International Journal of Scientific Modern Research and Technology
(Volume: 23, Issue: 03, Number: 01) Paper ID: IJSMRT - 26060301
ISSN: 2582-8150

Advanced Detection and Prevention of SQL Injection Attacks Using Machine Learning Techniques for Enhanced Web Security

Pooja Pandiya[1], Madhuvan Dixit[2]
[1]Research Scholar, [2]Head and Professor
Computer Department of IT, NIIST, Bhopal, Madhya Pradesh, India

Abstract: SQL Injection (SQLi) attacks remain one of the most critical and frequently exploited vulnerabilities in modern web applications. These attacks target database-driven systems by injecting malicious SQL queries into user inputs, leading to unauthorized access, data leakage, data manipulation, authentication bypass, and even complete system compromise. Traditional security mechanisms such as signature-based detection systems, Web Application Firewalls (WAFs), and rule-based filtering techniques are often insufficient against evolving and sophisticated SQL injection variants. Recent advancements in Machine Learning (ML) and Artificial Intelligence (AI) have introduced intelligent, adaptive, and scalable solutions for detecting and preventing SQL injection attacks. This review paper presents a comprehensive study of advanced SQL injection detection and prevention mechanisms using Machine Learning techniques for enhanced web security. The paper discusses the fundamentals of SQL injection attacks, various attack types, traditional mitigation strategies, and the limitations of conventional approaches.

Keywords: SQL Injection, Machine Learning, Web Security, Cybersecurity, Deep Learning, Intrusion Detection System, Artificial Intelligence, Web Application Firewall, Database Security.

***How to cite this article**: Pooja Pandiya, Madhuvan Dixit. (2026). Advanced Detection and Prevention of SQL Injection Attacks Using Machine Learning Techniques for Enhanced Web Security, International Journal of Scientific Modern Research and Technology (IJSMRT), ISSN: 2582-8150, Volume-23, Issue-03, Number-01, June-2026, pp.01-08, URL: https://www.ijsmrt.com/wp-content/uploads/2026/06/IJSMRT-26060301.pdf*

IJSMRT-26060301

I. INTRODUCTION

The rapid growth of internet-based services and web applications has transformed the digital landscape across industries such as banking, healthcare, education, e-commerce, and government sectors. Most modern web applications rely heavily on backend databases for storing and retrieving user information, transaction records, credentials, and sensitive organizational data. While database-driven applications improve functionality and accessibility, they also introduce serious security vulnerabilities.

Among the numerous cyber threats targeting web applications, SQL Injection (SQLi) attacks remain one of the most dangerous and persistent attack vectors. SQL injection occurs when attackers manipulate input fields or parameters to inject malicious SQL statements into application queries. If the application fails to properly validate or sanitize user input, the injected query may be executed by the database management system (DBMS), leading to severe consequences.

SQL injection (SQLi) attacks remain one of the most critical security threats to web applications, allowing attackers to manipulate databases by injecting malicious SQL code. Traditional defense mechanisms, such as input validation and web application firewalls, often fall short in detecting sophisticated SQLi attempts. To address this challenge, machine learning (ML) techniques have emerged as a powerful approach for enhancing web security by analyzing patterns and anomalies in web traffic. This paper

explores advanced detection and prevention strategies for SQL injection attacks using ML algorithms, including supervised and unsupervised learning models, to identify malicious queries with high accuracy. By leveraging natural language processing (NLP) and deep learning frameworks, the proposed system enhances the detection of SQLi attempts while minimizing false positives. Furthermore, real-time implementation of ML-based security solutions ensures adaptive protection against evolving cyber threats. This research aims to provide an effective and intelligent defense mechanism that significantly strengthens web application security against SQL injection attacks.

