ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 13 (2026) No. 6, pp. 1-16
DOI:10.33168/JLISS.2026.0601

Network Security Risk Management for Artificial Intelligence Service Platforms: An AI-Based Assessment Framework

Dexi Chen, Yongqiang Ma*, Wei Sun, Feng Liu

School of Computer and Big Data, Jining Normal University, Ulanqab 012000, Inner Mongolia, China

cdx_801019@163.com, nsd-myq@126.com (Corresponding author), sunw85@126.com, liufeng2026@126.com

Abstract. Purpose: In order to improve the level of network security assurance, and enhance network security situational awareness and risk prevention and control capabilities, this article studied the network security management of artificial intelligence application platforms as informatics-enabled service systems. By ensuring the security of these platforms, the research aims to enhance the reliability, continuity, and quality of AI-driven services. Method: This article analyzed the basic network security requirements and network security risks of artificial intelligence technology application platforms, designed a network security risk assessment process in the artificial intelligence application platform, identified assets, threats, and vulnerabilities of the network platform, and used the MRMR (minimum redundancy maximum reliability) feature extraction method for data feature extraction. Adopting an information security evaluation method based on attack graphs, an efficient and scientific network security risk assessment mechanism has been established, effectively managing network security. Result: The network security risk assessment method used in this article has high feature extraction accuracy and stability, with a feature extraction accuracy of over 80%. The predicted value of network risk is close to the true value of network risk, and can effectively monitor network security risks. Conclusion: The research results effectively improved the perception and early warning ability of network risks, realized the overall planning and monitoring of network security management, and provided new ideas for internet security management.

Keywords: network security management; artificial intelligence application platform; network security risk assessment; network attack

1. Introduction

As artificial intelligence (AI) increasingly becomes the backbone of modern informatics-enabled services, the security of AI application platforms has emerged as a critical concern. Unlike traditional information systems, AI platforms operate within data-intensive and service-oriented environments, which significantly increases system complexity and expands potential attack surfaces (Jimmy, 2021). Traditional network security mechanisms primarily rely on perimeter protection and signature-based detection methods. However, such approaches are often insufficient for addressing the dynamic and intelligent threats that arise within AI-driven service ecosystems.

A significant research gap exists in current security management frameworks. Existing studies typically focus on isolated technical detection tasks, such as intrusion classification or anomaly identification, without integrating these technical findings into a comprehensive, service-level risk prediction framework (Cheimonidis and Rantos, Islam et al. 2025). As a result, security monitoring systems often fail to provide actionable insights for service operators regarding the potential impact of security vulnerabilities on service continuity and system reliability.

Moreover, the lack of coordination between high-dimensional feature selection techniques and multi-stage attack path modeling further limits the effectiveness of current cybersecurity approaches. In particular, many intrusion detection models focus solely on improving detection accuracy while overlooking the broader implications of vulnerabilities within complex service environments. This limitation prevents system administrators from understanding how security threats propagate across service infrastructures and how they ultimately affect overall service availability (Lyu et al. 2023; Mallampati and Hari, 2023).

Artificial intelligence application platforms can be regarded as complex informatics-enabled service systems that integrate data processing, intelligent model deployment, and real-time service delivery (Naeem et al. 2025). Within such environments, network security management is not merely a technical protection mechanism but also a fundamental component of service governance and operational management (Awan & Alam, 2025). Effective security mechanisms are therefore essential for ensuring service reliability, maintaining system stability under potential cyber threats, and guaranteeing the quality of service (QoS) expected from modern AI infrastructures (Theodoropoulos et al. 2023; Sefati et al. 2025).

To address the limitations of existing research, this study proposes an integrated network security risk assessment framework for artificial intelligence service platforms. The proposed framework bridges the gap between low-level network data features and high-level service risk evaluation by integrating Minimum Redundancy Maximum Relevance (MRMR), attack graph modeling, and Hidden Markov Models (HMM). Through the combination of feature extraction, attack path analysis, and probabilistic risk prediction, the proposed approach enables a more comprehensive understanding of cybersecurity risks in AI-driven service environments, thereby supporting more effective security management and decision-making.

Main Contributions

The primary scientific contributions of this research are summarized as follows:

Integrated Security Management Framework for AI Service Platforms: This paper proposes a multi-stage security evaluation framework that bridges the gap between technical intrusion detection and service-level risk assessment.

Enhanced Feature Extraction using MRMR: Unlike traditional methods, the application of Minimum Redundancy Maximum Reliability (MRMR) in this framework significantly reduces data noise and redundancy, improving the accuracy of identifying sophisticated attacks (like R2U and U2R) in AI service environments.

Dynamic Risk Modeling via Attack Graphs and HMM: By integrating attack graphs with Hidden

Markov Models (HMM), the proposed approach moves beyond static vulnerability scanning to dynamic path prediction, allowing service operators to infer attacker intentions and prioritize defense strategies for critical service nodes.

Validation in Informatics-Enabled Environments: The methodology is validated through experimental data, demonstrating that AI-driven risk assessment can achieve over 80% accuracy in feature extraction and provide risk values close to real-world scenarios, thereby ensuring service continuity.

2. Literature Review

Network security has attracted increasing attention due to the rapid expansion of internet infrastructure and the growing sophistication of cyberattacks. Large-scale network attacks have posed serious security threats to individuals, organizations, and critical information infrastructures (Yang and Yang, 2021). With the continuous evolution of network technologies, malicious programs such as viruses and trojans have incorporated machine learning and pattern recognition techniques, which significantly enhance their ability to evade detection mechanisms. These malicious programs often disguise themselves as normal data files and remain hidden within network environments for extended periods, potentially causing severe damage once activated (Li et al. 2019).

In cybersecurity research, network intrusion is commonly defined as an attempt to compromise system integrity, availability, confidentiality, or service quality within a computing environment (Wang et al. 2021). To protect network systems and sensitive information resources, organizations typically implement multiple security measures, including password authentication mechanisms, firewall protection, and strict access control strategies (Wheelus et al. 2020). Although these defensive mechanisms provide a certain level of protection, traditional security approaches often struggle to cope with the increasing complexity and frequency of modern cyberattacks.

