

AI-Based Machine Learning Web Application Firewall (ML-WAF)

M. Fatima^{1*}, Ritik Lodhi², Ashika Bajpai³, Keshav Pachlaniya⁴, Ruchika Nagle⁵, Sourabh Kumar⁶

Abstract

This research investigates the use of deep learning techniques for the real-time detection of malicious activities in web traffic and proposes an intelligent, AI-driven Web Application Firewall (WAF) designed to provide automated and adaptive security. The system analyzes diverse components of HTTP requests, including request methods, URLs, headers, cookies, and payload content, to accurately identify and classify malicious behavior. The proposed model targets a wide range of common and critical web-based attacks, such as SQL Injection, Cross-Site Scripting (XSS), Path Traversal, and HTML Injection, which continue to pose serious threats to modern web applications. The primary objective of this work is to design, implement, and deploy a robust security mechanism capable of classifying incoming web requests as either benign or malicious in real time. Upon detection, the system actively mitigates threats by blocking, logging, or flagging suspicious traffic for further analysis. The training process relies on a labeled dataset composed of HTTP request logs collected from intentionally vulnerable web applications, ensuring exposure to realistic attack patterns. By learning behavioral relationships across multiple request attributes, the deep learning model demonstrates strong generalization capabilities across different attack vectors. When integrated into a live web environment, the proposed intelligent WAF provides dynamic protection that adapts to evolving attack techniques, thereby enhancing the overall resilience and security posture of web applications.

Keywords: AI, deep learning, HTTP traffic, machine learning, web application firewall, XSS

INTRODUCTION

In the digital era, web applications have become essential for businesses, governments, and individuals to facilitate seamless online interactions and data exchange. Despite the ease and flexibility offered by modern web technologies, they introduce substantial security risks. Attacks such as SQL injections, cross-site scripting (XSS), command injection, and path traversal represent major threats to the safety of web applications. Exploiting these weaknesses can result in unauthorized data exposure, monetary losses, damage to the organizational reputation, and interruptions to normal system operations [1].

*Author for Correspondence

M. Fatima

E-mail: mfcollge2050@gmail.com

¹Professor, Department of Artificial Intelligence & Machine Learning (AIML), Computer Science & Engineering - Cyber Security (CSE-CS), (SIRT) Sagar Institute of Research and Technology, Bhopal, Madhya Pradesh, India

²⁻⁶ Student, Department of Artificial Intelligence & Machine Learning (AIML), Computer Science & Engineering - Cyber Security (CSE-CS), (SIRT) Sagar Institute of Research and Technology, Bhopal, Madhya Pradesh, India

Received Date: January 15, 2026

Accepted Date: January 23, 2026

Published Date: February 10, 2026

Citation: M. Fatima, Ritik Lodhi, Ashika Bajpai, Keshav Pachlaniya, Ruchika Nagle, Sourabh Kumar. AI-Based Machine Learning Web Application Firewall (ML-WAF). Journal of Computer Technology & Applications. 2026; 17(1): 72–77p.

To overcome these issues, this project aims to design an artificial intelligence-driven web application firewall (WAF) by employing both sequential and scalar modeling approaches. The goal is to enhance cybersecurity by identifying and mitigating these common web vulnerabilities in real-time, thereby ensuring robust protection against malicious activities. Just as stock market prediction models aim to predict price movements amid uncertainties, this WAF project aims to anticipate and block potential security threats before exploiting vulnerabilities [2].

Objectives

Traditionally, rule-based systems and signature-based detection methods have been the primary approaches to web application security. Although these techniques are capable of stopping known attack types, they frequently fail to detect advanced and continuously evolving threats that do not conform to predefined signatures or rules [3]. This limitation has driven the adoption of AI-based approaches that can analyze complex and dynamic web traffic patterns more effectively.

Sequential and scalar models such as deep neural networks offer significant advantages in this context. They can learn intricate relationships within the data, identify anomalies, and adapt to new attack patterns over time, thereby providing a more comprehensive defense against web vulnerabilities. By leveraging these advanced AI techniques, this project aims to overcome the limitations of traditional methods and provide a robust and adaptable solution for real-time web application security [4].

