Information Security and Risk Control Model Based on Plan-Do-Check-Action for Digital Libraries

Yimin Yin

Library & Archives, North Sichuan Medical College, Nanchong, 637000, China
E-mail: yinyimin@nsmc.edu.cn

Received 22 October 2023; Accepted 08 November 2023;
Publication 13 February 2024

Abstract

To continuously and effectively control the potential risk factors of digital libraries, further improve the quality of digital library usage, ensure the smooth operation and information security of digital libraries, a security and risk factor control model for digital libraries based on the Plan-Do-Check-Action is constructed. A digital library security and risk identification model based on improved support vector machines is constructed. Based on the obtained risk identification results, a Plan-Do-Check-Action management model is implemented for information security in digital libraries. The R2L dataset in the network attack dataset is used for training and testing model performance. Then the proposed method is applied to two different digital libraries. The average accuracy of the risk identification model proposed in the research is 96.46%. The risk error of each level is maintained within 0.4%, which is within an acceptable range. The proposed method has a high recognition effect. It can be used to identify common security and risk issues encountered in digital libraries, and improve the information security performance.

Keywords: PDCA, digital library, information security, risk control model, SVM, PSO.

Journal of Cyber Security and Mobility, Vol. 13_2, 305–326.
doi: 10.13052/jcsm2245-1439.1326

1 Introduction

Internet technology has effectively promoted the transformation of various resources towards digitization. Digital libraries (DL) are one of the important derivative development fields. The digital library has effectively achieved resource sharing. It can achieve the maximum utilization of limited resources, while breaking the limitations of traditional libraries in time and location, and enhancing the core competitiveness of libraries [1, 2]. However, as libraries move towards digitization, many security issues and risk factors inevitably arise, such as virus infections, illegal intrusions, human error operations, and software and hardware failures. The existing methods identify potential risks that may exist in DL through each independent detection when controlling their risk factors. They cannot continuously identify and control potential risks in DL. The efficiency and quality of risk control need to be improved [3]. In existing research on DL, most of them focus on the service efficiency and usage of digital resources. The overall continuity management of DL and the risks and security issues they face are rarely mentioned. Therefore, to continuously and effectively control the potential risk factors of DL, reduce security failures and risk factors, and timely solve existing security problems, the security and risk factors need to be controlled and identified. Among them, the Plan Do Check Action cycle management model (PDCA) is an effective tool that can be used for planning, implementing, verifying, and improving measures. Both high-level strategic processes and simple homework activities can effectively play a role. Therefore, the PDCA cycle management model has been widely applied in many fields [4, 5]. In the research, PDCA is applied to information security and risk control management in DL. However, to better leverage the PDCA model, the first step is to construct an improved support vector machine based digital library security and risk control identification model. On the basis of identifying the risk issues faced by DL, a management method based on the PDCA cycle model is constructed. It is expected to achieve effective identification and control of digital library security and risk factors, ensure higher security, and improve service quality.

The contributions of the research are as follows. Firstly, the study analyses the potential risk factors faced by DL and constructs information security and risk indicators for DL. Secondly, a potential risk control management model for DL based on PDCA cycle management mode is constructed, achieving continuous identification and control of potential risk factors in DL.

The research consists of four parts. The first section summarizes the research on PDCA cycle management mode and DL. The second section first constructs a digital library risk identification model based on improved SVM. On this basis, a digital library security and risk factor control management model based on the PDCA cycle is constructed. The third section conducts experimental analysis on the performance of the digital library risk identification model and the PDCA risk management model. The fourth part summarizes the research.

2 Related Works

2.1 Overview of PDCA Cycle Management Methods

The PDCA cycle management model has been widely applied in areas such as product quality, information security, human resources, and healthcare. Numerous scholars have conducted in-depth research on it. Xue explored the application value of PDCA model in the safety management of radiology nursing. A retrospective analysis is conducted on 320 patients. Then the quality of patient care, management scores, nursing satisfaction, and incidence of adverse events are recorded and compared. The results show that the application of PDCA management in radiation department nursing safety management can effectively improve nursing quality. It indicates that PDCA management has great application value in future radiology management [6]. For this study, the PDCA cycle management model is applied to the management of patient clinical care to improve nursing outcomes. Sally et al. used the PDCA management method to manage the mental health of the population in a certain area. Researchers and community advisory committees have created processes and charts to reveal individual, environmental, political, and procedural factors that contribute to poor mental health outcomes. This experiment successfully demonstrates how to use the PDCA cycle to reveal the potential causes of adverse health outcomes [7]. For this study, the author applies the PDCA circular management model to the health management of the regional population, which effectively revealed potential risk factors in the population. Jie et al. applied PDCA to the management of hyperglycemic patients. The blood glucose levels of 1003 critically ill patients are used as experimental subjects for PDAC cycle management. According to the PDCA management method, solutions to potential causes are regulated and implemented. The results show that the PDCA can develop a reasonable insulin infusion plan for critically ill patients and reduce the incidence of hyperglycemia [8]. For this study, the author applies the PDCA cycle management model to the clinical care of patients with hyperglycemia, reducing the risk of disease occurrence. Amaral Vitória P et al. implemented internal logistics upgrades in the automotive industry based on the PDCA cycle management method. According to the PDCA cycle stage, the process is analyzed. This circular management system can integrate inventory management systems and improve quality management [9]. Antoaneta et al. proposed a food safety management model based on the PDCA cycle. After analyzing various aspects of food safety management, a model for risk analysis and assessment at the operational and organizational levels is proposed. The results show that the model can effectively meet the regulatory requirements of food safety management [10]. In the above research, scholars apply the PDCA circular management model to the disease care, public health, logistics, and food regulation, respectively, to manage and control potential risks faced by relevant objects, proving the feasibility of using the PDCA management model for risk control in this study.