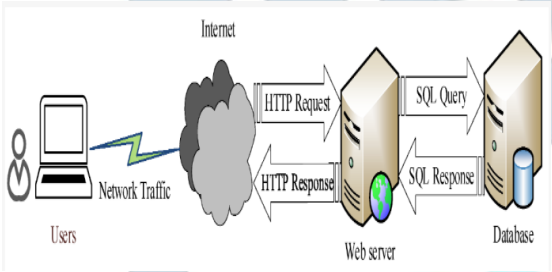

Figure 1: Detection and Prevention of SQL Injection Attacks

Enhanced Web Security

Enhanced Web Security refers to a comprehensive approach to safeguarding websites, web applications, and online transactions from cyber threats such as hacking, data breaches, malware, and phishing attacks. With the increasing reliance on digital platforms for communication, business, and personal use, ensuring web security has become a critical priority. Enhanced security measures include the use of advanced encryption protocols, multi-factor authentication, secure coding practices, firewall protection, and real-time threat detection systems. Additionally, implementing strict access controls, regular security audits, and compliance with cybersecurity standards help mitigate risks and protect sensitive information. By strengthening web security, organizations and individuals can prevent unauthorized access, maintain data integrity, and ensure a safer online experience for users.

Enhanced web security is essential in defending against evolving cyber threats that target vulnerabilities in websites and applications. Attackers continuously develop sophisticated techniques such as SQL injection, cross-site scripting (XSS), and distributed denial-of-service (DDoS) attacks, which can disrupt services and compromise sensitive data. To counteract these threats, organizations implement security frameworks like HTTPS, secure APIs, and web application firewalls (WAF) to filter and monitor traffic. Regular software updates, security patches, and vulnerability assessments are also crucial in maintaining a secure web environment. Additionally, user education and awareness programs play a vital role in preventing social engineering attacks, ensuring that individuals recognize phishing attempts and follow best security practices. By adopting a proactive and multi-layered security approach, businesses and individuals can significantly reduce cyber risks, build user trust, and create a safer digital ecosystem.

II. LITERATURE REVIEW

Hassan Shabani Mputu, et.al (2024), Author are Abdulrasheed Jimoh, et.al (2024) - Author are presented a This research offers insightful information about various algorism for the SQLI attacks detection within web-based applications. The results highlight the significance of choosing the right model to fortify web application security, with CNN, and GRU emerging as strong contenders. As the digital landscape continues to evolve, understanding the strengths and weaknesses of these models is vital for ensuring the robustness of web applications against SQLI attacks. To effectively neutralize SQLi attack, the adoption of advanced techniques Capsule Networks, Transformer-based Models is strongly advised in future while a larger dataset not only enhances the model's ability to detect and classify SQLI attacks but also improves its generalization capabilities. It allows the model to learn intricate patterns, variations, and anomalies associated with SQLI attacks, thereby boosting its accuracy and reliability in real world scenarios [01].

Bahman Arasteh, et. al, (2024) - Author are analysis SQL injection is one of the important security issues in web applications because it allows an attacker to interact with the application's database. SQL injection attacks can be detected using machine learning algorithms. The effective features should be employed in the training stage to develop an optimal classifier with optimal accuracy. Identifying the most effective features is an NP-complete combinatorial optimization problem. Feature selection is the process of selecting the training dataset's smallest and most

effective features. The main objective of this study is to enhance the accuracy, precision, and sensitivity of the SQLi detection method. In this study, an effective method to detect SQL injection attacks has been proposed. In the first stage, a specific training dataset consisting of 13 features was prepared. In the second stage, two different binary versions of the Gray-Wolf algorithm were developed to select the most effective features of the dataset. The created optimal datasets were used by different machine learning algorithms. Creating a new SQLi training dataset with 13 numeric features, developing two different binary versions of the gray wolf optimizer to optimally select the features of the dataset, and creating an effective and efficient classifier to detect SQLi attacks are the main contributions of this study [02].

Mohammed A M Oudah, et.al, (2024) - Author are study new SQL injection attacks are critical security vulnerability exploitation in web applications, posing risks to data, if successfully executed, allowing attackers to gain unauthorized access to sensitive data. Due to the absence of a standardized structure, traditional signature-based detection methods face challenges in effectively detecting SQL injection attacks. To overcome this challenge, machine learning (ML) algorithms have emerged as a promising approach for detecting SQL injection attacks. This paper presents a comprehensive literature review on the utilization of ML techniques for SQL injection detection. The review covers various aspects, including dataset collection, feature extraction, training, and testing, with different ML algorithms. The studies included in the review demonstrate high levels of accuracy in detecting attacks and reducing false positives [03].

