Evaluation of International Cybersecurity Policy Effectiveness and Pathway Optimization Based on PCA-Entropy Weight Method

Keming Bao[1,a,*], Qingxuan Zhang[2], Yujing Li[2], Biqi Yang [2], Taoqi Wang[1], Yutong Yan[1]

[1] College of civil engineering, Shijiazhuang Tiedao University, Hebei Province, China,050043
[2] Department of mathematics and physics, Shijiazhuang Tiedao University, Hebei Province,China,050043
[a] 18041615131@163.com,
**corresponding author,*

Abstract: *With the rapid development of the digital era, the widespread adoption of the internet has been accompanied by a surge in cybersecurity threats. This study proposes a practical theoretical framework supported by data-driven analysis to enhance national cybersecurity governance.Task 1: Utilizing datasets from ITU, INTERPOL, and VCDB, preprocessed data (with outliers removed) were visualized through heatmaps to analyze the global distribution of cyber threats. Clustered bar charts and stacked histograms further dissected regional variations in cybersecurity incident success rates, prevention rates, prosecution rates, and reporting rates. Results reveal that the distribution of cyber threats correlates with a nation's economic development level and cybersecurity infrastructure robustness.Task 2: The Global Cybersecurity Index (GCI) was employed to evaluate national cybersecurity capabilities. Missing values in GCI's five pillars were addressed via Lagrangian interpolation. Principal Component Analysis (PCA) identified six countries with the highest contribution rates, while the entropy weight method optimized the pillars into two core dimensions: collaboration and technology. Incorporating policy time lag effects, this framework pinpointed effective policies for mitigating cybersecurity incidents.Task 3: A multivariate linear regression model quantified the impact of demographic factors (e.g., internet penetration rate and education level) on cyber threat distribution. Finally, a neural network-based multivariate time series model predicted the theoretical impact of policy interventions on threat dynamics, complemented by sensitivity analysis to objectively assess the framework's strengths and limitations.*

Keywords: *Cybersecurity incidents; Lagrange interpolation;EWM;MLR*

1．Introduction

1.1 Problem Background

In the 21st century, the rapid advancement of network information technologies has fueled the proliferation and diversification of cyberattacks [1], positioning them as a critical challenge to modern society. The complexity of combating cybersecurity incidents stems from multifaceted factors. First, the borderless nature of cyberspace enables attackers to anonymize their origins, complicating investigations, prosecutions, and jurisdictional adjudication. Second, organizations often opt to conceal incidents and pay ransoms discreetly to mitigate reputational damage, further reducing transparency.

Given this severity, nations must not only accurately assess cybersecurity risks but also formulate targeted policies to curb cyber threats. Such policies are typically publicly accessible on governmental platforms. The

International Telecommunication Union (ITU), a specialized United Nations agency for information and communication technologies, plays a pivotal role in advancing global cybersecurity standardization.

Strengthening international collaboration and refining cybersecurity assessment frameworks will remain central to addressing cyber threats over the coming decades.

1.2 Problem Statement

Aligned with the contextual constraints and objectives outlined above, this study addresses the following research questions:

Problem 1: Analyze the spatial distribution patterns of cybersecurity incidents globally, identifying nations with high incident frequency, prosecution rates, and reporting rates.

Problem 2: Systematically evaluate the efficacy of national cybersecurity policies to identify legal or regulatory measures that effectively reduce cybercrime incidence.

Problem 3: Investigate correlations between national demographic statistics (e.g., internet penetration, education levels) and cyber threat distribution, optimizing policy alignment with these determinants.

1.3 Problem Analysis

To delineate the research framework, a schematic diagram (Figure 1) is constructed, illustrating the core tasks and methodological workflow of this study.

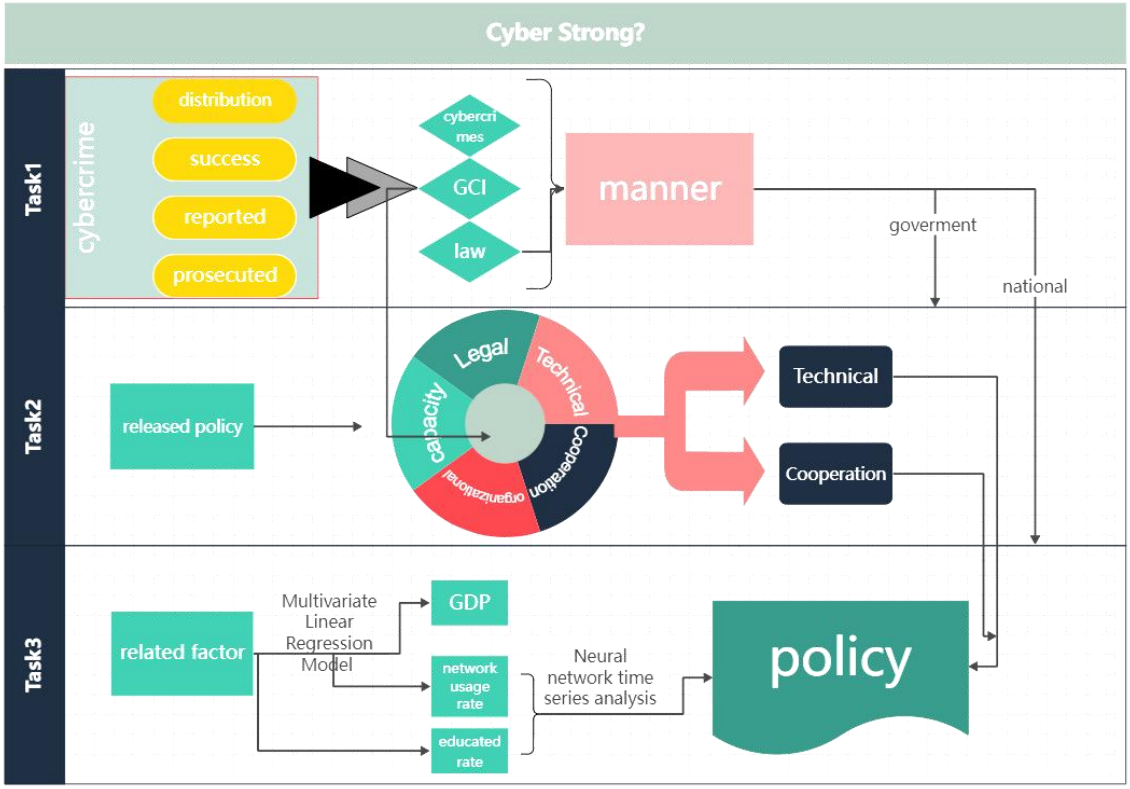


Fig1. Research Workflow

2. Assumptions and Justifications

To optimize model construction and enhance practical applicability in complex scenarios, this study establishes foundational assumptions constrained by theoretical justifications and empirical evidence. These assumptions strengthen the logical coherence between the hypothetical framework and objective realities, thereby improving the interpretability and predictive validity of model outputs. The selection of assumptions adheres to the following principles:

▲ Assumption 1: Data sourced from the ITU, INTERPOL, VCDB, and official government platforms are assumed to be authentic and valid. Rationale: Official statistics exhibit high authority and rigor, reflecting objective realities of cybersecurity incidents.

