Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2025.050111
ISSN 2790-0983 Vol. 5, Issue 1: 108-116

Research on Automated Risk Detection Methods in Machine Learning Integrating Privacy Computing

Mingjie Chen

Software and Societal Systems Department, School of Computer Science, Carnegie Mellon University, Pittsburgh 15213

Keywords: Machine learning, neural network, Privacy Computing, Risk Detection, attention mechanism

Abstract: As the digitalization process of enterprises accelerates, various application interfaces have become key hubs for system interconnection and data exchange. At the same time, they may also serve as entry points for attackers to obtain sensitive information. Without effective protection, they will face the risks of data leakage and business interruption. To address this issue, this paper proposes an automated risk identification method that integrates privacy computing and deep learning. By conducting semantic parsing and behavioral pattern mining on interface request data, a hybrid framework combining rule constraints and neural network feature extraction is constructed to achieve intelligent identification of abnormal requests and potential threats. This method utilizes vector representation, bidirectional recurrent networks, attention mechanisms, and multi-layer convolutional networks to optimize the output, effectively enhancing the model's risk prediction capability in complex environments. The developed interface security assessment system can dynamically identify attacks and sensitive data leaks, providing enterprises with efficient automated interface protection solutions. It also provides a reference for the application of privacy computing and deep learning in the security management of actual business.

1. Introduction

With the rapid development of information and communication technology, various intelligent applications and digital services have been rapidly popularized in fields such as finance, healthcare, agriculture, intelligent manufacturing, and smart environment. These systems are highly interconnected through Internet of Things (iot) devices, enabling efficient operation of functions such as remote management of household devices, monitoring of health indicators, and precise regulation of agricultural production At the same time, it has promoted the integration and functional sharing of cross-platform services, providing enterprises and users with flexible operation and efficient data interaction capabilities. Application programming interfaces, as the core interaction hub between different systems and services, play a crucial role in this process. The application interfaces that enterprises rely on in the process of informatization and digitalization not only undertake the core tasks of expanding system functions and achieving data transmission, but

also play an irreplaceable role in optimizing enterprise business processes, improving overall operational efficiency, and ensuring service interoperability and security among different platforms and systems. Enable various applications and services to maintain a stable and reliable operating state in a complex and ever-changing operating environment. With the continuous increase in the number of interfaces and the rapid rise in their usage frequency, potential security threats faced by enterprises are gradually emerging. There is an urgent need to improve it by integrating more intelligent and semantic-aware technical means. Deep learning, with its advantages in sequence modeling, feature extraction and pattern recognition, can automatically learn the potential patterns in the request data, improve the accuracy of abnormal behavior recognition, and enhance the system's adaptive ability while reducing manual intervention, making the detection results more reliable and timelier.

In view of the limitations of existing methods, this paper proposes an intelligent risk detection method integrating privacy computing mechanisms. By combining rule constraints with deep learning models, it can comprehensively analyze the semantic features, parameter correlations, and behavioral contexts of interface requests, achieving intelligent identification, dynamic prediction, and risk assessment of known and unknown threats. At the same time, protect the security of users' sensitive information during the data processing. Based on this method, a complete interface security assessment system has been developed. This system can collect, parse and preprocess HTTP requests in real time, and generate alerts by combining abnormal behavior recognition and sensitive data detection, providing efficient and actionable risk analysis and decision support for developers and security managers.

2. Relevant research

Against the backdrop of the rapid development of modern information technology, the security threats faced by cyberspace and physical systems exhibit highly complex, diverse, and continuously evolving characteristics. This poses unprecedented challenges to achieving intelligent and automated risk monitoring and protection, and also requires researchers to conduct more in-depth and systematic exploration in method design and system construction. Yan and his team proposed an automatic avoidance strategy that combines model extraction and migration attacks to address the security issues of network intrusion detection systems. This strategy can effectively modify network traffic samples in a black box environment where only tag feedback can be obtained, with an average attack success rate of over 75%. They also clearly pointed out the importance of establishing a dedicated defense mechanism, providing new theoretical and practical references for cyber security protection[1].

Mohanty and his team designed an automatic defect identification method. By collecting images with drones and integrating deep learning computer vision technology, they have achieved precise and non-destructive identification of potential defects in photovoltaic modules [2]. Omer and other scholars have developed an optimized machine learning framework that can automatically analyze network traffic and identify abnormal behaviors, enabling real-time monitoring and rapid response to potential attacks, and providing reliable technical support and practical experience for addressing increasingly complex network threats [3].