The key objectives of this project include:

- To develop an AI-driven WAF capable of detecting and preventing common web attacks including SQL injections, XSS, command injections, and path traversal.
- Leverage sequential and scalar models for accurate anomaly detection in real-time web traffic
- Improving the overall security posture of web applications by reducing false positives and minimizing potential attack surfaces.
- To create a scalable and adaptable system that can evolve with emerging cybersecurity threats

LITERATURE REVIEW

Sequential and scalar models have been extensively studied in machine-learning research for a wide range of applications, particularly in areas such as cybersecurity and anomaly detection. These models offer powerful tools for understanding patterns in time-series data, making them well suited for the development of intelligent Web Application Firewalls (WAFs) [5].

Recurrent neural networks (RNNs), which were first introduced by Elman in 1990, represent a fundamental method for handling sequential data. Their architecture incorporates feedback connections that enable the retention of information across time steps, making them suitable for modeling temporal relationships. However, standard RNNs face difficulties in learning long-range dependencies and often suffer from problems such as gradient vanishing, as highlighted by Hochreiter in 1991.

To address these shortcomings, Long Short-Term Memory (LSTM) networks were proposed by Hochreiter and Schmidhuber in 1997. LSTMs are designed to remember information for extended periods, making them ideal for applications that require sequence learning, such as speech recognition, machine translation, and cybersecurity. Scalar models, however, are designed to handle simpler, non-sequential data.

These models, including feedforward neural networks and logistic regression, have been widely used for anomaly detection owing to their simplicity and efficiency. However, they lack temporal awareness of sequential models, making them less effective for dynamic web traffic analysis.

In recent years, AI-based WAFs have gained attention as a flexible and adaptive approach to web security. Traditional firewalls rely heavily on rule-based detection, which, while effective against known attacks, struggles against zero-day vulnerabilities and evolving threat patterns (Sommer and Paxson) [6]. AI-driven approaches, including sequential and scalar models, can identify complex attack signatures and adapt to emerging threats, providing a more comprehensive security solution.

Motivation

This study aims to bridge the gap between static rule-based security and advanced AI defenses [7]. The system processes large volumes of traffic to instantly identify and mitigate threats. The scope includes:

- Building a WAF using Sequential and Scalar AI models.
- Focusing on critical threats: SQL injection, XSS, and path traversal.
- Providing low-latency, real-time protection suitable for small to medium applications.
- Covering the full pipeline: Data preprocessing, training, and deployment.
- Ensuring the system is user-friendly and scalable for future growth.

Problem Formulation

The major drawback of the existing approaches lies in their limited ability to adapt to rapidly evolving cyber threats. Traditional methods rely on predefined rules and static signatures, rendering them inflexible. This study aims to address these shortcomings by leveraging the learning capabilities of sequential and scalar models to detect unknown attack patterns and reduce high false positive rates [8].

AI-BASED WEB APPLICATION FIREWALL

Web Application Firewall

A Web Application Firewall (WAF) is a security solution that safeguards web applications by inspecting, monitoring, and controlling HTTP traffic exchanged between users and web servers at the application layer (layer 7) [9]. Unlike traditional firewalls, a WAF prevents application-level attacks such as SQL Injection and XSS.

SQL Injection

SQL injection (SQLi) allows an attacker to interfere with the queries that an application makes in its database. This allows unauthorized access to sensitive data such as passwords and credit card details.

- *Common payload:* ' OR '1'='1 --
- *Prevention:* Input Validation, Least Privilege Principle, and Machine Learning-based WAFs.

HTML Injection

HTML Injection occurs when an untrusted input is inserted directly into a webpage's HTML content. This can lead to content-spoofing and phishing interfaces.

- *Common payload:* Hacked! or
- *Prevention:* Use Content Security Policy (CSP) Headers and validate input types.

Flow Diagram

The flow diagram illustrates the working process of the proposed AI-based Web Application Firewall (WAF). It begins with the reception of an incoming HTTP request, followed by feature extraction from various request components, such as HTTP methods, headers, URL parameters, and payload data. The extracted features are then preprocessed through tokenization, encoding, and normalization to prepare them for analysis. The processed input is passed to a trained machine-learning model that classifies the request as either benign or malicious. Based on the classification results, benign requests are forwarded to the web server, whereas malicious requests are blocked and logged for security monitoring and analysis Figure 1.

Algorithms

We used a supervised machine-learning classifier for real-time threat detection. Steps include:

- *Data collection:* Collect benign and malicious HTTP requests.
- *Feature Extraction:* Extract URL patterns, input length, special characters, and keywords.
- *Training:* Train the model on the dataset using scikit-learn/TensorFlow.
- *Prediction & response:* Deploy via Flask to block or allow requests.