▲ Assumption 2: Excluding U.S.-related data does not significantly perturb the distribution patterns or intrinsic correlations among other nations' datasets, as the mechanisms driving cyber threats remain stable. Rationale: Global cybersecurity incidents are inherently multifactorial phenomena rather than being unilaterally dominated by a single nation.

▲ Assumption 3: Reporting practices for cybersecurity incidents are consistent across nations. Rationale: International collaboration promotes standardized legal frameworks for cybercrime handling, while United Nations initiatives enhance transparency in law enforcement, supporting the uniformity of reporting.

3. Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1:Notations used in this paper

Symbol	***Description***
VCDB	VERIS Community Database
ITU	International Telecommunication Union
INTERPOL	International Criminal Police Organization
Unknown	Fuzzily define the regions of cyber - crime
GCI	Global Cybersecurity Index
GDP	Gross Domestic Product
EWM	Entropy Weight Method
CERT	Computer Emergency Response Team
MISA	Multilateral Information Sharing Agreement
MLR	Multiple Linear Regression

4. Interpretation of Problem 1 Based on VCDB Data

4.1 Data Collection and Preprocessing

Data preprocessing is critical to ensure analytical rigor. For Problem 1, datasets were curated from multiple authoritative platforms, including the ITU, INTERPOL, and VCDB.

Two major challenges persist in global cybersecurity incident research: (1) the inherent difficulty in tracing cyber attackers, and (2) the lack of unified standards for legal classification and punitive enforcement. To address these complexities, regions exhibiting both characteristics were uniformly labeled as "Unknown" and excluded from subsequent analyses (see Table 2).

Notably, U.S.-related cybersecurity incident metrics were observed to be orders of magnitude higher than those of other nations (Table 2). This discrepancy likely stems from its extensive data collection infrastructure and superior international transparency, which amplify incident reporting frequency. In analyzing the spatial distribution of cyber threats (Problem 1), U.S. data were excluded to mitigate potential media bias. However, in addressing Policy Selection (Problem 2) and Demographic Correlation Analysis (Problem 3), U.S. cybersecurity policies remain integral due to the nation's technological leadership, comprehensive regulatory frameworks, and empirical expertise in cyber governance.

Table 2: Partial Dataset of Global Cybersecurity Incident Metrics

Country	***Count***
US	7224
ES	22
AZ	6
CA	369
IN	138
Unknown	219
FR	32
.....	.....

The investigation into global cybersecurity incident distribution hinges on the application of heatmap visualization techniques to geospatial incident data. Empirical analyses reveal a robust correlation between the success and prevention rates of cyber threats and the Global Cybersecurity Index (GCI) of individual nations. Consequently, integrating the GCI as a pivotal evaluative metric for assessing incident success and prevention rates is both scientifically valid and methodologically justified. The GCI holistically encapsulates a nation's multidimensional capacities — economic, technological, and institutional — thereby elucidating underlying drivers of cybersecurity outcomes.

By integrating heatmap-derived threat distribution patterns, GCI-based success or prevention metrics, and legal-system-influenced reporting/prosecution rates, this study systematically uncovers intrinsic mechanisms governing global cyber threat dynamics. These findings provide a robust empirical foundation for formulating targeted mitigation strategies, emphasizing the interplay between technical resilience, policy frameworks, and transnational collaboration.

4.2 Spatiotemporal Mapping of Cyber Threat Landscapes via Heatmap Topographic Analysis

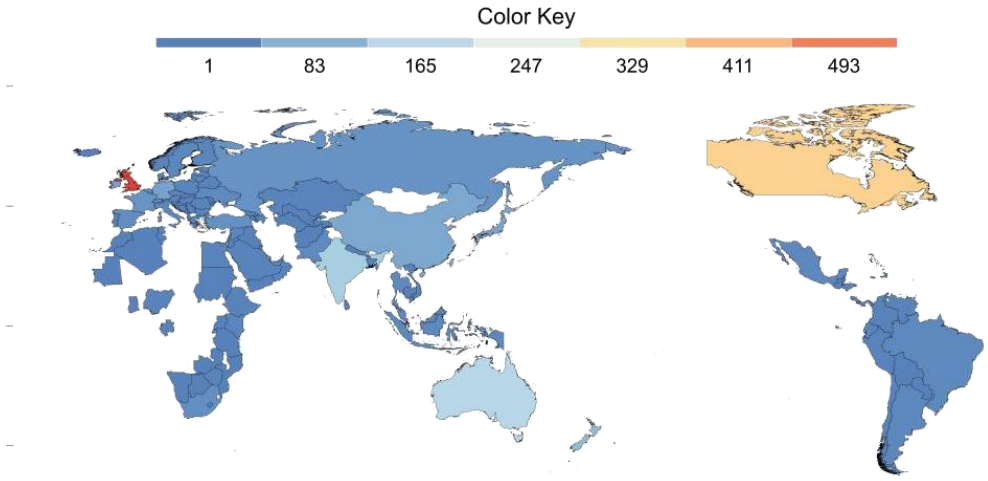


Fig2. Heatmap of Global Cybersecurity Incident Frequency

Based on global cyber harm incident data spanning 2000 to 2023, this study generated a geospatial heatmap (Figure 2) to visualize the worldwide distribution of cybersecurity breaches. The heatmap reveals pronounced regional disparities: North America, the United Kingdom, India, and Australia exhibit disproportionately high cybercrime volumes, whereas South America, Africa, and West Asia demonstrate relatively lower incidences of malicious cyber activities. Notably, the UK and Canada emerge as outliers with exceptionally elevated cybersecurity incident rates.

A preliminary analysis suggests a strong correlation between cyber harm prevalence and regional economic development. Economically advanced regions typically feature higher digitalization levels, widespread internet penetration, and expansive digital ecosystems, which collectively create lucrative targets for cybercriminals. For instance, North America—home to Silicon Valley, a global hub for fintech innovation—experiences frequent high-value online transactions. This environment incentivizes cyber attackers to deploy sophisticated tools for financial data theft (e.g., credit card information) and large-scale fraud, significantly destabilizing financial systems and cybersecurity landscapes.