With the rapid advancement of artificial intelligence and big data technologies, AI has become an important tool in cybersecurity management. Artificial intelligence techniques have been widely applied in network security monitoring, anomaly detection, and automated threat identification (Waqas et al. 2022). In practical network security maintenance tasks, security administrators frequently rely on intelligent monitoring systems to inspect firewalls, antivirus gateways, and various network security devices. Intrusion detection systems play a crucial role in preventing unauthorized access and protecting sensitive data from external attackers, thereby maintaining the overall security of network environments (Becue et al. 2021).

Despite these advances, existing studies still face several limitations in effectively managing the security of artificial intelligence platforms. Most current approaches primarily focus on specific detection algorithms without establishing comprehensive frameworks that integrate feature extraction, attack modeling, and risk prediction. Consequently, further research is required to develop more systematic and intelligent security management approaches that can effectively monitor, analyze, and predict network security risks in AI service platforms.

3. Research Methodology

3.1.Network Security Evaluation of Artificial Intelligence Application Platform

3.1.1 Basic requirements for network security

The information security framework is formed upon a reliable operational foundation, establishing protective mechanisms that maintain security boundaries throughout data transmission, storage, and processing stages, thereby preventing unauthorized access and information leakage（ Jeong et al. 2019）. The original state of information across its lifecycle remains preserved, ensuring consistency and

authenticity during data generation, storage, and transmission while preventing illegal modification that could compromise system logic and result accuracy（ Sun et al. 2021）. Authenticity within network interactions relies on identity verification and behavioral traceability mechanisms, where communication participants establish verifiable accountability relationships and operational activities possess auditability in both technical and managerial dimensions（ Liu et al. 2020）. These security attributes collectively constitute the foundational structure supporting network information system operation, enabling artificial intelligence application platforms to maintain stable operational order in complex service environments and providing a unified security baseline for subsequent risk assessment and governance processes（ Zhang et al. 2020）.

3.1.2 Network security risks

With the continuous expansion of artificial intelligence applications, the openness of network environments and the complexity of system architectures have increased, leading to progressively intensified security threats faced by artificial intelligence application platforms（ Liu et al. 2020）. Intrusion activities evolve in both technical approaches and attack strategies, while malicious programs penetrate network systems through concealed transmission paths and long-term latent mechanisms, imposing higher uncertainty risks on platform operating environments（ Zhang et al. 2020）. Network security risks mainly originate from information resources and network devices, where system operational states are influenced simultaneously by human factors and technological deficiencies, forming security challenges shaped by the interaction of multiple risk sources.

Unintentional operational behaviors exert persistent impacts on system security, as configuration deviations and management negligence introduce latent vulnerabilities that accumulate risks during system operation（ Zheng et al. 2021）. Malicious attacks targeting network systems constitute a primary source of information security threats, where attackers disrupt information authenticity and integrity through diverse technical paths or conduct covert information acquisition while the system appears operationally stable, resulting in exposure of sensitive data and imbalance in system control authority. Structural imperfections inherent in network software design and implementation generate vulnerabilities that serve as critical entry points for external attacks, and software environments supporting system operation continuously reveal emerging security weaknesses throughout their evolution（ Feng et al. 2020）.

While network platforms deliver information services and resource interaction capabilities, structural fragility and operational complexity amplify the propagation effect of security risks, allowing threats to evolve from isolated issues into systemic risks（ Zhu ，2021）. Risk factors within network environments interact dynamically, forming cascading effects that continuously influence platform service stability and information security conditions, thereby driving network security management toward a systematic risk assessment and governance framework oriented to the overall service ecosystem rather than isolated defensive measures.

3.1.3 Asset identification

Asset identification constitutes a central component of the network security risk assessment process, where risk analysis activities are oriented toward objects within the system that possess value and require protection, forming a foundational cognitive framework for network security governance through systematic classification of diverse resources (Ganin et al. 2020; Kandasamy et al. 2020). The operation of an artificial intelligence application platform relies on a multi-layered resource structure in which system assets form an interconnected whole across functional support, data operation, and service maintenance, and their security status directly influences service continuity and operational stability (Swetha et al. 2025).

Hardware resources supporting network system operation undertake computational execution and information acquisition tasks, while differences in operational conditions and deployment environments

continuously affect system stability (Maia et al. 2024). Internal computing equipment and external sensing terminals exhibit distinct security characteristics during risk assessment due to variations in physical environments and operating conditions. Software resources constitute the foundation for functional realization and service orchestration within the platform, where operating systems and application programs sustain normal service processes, and the status of software maintenance and log management directly affects security monitoring and fault response performance (Ndibe et al. 2025).

Data resources carry core value within the artificial intelligence service ecosystem, extending throughout the lifecycle of collection, storage, processing, and transmission. Data management mechanisms and access control strategies determine the stability of information security boundaries, while data integrity and confidentiality sustain the reliability of operational logic and the credibility of service outputs. Alongside tangible resources, intangible assets formed during system operation represent critical protection targets, where service quality performance, system reputation, and accumulated intellectual achievements embody long-term platform value. Internal management deviations and external environmental disturbances exert continuous influence on these assets and further shape the overall security posture of the platform. All categories of assets operate within a unified security framework and maintain interdependent relationships that provide essential support for subsequent threat identification and quantitative risk analysis.

3.2.Risk Assessment Pipeline

3.2.1Principles of network security risk assessment

Unlike traditional IT infrastructures, an Artificial Intelligence Application Platform is a specialized informatics-enabled service system designed for high-concurrency data processing, model training, and intelligent service delivery. Its uniqueness lies in its dependency on large-scale datasets, complex model deployment pipelines (MLOps), and the dynamic nature of AI-driven decision-making processes, which expand the attack surface beyond traditional network perimeters.

3.2.2 Overview of the assessment pipeline

To ensure a rigorous evaluation of the AI service platform, the proposed methodology follows a structured four-stage pipeline:

Stage 1: Data Pre-processing and Feature Extraction: Raw network traffic is processed using the MRMR method to filter out noise and retain the most relevant security indicators.

Stage 2: Asset and Vulnerability Identification: System assets are categorized, and vulnerabilities are identified using SVM and FCM-based diagnosis to determine potential entry points for attackers.