Software Tools

- *Python:* Core programming language.
- *Flask:* Web framework for the firewall interface.
- *TensorFlow/Scikit-learn:* For model training.
- *Pandas/NumPy:* For data manipulation.

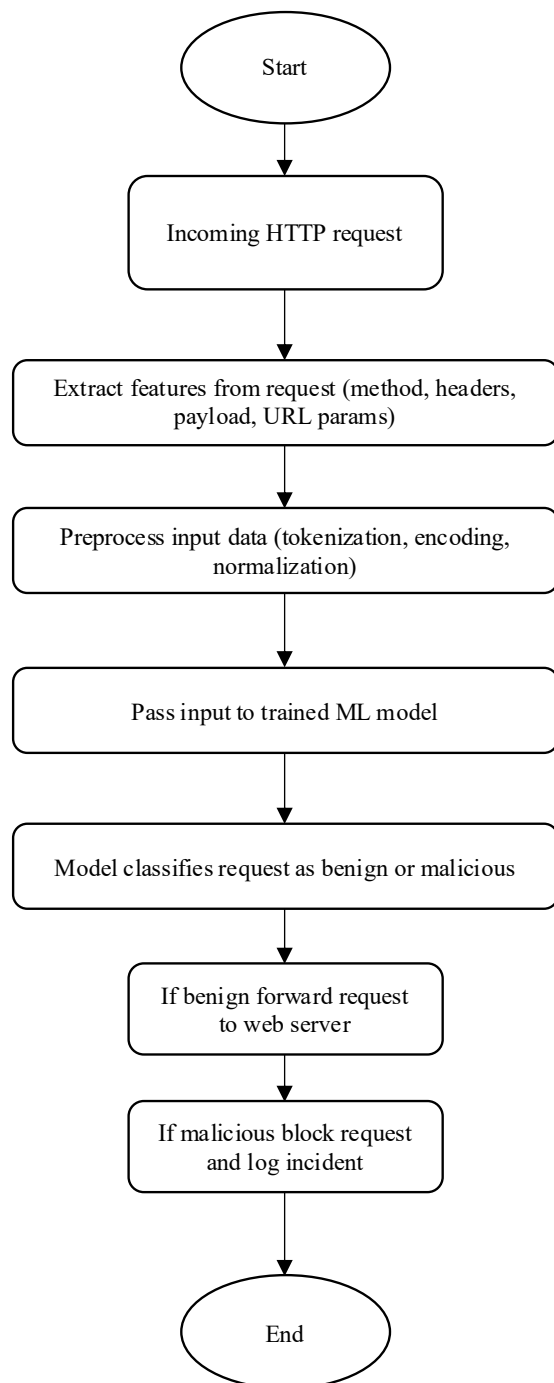


Figure 1. Flow diagram representing the real-time request processing and decision-making workflow of the proposed AI-based Web Application Firewall (WAF).

Dataset Generation

The dataset was created by collecting HTTP requests from simulated attack payloads (XSS and SQLi) and open datasets (OWASP logs) [10].

Efficiency Checksum

- *Model accuracy:* 99.2%.
- *Precision:* 92.1%.
- *Recall:* 95.0%. Real-time prediction speed: < 0.5 seconds/request.

Table 1. Case study.

Performance metrics for case study 1 (learning rate = 0.001)					
<i>S.N.</i>	<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
1	0 (Non-Malicious)	1.00	1.00	1.00	20100
2	1 (Malicious)	0.98	0.95	0.97	1972
3	Accuracy			0.99	22072
4	Macro Average	0.99	0.99	0.98	22072
5	Weighted Average	0.99	0.99	0.99	22072
ROC AUC Score: 0.9986					

EXPERIMENTAL RESULTS

Initial Values

This work utilizes libraries such as pandas, NumPy, sklearn, and tensorflow.keras. Final payload dataset csv was used, which contained feature-engineered values for various payloads.

Case Study 1(Learning Rate = 0.001)

A deep neural network model was trained with a learning rate of 0.001 using the Adam optimizer. ROC AUC Score: 0.9986

- *Accuracy: 99%*
- *Observation:* The model achieved consistently higher and more stable validation accuracy than higher learning rates.

Case Study 2

The model was retrained with a learning rate of 0.01.

- *ROC AUC score: 0.9967*
- *Accuracy: 99%*
- *Observation:* This model displayed more fluctuations and slightly lower accuracy, indicating that a higher learning rate caused instability.
- *Conclusion of Experiment:* Case Study 1 (LR=0.001) trained more reliably and consistently, avoiding overfitting issues found in Case Study 2.