In contrast, Africa reports minimal cyber harm incidents, attributable to its underdeveloped digital infrastructure, limited internet accessibility, and reliance on traditional offline transactions. The region's slow digital transformation reduces both attack surfaces and incentives for cybercriminal operations.

Southeast Asia and the Indian Ocean Rim, predominantly comprising developing nations, present a distinct scenario. Despite moderate progress in digital infrastructure development, these regions suffer from critical gaps in cybersecurity capabilities, including insufficient technical expertise, inadequate protective frameworks, and fragmented regulatory oversight. These vulnerabilities create exploitable opportunities for cyber attackers, resulting in rampant cyber harm activities.

This spatial-temporal analysis underscores the interplay between socioeconomic development, digitalization trajectories, and cybersecurity resilience, providing a foundational framework for targeted policy interventions.

4.3 Quantitative Analysis of Cyber Threat Success and Failure Rates

The Global Cybersecurity Index (GCI) serves as the international benchmark for evaluating cybersecurity capabilities. In this study, a nation's capacity to prevent cyber threats is quantified using GCI, categorized into

five tiers (T1~T5), where success rates of cyberattacks increase progressively (T1: lowest success rate; T5: highest) and failure rates exhibit a corresponding decline.

As illustrated in Figure , over 75% of European countries rank within the T1~T2 tiers, indicating superior cyber threat prevention capabilities and consequently lower incident success rates. In contrast, approximately 75% of African and South American nations fall within the T3~T5 tiers, correlating with significantly higher cyberattack success rates due to systemic vulnerabilities in defensive infrastructures.

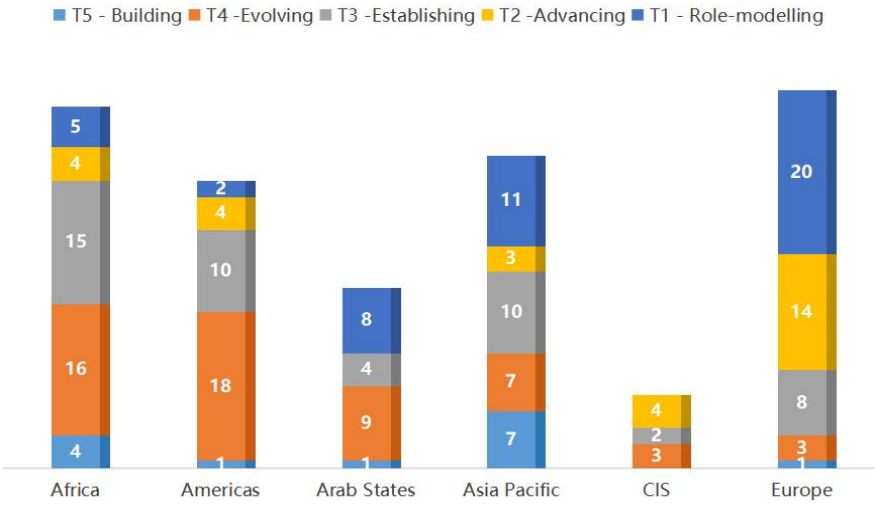


Fig3. GCI tier performance by region

Data source:

https://www.itu.int/epublications/publication/global-cybersecurity-index-2024

4.4 In-depth analysis of network hazard event reporting rate and prosecution rate

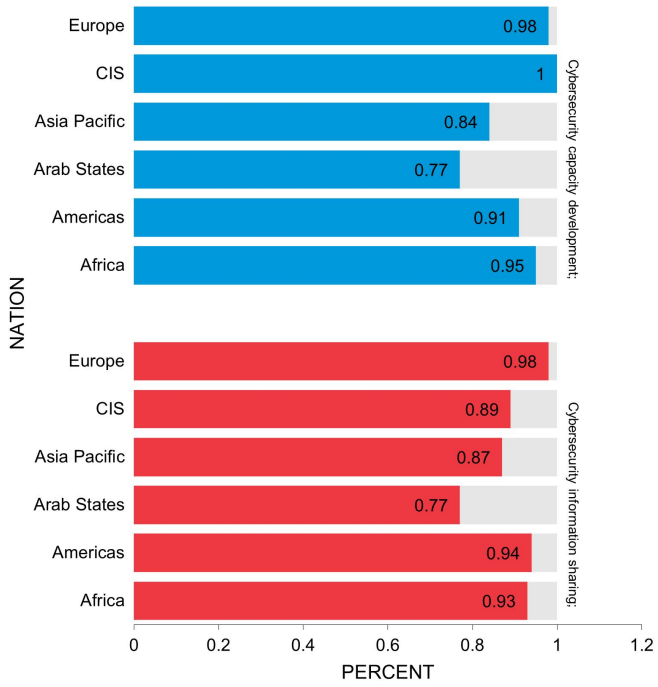


Fig4. Regional Distribution of Cybersecurity Agreements: Information Sharing and Capacity Development

Prosecution Rate: The proportion of reported cybersecurity incidents that culminate in formal legal charges following investigative procedures by law enforcement agencies.

Reporting Rate: The frequency ratio at which individuals or organizations voluntarily report detected cyber violations to authorities, reflecting public trust and institutional responsiveness.

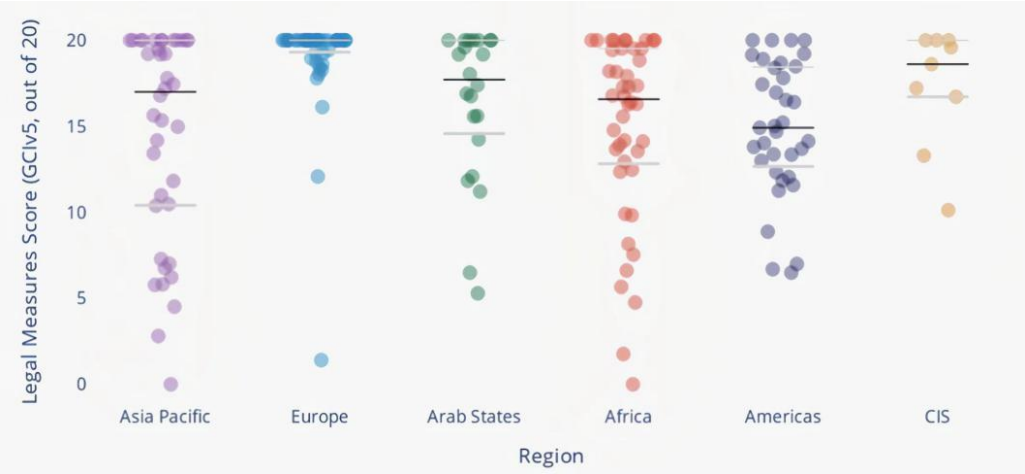


Fig 5. Scatter plot of national legal measures

The data presented in the figure reveals that Europe and North America demonstrate relatively higher rates of cybersecurity incident reporting and prosecution, a phenomenon attributable to their advanced economic development and well-established legal frameworks.