2.2 Overview of Security Management in Digital Libraries

With the development of information technology, various digital resources have been widely used. The information and resources based on DL have effectively expanded the storage capacity of library resources and provided richer resources. Many scholars have conducted research on it. Elizarov et al. discussed the digital technology development in scientific activities. The concept of metadata factory in digital science libraries is introduced as an element of the scientific digital platform ecosystem. This method solves multiple tasks related to the construction of Lobachevsky-DML metadata factory and the formation of electronic collections [11]. For this study, the author discusses the role of digital technology in the construction of DL. Kaba A et al. tested the return on investment (ROI) of academic libraries in Arab countries. This method can effectively measure the ROI of academic libraries. According to the findings, it provides support for the effective utilization of digital library resources in Arab countries [12]. For this study, the author explores the effective utilization value of digital library resources. Azam N et al. evaluated the usability and information architecture of the Iranian National DL website. An exploratory hybrid method is adopted to collect relevant data. The results show that only 30% of initiating users and 10% of expert users are satisfied with the website. With the increasing development of information technology, the importance of this research for better service provision is increasing [13]. For this study, the resource collection methods in digital libraries have been analyzed. Samaneh et al. explored relevant literature on personalized digital library services. The Systematic review method is applied to obtain different types of personalized indicators in the library context. A total of 103 indicators are extracted for different types of personalization. Among these indicators, 90 are considered important by experts. This result is effective for developing personalized digital library services [14]. For this study, the author discusses the service characteristics of DL. Xie I et al. investigated the cognitive differences among key stakeholders in evaluating the accessibility and usability guidelines for digital libraries (DLAUG). BVI users and developers have significant differences in the relevance, clarity, and usefulness of DLAUG [15]. In the above research, the author conducted in-depth discussions on the construction, application, and service characteristics of digital libraries. Few studies have conducted sufficient research on risk control in DL.

In summary, the resources and information of DL have been widely developed and utilized. Many scholars have also studied from various perspectives. However, from the above research, most of the existing research on DL is focused on the service efficiency and usage of digital resources. There is little mention for the overall management of DL and the risks and security issues. It is necessary to study the information security and risk control of DL. Therefore, a digital library risk identification model based on improved SVM is first constructed. Based on the advantages of the PDCA cycle management model, a management model for DL information security and risk control based on PDCA is constructed. It is expected that this method can effectively identify various potential security issues and risk factors faced by digital libraries. The advantages of the PDCA cycle management model can be better utilized to reduce information security and risk issues in DL.

3 Information Security and Risk Control Model Construction for DL Based on PDCA

The potential risk factors of information security and risk identification faced by libraries are analysed. Then, an information security and risk identification model based on the improved SVM model is constructed. Finally, a DL information security and risk control model based on PDCA is constructed.

3.1 Risk Factors Analysis in Digital Libraries

The risk factors of libraries are the threats it faces and various security incidents caused by their vulnerability. The risk elements of information security in DL include assets, threats, and vulnerabilities. Among them, assets are all the resources possessed by DL that can bring social and economic benefits to DL, including confidentiality, availability, and integrity [16]. The influencing factors of assets are shown in formula (1).

$$A = A(A_I, A_C, A_U, W) \tag{1}$$

In formula (1), A_I stands for the confidentiality of the asset. A_C stands for the integrity of the asset. A_U represents the asset value in availability. W represents the importance of confidentiality, integrity, and availability to asset value. The threat mainly refers to various system damages suffered by DL. The impact factors of the threat are shown in formula (2).

$$T = T(T_M, T_A) = T(T_M, L(T_I, T_C, T_U, W)) \tag{2}$$

In formula (2), T_M is the frequency of threat occurrence. T_A represents the loss caused by the threat to the asset. T_I, T_C and T_U respectively represent the losses caused to the confidentiality, integrity, and availability of assets after the threat occurs. Vulnerabilities are easily exploited by threats in DL. The influencing factors are extensive, including the organizational management structure, personnel, procedures, and defects in the assets of DL [17]. The influencing factors of vulnerability are shown in formula (3).

$$V = V(V_I, V_Q, V_T) \tag{3}$$

In formula (3), V_I and s represent the vulnerability assignments of DL obtained through background checks and tool scans, respectively. V_Q represents the vulnerability assignment obtained from statistical analysis. Table 1 depicts the information security and risk indicators of the constructed DL [18].

From Table 1, DL faces numerous security and risk factors. Although they are generally divided into assets, threats, and vulnerabilities, there is a close connection between the three. If there is a problem in one of the links, it will have a direct or indirect impact on other factors. In addition, each secondary indicator contains multiple tertiary indicators, namely the specific risks and security issues faced by DL [19]. The Analytic Hierarchy Process is applied to calculate the weight of indicators. The specific calculation process is as follows. Firstly, the corresponding judgment matrix is constructed. $M_i(i = 1, 2,$

Table 1 Common risk factor indicators for digital libraries

Primary Indicator	Secondary Indicators	Indicator Code
Property	Electronic resources	X_1
	Data document	X_2
	Physical assets	X_3
	Software assets	X_4
	Service	X_5
	Various maintenance personnel	X_6
Threatens	System	X_7
	Environment	X_8
	Nature	X_9
	External personnel	X_{10}
Vulnerability	Physical environment	X_{11}
	Network structure	X_{12}
	Systems software	X_{13}
	Database software	X_{14}
	Application Middleware	X_{15}
	Application system	X_{16}
	Technical management	X_{17}
	Organizational management	X_{18}

$3, \ldots, n)$ represents a judgment matrix of n element. The judgment matrix obtained from the indicator scoring results is shown in formula (4).