Case Study Data

Input Dataset

The same dataset and preprocessing pipeline as in Case Study 1 were used to ensure a fair comparison.

Model Configuration

- Same neural network architecture
- Learning rate adjusted to 0.01.

Outcomes

Model trained with a learning rate of 0.01 under the same conditions Table 1.

CONCLUSION

This study successfully demonstrates the effectiveness of an AI-driven Web Application Firewall (WAF) built using deep learning techniques to enhance modern web application security. By leveraging a TensorFlow Sequential model trained on real-world HTTP request data, the proposed system can accurately identify and mitigate common web attacks such as SQL injections (SQLi) and Cross-Site Scripting (XSS). Unlike traditional rule-based WAFs, which rely on predefined signatures and static rules, the proposed approach learns behavioral patterns from traffic data, thereby allowing it to detect obfuscated and previously unseen attack variations. The results highlight the superiority of machine-

learning-based detection in handling complex and evolving threat landscapes. The ability of the system to operate in real time and adapt to diverse request parameters makes it a practical and scalable solution for deployment in live web environments. Overall, this study confirms that integrating artificial intelligence into web security mechanisms significantly improves detection accuracy, adaptability, and resilience against sophisticated cyber threats.

Future Scope

Although the proposed AI-based WAF demonstrates promising results, several enhancements can further improve its effectiveness and applicability. One major extension involves implementing multiclass classification to enable the system to distinguish between specific attack types, such as SQL injection, XSS, path traversal, and other web-based threats, rather than performing binary classification. Additionally, incorporating real-time or online learning mechanisms would allow the model to continuously update itself using new traffic data, eliminating the need for frequent retraining, and improving responsiveness to emerging attack patterns. Integration of security information and event management (SIEM) systems can provide better incident correlation, centralized monitoring, and forensic analysis. Furthermore, adopting advanced deep learning architectures such as long short-term memory (LSTM) or gated recurrent units (GRU) could improve temporal pattern recognition in sequential request data. These improvements can significantly strengthen the adaptability, intelligence, and overall robustness of the proposed security framework.

REFERENCES

1. Hernández JA, Solís-Pérez JE, Parrales A, Mata A, Colorado D, Huicochea A, Gómez-Aguilar JF. A conformable artificial neural network model to improve the void fraction prediction in helical heat exchangers. *International Communications in Heat and Mass Transfer*. 2023 Nov 1;148:107035. 10.1016/j.icheatmasstransfer.2023.107035
2. Elman JL. Finding structure in time. *Cognitive Science*. 1990 Mar;14(2):179–211. https://doi.org/10.1207/s15516709cog1402_1
3. Monner D, Reggia JA. A generalized LSTM-like training algorithm for second-order recurrent neural networks. *Neural Networks*. 2012 Jan 1;25:70–83. doi:10.1016/j.neunet.2011.07.003.
4. Alazmi S, De Leon DC. A systematic literature review on the characteristics and effectiveness of web application vulnerability scanners. *IEEE Access*. 2022 Mar 22;10:33200–19. 10.1109/ACCESS.2022.3161522
5. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. 1st edition. Cambridge, USA: MIT Press; 2016. pp. 1–800.
6. Sommer R, Paxson V. Network Intrusion Detection. In: IEEE, editors. 2010 IEEE Symposium on Security and Privacy. Oakland, USA: IEEE; 2010. pp. 305–316.
7. Ito M, Iyatomi H. Web application firewall using character-level convolutional neural network. In: 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA). Penang, Malaysia: IEEE; 2018. pp. 103–106.
8. Moradi Vartouni A, Teshnehlab M, Sedighian Kashi S. Leveraging deep neural networks for anomaly-based web application firewall. *IET Information Security*. 2019 Jul;13(4):352–361.
9. Sharma S, Zavorsky P, Butakov S. Machine learning based intrusion detection system for web-based attacks. In: 2020 IEEE 6th International Conference on Big Data Security on Cloud. Baltimore, USA: IEEE; 2020. pp. 227–230.
10. Scano C, Floris G, Montaruli B, Demetrio L, Valenza A, Compagna L, Ariu D, Piras L, Balzarotti D, Biggio B. ModSec-Learn: Boosting ModSecurity with Machine Learning. In: *Distributed Computing and Artificial Intelligence*. Cham, Switzerland: Springer Nature Switzerland; 2024. pp. 23–33.