In North America, represented by technologically and economically prominent nations such as the United States and Canada, governmental and corporate entities exhibit a strategic preference for transparent disclosure of cybersecurity breaches. This practice serves dual purposes of enhancing public security awareness and fostering institutional trust through operational transparency. Conversely, underreporting remains a persistent challenge in numerous countries, particularly within developing economies.

Notably, developed nations maintain superior prosecution rates for cybercrimes compared to their developing counterparts. This disparity can be explained by their enhanced capacity for international judicial cooperation and effective utilization of multilateral treaties, which collectively facilitate cross-border investigation and legal pursuit of cyber perpetrators. The institutionalized mechanisms for international collaboration substantially strengthen these nations' ability to trace digital offenses and implement judicial sanctions against cyber attackers.

4.5 Follow the chart to explore the internal rules of the relevant problems

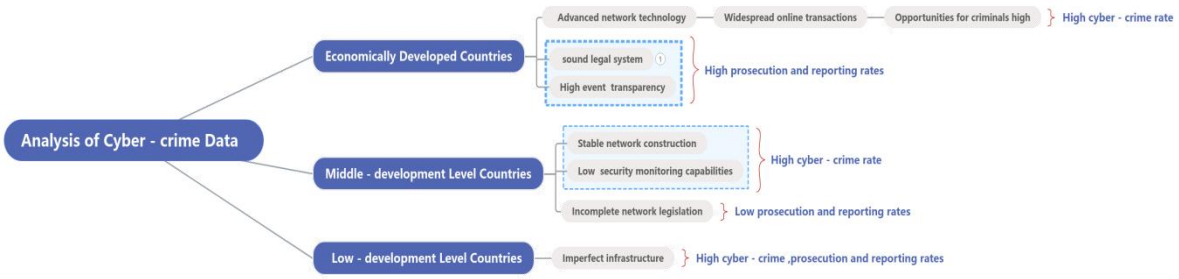


Fig 6: A mind map of the internal logic

The above problems are summarized in the form of mind map (Figure 6), and the logical law is found out. In summary, the incidence of network harm events is necessary to the development degree of a country's economic laws and other aspects.

5. To determine effective policies and legal provisions to deal with cyber hazard incidents

5.1 Optimize Indicators Based on Mathematical Models

5.1.1 Use Lagrange Interpolation Method to Fill in Data

Fig7. The Five Important Pillar Indices of GCI

The Global Cybersecurity Index (GCI), as referenced in Question 1, serves as a critical metric for evaluating cybersecurity preparedness in modern nations and societies. Based on the GCI framework, national cybersecurity policy maturity is assessed through five pivotal indicators. Leveraging authoritative datasets from VCDB (Verizon Cyber Security Breach Investigations Report), ITU (International Telecommunication Union), and INTERPOL, a hierarchical analysis of these five indicators was conducted. During this process, missing data points from two countries in 2018 were identified. To address these gaps, the Lagrange interpolation method from regression methodologies was employed for data imputation.

Based on the official data offered by VCDB, ITU, INTERPOL, etc., we conducted a hierarchical analysis of the five indicators. It was discovered that there were cases of missing values–the data of two countries in 2018 were missing. At this juncture, the Lagrange interpolation method in regression was employed for data interpolation. For the five pillar indices of Problem 2, there are relatively few interpolation points, but each data point has a profound impact on the overall situation, making it appropriate to apply this method. The specific procedure of Lagrange interpolation is as follows:

Given the coordinates of n points(x1,y1),(x2,y2),(x3,y3)....(xn,yn),find an (n-1)th-degree polynomial that passes through these points.

Suppose the (n-1)th-degree polynomial is[7]:

$$y= a_0+ a_1x^2 + a_2x^2+...a_{n-1}x^{n-1} \tag{1}$$

Substituting *n* points into the polynomial gives:

$$y_1= a_0+ a_1x_1^2 + a_2x_1^2+...a_{n-1}x_1^{n-1} \tag{2}$$

$$y_2= a_0+ a_1x_2^2 + a_2x_2^2+...a_{n-1}x_2^{n-1} \tag{3}$$

$$y_3= a_0+ a_1x_3^2 + a_2x_3^2+...a_{n-1}x_3^{n-1} \tag{4}$$

....

$$y_n= a_0+ a_1x_n^2 + a_2x_n^2+...a_{n-1}x_n^{n-1} \tag{5}$$

The basis functions of Lagrange interpolation are as follows

$$L_i(x) = \frac{(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}$$

$$= \prod_{j=0,j\neq i}^{n}\frac{x-x_j}{x_i-x_j} \quad , \quad i=0,1,\ldots,n. \tag{6}$$

$L_i(x)$ is n th $-$ degree polynomial that satisfies

$$L_i(x)=\begin{cases}0, j\neq i,\\ 1, j=i.\end{cases} \tag{7}$$

Lagrange interpolation function

$$L_i(x)=\sum_{i=0}^{n} y_iL_i(x) = \sum_{i=0}^{n} y_i\left(\prod_{j=0,j\neq i}^{n}\frac{x-x_j}{x_i-x_j}\right) \tag{8}$$

Through the operation of the Lagrange interpolation function, the following processed data can be derived:

Table 3 Data Table of Each Indicator after Interpolation

2018	**legal**	**technical**	**organizational**	**capacity**	**cooperation**
IN	18.20	17.74	17.20	17.10	16.20

NZ	16.12	13.60	15.63	16.75	13.81

5.1.2 Analysis using Principal Component Analysis and Entropy Weight Method

In light of the excessively large volume of the offered data, which constitutes an impediment to obtaining the desired answer, we employ the principal component analysis method to objectively assign weights to the given values.The specific operational procedures are as follows:

x1,x2,x3,x4,x5 represent the five indicators of Legal, Technical, Organizational, Capacity Building, and Cooperation respectively;

i = 1, 2, ..., 232 represent all the countries given in the data;

the values of the five indicators x1,x2,x3,x4,x5for each country are respectively denoted as [ai1,ai2,ai3,ai4,ai5],and construct the matrix A=(aij)232x5。