Ankita Ghosh et.al (2024) - Deep Learning models, more specifically convolutional neural networks and long short-term memory networks, are more effective at repelling SQL Injection attacks. We found that Convolutional Neural Networks and Long Short-Term Memory consistently did better in our tests. This is because they can find complex patterns and relationships in SQL queries. This skill helped them find SQL Injection tries in online applications more accurately and regularly. Given the efficacy of Long Short-Term Memory and Convolutional Neural Networks, they may be beneficial in ensuring that online applications are safeguarded from SQL Injection assaults. Even though they are in a state of perpetual flux, these models have the ability to adapt to cyber threats by utilizing an active defensive system that is facilitated by Deep Learning and can learn from vast amounts of data [04].

Hilmi Salih Abdullah et.al (2024) - The ever-changing world of cybersecurity, it is becoming more important to ensure integrity of web applications as well as securing sensitive data. Among a variety of vulnerabilities, SQL injection is considered a significant risk with severe consequences. Addressing this crucial threat has always attracted the researchers to explore various approaches to identify and detect SQL injection attacks. The machine learning has captured the attention of the researchers to explore its potential due to its success in several different fields and the limitation of other rule-based approaches. This study provides a comprehensive review on a variety of the most recent researches that have been carried out using supervised learning algorithms. The study reveals that machine learning has a huge potential in the process of identification and detection of SQL injection attacks [05].

Abdulrasheed Jimoh et.al (2024) - In the realm of cybersecurity, safeguarding web applications against SQLi attacks is important; it remains a severe threat to web security, necessitating robust detection methods. This study presents a comparative study of the efficacy of various deep learning and machine learning approaches in detecting SQLI attacks. The models evaluated in this research include ResNet, LSTM, CNN, RNN, GRU, Decision Tree, SVM, Random Forest, Naïve Bayes and Logistic Regression based on critical performance metrics. The research leveraged a Kaggle dataset containing 33,721 SQLi queries and normal texts. Our findings reveal that CNN emerges as a standout performer with an impressive Accuracy of 97.86% and a high Precision of 99.56%. RNN and LSTM exhibit commendable performance, with Accuracy and F1 Score exceeding 94%, emphasizing their adaptability to SQLi detection. Logistic Regression, Random Forest, and ResNet exhibit notable precision, while SVM achieves perfect Recall, indicating its strength in identifying harmful SQL queries. However, Naive Bayes demonstrates limitations in detection effectiveness, with an Accuracy of 58.94%. These findings underscore the efficiency of various models in SQLi detection, with unique advantages and limitations. This research provides valuable insights for enhancing web security through optimized SQLi detection methodologies [06].

Fawaz Khaled Alarfaj et.al, (2023) - Researcher are Comparative analysis is presented here of SQL injection attack is considered one of the most dangerous vulnerabilities exploited to leak sensitive information, gain unauthorized access, and cause financial loss to individuals and organizations. Conventional defense approaches use static and heuristic methods to detect previously known SQL injection attacks. Existing research uses machine learning techniques that have the capability of detecting previously unknown and novel attack types. Taking advantage of deep learning to improve detection accuracy, we propose using a probabilistic neural network (PNN) to detect SQL injection attacks. To achieve the best value in selecting a smoothing parament, we employed the BAT algorithm, a metaheuristic algorithm for optimization. In this study, a dataset consisting of 6000 SQL injections and 3500 normal queries was used. Features were extracted based on tokenizing and a regular expression and were selected using Chi-Square testing. The features used in this study were collected from the network traffic and SQL queries. The experiment results show that our proposed PNN achieved an accuracy of 99.19% with a precision of 0.995%, a recall of 0.981%, and an F-Measure of 0.928% when employing a 10-fold cross-validation compared to other classifiers in different scenarios [07].