Sumalatha and his team proposed a malware detection method based on deep integrated neural networks. By extracting key features from malware samples, graphically representing them and integrating ensemble learning strategies, automated and high-precision malware identification has been achieved, which can significantly enhance the overall security of mobile terminals and computer network systems [4]. Sun and his team proposed a machine learning method based on graph embedding, which captures high-dimensional feature information by simulating potential

access paths in the network structure and establishes a risk assessment and situation awareness framework for the system [5], thereby significantly improving the accuracy and protection capability of risk analysis in complex power systems[6].

These studies not only demonstrate the application potential of machine learning and deep learning in fields such as network intrusion detection, energy system monitoring, malware identification, and power system risk assessment[7], It provides a solid theoretical foundation and rich practical experience for building an automated risk detection method that integrates privacy computing[8], offers a feasible solution for high-precision risk identification across domains[9], and also ensures the security and privacy of data[10].

3. Risk Detection Methods

In this study, after completing the initial protection at the rule level, a deep learning model is introduced to enhance the system's ability to identify complex attack patterns and hidden risk behaviors. The processing flow first normalizes the API request text for preprocessing and then uses the word embedding method to convert the text into continuous low-dimensional feature vectors. Thus, while retaining the original semantic information, the contextual relationships and potential correlations between words can be captured. Subsequently, the system adopts a bidirectional long short-term memory network to model the feature vectors. This network can simultaneously integrate the forward and backward information flows, enabling the model to fully understand the internal semantic structure and complex dependencies of the HTTP request text, and assign higher weights to key features through the attention mechanism.

The security of interfaces and the reliability of data exchange directly affect the overall operational stability of the system and the level of user privacy protection. Therefore, interface risk detection has become an important direction in network security research. Interfaces exist in various forms, including function calls embedded in programs, dependency library calls, and data requests transmitted over the network. In practical applications, the form of network requests is the easiest for users to access and analyze. Therefore, this study takes it as the core research object to reflect the potential security threat characteristics in actual scenarios. The data used in the experiment is derived from a public network request log dataset, which contains a large number of normal and abnormal requests. Among them, the abnormal requests cover various attack methods such as injection attacks, cross-site scripting, file access anomalies, and sequence overflows, providing a complex and rich data environment for model training and performance verification.

Firstly, the parameters of the HTTP request were merged and normalized. Null values, redundant information and non-critical payloads were eliminated to form a unified text sequence for input, so that the deep learning model could efficiently extract features. Subsequently, a comprehensive decoding and standardization of the included encoded characters were carried out, restoring URL escape characters, Unicode encodings, and special characters to their original forms. Subsequently, all characters are uniformly converted to lowercase, and the numeric and time types are identified. At the same time, delimiters and repeated keywords are cleaned up to ensure the structural standardization and information completeness of the input data. The processed data is divided into a training set and a validation set for model training and performance evaluation.

The experiment first adopted a rule-based method for risk detection. The results showed that this method had a relatively low false alarm rate when identifying regular abnormal requests. However, for highly concealed or abnormal requests related to specific business logic, its identification ability was significantly insufficient. To enhance the detection effect, this study introduces deep learning methods to convert HTTP request sequences into vector representations. It captures the key features in the request sequences through long short-term memory networks and bidirectional long short-

term memory networks, and combines the attention mechanism to enhance the model's sensitivity to important information. The experimental results show that the deep learning method is significantly superior to the traditional rule-based method in terms of accuracy, precision, recall rate and F1 value. The bidirectional LSTM performs well in dealing with long sequence dependency problems, while the attention mechanism further enhances the model's ability to recognize core features.

The finally proposed integrated model MO_BLA combines the efficient feature extraction ability of the deep learning model with the low false positive advantage of the rule method, achieving automated interface risk detection. The overall detection accuracy has been significantly improved, and the performance for identifying abnormal requests is stable. As shown in Figure 1、2, compared with the existing methods, the data indicates that the model has obvious advantages both in comprehensive performance and practical application effects.

