Ministry of Education of the Republic of Belarus
Educational institution
BELARUSIAN STATE UNIVERSITY OF INFORMATICS AND RADIOELECTRONICS

UDC 004.056:004.032.26

Qiao Xinru

A machine learning-based software module for intrusion detection

ABSTRACT

For master's degree

Specialty 7-06-0611-02 Information security

Scientific supervisor

Sergei Nikolaevich Petrov
Candidate of Engineering Sciences,
Assistant Professor

Minsk 2026

INTRODUCTION

Cross-Site Scripting (XSS) is one of the most common and persistent security vulnerabilities in modern web applications. It allows attackers to inject malicious scripts into trusted web pages, thereby compromising the confidentiality of user information, session integrity, and the security of web platforms. With the rapid growth of dynamic web systems, cloud services, and online interactive applications, XSS attacks have become increasingly diverse in form and more difficult to detect using traditional security mechanisms.

Conventional XSS detection approaches mainly rely on manually designed rules, signature matching, blacklist filtering, or pattern-based analysis. These methods are effective for previously known attack payloads, but they often show poor adaptability when facing obfuscated, mutated, or previously unseen malicious scripts. In practice, attackers frequently modify payload structures through encoding, event injection, mixed tags, and script fragmentation, which significantly reduces the effectiveness of traditional defense techniques.

The rapid development of deep learning has provided a new direction for web security detection tasks. Unlike traditional methods, deep learning models can automatically learn semantic and structural features from large-scale input data and demonstrate strong generalization capability in complex classification tasks. In particular, neural network architectures such as CNN, BiLSTM, and hybrid models can effectively capture local patterns, contextual dependencies, and sequential behavior in malicious payloads, making them highly suitable for XSS attack detection.

This thesis systematically studies the application of deep learning methods in XSS attack detection. By designing and implementing multiple neural network models and comparing their performance on the same dataset, the thesis verifies the superiority of intelligent detection methods over conventional approaches in terms of classification accuracy, robustness, and generalization ability. The study also emphasizes the practical implementation of the proposed system and discusses its applicability in real-world web security defense scenarios.

An examination of the dissertation for the correctness of the use of borrowed materials using the «Antiplagiat» network resource in on-line mode on June 11, 2026 showed the correctness of the use of borrowed materials (originality is 95%).

GENERAL DESCRIPTION OF WORK

Objective and tasks of the research

The objective of this research is to design and implement an intelligent XSS attack detection system based on deep learning methods and to verify the effectiveness of deep learning models in identifying malicious script payloads under complex web input conditions.

To achieve the above objective, the thesis accomplishes the following tasks:

Collect and preprocess XSS-related datasets, and construct the training and testing samples;

Design and implement deep learning models for XSS payload classification, including CNN, BiLSTM, and hybrid neural network architectures;

Complete the preprocessing pipeline, including normalization, tokenization, vectorization, and class balancing;

Design comparative experiments and evaluate the detection performance of different models using accuracy, precision, recall, F1-score, and confusion matrix analysis;

Analyze the experimental results and summarize the advantages of hybrid deep learning methods in XSS attack detection;

Discuss the practical implementation value of the proposed system in web security protection scenarios.n.

Connection with the priority areas of research and requests of the real sector of the economy

The topic of the thesis corresponds to paragraph 6 Ensuring the security of humans, society and the state (means of technical and cryptographic information protection, cryptology and cybersecurity) of the list of priority areas of scientific, scientific-technical and innovative activities for 2021–2025, approved by the Decree of the President of the Republic of Belarus dated 07.05. 2020 No. 156.

Personal contribution of Master's degree student

My specific contributions in this research are as follows:

1. Model construction: completed the design and implementation of deep learning models for XSS detection, including CNN, BiLSTM, and hybrid architectures;

2. Algorithm implementation: wrote the complete code flow for dataset preprocessing, payload normalization, token encoding, model training, and testing;

3. Comparative experiments: designed and implemented comparison experiments among multiple models to verify the advantages of the proposed approach;

4. Data analysis: conducted systematic analysis and visual presentation of the experimental results and drew persuasive conclusions;

5. Cooperation with the supervisor: discussed the research idea, experimental design, and result analysis with the supervisor to ensure the scientific and practical value of the dissertation.

Dissertation advisor S.N. Petrov, doctor of philosophy, assistant professor, is a co-author of main publications. In-depth discussion with the supervisor on the research idea, experimental design and result analysis to ensure the scientific and innovative nature of the research work.