Nanang Cahyadi, et. al (2023) - Authors presented SQL injection attacks (SQLIAs) pose increasing threats as more organizations adopt vulnerable web applications and databases. By manipulating queries, SQLIAs access and destroy confidential data. This paper delivers three contributions around improving SQLIA detection research: first, a literature review assessing current detection/prevention systems to produce an SQL injection detection framework; second, specialized deep learning models optimizing session pattern analysis and feature engineering to enhance performance; third, comparing proposed models against previous defenses to surface promising research directions. Results highlight opportunities like real-time systems generalizing across attack variants through emerging techniques. Additionally, with attack complexity rising, systematized SQLIA investigation is warranted. Despite extensive study, current perspectives lack cohesive guidance informing mitigation strategies. Therefore, a framework is proposed holistically mapping knowledge gaps around contemporary SQLIAs, seminal threats in web applications, and security solutions. Furthermore, a multi-faceted framework examines research trends divided into hardening existing apps, detecting attacks on production systems, and integrating secure development practices. Literature suggests comprehensive resilience requires concurrent strength across these areas. Finally, future work remains in integrated frameworks, deep reinforcement learning adoption, automated AI auditing, and differential privacy to advance real-world SQL injection detection and prevention [08].

III. METHOD

The goal of this research is to develop a system capable of detecting SQL Injection (SQLi) attacks using machine learning. SQLi attacks are one of the most common and dangerous types of attacks in web applications, where malicious SQL queries are executed to manipulate a database. Machine learning models, when trained with the right data, can automatically identify these attacks and classify SQL queries as either malicious or benign.

About dataset

The dataset used for SQL Injection (SQLi) detection consists of SQL queries that are categorized as either benign or malicious. It is designed to enable machine learning models to learn the patterns and characteristics of SQL queries, helping to distinguish between normal SQL commands (benign) and potentially harmful ones (malicious, often representing SQL injection attempts).

Here's a detailed description of the dataset:

1. Structure of the Dataset

The first step in the methodology is data collection. A comprehensive dataset containing both benign (normal) SQL queries and malicious (SQL injection) queries is crucial for training an effective machine learning model. The dataset used in this project includes labeled SQL queries where:

- Benign queries represent legitimate database commands.
- Malicious queries represent SQL injection attacks intended to exploit vulnerabilities in SQL databases.

Each SQL query is labeled either as "Benign" (0) or "Malicious" (1), making it suitable for supervised

learning. The dataset may consist of several thousand rows of SQL queries, with each row containing a query string and a corresponding label. The dataset contains SQL queries, each labeled with one of the two categories:

Benign Queries: These represent regular, non-malicious SQL queries used in legitimate database operations, such as retrieving, inserting, or updating records.

Malicious Queries: These represent SQL injection attempts. These queries often contain dangerous SQL commands or manipulation techniques intended to compromise the database.

Each record in the dataset typically includes:

SQL Query: The actual SQL command (a string of characters).

Label: The class label indicating whether the query is benign or malicious. This label allows the machine learning model to distinguish between the two types of queries.

Algorithm for SQL Injection Detection

1. Start

2. Collect Dataset: Gather labeled SQL queries categorized as "benign" or "malicious."

3. Preprocess Data:

- Remove duplicates and handle missing values.
- Normalize query text, tokenize, and extract features.

4. Extract Features:

Identify SQL keywords, logical operators, escape characters, and calculate query length.

5. Model Selection:

Test different machine learning models (e.g., Logistic Regression, Decision Trees, SVM).

6. Train Model:

- Split the dataset into training (70-80%) and testing (20-30%).
- Train the selected model using extracted features.

7. Hyperparameter Tuning:

Optimize model parameters using techniques like Grid Search or Random Search.

8. Evaluate Model:

- Use metrics like accuracy, precision, recall, and F1-score to assess performance.
- Select the model with the best balance of these metrics.

9. Deploy Model:

- Implement the trained model in a real-world system to classify incoming SQL queries.
- Monitor performance and retrain with updated data as needed.

IV. SIMULATION AND RESULT

Model Performance

The models used in this study included Logistic Regression, Decision Trees, and Support Vector Machine (SVM). Each was evaluated for its suitability in identifying SQL injection attacks.

```
Accuracy: 0.9074778200253485
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.99      0.94       587
           1       0.97      0.66      0.78       202

    accuracy                           0.91       789
   macro avg       0.93      0.83      0.86       789
weighted avg       0.91      0.91      0.90       789
```

Figure 2: Classification Report for logistic regression

Support Vector Machine (SVM)

This image displays the classification report for an SVM model. The model's accuracy is 90.87%. For class 0, the precision is 89%, recall is 100%, and the F1-score is 94%. For class 1, the precision is 100%, recall is 64%, and the F1-score is 78%. The macro average precision is 95%, recall is 82%, and F1-score is 86%.