$$M = \begin{bmatrix} m_{11} & \ldots & m_{1j} \\ \ldots & \ldots & \ldots \\ m_{i1} & \ldots & m_{ij} \end{bmatrix} \tag{4}$$

In formula (4), $m_{ij} > 1$, $m_{ij} = 1/m_{ji}(i \neq j)$, and $m_{ij} = 1(i = j)$. Then the corresponding weight vector is calculated. The maximum characteristic value is represented as $\lambda_{\max}$. The product of each row element is R_i. The n-th root of R_i is w_i. After normalization, w_i is shown in formula (5).

$$\begin{aligned} w_i &= \frac{w_i'}{\sum_{j=1}^{n} w_i'}, \quad i = 1, 2, 3, \ldots, n \\ w_i' &= \sqrt{\prod_{j=1}^{n} m_{ij}}, \quad i = 1, 2, 3, \ldots, n \end{aligned} \tag{5}$$

The obtained weight vector is depicted in formula (6).

$$w = (w_1, w_2, w_3, \ldots, w_n) \tag{6}$$

The obtained weight vector can be used to calculate the weight value of each risk factor. Therefore, the likelihood of the risk factor occurring is judged.

3.2 Construction of Digital Library Risk Identification Model Based on Improved SVM

Based on the risk factors analysis in DL mentioned above, to better implement the PDCA management model for library information security and risk control, a risk identification model for DL is constructed. Support Vector Machine (SVM) has unique advantages in generalized linear classification for data [20]. The classification goal is to find a Hyperplane that can segment samples to the maximum extent by learning samples. The basic working principle is shown in Figure 1.

Figure 1 presents two sample points. The purple triangle stands for a positive sample. The red triangle stands for a negative sample. SVM expects to find a straight line. w represents the slope of the line. b represents the intercept of the line, which maximizes the distance d between the two support vectors and the line. The larger the distance d, the smaller the impact of sample disturbance on classification accuracy. The generalization ability is also stronger. In practical applications of support vector machine data classification, data samples with completely linear separability are rarely found. For the vast majority of samples that cannot be completely linearly separable, the implementation process is shown in Figure 2.

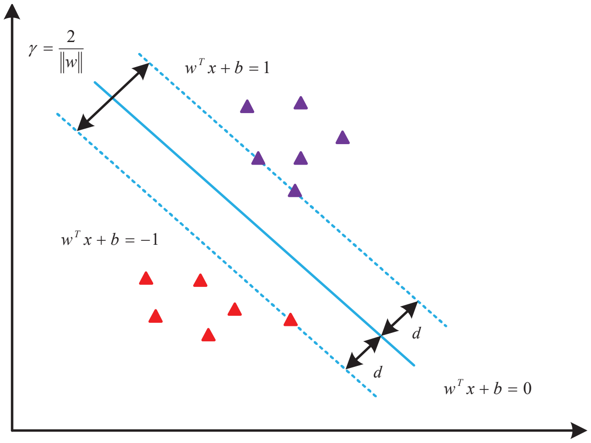


Figure 1 Basic principle of SVM.

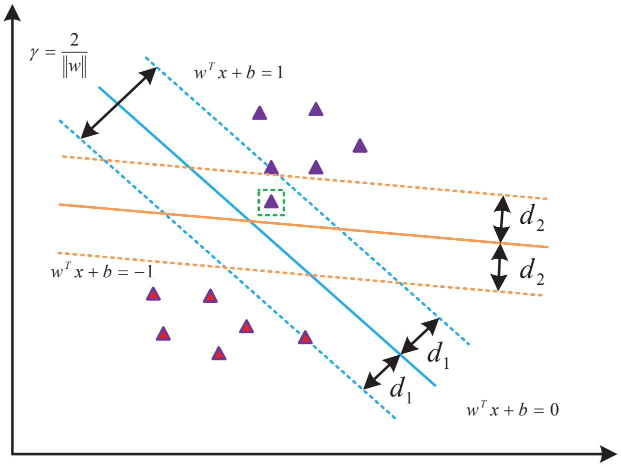


Figure 2 The classification process of nonlinear samples.

The split Hyperplane is expressed as (7).

$$f(x) = w\Phi(x) + b \tag{7}$$

In formula (7), x represents the sample point. $\Phi(x)$ stands for the vector after x is mapped to the new feature space. The original solving equation can be transformed into formula (8).

$$\min \frac{1}{2}\|w\|^2 \quad s.t. \quad y_i(w^T\Phi(x_i) + b) \geq 1 \tag{8}$$

In the existing experience of applying SVM to information security and risk control, the kernel functions of SVM are mostly designed based on human experience. As a result, SVM has problems such as low parameter accuracy and low accuracy in risk identification. Meanwhile, the correlation between manually set parameters and sample data features is weak, making it difficult to better learn the features of the sample data. Therefore, in response to the above issues, the PSO is introduced to improve SVM. PSO is one of the most popular metaheuristic algorithms, derived from simulating the evolution of bird flight and foraging processes [21]. In the standard PSO process, the speed of each particle is shown in formula (9).

$$v_i(t+1) = w \cdot v_i(t) + c_1r_1(p_i(t) - x_i(t)) + c_2r_2(p_g - x_i(t)) \tag{9}$$

In formula (9), w stands for the inertia weight. c_1 and c_2 represent learning factors. r_1 and r_2 are random numbers within $[0, 1]$. $x_i(t)$ represents the position of the i-th particle during the t-th iteration. $v_i(t)$ represents the velocity at this time. $p_i(t)$ represents the individual optimal value at this time. p_g represents the global optimal value. Formula (10) is applied to update the position.