Practical approval and publication of Dissertation results

Information resource management: materials of the XXI International Scientific and Practical Conference, Minsk, March 28, 2025.

Technical means of information security: materials of the XXIV International Scientific and Technical Conference, Minsk, April 8, 2026

Publication of dissertation results

Based on the research results presented in the paper, two publications were published in the collection: Materials of the XXI International Scientific and Practical Conference « Studying the efficiency of machine learning methods for intrusion detection»; materials of the XXIV International Scientific and Technical Conference « Detection of xss attacks and sql injections using convolutional neural networks».

Structure and size of the work

The thesis consists of an introduction, General characteristics of the work, the main part of the three chapters, conclusion, references. The full volume of the thesis is 66 pages, including 30 illustrations, 15 tables, list of references of 22 titles, including 2 publications of the author.

SUMMARY OF THE WORK

The introduction contains a brief description of the work and the reasons for conducting the research. This paper implements multiple deep learning methods for XSS attack detection and compares their performance, concluding that the hybrid deep learning model has the most significant advantage.

In the first part, the theoretical basis of XSS attacks and existing detection methods is studied. The work analyzes the typical forms of malicious payload injection in modern web applications and summarizes the limitations of traditional rule-based, signature-based, and filtering-based approaches. It is shown that conventional detection methods are often ineffective against obfuscated, encoded, or previously unseen XSS payloads and therefore cannot fully satisfy the requirements of intelligent web security defense.

In the second part, the dataset preprocessing pipeline is implemented. The program completes data cleaning, normalization, tokenization, vector representation, and category balancing for malicious and benign input samples. Through preprocessing, raw payload sequences are transformed into model-readable representations suitable for deep learning training and evaluation. This part establishes the data foundation for the subsequent intelligent classification task.

In the third part, several deep learning models are constructed and tested for XSS attack detection. A convolutional neural network (CNN) is designed to extract local structural features from the payload sequences. A bidirectional long short-term memory network (BiLSTM) is implemented to capture contextual and sequential dependencies in the input data. In addition, a hybrid architecture combining convolutional feature extraction and sequence modeling is proposed in order to improve classification accuracy and robustness. Each model is trained and evaluated on the same dataset to ensure the fairness of comparison.

In the fourth part, comparative experiments are carried out on the same test set, and the results are analyzed in a visual format. Multiple evaluation metrics, including accuracy, precision, recall, F1-score, and confusion matrices, are used to assess the performance of each model. The experimental results demonstrate that the hybrid model achieves the best overall detection effect among the compared methods. Compared with single-architecture models, it shows stronger generalization ability and better robustness when identifying diverse XSS payload patterns.

From the practical perspective, the proposed XSS attack detection system can be integrated into web application firewalls, secure gateways, or server-side input validation modules to improve automated malicious script recognition capability.

The modular design of the system also provides a technical basis for future extension toward real-time inference and deployment in large-scale production environments.

Conclusion

This paper systematically designs and implements an intelligent XSS attack detection system based on deep learning methods. Several neural network architectures, including CNN, BiLSTM, and a hybrid model, were constructed and comparatively evaluated on the same dataset. The experimental results show that the hybrid deep learning model significantly outperforms single-model approaches in terms of detection accuracy, precision, recall, F1-score, and generalization ability.

Compared with traditional rule-based and signature-based methods, the proposed approach demonstrates stronger adaptability to obfuscated, variant, and previously unseen XSS payloads. The research validates the superiority of deep learning techniques in intelligent web security detection tasks and confirms their practical value for modern web application protection.

The results of this thesis provide both theoretical support and engineering reference for the development of intelligent web security defense systems. The proposed method has strong application potential in practical cybersecurity scenarios and can be further extended to real-time detection and broader web attack classification tasks in the future..

LIST OF OWN PUBLICATIONS

1-A. Qiao X. Studying the efficiency of machine learning methods for intrusion detection / Qiao X., Kedo E.S. // Управление информационными ресурсами: материалы XXI Международной научно-практической конференции, Минск, 28 марта 2025 г. / Академия управлени я при Президенте Республики Беларусь. – Минск, 2025. – С. 304–305.

2-A. Qiao, X. Detection of xss attacks and sql injections using convolutional neural networks / X. Qiao, E. S. Kedo // Технические средства защиты информации : материалы XXIV Международной научно-технической конференции, Минск, 8 апреля 2026 года / Белорус с к и й государственный университет информатики и радиоэлектроники [и др.] ; редкол.: О. В. Бойправ [и др.]. – Минск, 2026. – С. 212–215..