Stage 3: Threat Modeling: An attack graph is generated to visualize all possible multi-stage attack paths based on the identified vulnerabilities.

Stage 4: Dynamic Risk Calculation: The HMM is applied to the attack graph to calculate the probability of specific attack sequences and determine the final risk value.

Information security risk refers to a potential, unknown, and negative type of information that exists in the network. Evolving network events are network systems that present, have occurred and are in a negative state.

Network security risks are composed of six aspects, namely origin, mode, receptor, pathway, consequence, and prevention. The related attributes from high to low are threat subject, threat behavior, asset, vulnerability, impact, and defense subject.

Network security risk assessment is divided into three parts: asset part, threat part, vulnerability part. The basic principles of network security risk assessment are:

(1) Asset identification is the measurement of the cost, value, and importance of security of assets in a network system;

(2) Threat identification is the identification of various types of hazards and the statistical analysis of the number of occurrences of each type of hazard;

(3) Vulnerability identification is the identification and quantitative evaluation of vulnerabilities;

(4) The probability of network attacks is calculated based on the probability of successful exploitation of vulnerabilities and the probability of possible network attacks, and this index is quantified;

(5) The calculation of the losses caused by the attack event is quantified based on the vulnerability being attacked and the benefits obtained by the attacker;

(6) The evaluation of network security risks is based on the previous 5 steps to obtain relevant indices and calculate the adverse effects that a network attack would have on the entire network system.

The specific implementation process of network security risk assessment is shown in Figure 1, and the implementation steps are:

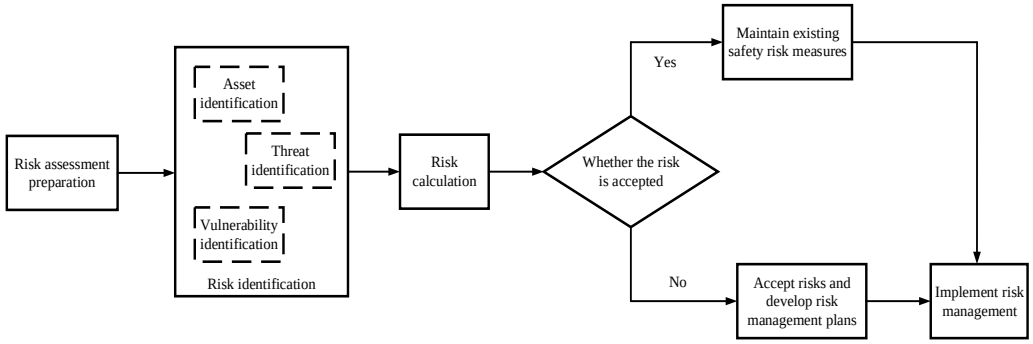

Fig.1: Specific implementation process of network security risk assessment.

(1) The preparation for risk assessment is to establish the objectives and scope of risk assessment, organize evaluation teams, conduct research on network systems, and select evaluation methods and implementation plans;

(2) Asset identification is the process of identifying various resources present in a network system and measuring their confidentiality, integrity, and availability;

(3) Threat identification is the process of dividing threats into several categories and assigning values to them based on the threats discovered by intrusion detection tools and logs, as well as the threats discovered by a specific industry published by international organizations, and their frequency of occurrence;

(4) Vulnerability identification is the identification of vulnerabilities in resources through two levels of technology and management, using a hierarchical approach to assign importance to identified vulnerabilities;

(5) Risk analysis is to calculate the adverse consequences of an attack based on the severity of the vulnerability and the value of the asset after completing the first four steps, which is the risk value. Based on the results of the risk assessment, the danger levels are divided and corresponding security and defense strategies are adopted to raise the remaining risks to a level of “moderate safety”;

(6) The risk assessment document records a series of process and result archives formed during the evaluation process. For example, risk assessment plans, asset identification, vulnerability lists, threat lists, and risk assessment reports are stored, retrieved, and protected.

The logical integration of these methods follows a sequential data flow to ensure a comprehensive risk assessment. First, SVM and FCM are employed in a hybrid manner to identify and categorize vulnerabilities (Section 3.4). Specifically, FCM clusters the raw vulnerability data to identify potential threat patterns, while SVM provides precise classification of these vulnerabilities. The identified vulnerability states then serve as the contextual basis for MRMR feature selection (Section 4), where

the most representative network traffic indicators associated with these vulnerabilities are extracted. Finally, these features inform the state transitions in the Attack Graph and HMM model (Section 5) to calculate the cumulative risk.

3.2.3 Threat identification

In order to effectively evaluate the security of information systems, it is necessary to first determine which risk factors would affect the security of the network system.

TMSRA (targeted mass spectral ratio analysis) uses the Delphi group discussion method, combined with historical data and system vulnerability scanning methods, to effectively identify information systems. Among them, the analysis of historical data mainly focuses on the data within the information system, with the data within the network system as the main content, to organize and analyze the data. On this basis, by identifying threats, the main threats that affect the information system are identified, and threat set $T:\{t_i|i=1,2,...,n\}$ is constructed. Among them, t_i is the i-th type of threat and n is the threat type.

How to measure the harm that threats bring to information systems, especially to quantify their impact, is a very important and difficult issue. In real life, certain specific threats can bring multiple levels of threats to information systems. The magnitude of individual outcomes caused by a specific threat varies, and the importance of the harm sustained by a specific information system also varies. On this basis, a multi-factor-based threat consequence analysis method is used, which defines the damage of a risk to an information system as a multi-factor threat consequence. Moreover, threat outcomes have weights, which are determined by the importance and tolerance of a certain information system towards various threat outcomes.

Therefore, it is necessary to judge the nature of threat outcomes from a practical perspective, that is, what kind of security hazards these properties would bring to the information system. In order to better reflect the true situation of the evaluated target, it is necessary to conduct a comprehensive assessment of the harm caused by the threat to the information system. The types of threat result attributes are described in the following way: $X:\{x_j|j=1,2,...,m\}$. Among them, x_j is the j-th result attribute and m is the number of result attributes. The value of the threat result attribute is $W:\{w_j|j=1,2,...,m\}$. Among them, w_j is the weight of the j-th result attribute.