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{10}$$

In formula (10), $x_i(t)$ represents the position of the i-th particle during the t-th iteration. F stands for the latest velocity. In response to the insufficient search ability of PSO algorithm, genetic algorithm (GA) is used to optimize it. The particle update module is introduced into the standard PSO. The essence is to use the crossover and mutation operations in the GA to locally update the particle positions. Then, the fitness of the particles is recalculated. The fitness values before and after are compared. Better fitness values are used for module updates on particles. The process of improved PSO algorithm based on GA is shown in Figure 3.

After improving PSO using GA algorithm, the search and development capabilities of PSO algorithm can be effectively improved. The risk detection and identification framework of SVM optimization model based on improved PSO algorithm is shown in Figure 4.

In Figure 4, the SVM algorithm optimized by the PSO algorithm is used as the data training subject. The data consists of training set and a testing set. The training dataset used in the study is Remote to Local (R2L), which involves unauthorized remote users illegally obtaining user privileges on the local host. The R2L dataset comes from the DARPA dataset, which is the standard dataset in the field of network intrusion detection to date. This dataset includes three datasets, DARPA 1998, DARPA 1999, and DARPA 2000.

3.3 The Implementation Process of PDCA Security Management in DL

After the construction of the SVM risk detection and identification algorithm based on the improved PSO algorithm, the study combines the actual situation of digital library information security and risk control. The PDCA management mode is adopted to comprehensively manage and control the security and risk of DL based on the GA-PSO-SVM digital library risk identification model. The PDCA model is derived from the national information security standards. It is applicable to the entire process of the information security system. Therefore, a PDCA DL information security and risk control model based on GA-PSO-SVM risk identification is established.

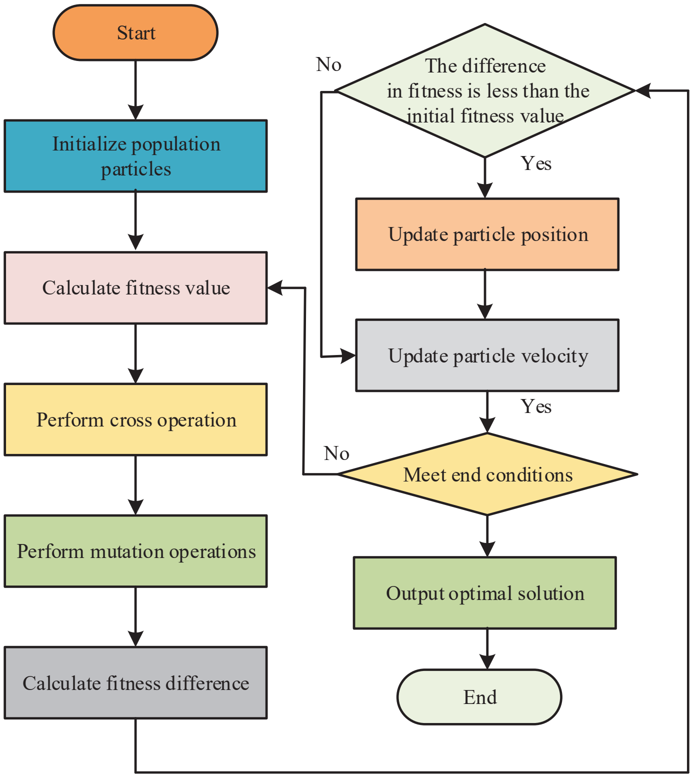


Figure 3 Improved PSO algorithm flowchart.

The information security and risk control of DL is a continuous process. The entire information security service will not stop working just there is a problem in a certain link. Therefore, the DL information security management based on the PDCA model has formed a complete cycle model. A problem is solved, the information security and risk management model will discover new unresolved issues, thereby achieving continuous improvement of the model in DL. The DL information security and risk control model based on PDCA is shown in Figure 5.

In the implementation process of Figure 5, Plan is the starting point of the PDCA cycle. It is also the initial step of the DL information security management system. At this stage, the necessary resource system for information security in DL is constructed. The scope of information security management is determined. The overall information security and risks in DL are identified and appropriate processing plans are developed.

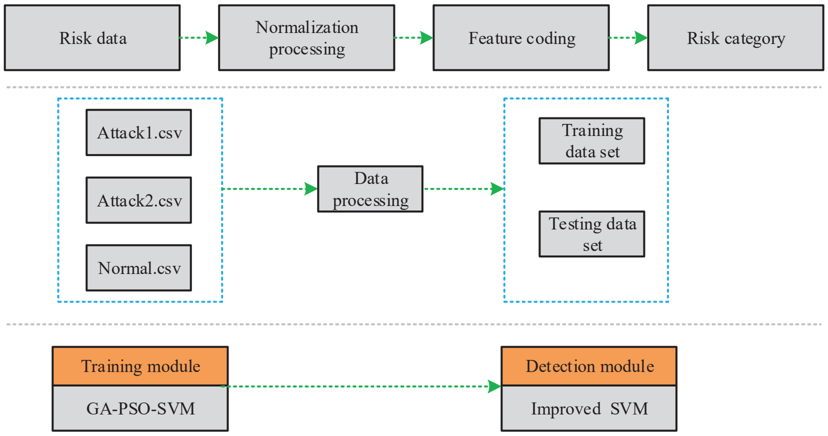


Figure 4 Optimizing risk detection algorithms for SVM models.

