Workshop on Big Data and Information Security (IWBIS)*, Depok, Indonesia, 2021, pp. 87-90, doi: 10.1109/IWBIS53353.2021.9631851.
- [18] Shaheed, A., & Kurdy, M. B. (2022). Web application firewall using machine learning and features engineering. *Security and Communication Networks*, **2022**(1), 5280158.
- [19] Kalariya, P., Jethva, M., & Alginahi, Y. (2024, April). ML Assisted Web Application Firewall. In *2024 12th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-6). IEEE.
- [20] Li, P., Wang, Y., Li, Q., Liu, Z., Xu, K., Ren, J,... & Lin, R. (2023, November). Learning from Limited Heterogeneous Training Data: Meta-Learning for Unsupervised Zero-Day Web Attack Detection across Web Domains. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1020-1034).
- [21] Román-Gallego, J. Á., Pérez-Delgado, M. L., Viñuela, M. L., & Vega-Hernández, M. C. Artificial Intelligence Web Application Firewall for advanced detection of web injection attacks. *Expert Systems*, e13505. (2023).
- [22] Demetrio, L., Valenza, A., Costa, G. and Lagorio, G., (2020) WAF-A-MoLE: evading web application firewalls through adversarial machine learning. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing* (pp. 1745-1752).
- [23] Román-Gallego, J. Á., Pérez-Delgado, M. L., Viñuela, M. L., & Vega-Hernández, M. C. Artificial Intelligence Web Application Firewall for advanced detection of web injection attacks. *Expert Systems*, e13505. (2023).
- [24] Čisar, P., Popović, B., Kuk, K., Čisar, S. M., & Vuković, I. (2022). Machine Learning Aspects of Internet Firewall Data. In *Security-Related Advanced Technologies in Critical Infrastructure Protection: Theoretical and Practical Approach* (pp. 43-59). Dordrecht: Springer Netherlands.

- [25] X. D. Hoang, "Detecting common web attacks based on machine learning using web log," in Proceedings of the International Conference on Engineering Research and Applications, pp. 311–318, Springer, ai Nguyen, December 2020.
- [26] Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], 9(1), 381-386.
- [27] G. T. Reddy, S. Bhattacharya, S. S. Ramakrishnan et al., "An ensemble based machine learning model for diabetic retinopathy classification," in Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE), pp. 1–6, IEEE, Vel lore, India, Feb 2020.
- [28] J.J. Praise, R.J Raj, J.V. Benifa, "Development of Reinforcement Learning and Pattern Matching (RLPM) Based Firewall for Secured Cloud Infrastructure", *Wireless Personal Communication*, Springer, vol.115, p.p. 993–1018, 2020.
- [29] Khalid, M. N., Farooq, H., Iqbal, M., Alam, M. T., & Rasheed, K. (2019). Predicting web vulnerabilities in web applications based on machine learning. In *Intelligent Technologies and Applications: First International Conference, INTAP 2018, Bahawalpur, Pakistan, October 23-25, 2018, Revised Selected Papers 1* (pp. 473-484). Springer Singapore.
- [30] Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. *IEEE Communications Surveys & Tutorials*, 22(3), 1646-1685.
- [31] S. Sharma, P. Zavorsky, and S. Butakov, "Machine learning based intrusion detection system for web-based attacks," in Proceedings of the 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), pp. 227–230, IEEE, Baltimore, MD, USA, May 2020.
- [32] Appelt, D., Nguyen, C. D., Panichella, A., & Briand, L. C. (2018). A machine-learning-driven evolutionary approach for testing web application firewalls. *IEEE Transactions on Reliability*, 67(3), 733-757.
- [33] Ho, T. P., Nam, H. T., & Thang, N. M. (2022). A new approach to improving web application firewall performance based on support vector machine method with analysis of Http request. *Hội thảo nghiên cứu ứng dụng Mật mã và An toàn thông tin*, 1(15), 62-73.
- [34] Annas, M., Adek, R. T., & Afrillia, Y. (2024). Web Application Firewall (WAF) Design to Detect and Anticipate Hacking in Web-Based Applications. *Journal of Advanced Computer Knowledge and Algorithms*, 1(3), 52-58.

Citation of this Article:

Aya A. Zaki, & Saja J. Mohammed. (2024). Artificial Intelligence for Web Application Firewall (WAF): A Comprehensive Review. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 8(11), 219-224. Article DOI: <https://doi.org/10.47001/IRJIET/2024.811027>