After identifying the evaluated system, each threat is likely to arise and its potential impact is analyzed. By investigating and collecting such threats, and analyzing relevant historical information and data, a reliable basis is provided for their risk assessment. A set of threat occurrence probabilities $P:\{p_i|i=1,2,...,n\}$and its corresponding set of consequence attributes $V:\{v_{ij}|i=1,2,...,n;j=1,2,...,m\}$ are determined. Among them, p_i is the occurrence probability of the i-th threat in the set of threats, and $4v_{ij}$ is the possible impact value of the threat on the result attribute.

$$v_{ij}^{*}=\frac{v_{ij}}{\max_{k=1}^{n}\{v_{kj}\}} \quad (1)$$

$\max_{k=1}^{n}\{v_{kj}\}$ refers to the maximum impact value of the results generated by all threats in the threat set T of the result attribute.

3.3.Vulnerability identification

By using support vector machine algorithms to mine randomly sampled data, the sampling process can be seen as the process of classifying the sampled objects. Firstly, a set of small sample data is statistically analyzed and divided into two types, namely large interval and small interval. The optimization method is to classify discrete points on a plane, and the steps are:

$$y=\delta^{T}\eta(x)+h \quad (2)$$

In the formula, y represents the security data level of the network; δ is the weight vector for data security; T is the time required for data transmission; η is the relaxation factor; x is used to represent the dataset of network security; h is used to represent the deviation of data.

A network fault diagnosis method based on Fuzzy C-means (FCM) is used. In the process of clustering, the data of the whole network is discretization, so that the data of the network is distributed in each layer of the network in a random manner. This method first establishes a clustering center in the data layer, and then determines the clustering center based on its weight. By using this method, abnormal data can be trained and the trained data can be guaranteed to have extreme values.

Assuming the total number of samples is N and the number of clusters is c, then the value of c is [2~N-1]. On this basis, an optimal classifier based on fuzzy set theory is used.

$$Q=\sum_{C>1}^{C}\sum_{k>1}^{K}w_{ck} \quad (3)$$

Among them, Q is the membership relationship between the samples in the cluster and the cluster center; k is the cluster center; w is the extreme point of the target point.

On this basis, the spatial distance between each discrete point and the cluster center is calculated.

$$f(x)=\mathrm{sgn}(\alpha\cdot h\cdot\beta+\frac{\|x\|}{2\tau^2}) \quad (4)$$

f(x) is used to represent the geometric relationship between network security data samples and cluster centers; α is a support vector machine; β is a dot product operation; τ is the cross testing process.

In Equation (2), the variable δ represents the weight vector of specific vulnerability attributes (e.g. exploitability and impact), which determines the influence of a vulnerability on the overall service state. η is the relaxation factor used to balance the classification margin in SVM, reflecting the system's tolerance for noise in security logs. The deviation variable h quantifies the distance between the observed network behavior and the baseline security profile. In Equations (3) and (4), the threshold τ serves as a decision boundary; if the calculated vulnerability score exceeds τ, the service node is flagged for immediate mitigation to prevent potential service interruption.

This method achieves the goal of clustering abnormal data by clustering data with small spatial distance differences into a cluster. On this basis, abnormal data is clustered and the network security status is evaluated based on this. Before evaluation, evaluation indicators such as the threat level of the network after being attacked, the frequency of remote attacks, network importance, network bandwidth utilization and occupancy rate, and system host importance coefficient are selected. The various indicators that can be used to evaluate network security are combined and quantified, which can be expressed as:

$$\varepsilon=\frac{\varepsilon_1}{1+\varepsilon_1} \quad (5)$$

ε is a quantitative evaluation of the network security situation; ε_1 is the weight of the evaluation index.

On the basis of ε calculation, the network security of the system is classified. It is quantified. It is also divided into extremely high level, high level, medium level, low level, and extremely low level. After completing the above operations, the system network security situation is evaluated.

The system network is divided into multiple modules and the correlation coefficients between each module are calculated using the formula:

$$\mu=\frac{\min+\phi\cdot\max}{\Delta\phi} \quad (6)$$

μ is the correlation degree between the order of each block network; min is the minimum correlation coefficient; max is the maximum correlation coefficient; ϕ is the system network block.

The correlation coefficient defined in Equation (6) measures the statistical dependence between network traffic features and specific attack types. A higher coefficient indicates that a feature is a strong indicator of an ongoing attack (e.g. sudden spikes in SYN packets correlating with DoS attacks). In this framework, we set a heuristic threshold based on empirical testing; features with a correlation value above 0.6 are prioritized for the MRMR process. This selection logic ensures that the assessment focuses on the most 'informative' signals, thereby reducing the false alarm rate and providing more

reliable decision support for service operators.

After understanding the correlation between each network block, network sequences with low or unrelated correlation between each network block are identified and considered as security risks to the system.

3.4.MRMR filtering

Building upon the vulnerability categories identified, the MRMR method is applied to extract features that specifically characterize the exploitation of these vulnerabilities. This transition ensures that the feature selection process is not just a statistical exercise but is grounded in the platform's identified security weaknesses.

MRMR is a kind of feature subset that can filter out the best feature subset on the basis of mutual information. This feature subset is closely related to the object to be identified, and has nothing to do with other features. In information theory, mutual information is used to measure the correlation between features. Assuming two random variables x and y, their mutual trust I(X, Y) is determined by the following formula:

$$I(X,Y)= \iint P(x,y) \log \frac{P(x,y)}{P(x)P(y)} dxdy \quad (7)$$

The greater the correlation, the stronger the correlation between the selected features and the selected indicators, and the selected features can better reflect the information of the sample. Minimum redundancy refers to the low degree of correlation between selected features, which is a method of eliminating redundant features. The calculation formulas are as follows:

$$\max D(S,c), D=\frac{1}{|S|}\sum_{x_i \square S} I(x_i;c) \quad (8)$$

$$\min R(S), R=\frac{1}{|S|^2}\sum_{x_i,x_j \square S} I(x_i,x_j) \quad (9)$$

Among them, S is a subset of features; c is the classification vector; $I(x_i;c)$ represents the category vector of mutual information between feature x_i and class c. $I(x_i,x_j)$ represents the mutual information of features x_i and x_j.