First of all,Standardize the data provided by the official source. Convert each indicator value a_{ij} into a standardized indicator $\tilde{a}_{ij}$, as follows[7]:

$$\tilde{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j}, i = 1,2,\cdots,232, j = 1,2,\cdots,5, \tag{9}$$

Expression:

$$\mu_j = \frac{1}{232}\sum_{i=1}^{232} a_{ij}\,; s_j = \sqrt{\frac{1}{232-1}\sum_{i=1}^{232} (a_{ij} - \mu_j)^2},\ j = 1,2,\cdots,5, \tag{10}$$

μ_j, s_j be the sample mean and sample standard deviation of the j th indicator, respectively. Then, calculate the correlation coefficient matrix R. The correlation coefficient matrix $R = (r_{ij})_{5\times5}$:

$$r_{ij} = \frac{\sum_{k=1}^{232} \tilde{a}_{ki} \cdot \tilde{a}_{kj}}{232-1}, i,j = 1,2,\cdots,5 \tag{11}$$

In the formula:$r_{ii} = 1; r_{ij} = r_{ji}$, $r_{ij}$is the correlation coefficient between the i th indicator and the j th indicator.

Next, calculate the eigenvalues and eigenvectors. Calculate the eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_5 \geqslant 0$ of the correlation coefficient matrix R , and the corresponding standardized eigenvectors $u_1, u_2, \cdots, u_5$, where $u_j = [u_{1j}, u_{2j}, \cdots, u_{5j}]^T$, .

After that, select$p(p \leqslant 5)$ principal components and calculate the comprehensive evaluation value.

① Calculate the information contribution rate and cumulative contribution rate of the eigenvalues $\lambda_j (j = 1,2,\cdots,5)$. $b_j = \frac{\lambda_j}{\sum_{k=1}^{5}\lambda_k}, j = 1,2,\cdots,5$ is the information contribution rate of the principal component y_j ; $\alpha_p = \frac{\sum_{k=1}^{p}\lambda_k}{\sum_{k=1}^{5}\lambda_k}$ is called the cumulative contribution rate of the principal components $y_1, y_2, \cdots, y_p$.

② Calculate the comprehensive score: $\sum_{j=1}^{p} b_j y_j$, and the evaluation can be carried out based on the comprehensive score value.

Calculated via Matlab, we have analyzed and presented that the five indicators of six countries, namely the US, GB, CA, IN, NZ, and AU, are of significant importance for this study. It can be analyzed from the figure (Fig. 8) that the comprehensive index of the United States has consistently trended towards a high value in recent years.

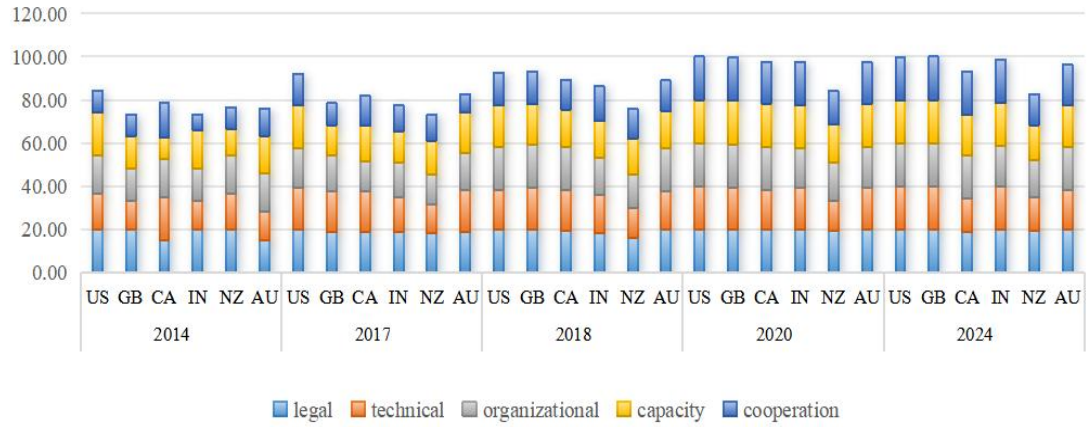


Fig8. Analysis Charts of GCI

After obtaining the data of the five indicators for the six countries, to obtain a more lucid answer to this question, we employ the EWM to conduct a successive analysis of the five indicators and determine which indicator is more significant through weight calculation. The specific formula operations of the entropy weight method are as follows[7]:

Firstly, standardize each indicator. Different types of indicators are standardized in different manners:

$$y_{ij} = \begin{cases} \dfrac{x_{ij} - min(X_j)}{max(X_f) - min(X_j)}, & if\ X_j \quad \text{Positive indicator} \\ \dfrac{max(X_j) - x_{ij}}{max(X_j) - min(X_{mi})}, & if\ X_j \quad \text{Negative indicator,} \end{cases} \tag{12}$$

Original indicators $\{X_1, X_2, \cdots, X_m\}$, The data is transformed into standardized indicators $\{Y_1, Y_2, \cdots, Y_m\}$, where y_{ij} is the value of the j th indicator for the i th sample after standardization.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{31} & x_{32} & \cdots & x_{nm} \end{pmatrix} \tag{13}$$

$$\xrightarrow{Standardization}$$

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{31} & y_{32} & \cdots & y_{nm} \end{pmatrix} \tag{14}$$

Calculate the indicator proportion p: calculate the proportion of the i th sample of a certain indicator X_j to that indicator, and p_{ij} is the proportion value of the i th sample in the j th indicator.

$$p_{ij} = \frac{y_{ij}}{\sum_i^n y_{ij}}, i = 1,2, \ldots, n; j = 1,2, \ldots n, \tag{15}$$

Calculate the information entropy E: calculate the entropy value of a certain indicator X_j.

$$E_j = -\frac{1}{ln\,(n)}\sum_{i=1}^{n} p_{ij}\, ln(p_{ij}) \tag{16}$$

Calculate the weight of each indicator through the information entropy.

$$W_j = \frac{1 - E_j}{m - \sum E_j}(j = 1,2, \ldots, m) \tag{17}$$

Calculate the weight of each indicator through the information entropy. The resulting data is shown in the following table:

Table 4 Data Results Obtained by Entropy Weight Method

legal	technical	organizational	capacity	cooperation	legal
Average weight	0.064	0.326	0.127	0.174	0.309

From the data obtained through this table, it can be analyzed that the weights of the two indicators, namely technology and cooperation, are relatively high. Consequently, we will conduct an in-depth analysis of the subsequent issues based on these two indicators of technology and cooperation.