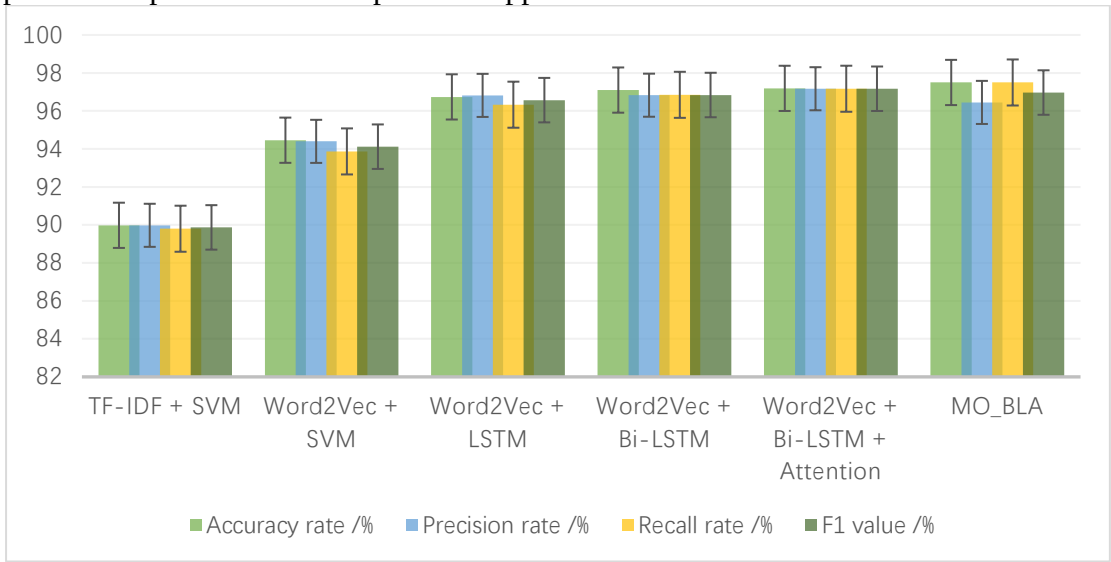

Figure 1 Comparison of results from various experimental methods

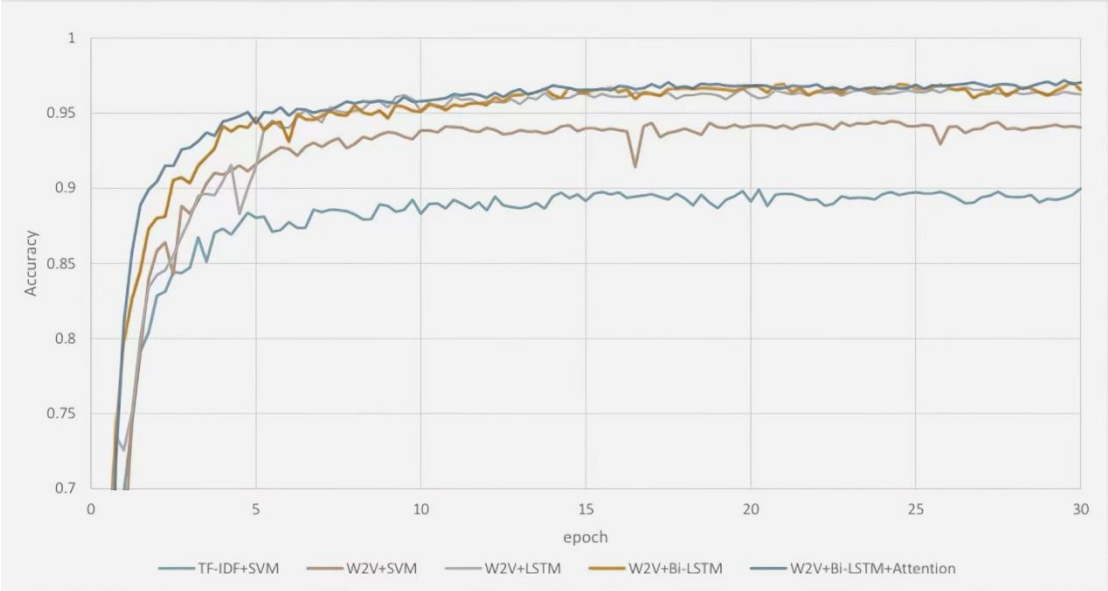

Figure 2 Changes in Accuracy Rate

4. Context and BERT Feature Detection

Interfaces in modern information systems are not only important hubs for interconnection and data exchange among different businesses, but also undertake the significant tasks of system function expansion and business collaboration. But at the same time, it also brings multi-level and multi-faceted security risks. Because HTTP requests usually have limited information and short structures on the surface, attackers can easily bypass traditional rule detection by constructing different requests, making it difficult for protection measures that rely solely on static rules or empirical judgments to deal with increasingly complex and covert attack behaviors. When confronted with a large number of interface requests, the risk detection model must be capable of simultaneously capturing local semantic features and understanding the overall context relationship to accurately identify those abnormal operations that are disguised or hidden, as shown in Figure 3. The capabilities of deep learning in semantic understanding and feature extraction provide reliable technical support for solving this problem.

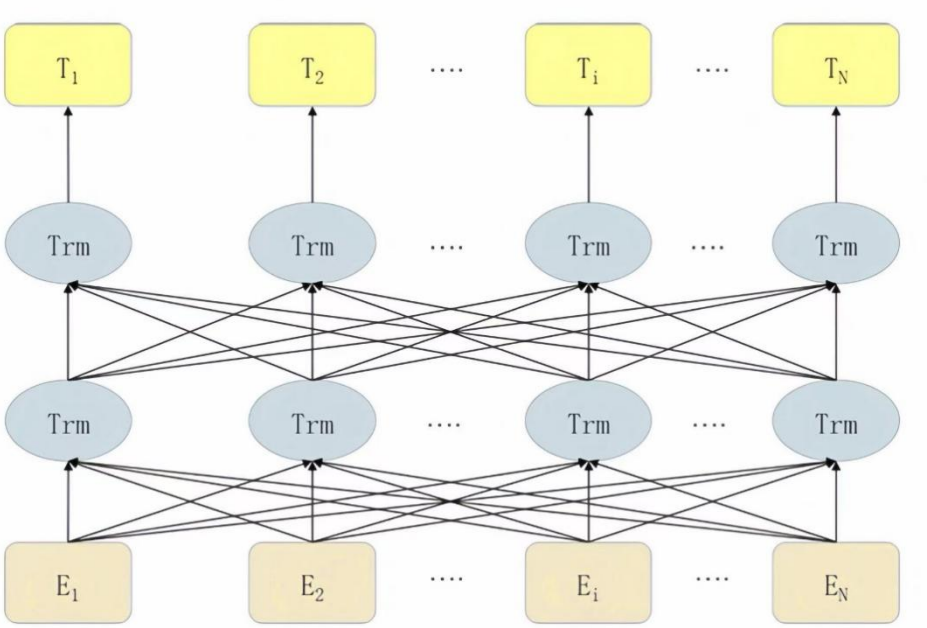

Figure 3 Pre-trained model

In the specific implementation of this study, the original interface request sequence is transformed into a high-dimensional semantic vector with context sensitivity, so that the model can better identify the deep dependencies and logical connections between requests, thereby improving the accuracy and stability of abnormal behavior detection. High-frequency features of local fragments are extracted through convolutional neural networks, and bidirectional recurrent neural networks are also used to model long-distance semantic dependencies and complex logical relationships, enabling the model to retain local information while grasping global structural features. In the feature fusion stage, the attention mechanism is introduced to perform weighted allocation based on the correlation between each feature and potential risks, enabling the model to focus on analyzing the most critical risk information and ultimately complete the intelligent classification and judgment of abnormal requests through the fully connected layer and the Softmax function. The model has achieved significant improvements in detection accuracy, recall rate and robustness by applying a multi-module collaborative strategy, taking into account both local details and overall dependencies. This enables it to maintain excellent generalization ability and practical

application value when facing new, unknown or highly concealed attack behaviors.

More importantly, this research integrates the concept of privacy computing into the detection process, completing data encryption and feature processing in the local environment, and then conducting modeling and reasoning through secure aggregation methods. Thus, it maintains the efficiency and reliability of the detection system while protecting sensitive information. This automated interface risk detection method that integrates deep learning with privacy protection has verified the feasibility of achieving a balance between high performance and compliance in actual network environments, and provides a technical solution worth promoting for future intelligent security management.