$\varnothing(D,R)$ represents the maximum correlation and minimum redundancy between features, providing a basis for selecting feature subsets. Here is the formula:

$$\max \varnothing(D,R), \varnothing = D-R \quad (10)$$

3.5.Attack graph generation

This article adopted an information security evaluation method based on attack graphs, which calculated the situational value of the path to identify the optimal path for the attacker and infer their attack intention.

In general, the security elements used in attack graphs include vulnerabilities, network services, network connections, network configuration, access permissions, etc. Therefore, before constructing a network attack graph, it is necessary to collect the topology of the network and system vulnerabilities. These elements constitute the initial network state and represent all the resources currently available for conducting attacks. These security elements are the input parts of the attack graph generation system. In order to facilitate system recognition and operation, these input files are written in Datalog language, usually represented by tuples.

The attack graph consists of fact nodes, inference process nodes, and their boundaries. A fact node is associated with one or more inference process nodes, while another or more inference process nodes are associated with one or more fact nodes. MulVAL (multihost, multistage, vulnerability analysis) is based on logical reasoning and can effectively describe an attack behavior by describing the attack graph. It matches existing security elements with inference rules to infer specific attacks that may occur. When all existing security elements meet the prerequisites of a certain inference rule, an attack would occur and bring corresponding results to the next attack.

The platform adopts a forward generation method based on “depth first”. If all prerequisites are met, the attack is effective, generating corresponding boundary values, and obtaining corresponding results. The completed result would be used as a new security element to participate in the next round of pairing until the attack goal is achieved.

3.6.Network Risk Assessment Model

HMM (Hidden Markov Model) is a model based on two stochastic processes. It is a model based on two stochastic processes: one is a model based on Markov chain, which is used to describe the order of states in the model; another method is to study the relationship between state variables and observed variables. It is assumed that the state of the host system is a Markov chain, and the vulnerability and nuclear attack used in the attack are observable variables. Based on the product of attack probability and success probability, the situation value of the final route is determined. HMM is adopted due to its superior ability to model stochastic processes where the internal state of the host is hidden, but the sequence of attack observations (vulnerabilities exploited) is visible.

The state set is $S=\{S_1,S_2,...,S_n\}$. S represents the attack graph state nodes, and n represents the number of node state nodes in the attack graph. The observation set is $V=\{V_1,V_2,...,V_m\}$. V represents the hole or atom being attacked, and m represents the number of attacks. The status sequence is $Z=\{Z_1,Z_2,...,Z_t\}$, and Z_t represents the node status of the host at the specified time t. The observation sequence is $O=\{O_1,O_2,...,O_t\}$. O_t represents a vulnerability that has been used or an atomic attack that has been used at time t. The initial state distribution is $\pi=\{\pi_i\}$, which represents the possibility of a state of S_i at the beginning.

On the basis of situational values, the probability of attack route occurrence is calculated using situational values. In hybrid models, the Viterbi algorithm is widely used to calculate the maximum possible state path. The basic idea is that in HMM, there are N states that correspond to a specific observable state. Since the state transition sequence is not visible, each iteration must consider N states and select the most likely one from N states to obtain the optimal solution. However, for each attack route, each observation route has a unique state transition sequence, that is, each attack route has and only has a unique host state transition sequence.

Because one need to use the results of the previous moment to calculate the probability of each moment, when the state of the t-th moment is S_i, the probability of observing $O_1,O_2,...,O_t$ in the previous moment is defined as:

$$\beta_i(t)=P(O_1,O_2,...,O_t,Z_t=S_i|\lambda) \quad (11)$$

The probability of all states appearing at the first moment has been determined, which is not only related to the state itself, but also to the observed state. Because this depends on the first element in the offensive route, there is only one condition:

$$\beta_1(i)=\pi(i)b_{io_1} \quad (12)$$

$\pi(i)$ represents the probability of starting time $Z_1=S_i$, and b_{io_1} represents the probability of observing o_1 under S_i.

Assuming that it is t at this time and the state is S_k, through iterative calculation, using the results at time t-1, the result can be obtained:

$$\beta_t(k)=\beta_{t-1}(i)\times\beta_{ik}\times b_{ko_t} \quad (13)$$

$\beta_{t-1}(i)$ is the probability of a state of S_i at t-1. Multiplying the transition probability from S_i to S_k is the probability of a state of S_k at t, and then multiplying it with the corresponding observation probability is the obtained result.

By using this method, high-risk vulnerabilities and their corresponding state transition sequences can be obtained, thus developing the best defense strategy.

4. Results and Discussion

4.1.Experimental design and evaluation

The most common feature selection method is to select n variables closely related to categorical variable. These feature variables can have strong correlations, and these features are redundant features. In this way, the optimal feature subset needs to eliminate both irrelevant and redundant feature subsets. Therefore, MRMR is selected because it maximizes the correlation between features and target variables while minimizing redundancy, which is critical for handling the high-dimensional data typical of AI platforms.

In this study, the KDD 99 dataset is utilized for experimental validation. While it remains a widely recognized benchmark in network intrusion research, the authors acknowledge its limitations, particularly regarding its age and the absence of some modern sophisticated attack patterns. However, it is selected here because of its comprehensive labeling of R2U and U2R attacks, which provide a foundational baseline for testing the feature extraction stability of the MRMR method in handling imbalanced data.

To address the extreme class imbalance in the original KDD 99 dataset, where R2U and U2R attacks are significantly underrepresented, we employed the Synthetic Minority Over-sampling Technique (SMOTE). Specifically, we increased the sample size of R2U and U2R categories by generating synthetic examples based on the feature space of existing instances, ensuring a more balanced training environment for the SVM classifier. This transparency in data synthesis allows for a more rigorous evaluation of the MRMR method's ability to extract features from rare but critical service-disrupting attacks.