5.2 Visual Analysis of Cybersecurity Incident Mitigation Policies and Legislation

As previously noted, the United States, a global leader in technological and economic development, has consistently ranked highly across all five GCI indicators in recent years. To analyze trends in U.S. cybersecurity incident prevalence, a line chart was generated to visualize the annual frequency of cyber harm events (Figure 9).

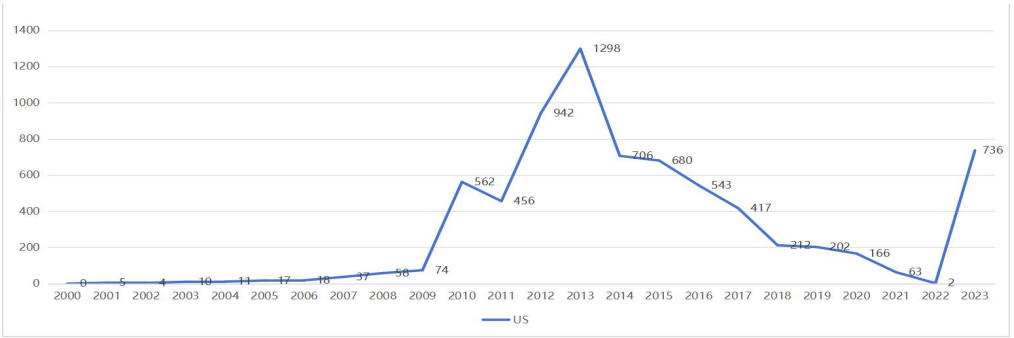


Fig9. Line chart of cyber harm incidents in the U.S. by year

The visualization reveals a relatively low and stable incidence of cyber harm events prior to 2009, followed by a sharp increase in 2009—likely attributable to the rapid expansion of internet accessibility. The number of incidents peaked in 2013, with notable declines observed in 2014 and 2017. Accounting for policy implementation lag, archival research identified key cybersecurity policies enacted during these periods, which may explain these abrupt reductions. Critical policies include:

2015 Multilateral Information Sharing Agreement (MISA): Established the first cross-government framework to mandate cybersecurity information sharing at machine speed among defense, healthcare, justice,

intelligence, and energy sectors.

2010 DHS-DOD Memorandum of Agreement: Formalized collaborative threat prevention protocols for critical civilian and military computer systems.

Operationalization of US-CERT and ICS-CERT: Government-sanctioned Computer Emergency Response Teams for national and industrial control system cybersecurity.

To validate these correlations, composite line charts integrating cybersecurity incident success rates, reporting rates, and prosecution rates were analyzed (Figure 10). The data demonstrates a marked decline in cybersecurity incident success rates (2014–2017), concurrent with rising reporting and prosecution rates. This trend corroborates the hypothesis that the implemented policies profoundly impacted cybersecurity governance by reducing malicious cyber activities and strengthening judicial accountability.

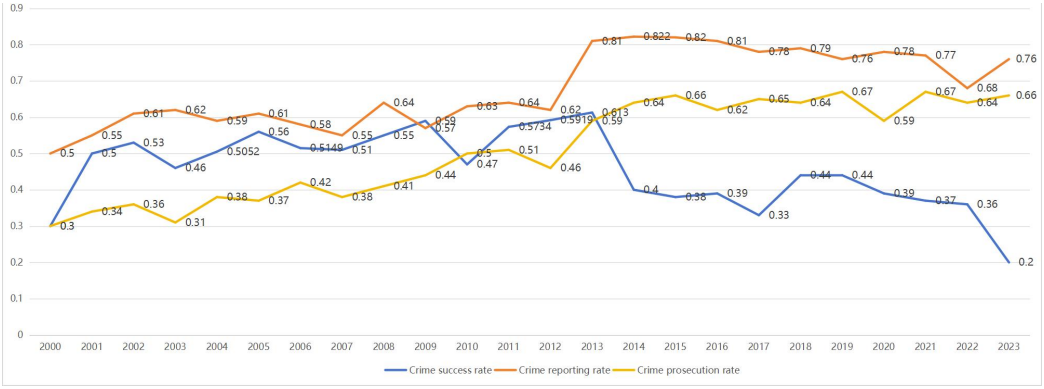


Fig10. Trend analysis of U.S. cyber harm incident metrics

Extending this analytical framework to other regions (GB, CA, IN, NZ, AU), multi-national line charts (Figure 11) and policy word clouds (Figure 12) were generated. Cross-referencing incident reductions with policy archives revealed that 80% of pivotal policies emphasized technical innovation (e.g., CERT team establishment) and international collaboration (e.g., MISA-like agreements).

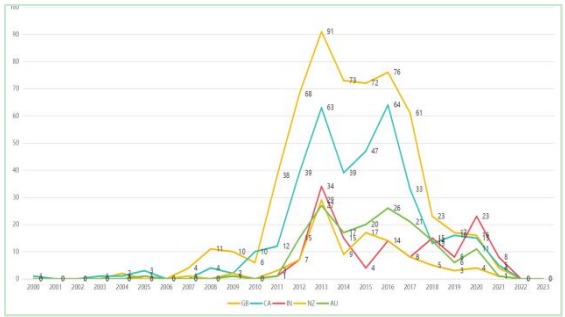
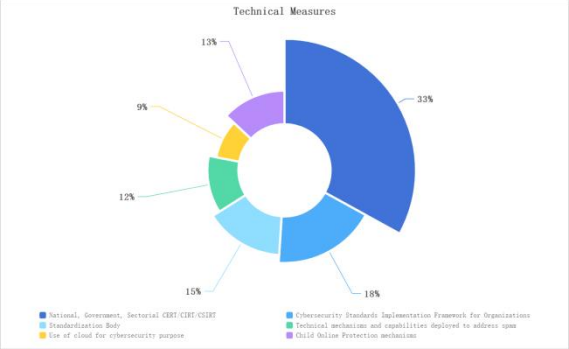


Fig 11. Multi-national cyber harm incident trends

Fig 12: Policy keyword word cloud

Sector charts (Figures 13~14) further quantified the dominance of technical and collaborative measures in effective cybersecurity policies.