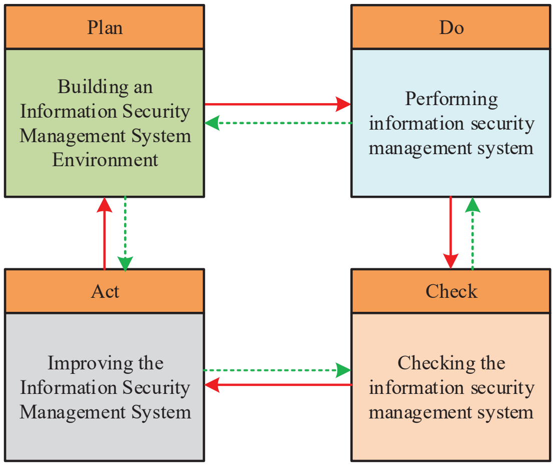

Plan
Building an Information Security Management System Environment
Do
Performing information security management system
Act
Improving the Information Security Management System
Check
Checking the information security management system

Figure 5 PDCA circular management mode of digital library.

Next is the operation of the information security management system (Do stage), which is the implementation stage of PDCA cycle management. During the implementation process, various resources, including librarians, funds, and various facilities and equipment, should be reasonably allocated to ensure the orderly operation of this stage. At the same time, corresponding acceptable levels and solutions should be set for relevant risk issues.

The third stage is to monitor and review the information security system (Check stage). Managers can regularly review the effectiveness of information security management systems and audit execution procedures through rapid detection and identification. Unreasonable and imperfect parts are promptly improved to ensure that the risks of DL are acceptable.

Finally, there is the Act stage. This stage is to improve the information security management system, which is the summary part of the PDCA cycle [22]. After the three stages of planning, implementation, and inspection, a relevant conclusion is drawn on the implementation of the entire PDCA cycle. The system is judged whether to continue execution, improve execution, or terminate the loop. The circular management model based on risk identification can effectively improve the information security and risk control in DL, and enhance the security performance and risk prevention control capabilities.

4 Performance Analysis of DL Information Security and Risk Control Model Based on PDCA

4.1 Weighting Analysis of Risk Factors in Digital Libraries

To better validate the performance of the proposed digital library information security identification model, corresponding experiments are designed to verify the performance. Firstly, the weight values of security and risk factors faced by DL are analysed. The weight analysis results of common risk factor indicators are shown in Table 2. The indicators include three primary indicators and 18 secondary indicators. The average weight of the electronic resources, system software, and technical management reaches 0.08. The information security and risks of DL account for a significant proportion in these three aspects, which have a significant impact on the information security of DL. In addition, the weight of physical assets and network structure also reaches 0.07. This indicates that improving the information security of DL mainly starts from aspects such as resource management and system maintenance. The risk prevention capability of the system has been continuously strengthened to enhance the information security of DL and reduce the various risk accidents.

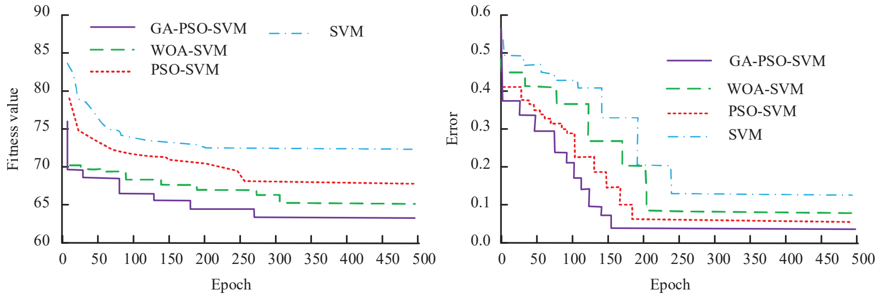
4.2 Performance Analysis of Digital Library Risk Identification Model Based on Improved SVM

To verify the availability of the designed digital library security and risk identification method, experiments are designed. Firstly, the model is trained and tested using data from the R2L dataset. 8000 software samples are

Table 2 Comparison of risk factor weights

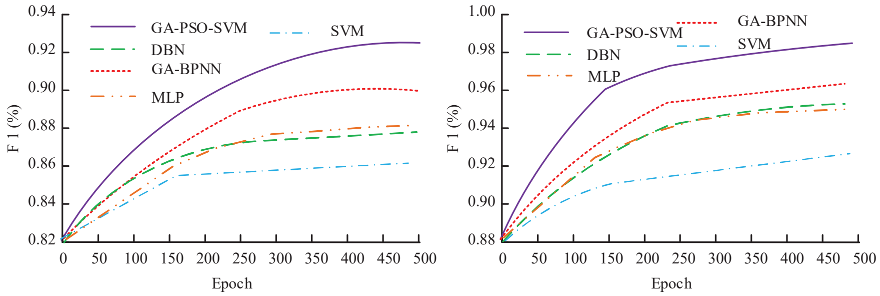
Primary Indicator	Secondary Indicators	Indicator Code	Weight
Property	Electronic resources	X_1	0.08
	Data document	X_2	0.06
	Physical assets	X_3	0.07
	Software assets	X_4	0.06
	Service	X_5	0.06
	Various maintenance personnel	X_6	0.05
Threatens	System	X_7	0.06
	Environment	X_8	0.03
	Nature	X_9	0.05
	External personnel	X_{10}	0.06
Vulnerability	Physical environment	X_{11}	0.05
	Network structure	X_{12}	0.07
	Systems software	X_{13}	0.08
	Database software	X_{14}	0.05
	Application Middleware	X_{15}	0.04
	Application system	X_{16}	0.05
	Technical management	X_{17}	0.08
	Organizational management	X_{18}	0.03

selected from the R2L dataset and divide them into training and testing sets in a 7:3 ratio. The whale optimization algorithm (WOA), PSO and the improved PSO (GA-PSO) proposed by the research are used to optimize the SVM. The experimental parameters for the PSO algorithm are set as follows. The number of iterations is set to 500. The population size is set to 5. The weight value is 0.7. The learning factor is 2. The experimental parameters of the GA algorithm are set as follows. The population size is set to 50, the number of generations is set to 500, the crossover probability is 0.5, and the mutation probability is 0.0001. The parameter settings for SVM are as follows. The learning rate is 0.01. The optimization range for SVM kernel function parameters C and gamma is (0.01, 200). The results are shown in Figure 6. The convergence of IPSO-SVM algorithm, DBN algorithm, GA-BPNN, and SVM is shown in Figure 6(a). The GA-PSO-SVM algorithm has iterated 260 times to reach the optimal fitness value of 63. Figure 6(b) shows the training errors during the training process after optimizing the four algorithms. GA-PSO-SVM has the fastest convergence rate and the highest precision. The error remains below 0.05. Based on the above results, among the commonly used optimization schemes, the GA-PSO algorithm proposed in the study has the best optimization effect.