Using the KDD 99 (Thrid International Knowledge Discovery and Data Mining Tools Competition) dataset, the KDD 99 data types are shown in Table 1:

Table 1. KDD 99 data types

Category	Full name	Meaning	Proportion (%)
Normal	-	Normal data	19.96
DoS	Disk Operating System	Illegal attempt to interrupt or interfere with the normal operation of the host or network	78.84
Probe	-	Illegal scanning of hosts or networks, searching for vulnerabilities, searching for system configurations or network topology	0.79
R2U	Remote to User	Remote unauthorized users illegally obtain user privileges on the local host	0.28
U2R	User to Root	Local unauthorized users illegally obtain privileges of local super users or administrators	0.13

Due to the small number of R2U and U2R samples, it is difficult to configure classifiers for these attack classes, resulting in a high error detection rate. Most of these two categories were classified as normal data. In order to provide basic classification characteristics, this article added the number of R2U and U2R classes to the dataset.

4.2.Feature Selection Performance

The feature selection methods of CFS (Correlation Feature Selection), Pearson correlation coefficient, and MRMR were selected. The experimental results of the accuracy of the feature selection methods are shown in Figure 2.

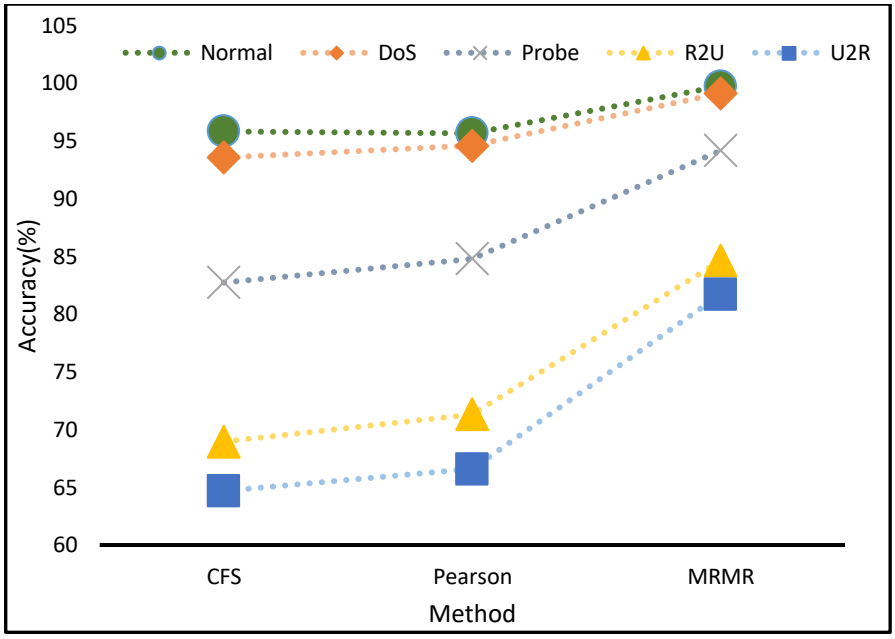

Fig. 2: Accuracy of feature selection method.

The CFS and Pearson correlation coefficient feature selection methods had higher accuracy in selecting normal and DoS data. The accuracy of selecting Probe type data was second, but the accuracy was above 80%. The selection accuracy of R2U and U2R data was poor. Although the proportion of R2U and U2R data has been increased during the experiment, the selection accuracy was still below 75%. Compared with the selection results of CFS and Pearson correlation coefficient feature selection methods, the MRMR feature selection method had significantly better selection results, with improved selection accuracy for all five types of data. The selection accuracy for Normal and DoS types was above 99%; the selection accuracy for Probe type data was above 90%; the selection accuracy for R2U and U2R type data was above 80%; the overall accuracy of data feature selection was above 80%, which had good data feature selection and classification performance.

The significant improvement in accuracy shown in Figure 2, particularly for R2U and U2R attacks, has direct implications for service-level security governance. In a high-concurrency AI service environment, traditional methods often produce 'noise' that masks subtle unauthorized access attempts. By achieving higher precision through MRMR, the system reduces the risk of 'stealthy' privilege escalation, which is often the precursor to data exfiltration. For service managers, this means a more reliable automated response system that only triggers high-latency deep inspections when high-probability threats are detected, thus optimizing the balance between security and service responsiveness.

The stability experimental results of the feature selection method are shown in Table 2.

Table 2. Stability of feature selection methods

Category	CFS	Pearson	MRMR
Accuracy (%)	91.36	92.48	99.07
Misreporting rate (%)	0.46	0.57	0.19
False alarm rate (%)	0.21	0.26	0.11
Modeling time (s)	9.28	8.76	11.54

The MRMR feature selection method had higher accuracy in feature selection, lower false positives and false positives in feature selection. Although MRMR had a higher modeling time, compared to the data, it can be seen that the MRMR feature selection method was more stable and the feature selection effect was better.

4.3.Risk assessment experiment

The practical applicability of this framework extends beyond static datasets to modern informatics-enabled service infrastructures. For instance, in the campus network experiment, the four servers (NS, MS, FTP, and DS) represent essential nodes in an educational service system. By applying this AI-driven risk assessment, administrators can ensure high service availability and protect sensitive user data within contemporary cloud-based AI services or logistics information systems. Such predictive maintenance of security prevents service disruptions and maintains the quality of service (QoS) in large-scale digital environments. The network security risk assessment results are shown in Figure 3.