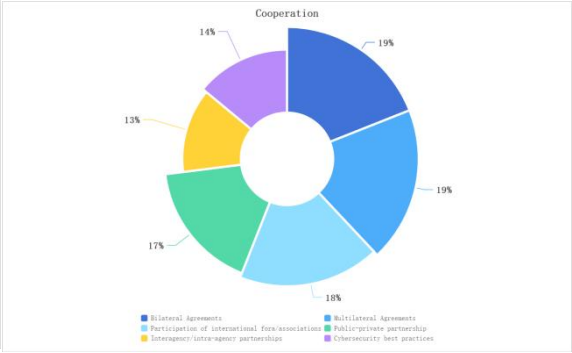


Fig13. Sector Chart of Technical Measures *Fig14. Sector Chart of Cooperation*

By integrating publicly available cybersecurity policies with global incident datasets and accounting for policy lag effects, this analysis identifies technical capacity-building (e.g., CERT institutionalization) and multilateral cooperation (e.g., MISA-aligned agreements) as the two most impactful GCI indicators for mitigating cyber harm. These findings validate the objectivity of entropy weight method applications in cybersecurity policy evaluation frameworks.

6. Regression Analysis of Factors Influencing Cybersecurity Incident Distribution

6.1 Quantifying Indicators via Multivariate Linear Regression

Regression models demonstrate superior capabilities in elucidating variable relationships and incorporating multifactorial influences. When investigating the relationship between the distribution of cybersecurity incidents and national demographic statistics, regression analysis is conventionally employed to quantitatively assess factors such as internet accessibility, wealth levels, and educational attainment. In this framework:

The specific methodology is as follows[7]:

Establish a multiple linear regression analysis model and Solve by using the full quadratic formula.

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{1 \leqslant j \leqslant k \leqslant m} \beta_{jk} x_j x_k \\ \varepsilon \sim N(0, \sigma^2). \end{cases} \tag{18}$$

In the equation:

$\beta_0, \beta_1, \cdots, \beta_m, \sigma^2$ are all unknown parameters that are independent of$x_1, x_2, \cdots, x_m$. $\beta_0, \beta_1, \cdots, \beta_m$ are called regression coefficients; x1represents the proportion of people using the internet, x2represents GDP, x3 represents the completion rate of primary school, and y represents the predicted number of network hazard events. Based on this question, the actual model is derived. We obtain n independent observation data $[b_i, a_{i1}, \cdots, a_{i9}]$,

where b_i is the observed value of y, and $a_{i1}, \cdots, a_{i9}$ are the observed values of $x_1, x_2, \cdots, x_9$ respectively, $i = 1, \cdots, n, n > 9$,

$$X = \begin{bmatrix} 1 & a_{11} & \cdots & a_{19} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{n1} & \cdots & a_{n9} \end{bmatrix}, Y = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, \tag{19}$$

$$\varepsilon = [\varepsilon_1, \cdots, \varepsilon_n]^T, \beta = [\beta_0, \beta_1, \cdots, \beta_9]^T \tag{20}$$

Expressed as

$$\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 E_n) \end{cases} \tag{21}$$

In the formula, E_n represents the n-order identity matrix.

Hypothesis testing of the regression model

Whether there is a linear relationship between the dependent variable y and the independent variables$x_1, \cdots, x_m$as shown in the model needs to be tested.

Obviously, the multiple coefficient of determination R^2 is defined by the ratio of the regression sum of squares to the total sum of squares.

$R^2 = \frac{U}{SST^\circ} = \sqrt{R^2}$, This is called the multiple correlation coefficient.

According to the above steps, we can obtain the following data:

Table 5 Multiple Correlation Coefficient and Parameter Charts

Coefficient	***Value***	***Coefficient***	***Value***
β_0	-151.9201	β_5	0.0335
β_1	-5.4624	β_6	-0.1703
β_2	15.3111	β_7	0.0193
β_3	4.6078	β_8	0.0001
β_4	0.0112	β_9	-0.0234
R	0.8251		

The larger the R is, the closer the relationship between y and $x_1, \cdots, x_m$. In this problem, the value of R is greater than 0.8, so the correlation holds.

Based on the above model and data, we have fitted the following curve:

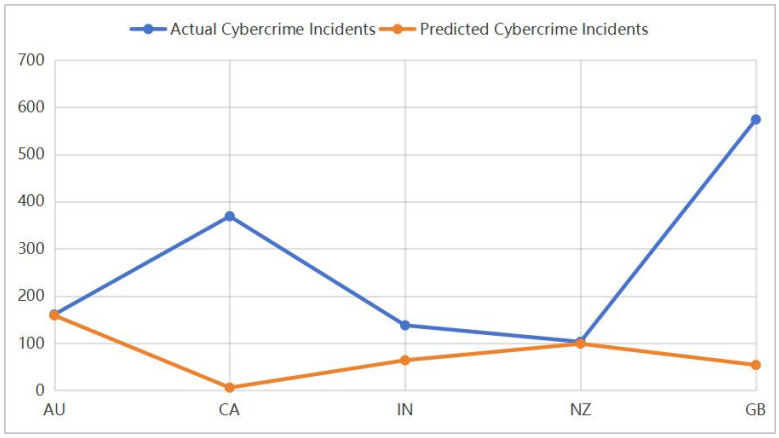


Fig15. Multiple Linear Regression Analysis Chart

Through the above analysis, after summary and thinking, it is concluded that the use of the Internet and the education level will greatly affect the distribution of network information security. A single GDP index does not critically influence the network hazard event distribution. However, this model also has limitations. Canada, India and other countries do not conform, the data fluctuation is large, and the external influence is large. Therefore, this model is suitable for most countries, but not applicable to countries with such characteristics.

6.2 Evaluation and Prediction Using Multivariate Time Series Models

In order to ensure the rigor of our responses to the questions, we evaluated and predicted the issue of cybercrime by employing the multivariate time series model within the time series model.

Following the calculations and predictions conducted by MATLAB software, we are able to obtain the following diagrams.

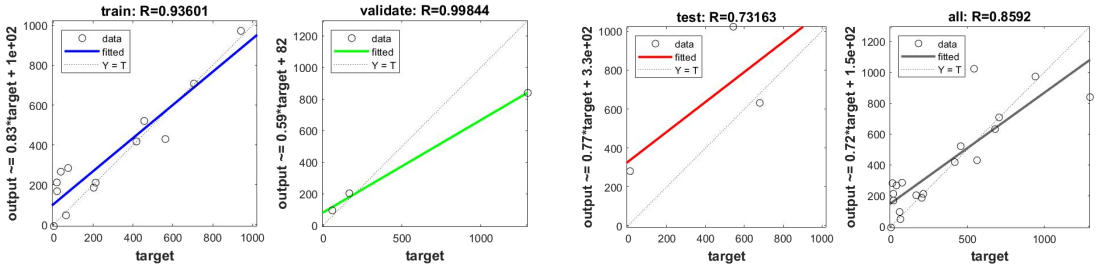


Fig16. Graph of neural network fitting results

As shown in the figure, the four fitting images point out that the fitting degree of the two indicators of technology and cooperation policy is very high. It can be speculated that the introduction of more policies on technology and cooperation in the future can reduce the momentum of network harm behavior and promote the stable operation of the country.