(a) Fitness value (b) Error

Figure 6 Optimization performance analysis of several algorithms.



(a) Training test (b) Testing set

Figure 7 F1 value of four models.

Commonly used risk identification models include Deep belief networks (DBN), GA optimized BPNN models (GA-BPNN), Multilayer Perceptron (MLP), SVM, and the proposed GA-PSO-SVM. The performance of the above recognition algorithms is verified. The results obtained are depicted in Figure 7. In Figure 7(a), on the training set, the F1 of the GA-PSO-SVM reaches 0.931, which is 0.061, 0.071, 0.63 and 0.068 higher than the DBN, GA-BPNN, MLP and SVM. In Figure 7(b), the F1 of the GA-PSO-SVM reaches 0.985, significantly higher than other methods. This indicates that the proposed method has better performance. When conducting fault identification, this method has more advantages. It can more accurately identify various different faults and potential risk factors.

The accuracy of the obtained risk identification model is analyzed. The results are depicted in Figure 8. In Figure 8, in multiple experiments, the accuracy of the GA-PSO-SVM method is higher than other methods, with an average accuracy of 96.46%. The average accuracy of DBN and GA-BPNN is 90.85% and 89.37%. In the identification of risk factors in DL,

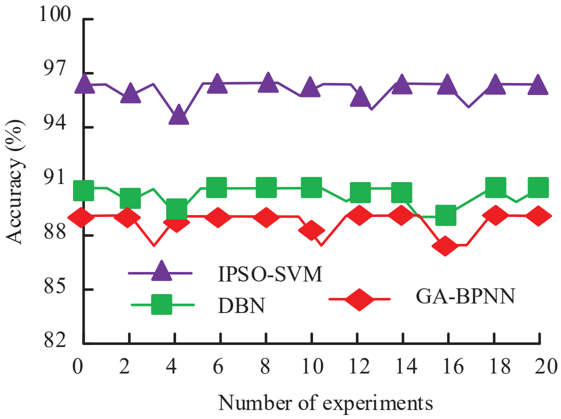
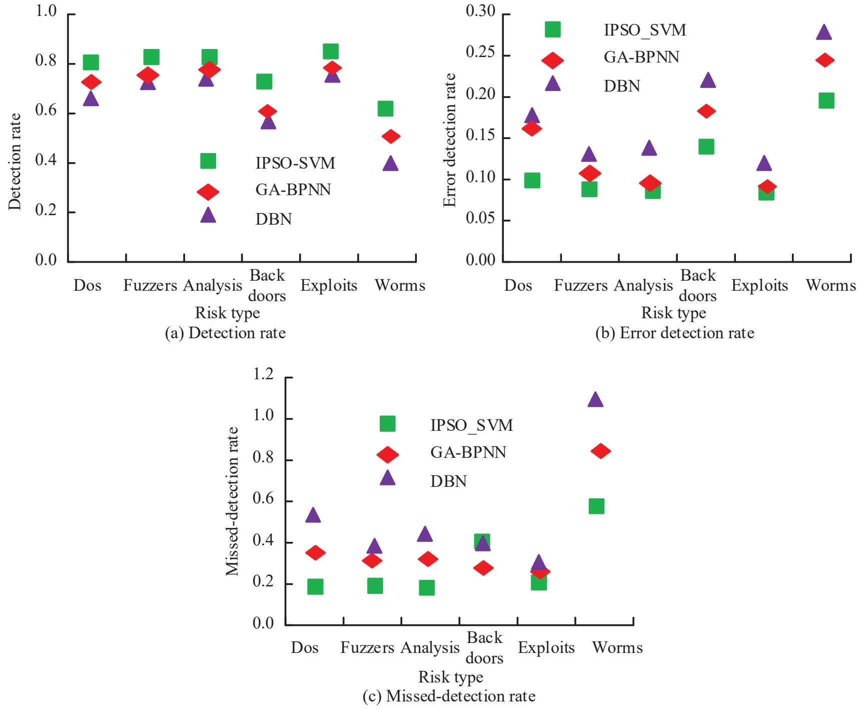


Figure 8 Accuracy comparison of multiple experiments.

the methods proposed in the study have better accuracy. It can meet more needs for identifying faults and risk factors.