When conducting a comprehensive performance evaluation of the automated risk identification method, this study incorporates four key indicators from the classification task into a unified analytical framework and deduces relevant parameters through the statistics of true examples, true counterexamples, false positive examples and false counterexamples, thereby ensuring that the model's performance can be evenly reflected in multiple dimensions. With the continuous advancement of training rounds, although traditional methods can capture sequence features to a certain extent, they show obvious limitations when modeling long-range dependencies and context semantics.

In contrast, the structure based on context embedding demonstrates more significant advantages in the convergence of loss functions and the improvement of accuracy. When the local feature extraction ability of convolutional networks is combined with the global dependency modeling of bidirectional cyclic mechanisms, the classification performance is further enhanced. With the support of privacy computing mechanisms, the integrated model proposed in this study achieves the optimal level in multiple dimensional indicators. The test results in Figure 4 show that this method not only outperforms the comparison models in terms of accuracy, precision, recall rate and F1 score, but also achieves a relatively ideal balance between maintaining privacy protection and risk identification accuracy. This advantage enables it to demonstrate stronger adaptability and practicality in complex real network environments.

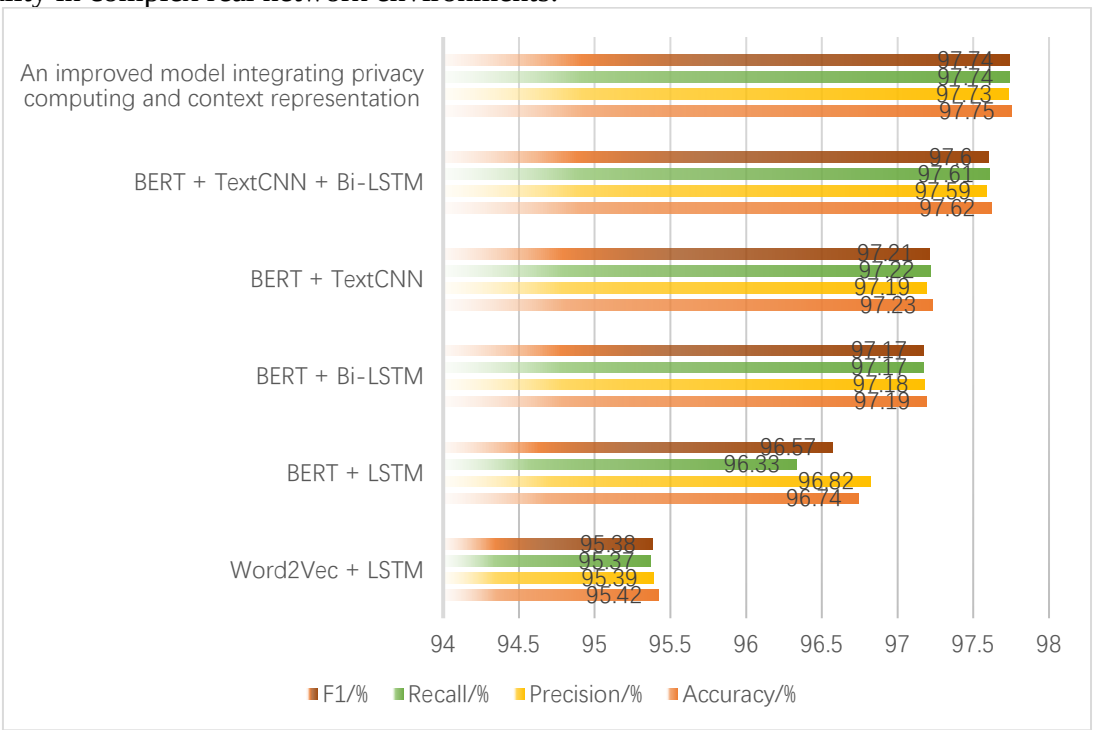

Figure4 shows the performance of each model on the test set

5. API Security Assessment System

In this study, the design objective of the system was established as achieving unified interaction among data collection, protocol parsing, risk identification, privacy protection and security alerts. The detection results are stored and persisted through the log system and fed back to the security management end through the real-time alarm mechanism, forming a technical system integrating identification, evidence storage and response, which significantly enhances the initiative and controllability of the system in security operation and maintenance.