 ISSN: 2582-8150

Analysis: SVM demonstrated strong performance in high-dimensional spaces. By finding an optimal hyperplane to separate benign and malicious queries, it achieved high precision and recall. However, its computational complexity posed challenges with larger datasets.

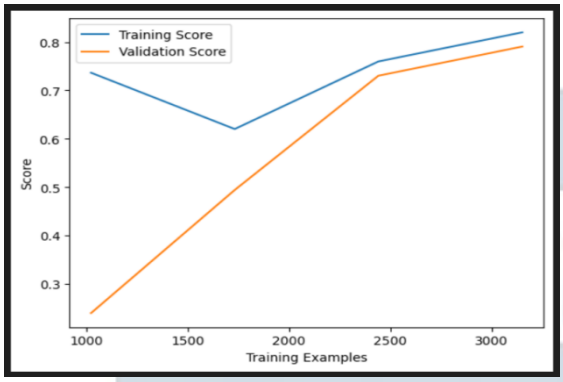

Figure 3: Learning Curve for Model Training and Validation

This image represents a learning curve, which is essential for evaluating the model's performance as the number of training examples increases. It is used to analyze how well the model generalizes to unseen data, ensuring its effectiveness in detecting SQL injection attacks.

Axes and Metrics

X-Axis (Training Examples):

The X-axis represents the number of training examples used during the training process. The data points on this axis range from 1,000 to 3,000 training examples, showcasing the incremental increase in dataset size.

Y-Axis (Score):

The Y-axis represents the model's performance, measured in terms of a score metric, such as accuracy. This allows for comparison between the training score and the validation score.

Table -1 Comparative Analysis of ML Techniques

Technique	Accuracy	Advantages	Limitations
Decision Tree	Moderate	Simple and interpretable	Overfitting
Random Forest	High	Robust and accurate	Computational overhead
SVM	High	Effective in complex spaces	Slow training
Naive Bayes	Moderate	Fast and lightweight	Assumes independence
ANN	High	Learns complex patterns	Requires large data
CNN	Very High	Automatic feature extraction	Resource intensive
LSTM	Very High	Sequence learning	Complex implementation
Ensemble Models	Excellent	Improved performance	High complexity

V. CONCLUSION

SQL Injection Attacks (SQLIAs) continue to be one of the most critical threats to modern web applications due to their ability to compromise sensitive data, bypass authentication mechanisms, and disrupt organizational systems. Traditional security approaches such as signature-based detection, input validation, and rule-based filtering provide a basic level of protection, but they often fail to detect sophisticated and evolving attack patterns. To overcome these limitations, Machine Learning (ML) techniques have emerged as an effective and intelligent solution for advanced SQL injection detection and prevention.

This review paper analyzed various machine learning approaches including Decision Trees, Random Forests, Support Vector Machines (SVM), Naive Bayes, Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) models. The study highlights that deep learning and ensemble-based methods achieve higher accuracy and better adaptability in identifying complex and previously unseen SQL injection patterns. Feature extraction methods, dataset quality, and real-time monitoring capabilities also play a significant role in improving detection performance.

VI. FUTURE SCOPES

Future studies should also focus on adversarial machine learning because attackers may attempt to bypass ML-based detectors using carefully crafted malicious queries. Developing adversarial robust models and secure Web Application Firewalls (WAFs) can significantly improve resilience against evolving threats.

Real-time deployment remains another major challenge. Future systems should emphasize lightweight and computationally efficient models suitable for cloud environments, IoT systems, and high-speed web applications. Hybrid detection frameworks combining machine learning, deep learning, and rule-based security mechanisms may provide better performance with lower false positives.

Future research can explore explainable artificial intelligence (XAI) techniques to make ML-based SQL injection detection systems more transparent and interpretable for cybersecurity analysts. Explainable models can improve trust, debugging, and practical implementation in enterprise environments.