4.3 The Application Effect Analysis of PDCA Based DL Information Security and Risk Control Model

The sample data in UNSW-NB15 is selected to validate the performance of the proposed method in the study. This dataset is about intrusion detection. This dataset has nine types of attacks, namely, DoS, Fuzzers, Analysis, Backdoors, Exploits, Generic, Reconnaissance, Shellcode and Worms. The number of records in the training set is 175,341 records and the testing set is 82,332 records from the different types, attack and normal. In this dataset, 14000 and 6000 pieces of data were randomly selected in a 7:3 ratio for experiments. The experimental datasets used include DoS, Fuzzers, Analysis, Backdoors, Exploits and Worms. Specifically, DoS: A malicious attempt to make a server or a network resource unavailable to users, usually by temporarily interrupting or suspending the services of a host connected to the Internet. Fuzzers: It attempts to cause a program or network suspended by feeding it the randomly generated data. Analysis: It contains different attacks of port scan, spam and html files penetrations. Backdoors: A technique in which a system security mechanism is bypassed stealthily to access a computer or its data. Exploits: The attacker knows of a security problem within an operating system or a piece of software and leverages that knowledge by exploiting the vulnerability. Worms: Attacker replicates itself in order to spread to other computers. Often, it uses a computer network to spread itself, relying on security failures on the target computer to access it. The application effectiveness of the proposed method is evaluated based on the detection rate (DR), false detection rate (FDR), and missed detection rate (MDR) of different algorithms. DR refers to the percentage of correctly detected quantities in the total number of tests. MDR refers to the percentage of qualified products detected as unqualified compared to the actual number of qualified products. The FDR refers to the percentage of non-conformities not detected during testing compared to the actual number of non-conformities. The DR, FDR, and MDR of different methods are shown in Figure 9. In different datasets, the DR, FDR, and MDR of the proposed method are higher than other comparative methods. Taking the results from the Dos dataset as an example, the detection rate, FDR, and MDR of GA-PSO-SVM are 81.56%, 10.11%, and 19.85%, respectively. This numerical result is significantly superior to other comparison methods. This indicates that the proposed method has better adaptability, which has good performance



(a) Detection rate

(b) Error detection rate

(c) Missed-detection rate

Figure 9 The detection performance comparison of risk identification models in different attack scenario.

Table 3 Risk intrusion detection time for attack scenario

	Time(s)				
Models	Dos	Fuzzers	Analysis	Backdoors	Exploits
DBN	11.34	3.49	15.99	5.61	11.57
GA-BPNN	6.52	1.08	8.87	2.03	7.72
GA-PSO-SVM	3.76	0.42	3.95	1.24	2.63

in different experimental datasets. This method can have significant effects in practical use, meeting different risk identification and control needs of DL.

The risk identification and detection time of the proposed method in different risk intrusion datasets are statistically analyzed. The efficiency improvement is mainly reflected in the improvement of hardware and software processing speed. The algorithm is optimized to reduce runtime. When evaluating performance, it is necessary to consider the usage background of the algorithm, such as power consumption, and throughput. Therefore, when the algorithm has good performance, the efficiency is also high. However, when the algorithm efficiency is good, the performance may not necessarily be optimal. Table 3 depicts the results. From Table 3, the time cost of the PIPSO-SVM risk identification model proposed in the research is the smallest in different risk intrusion datasets. Specifically, in the Dos dataset, the time overhead of DBN, GA-BPNN, and GA-PSO-SVM is 11.34 s, 6.52 s, and 3.76 s, respectively. In the Probe dataset, the time overhead of DBN, GA-BPNN, and GA-PSO-SVM is 15.99 s, 8.87 s, and 3.95 s, respectively. Based on the detection rate results shown in Figure 9, this indicates that the model proposed in the research has better efficiency and performance, with significantly lower time consumption than other commonly used methods.

To verify the feasibility and actual evaluation effect of the risk assessment method, two DL A and B are selected as the research objects for analysis. Through statistical analysis, A Library has a total of 3245 asset, threat, and vulnerability related items. B Library has a total of 2736 asset, threat, and vulnerability related items. The IPSO-SVM model is used to identify the risk value. The experiment is repeated 5 times. The results obtained are depicted in Table 4. The risk error of each level is maintained within 0.4%. After communicating with the relevant digital library management personnel, the scope is within an acceptable range. This indicates that the proposed method is more in line with the actual needs. The proposed method is feasible, which can play the role in risk identification and control in practical use, providing effective support for the safe and stable operation of DL.

Table 4 Risk level distribution

Risk Level		No		Low		Medium		High	
/		Recognition Result	Actual Result	Recognition Result	Actual Result	Recognition Result	Actual Result	Recognition Result	Actual Result
A	Number	78	80	604	615	1540	1536	1023	1014
	Ratio	2.4%	2.5%	18.6%	19.0%	47.4%	47.3%	31.6%	31.2%
B	Number	149	153	482	475	899	908	1206	1200
	Ratio	5.4%	5.7%	17.6%	17.3%	32.9%	33.1%	44.1%	43.9%

5 Conclusion

The information security and risk control of DL can effectively improve their resource utilization efficiency, ensuring the security of user information and digital library resources. However, existing research on information security and risk control in DL is relatively insufficient. Therefore, to continuously and effectively control the potential risk factors of DL, further improve the quality of digital library usage, ensure the smooth operation and information security of DL, a digital library security and risk identification model based on improved SVM is constructed. It is expected to achieve effective identification and control of digital library security and risk factors, ensure higher security, and improve service quality. Based on the risk identification results obtained, a DL information security and risk control model based on PDCA is constructed to achieve effective control of digital library information security. According to the findings, the F1 value of the improved SVM risk identification model reaches 0.931, which is 0.061, 0.071, 0.63 and 0.068 higher than the DBN, GA-BPNN, MLP and SVM, respectively. In the Dos dataset, the time overhead of DBN, GA-BPNN, and GA-PSO-SVM is 11.34 s, 6.52 s, and 3.76 s, respectively. The average accuracy of GA-PSO-SVM is 96.46%. The average accuracy of DBN and GA-BPNN is 90.85% and 89.37%, respectively. Finally, the collected digital library risk data is evaluated. The risk error for each level is maintained within 0.4%, which is in line with actual needs. Form the result, it is of great significance. The construction method can play an effective role in the security and risk control of DL. It has feasibility in practical applications, which can improve the security of digital library use. Based on safer and more effective digital library management methods, it can effectively ensure the security and experience during the use, maximizing the role in knowledge transfer and learning. The goals proposed in the study have been achieved to a certain extent. However, information security and risks are constantly changing and evolving. Therefore, further optimization is needed in the future. There are still shortcomings in the research. The risk factors and security issues faced by DL go far beyond the aspects covered in research. Therefore, in subsequent research, the types of information security that may exist in DL can be further studied. In addition, the study only analyzes the application effectiveness of information security faced by DL. In the future, in-depth analysis can be conducted on the performance of DL in controlling physical risks. In future research, experimental data can be further improved and expanded to further optimize model performance.