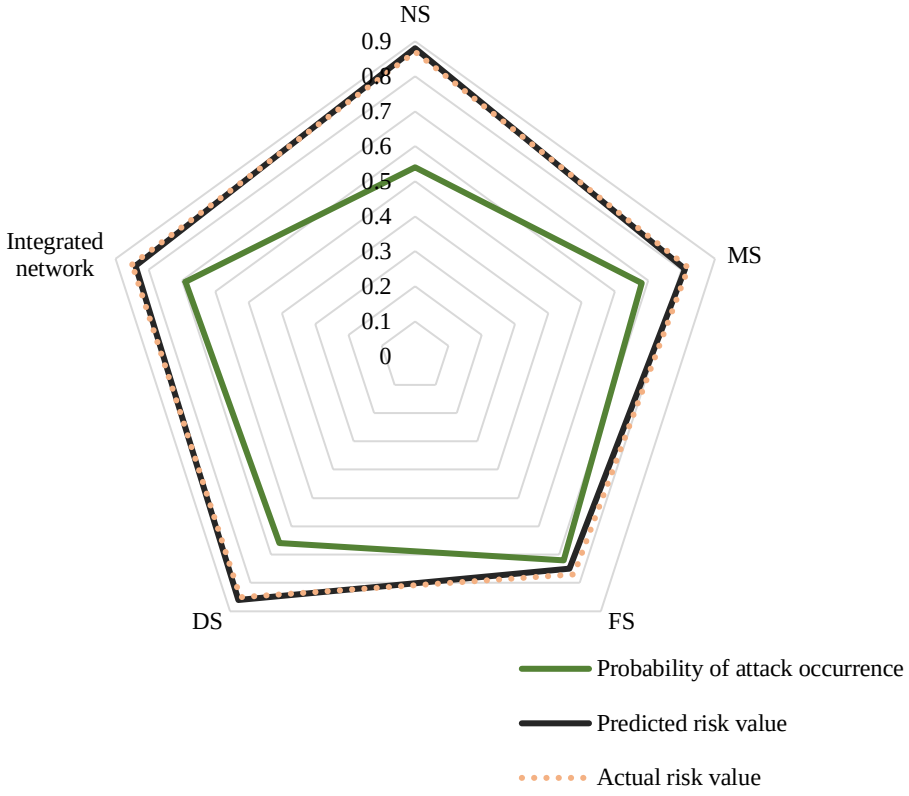

Fig. 3: Network security risk assessment results.

Figure 3 evaluated the probability of attacks on four servers in the campus network and the overall probability of attacks on the campus network, predicted the risk value of the campus network, and compared it with the actual risk value. The experimental results showed that the predicted network risk value using this network risk assessment method is relatively close to the actual network risk value, and can effectively predict network risk, manage network risk in a timely manner, and maintain network security.

The CFS and Pearson correlation coefficient feature selection methods had higher accuracy in selecting normal and DoS data. The accuracy of selecting Probe type data was second, but the accuracy was above 80%. The selection accuracy of R2U and U2R data was poor. Although the proportion of R2U and U2R data has been increased during the experiment, the selection accuracy was still below 75%. Compared with the selection results of CFS and Pearson correlation coefficient feature selection methods, the MRMR feature selection method had significantly better selection results, with improved selection accuracy for all five types of data. The selection accuracy for Normal and DoS types was above 99%; the selection accuracy for Probe type data was above 90%; the selection accuracy for R2U and U2R type data was above 80%; the overall accuracy of data feature selection was above 80%, which had good data feature selection and classification performance.

4.4.Discussion

The proposed network security risk assessment framework integrates MRMR feature extraction, attack graph modeling, and Hidden Markov Model-based dynamic prediction to address the limitations of conventional security evaluation methods applied to artificial intelligence service platforms. The experimental results demonstrate that the MRMR method achieves superior feature selection accuracy across all attack categories, particularly for R2U and U2R attacks, which are frequently misclassified by traditional correlation-based techniques. This improvement stems from the method's capacity to simultaneously maximize relevance to target variables while minimizing inter-feature redundancy, thereby preserving discriminative information critical for identifying low-frequency yet high-impact threats in AI-driven service environments. The integration of SVM and FCM for vulnerability identification establishes a structured foundation for subsequent attack path modeling, ensuring that the extracted features are contextually aligned with the platform's security posture. The attack graph generation mechanism enables comprehensive visualization of multi-stage attack paths, capturing the logical dependencies among vulnerabilities and service assets. The application of HMM to these attack graphs facilitates probabilistic inference of attacker behavior, allowing for the quantification of dynamic risk based on observable intrusion indicators. The close alignment between predicted and actual risk values in the experimental validation confirms the framework's effectiveness in capturing the stochastic nature of network attacks and their progression across service nodes. The high accuracy and stability of the feature extraction process reduce the incidence of false positives and false negatives, enhancing the reliability of risk alerts and supporting timely security interventions. The framework's ability to model the cascading effects of vulnerabilities across interconnected system components reflects the complex interdependencies inherent in AI service platforms, where security incidents affecting one node may propagate and degrade overall service continuity. By embedding risk assessment within a structured pipeline that spans data preprocessing, asset identification, threat modeling, and dynamic calculation, the proposed approach provides a cohesive mechanism for translating low-level network events into actionable insights for service-level security governance. The methodological consistency across these stages ensures that each component contributes to a unified representation of risk, enabling system administrators to prioritize defense strategies based on the predicted impact on service availability and data integrity. The framework advances network security management for artificial intelligence platforms by shifting the focus from isolated intrusion detection toward integrated risk prediction, thereby reinforcing the resilience of informatics-enabled service systems against evolving cyber threats.

5. Conclusions

The proposed AI-based assessment framework provides a strategic tool for service governance in artificial intelligence environments. By enabling precise risk prediction and minimizing potential service disruptions, the framework supports more effective decision-support systems for service operators. This study contributes to the field of informatics-enabled services by ensuring that the reliability and quality of AI-driven services are not compromised by evolving cyber threats, thereby fostering a more resilient service ecosystem. Beyond the technical validation, this study offers several practical implications for informatics-enabled service management:

Proactive Service Governance: The transition from static detection to dynamic risk prediction allows service providers to shift from a "reactive" defense posture to "proactive" governance. By anticipating attack paths, organizations can implement preemptive measures that ensure service continuity and minimize the economic impact of potential downtime.

Resource Optimization in Service Systems: The high efficiency of the MRMR method demonstrates that security managers can achieve superior protection with lower computational overhead. This is particularly vital for real-time AI services where low latency and high resource availability are critical for Service Quality (QoS).

Standardization of AI Service Security: The proposed four-stage pipeline provides a reproducible

blueprint for auditing the security of diverse AI-enabled infrastructures, from campus networks to cloud-based logistics systems.

Future Research will focus on extending this framework to edge computing environments and exploring the integration of automated response mechanisms to further enhance the resilience of informatics-enabled services.

Funding

This work was supported by PhD Innovation Research Fund Project of Jining Normal University (Numberjsbsjj2413). Intelligent Recognition and Image Processing Research Center (Number: jskypt2436)

References

Awan, M. & Alam, A. (2025). Cybersecurity threats and defensive strategies for small and medium firms: a systematic mapping study. *Administrative Sciences*, 15.12, 481-488.

Becue, A. Isabel, P. & Joao, G. (2021). Artificial intelligence, cyber-threats and Industry 4.0: Challenges and opportunities. *Artificial Intelligence Review*, 54.5, 3849-3886.