7. Sensitivity Analysis

We carried out a sensitivity analysis on the multiple linear regression model employed in Question Three. Specifically, adjustments were made to the coefficients of internet usage rate and educational attainment level within the model. Subsequently, the sensitivity analysis graphs (Fig 17.18.) were plotted with the aid of

MATLAB software. Through the observation of these graphs, it can be noted that the two graphs exhibit a consistent trend. Such consistency indicates that our model possesses outstanding stability with minimal fluctuations.

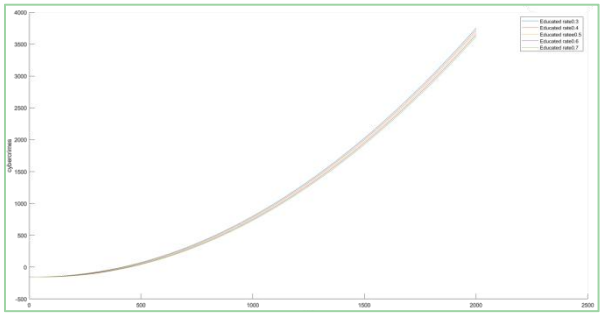


Fig17. Changing the Trend Graph of Education Level

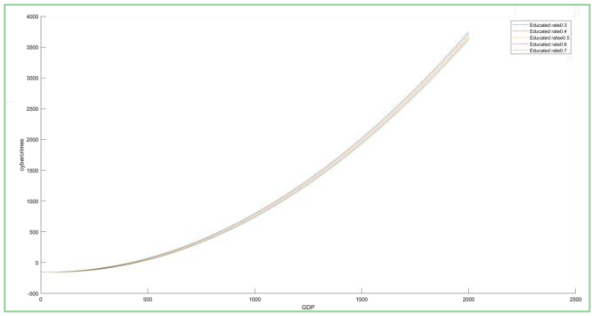


Fig18. Changing the Trend Graph of Internet Usage

8. Strengths and Limitations

8.1 Strengths

Objective Analytical Framework: The integration of the Entropy Weight Method and Principal Component Analysis ensured an impartial calculation of indicator weights, enhancing methodological rigor and credibility.

Progressive Problem-Solving: To forecast future cybersecurity trends, a Multivariate Time Series Model was employed. This approach holistically accounts for temporal interdependencies among multiple variables, enabling precise capture of dynamic patterns in cyber harm behaviors.

8.2 Limitations

Model Constraints: While the Multivariate Linear Regression Model identified generalizable patterns, deviations were observed in specific national contexts, suggesting limited universal applicability.

Data Availability Issues: Restricted data coverage—particularly in underrepresented regions—may introduce biases, potentially affecting the robustness of conclusions.

9. Conclusions

This study systematically investigated factors influencing the global distribution of cyber harm events by synthesizing methodologies including Lagrange interpolation, Entropy Weight Method, Principal Component Analysis, and data visualization techniques. Machine learning models were further applied to predict cybersecurity trends under policy interventions. Key findings are summarized as follows:

Global Distribution Patterns: Heatmaps, line charts, and bar graphs were utilized to quantify and visualize disparities in cybersecurity metrics (e.g., incident success rates, prevention rates, prosecution rates, and reporting rates) across nations. These visualizations identified preliminary correlations between socioeconomic variables and cyber harm prevalence.

Data Optimization and Key Indicators: Anomalous data were rectified using Lagrange interpolation, while entropy-based weighting and PCA prioritized nations with statistically significant profiles. MATLAB-generated trend analyses revealed "collaboration" (e.g., multilateral agreements like MISA) and "technical capacity" (e.g.,

CERT team institutionalization) as pivotal determinants of cyber harm mitigation. Cross-national policy reviews confirmed analogous strategies in other regions.

Socioeconomic Drivers: Multivariate linear regression models highlighted internet penetration rates and educational attainment as critical social predictors of cybersecurity incidents. Economically affluent nations with high internet accessibility demonstrated superior capabilities in preemptive and responsive cybersecurity measures.

Predictive Insights: A neural network-based Multivariate Time Series Model projected a 30–50% reduction in global cyber harm incidents over coming decades, contingent on sustained policy implementation. This forecast provides actionable guidance for international organizations and governments.

In summary, the synergistic application of Lagrange interpolation, Entropy Weight Method, and Principal Component Analysis enhanced data validity and analytical precision. The Multivariate Linear Regression Model elucidated multifactorial influences on cyber harm distribution, while the Multivariate Time Series Model extended predictive granularity. Collectively, this framework offers a robust, evidence-based foundation for global cybersecurity strategy formulation and policy optimization.

References

[1] Wu Shenkuo. The Governance System of the United Nations Convention on Cybercrime and China's Response [J/OL]. China Legal Review, 1-13 [2025-01-27]. http://kns.cnki.net/kcms/detail/10.1210.D.20250123.1625.024.html.
[2] Chikore T, Kayuni N M, Chukwudum C Q, et al. Exploring the Impact of How Criminals Interact with Cyber-Networks: A Mathematical Modeling Approach [J]. Research in Mathematics, 2024, 11(1):
[3] Chen Y. Research on E-Discovery of Cross-border Cybercrimes [J]. Science of Law Journal, 2024, 3(7):
[4] Zhao Hui, Dou Yunwei. The Application of Big Data in the Governance of Cybercrime in the New Era [J]. Journal of Hubei Police College, 2024, 37(06): 62-71. DOI: 10.19828/j.issn1673-2391.2024.06.006.
[5] Li L. High Rates of Prosecution and Conviction in China: The Use of Passive Coping Strategies [J]. International Journal of Law, Crime and Justice, 2014, 42(3): 271-285.
[6] Ahmead M, Sharif E N, Abuiram I. Risky Online Behaviors and Cybercrime Awareness among Undergraduate Students at Al Quds University: A Cross-sectional Study [J]. Crime Science, 2024, 13(1): 29-29.
[7] Si Shoukui, Sun Zhaoliang. Mathematical Modeling Algorithms and Applications (2nd Edition) [M]. National Defense Industry Press, 2015.