The system achieves low-latency processing and high-throughput support for large-scale concurrent requests through distributed message queues and multi-layer caching mechanisms, ensuring that data transmission and task scheduling between different modules remain efficient and stable. Message queues play a buffering and diversion role during peak traffic periods, enabling attack detection and sensitive data identification to be executed in parallel. Meanwhile, the collaboration between caching and databases provides a balance between fast access and long-term storage. The overall architecture is built on the integration of the Spring ecosystem and the deep learning computing framework, which not only ensures rapid iteration and flexible expansion, but also lays a technical foundation for the introduction of more complex models and privacy computing mechanisms in the future, thereby ensuring that the system can maintain continuous updates and adaptability in an environment where attack methods are constantly evolving.

The data collection and analysis stage has been given a core position to support subsequent analysis and modeling. The R&D team injected lightweight scripts into the browser environment, enabling the front end to continuously capture the request and response data generated by network interactions without interrupting normal communication. After being collected, these data are not directly sent to the back end. Instead, it first completes an aggregation and preliminary processing locally. By uniformly merging parameters, precisely restoring symbols, formatting and marking numerical values, and adjusting the consistency of keywords, the complex and disordered original traffic gradually transforms into structured input vectors. In the data transmission process, the system applies multiple encryption and desensitization processes to sensitive information, enabling the data to maintain high analyzability while minimizing the risk of privacy leakage. Thus, a dynamic balance is established between system operational efficiency and compliance requirements. The system does not rely on a single technical means in the threat identification process, but achieves an organic combination of refined processing and efficient response through a multi-level discrimination mechanism.

The rule-driven filtering engine conducts an initial review of all requests. The engine quickly scores each request with preset modes and thresholds and can directly intercept obviously abnormal traffic at the millisecond level, thus ensuring that the system can respond promptly to real-time attacks. For potential threats that are difficult to identify through static rules, the system will submit relevant requests to the deep learning model for in-depth analysis. The model uses semantic embedding and sequence feature extraction techniques to globally model the context information of requests and dynamically weights key features through an attention mechanism. It can still maintain high accuracy and stability when dealing with complex inputs and long-distance dependencies, significantly enhancing the system's risk identification ability and protection level in complex network environments. The system introduces a privacy computing framework in the model training and inference stages, combining secure aggregation with local inference. This meets the efficiency requirements of parallel computing and protects users' original data from being leaked during cross-domain data transmission, enabling the detection module to maintain stable adaptability in response to constantly evolving attack patterns.

The system will generate security event records that can be used for decision-making based on

the detection results and identify potential sensitive content through refined annotation methods, providing a reliable basis for subsequent security management and protective measures. Specifically, the system comprehensively utilizes multiple mechanisms such as regular expression matching, feature dictionaries, and deep semantic analysis to hierarchically identify potential private data in the response body and trigger different automated handling strategies based on risk levels, enabling multi-level measures from field masking to traffic blocking to be dynamically activated according to actual conditions. After all the detection is completed, the system will solidify and store the events in temporal order. The recorded content covers elements such as the request path, protocol type, trigger condition, model score, and execution strategy, enabling operation and maintenance personnel to quickly reproduce the attack scenario when they need to conduct traceability analysis and evidence collection. The system is equipped with cross-channel linkage functionality. Once a high-risk event is detected, it will immediately send an alert to the management end and update the blacklist or restrict traffic according to the preset security policy. In this way, detection and response can closely cooperate, ensuring data security and privacy while also guaranteeing the convenience and efficiency of system operation.

6. Conclusions and Prospects

Intrusion detection technology, as a core supporting means of modern database security systems, has been widely applied. The system achieves all-round real-time monitoring and dynamic response when facing potential threats by establishing multi-layer protection mechanisms in key links such as data storage, transmission, and application services. A single defense strategy is no longer effective in dealing with multiple risks as attack methods continue to evolve and become more complex. This study innovatively combines feature-based detection methods with behavior analysis techniques to design a comprehensive intrusion detection framework that integrates predictive capabilities and immediate response functions. This enables the system to maintain a high degree of adaptability and accuracy when facing unknown or hidden threats. The system also integrates mechanisms such as encrypted transmission, identity authentication, and log tracking, which can quickly intercept abnormal or illegal operations, accurately record and analyze normal behaviors, and provide reliable data support for security management decisions and event traceability. By integrating machine learning models to intelligently predict and judge abnormal behaviors, a transformation from traditional passive defense to active risk management has been achieved. On the basis of ensuring the efficient and stable operation of the database, the overall security protection capability has been significantly enhanced.