REFERENCES

[1] Abdulrasheed Jimoh, Muhammed Kabir Ahmed, Suraj Salihu, Bala Mod, and Mohammed Nasir Salihu. "Enhancing Web Security Through Comprehensive Evaluation of SQL Injection Detection Models." 23–25 May 2024.

[2] Bahman Arasteh1, Babak Aghaei, Behnoud Farzad Keyvan Arasteh1 Farzad Kiani4 Mahsa Torkamanian-Afshar. "Detecting SQL injection attacks by binary gray wolf optimizer and machine learning algorithms." Volume 36, pages 6771– 6792, (2024) 27 February 2024.

[3] Mohammed A M Oudah and Mohd Fadzli Marhusin. "SQL Injection Detection using Machine Learning." Volume 10,Issue No. 1 eISSN: 2601-0003, 5 April 2024.

[4] Ankita Ghosh, Sudip Diyasi, Siddhartha Chatterjee "Enhancing SQL Injection Prevention: Advanced Machine Learning and LSTM-Based Techniques" Volume 78, July 202 Vol. 01, Iss. 01, S. No. 002, pp. 20-31, July 2024, ISSN (E): 3048-8516.

[5] Hilmi Salih Abdullah, Adnan Mohsin Abdulazeez "Detection of SQL Injection Attacks Based on Supervised Machine Learning Algorithms" 25 Apr 2024.

[6] Fawaz Khaled Alarfaj and Nayeem Ahmad Khan. "Enhancing the Performance of SQL Injection Attack Detection through Probabilistic Neural Networks." Volume 13, Issue 7, 29 March 2023.

[7] Nanang Cahyadi, Syifa Nurgaida Yutia , Pietra Dorand. "Enhancing SQL Injection Protection Through Integration, Automation, and Privacy." 19-12-2023

[8] Wazir Muhammad, Supavadee Aramvith ,Takao Onoye "SQL Injection Detection using Machine Learning" 5 July 2023.

[9] Fredrick Ochieng Okello, Dennis Kaburu, & Ndia G. John "Automation-Based User Input Sql Injection Detectionand Prevention Framework" Vol. 16, No. 2; 2023 ISSN 1913-8989 E-ISSN 1913-8997, May 2, 2023.

[10] Hao Sun, Yuejin Du and Qi Li "Deep Learning-Based Detection Technology for SQL Injection Research and Implementation" Journals Applied Sciences,Volume 13, Issue 16, 21 August 2023.

[11] Nisrean Thalji, Ali Raza, Mohammad Shariful Islam , Nagwan Abdel Samee, And Mona M. Jamjoom "AE-Net: Novel Autoencoder-Based Deep Features for SQL Injection Attack Detection" Volume 11, 135509, 28 November 2023.

[12] Yuting Guan , Junjiang He, Tao Li, Hui Zhao and Baoqiang Ma "A Black-Box Adversarial Attack Method for SQL Injection Based on Reinforcement Learning" 2023, 15, 133, Volume 15, Issue 4, 30 March 2023.

[13] Fawaz Khaled Alarfaj and Nayeem Ahmad Khan "Enhancing the Performance of SQL Injection Attack Detection through Probabilistic Neural Networks" Appl. Sci. 2023, 13, 4365, Volume 13 Issue 7, 29 March 2023, Journals MDPI.

ISSN: 2582-8150

[14] Dongzhe Lu, Jinlong Fei and Long Liu "A Semantic Learning-Based SQL Injection Attack Detection Technology" Electronics 2023, 12, 1344, Electronics 2023, 12, 1344, 12 March 2023, https://doi.org/10.3390/electronics12061344.

[15] Taseer Muhammad, Hamayoon Ghafory "SQL Injection Attack Detection Using Machine Learning Algorithm," Vol.2022, pp. 5–17, 25 Feb 2022.

[16] Maha Alghawazi , Daniyal Alghazzawi and Suaad Alarifi "Detection of SQL Injection Attack Using Machine Learning Techniques" Cybersecur. Priv. 2022, 2, 764–777, MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations, Volume 2, Issue 4, 20 September 2022.