Funding

The research is supported by: The 2022 key project of the "14th Five Year Plan" of Nanchong Social Sciences Research "Research on the Role of Libraries in the View of overall national security concept" (Project NO: NC22A016).

References

[1] Shandilya S, Arora J, Vinayak K. PMDECS approach of red bin analysis – the art of problem solving in manufacturing industry. International Journal of Six Sigma and Competitive Advantage, 2023,14(3):300–329.

[2] Wolverton Jr. E R, Karen D, Dorkhosh M. Digital library of University of Tehran: opportunities and challenges. Journal of Electronic Resources Librarianship, 2022,34(4):332–336.

[3] Juan S, Wei C, Yang S, Yuan C M. Plan, Do, Check, Act (PDCA) Cycle Nursing Model Reduces the Risk of Hemangioma in Hemodialysis Patients. Iranian journal of public health, 2021,50(12):2560–2566.

[4] Yanez P C A, Macias S G, Clavijo V M M, Patino R C E, Carazas F J G. Excellence model for the maintenance area in Heavy-Duty Truck Company. IFAC PapersOnLine, 2022,55(19):163–168.

[5] Miriam M. Metadata in the digital library: building an integrated strategy with XML. Journal of the Australian Library and Information Association, 2023,72(1):110–111.

[6] Xue B. Application of the PDCA Cycle for Nursing Safety Management in Radiology Department. Journal of Radiology Nursing, 2023,42(2):241–244.

[7] Sally M, David C, Maria V. Using the PDCA cycle to uncover sources of mental health disparities for Hispanics. International Journal of Mental Health Nursing, 2022,32(2):556–566.

[8] Jie C, Cai W H, Feng L, Chen X C, Chen R, Ruan Z W. Application of the PDCA Cycle for Managing Hyperglycemia in Critically Ill Patients. Diabetes therapy: research, treatment and education of diabetes and related disorders, 2022,14(2):293–301.

[9] Amaral Vitória P A F, Bruna R. Internal Logistics Process Improvement using PDCA: A Case Study in the Automotive Sector. Business Systems Research Journal, 2022,13(3):100–115.

[10] Antoaneta S, Velichka M, Daniel S, et al. Food Safety Management System (FSMS) Model with Application of the PDCA Cycle and Risk Assessment as Requirements of the ISO 22000:2018 Standard. Standards, 2022,2(3):329–351.

[11] Elizarov A. M, Lipachev. E. K. Obachevskii Digital Library in the Scientific Space of Mathematical Knowledge. Scientific and Technical Information Processing, 2023,50(1):35–39.

[12] Kaba A, Refae E A G, Eletter S, Yasmin T. Testing a proposed ROI for academic libraries. Library Hi Tech News, 2023,40(4):15–18.

[13] Azam N. User experience and information architecture of the National Digital Library and archives of Iran: a usability investigation and card sorting. The Electronic Library, 2023,41(1):30–44.

[14] Samaneh K, Saeed S R, Amir G. Services personalization in digital academic libraries: a Delphi study. Digital Library Perspectives, 2023,39(1):39–61.

[15] Xie I, Babu R, Wang S, et al. Assessment of digital library design guidelines to support blind and visually impaired users: a study of key stakeholders' perspectives. The Electronic Library, 2022,40(6): 646–661.

[16] Sasmita P, Jyotshna S. A Literature Review on Digitization in Libraries and Digital Libraries. Preservation, Digital Technology & Culture, 2022,51(1):17–26.

[17] Xu F, Du T J. Research on the drivers of undergraduates' intention to use university digital libraries: affinity theory as an additional construct of information system success mode. Library Hi Tech, 2022,40(6): 1627–1641.

[18] Peter H, Chris B, Caroline G. Turning the lean world upside down. International Journal of Lean Six Sigma, 2022,13(5):989–1024.

[19] Wang X, Cheng M, Eaton J, et al. Fake node attacks on graph convolutional networks. Journal of Computational and Cognitive Engineering, 2022,1(4):165–173.

[20] Andrea B, Stefano C. Leonardo's Library: A Digital Library for Studying Leonardo da Vinci. International Information & Library Review, 2022,54(2):182–187.

[21] Maria A T, Michele T, Monica M, et al. What Do Users Think of the Digital Library? Analyzing Community's Sentiment during Covid-19. International Information & Library Review, 2023,55(1):66–76.

[22] Harry A. Book Review: Law, Insecurity and Risk Control: Neo-Liberal Governance and the Populist Revolt. Social & Legal Studies, 2022,31(3):501–505.

Biography

Yimin Yin, an associate researcher and deputy director of the Library and Archives of North Sichuan Medical College, and a member of the Scientific Research Integrity Construction Working Committee of China Society for Scientific and Technical Information (CSSTI), received his bachelor's degree in Information Management and Information Systems from Southwest Normal University in 2004 and master's degree in Software Engineering from the University of Electronic Science and Technology of China in 2012, and has published more than 30 papers and edited 1 textbook. He interested in the study of library resource construction, information services, and information security.