References

[1] *Yan H, Li X, Zhang W, et al. Automatic Evasion of Machine Learning-Based Network Intrusion Detection Systems[J]. Dependable and Secure Computing, IEEE Trans. on (T-DSC), 2024, 21(1):15. DOI:10. 1109/TDSC. 2023. 3247585.*

[2] *Mohanty S R, Maruf M U, Singh V, et al. Machine learning approaches for automatic defect detection in photovoltaic systems[J]. Solar Energy, 2025, 298. DOI: 10. 1016/j.solener.2025. 113672.*

[3] *Omer N, Samak A H, Taloba A I, et al. Cybersecurity Threats Detection Using Optimized Machine Learning Frameworks[J]. Computer Systems Science & Engineering, 2024, 48(1). DOI:10. 32604/csse. 2023. 039265.*

[4] *Sumalatha P, Mahalakshmi G S. DEF: DEEP ENSEMBLE NEURAL NETWORK CLASSIFIER FOR ANDROID MALWARE DETECTION[J]. international journal of computer networks and communications, 2024, 16(2):59-69.*

[5] Jing, X. (2025). Research on the Application of Machine Learning in the Pricing of Cash Deposit Products. European Journal of Business, Economics & Management, 1(2), 150-157.
[6] Wu X, Bao W. Research on the Design of a Blockchain Logistics Information Platform Based on Reputation Proof Consensus Algorithm[J]. Procedia Computer Science, 2025, 262: 973-981.
[7] Sun S, Huang H, Payne E, et al. A graph embedding-based approach for automatic cyber-physical power system risk assessment to prevent and mitigate threats at scale[J]. IET Cyber-Physical Systems: Theory & Applications, 2024, 9(4). DOI:10. 1049/cps2. 12097.
[8] Li, W. (2025). Discussion on Using Blockchain Technology to Improve Audit Efficiency and Financial Transparency. Economics and Management Innovation, 2(4), 72-79.
[9] Xu Q. AI-Based Enterprise Notification Systems and Optimization Strategies for User Interaction[J]. European Journal of AI, Computing & Informatics, 2025, 1(2): 97-102.
[10] Su H, Luo W, Mehdad Y, et al. Llm-friendly knowledge representation for customer support[C]//Proceedings of the 31st International Conference on Computational Linguistics: Industry Track. 2025: 496-504.
[11] Z Zhong. AI-Assisted Workflow Optimization and Automation in the Compliance Technology Field [J]. International Journal of Advanced Computer Science and Applications (IJACSA), 2025, 16(10): 1-5.
[12] Huang, J. (2025). Research on Resource Prediction and Load Balancing Strategies Based on Big Data in Cloud Computing Platform. Artificial Intelligence and Digital Technology, 2(1), 49-55.
[13] Li W. Building a Credit Risk Data Management and Analysis System for Financial Markets Based on Blockchain Data Storage and Encryption Technology[C]//2025 3rd International Conference on Data Science and Network Security (ICDSNS). IEEE, 2025: 1-7.
[14] Wu Y. Graph Attention Network-based User Intent Identification Method for Social Bots[J]. Advances in Computer and Communication, 2025, 6(4).
[15] Wu Y. Software Engineering Practice of Microservice Architecture in Full Stack Development: From Architecture Design to Performance Optimization[J]. 2025.
[16] Sun Q. Research on Accuracy Improvement of Text Generation Algorithms in Intelligent Transcription Systems[J]. Advances in Computer and Communication, 2025, 6(4).
[17] Liu X. Emotional Analysis and Strategy Optimization of Live Streaming E-Commerce Users Under the Framework of Causal Inference[J]. Economics and Management Innovation, 2025, 2(6): 1-8.
[18] Lai L. Risk Control and Financial Analysis in Energy Industry Project Investment[J]. International Journal of Engineering Advances, 2025, 2(3): 21-28.
[19] Chen X. Research on Architecture Optimization of Intelligent Cloud Platform and Performance Enhancement of MicroServices[J]. Economics and Management Innovation, 2025, 2(5): 103-111.
[20] Yuan S. Application of Network Security Vulnerability Detection and Repair Process Optimization in Software Development[J]. European Journal of AI, Computing & Informatics, 2025, 1(3): 93-101.