Cheimonidis, P. & Rantos, K. (2023). Dynamic risk assessment in cybersecurity: A systematic literature review. *Future Internet,* 15.10, 324-331.

Duxbury, S. W. & Dana, L. H. (2019). Criminal network security: An agent-based approach to evaluating network resilience. *Criminology*, 57.2, 314-342.

Feng, W. Wu, Y. & Fan, Y. (2020). A new method for the prediction of network security situations based on recurrent neural network with gated recurrent unit. *International Journal of Intelligent Computing and Cybernetics*, 13.1, 25-39.

Ganin, A. A. Quach, P. Panwar, M. Collier, Z. A. Keisler, J. M. Marchese, D. & Linkov, I. (2020). Multicriteria decision framework for cybersecurity risk assessment and management. *Risk Analysis*, 40.1, 183-199.

Islam, S. Basheer, N. Papastergiou, S. Ciampi, M. & Silvestri, S. (2025). Intelligent dynamic cybersecurity risk management framework with explainability and interpretability of AI models for enhancing security and resilience of digital infrastructure. *Journal of Reliable Intelligent Environments*, 11.3, 12-18.

Jeong, H. J. Sharma, P. K. Sicato, J. C. S. & Park, J. H. (2019). Emerging technologies for sustainable smart city network security: Issues, challenges, and countermeasures. *Journal of Information Processing Systems*, 15.4, 765-784.

Jimmy, F. (2021). Emerging threats: The latest cybersecurity risks and the role of artificial intelligence in enhancing cybersecurity defences. Valley International *Journal Digital Library*, 1.2, 564-574.

Kandasamy, K. Srinivas, S. Achuthan, K. & Rangan, V. P. (2020). IoT cyber risk: A holistic analysis of cyber risk assessment frameworks, risk vectors, and risk ranking process. *EURASIP Journal on Information Security*, 2020.1, 8-15.

Kou, G. Shuo, W. & Tang, G. (2019). Research on key technologies of network security situational awareness for attack tracking prediction. *Chinese Journal of Electronics*, 28.1, 162-171.

Li, W. Meng, W. Liu, Z. & Au, M. (2020). Towards blockchain-based software-defined networking: security challenges and solutions. *IEICE Transactions on Information and Systems*, 103.2, 196-203.

Li, Y. Huang, G. Wang, C. & Li, Y. (2019). Analysis framework of network security situational awareness and comparison of implementation methods. *EURASIP Journal on Wireless Communications and Networking*, 1 (2019), 1-32.

Liu, Y. Li, Z. & Feng, L. (2020). Design of multimedia education network security and intrusion detection system. *Multimedia Tools and Applications*, 79.12, 18801-18814.

Lyu, Y. Feng, Y. & Sakurai, K. (2023). A survey on feature selection techniques based on filtering methods for cyber attack detection. *Information*, 14.3, 191-198.

Maia, A. Boutouchent, A. Kardjadja, Y. Gherari, M. Soyak, E. G. Saqib, M. et al. (2024). A survey on integrated computing, caching, and communication in the cloud-to-edge continuum. *Computer Communications*, 219.1, 128-152.

Mallampati, S. B. & Hari, S. (2023). Fusion of feature ranking methods for an effective intrusion detection system. *Computers, Materials & Continua*, 76.2, 1721-1745.

Naeem, R. Kohtamäki, M. & Parida, V. (2025). Artificial intelligence enabled product–service innovation: past achievements and future directions: R. Naeem et al. *Review of Managerial Science*, 19.7, 2149-2192.

Ndibe, O. S. (2025). AI-driven forensic systems for real-time anomaly detection and threat mitigation in cybersecurity infrastructures. *International Journal of Research Publication and Reviews*, 6.5, 389-411.

Sefati, S. S. Arasteh, B. Halunga, S. & Fratu, O. (2025). A comprehensive survey of cybersecurity techniques based on quality of service (QoS) on the Internet of Things (IoT). *Cluster Computing*, 28.12, 792-799.

Sun, N. Li, T. Song, G. & Xia, H. (2021). Network security technology of intelligent information terminal based on mobile internet of things. *Mobile Information Systems*, 2021.25, 1-9.

Swetha, T. Kumaran, U. Meena, V. P. & Hameed, I. A. (2025). Leveraging AI for enhanced cybersecurity: a comprehensive review. *Discover Applied Sciences*, 7.6, 584-591.

Theodoropoulos, T. Rosa, L. Benzaid, C. Gray, P. Marin, E. Makris, A. et al. (2023). Security in cloud-native services: A survey. *Journal of Cybersecurity and Privacy*, 3.4, 758-793.

Wang, Y. Ma, J. Sharma, A. Singh, P. K. Gaba, G. S. Masud, M. et al. (2021). An exhaustive research on the application of intrusion detection technology in computer network security in sensor networks. *Journal of Sensors*, 29 (2021), 1-11.

Waqas, M. Tu, S. Halim, Z. Rehman, S. U. Abbas, G. & Abbas, Z. H. (2022). The role of artificial intelligence and machine learning in wireless networks security: Principle, practice and challenges. *Artificial Intelligence Review*, 55.7, 5215-5261.

Wheelus, C. & Zhu, X. (2020). IoT network security: Threats, risks, and a data-driven defense framework. *IoT*, 1.2, 259-285.

Yang, B. & Yang, M. (2021). Data-driven network layer security detection model and simulation for the Internet of Things based on an artificial immune system. *Neural Computing and Applications*, 33.2, 655-666.

Zhang, Z. & Jing, J. Choo, X. W. Kim-Kwang, R. Gupta, B. B. (2020). A crowdsourcing method for online social networks security assessment based on human-centric computing. *Human-centric Computing and Information Sciences*, 10.23, 1-19.

Zheng, G. Be,i G. & Yu, Z. (2021). "Dynamic network security mechanism based on trust management in wireless sensor networks. *Wireless Communications and Mobile Computing*, 2021.28, 1-10.

Zhu, X. (2021). Self-organized network management and computing of intelligent solutions to information security. *Journal of Organizational and End User Computing (JOEUC)*, 33.6, 1-16.