End-to-End Network Security Prediction via Dual Attention-Enhanced Situation Assessment and IHHO-GResNeSt Integration

Qingjun Ni
Department of Artificial Intelligence, Laiwu Vocational and Technical College, Ji'nan 271100, China
E-mail: lwzynqj@163.com

Keywords: Dual attention, network security prediction, improved Harris Hawks optimization algorithm, feature extraction, network security status assessment

Received: January 9, 2026

To address the limitations of existing cybersecurity prediction methods in terms of insufficient feature extraction and low parameter optimization efficiency, this study proposes an end-to-end cybersecurity prediction method based on Channel-Spatial Dual Attention (CS-DA) to enhance situation assessment and integrating an improved Harris Hawks Optimization algorithm and a Global Context Residual Split-Attention Network (IHHO-GResNeSt). This method first introduces CS-DA to enhance the extraction of key threat features during the cybersecurity status assessment stage, and then uses IHHO to globally optimize the assessment model parameters. Subsequently, quantified situation values are used as prior input, and GResNeSt is introduced for temporal feature learning. The Sparrow Search Algorithm (SSA) is then combined with the GResNeSt algorithm to optimize the prediction hyperparameters, achieving integrated end-to-end security situation prediction that combines assessment and prediction. Experimental results on three public benchmark datasets-NSL-KDD, UNSW-NB15, and CICIDS2017 show that the proposed method achieves a false negative rate of 0.05 and a false positive rate of 0.03 in the security assessment stage, with a situational assessment bias of 0.04 and an inference latency of only 32.16 ms, demonstrating good assessment accuracy and real-time performance. In the prediction stage, the RMSE (0.08) of the SSA-tuned model is superior to CNN-LSTM (0.17), XGBoost (0.20), and Transformer (0.15). These results demonstrate that the proposed method can achieve high-precision, robust, and efficient cybersecurity prediction in complex and dynamic network environments, providing reliable support for security early warning and decision-making.

Povzetek: Članek predlaga end-to-end napoved kibernetske situacije, ki z dvojnim kanalno-prostorskim mehanizmom pozornosti okrepi izluščanje groženj, z izboljšanim Harris Hawks optimizatorjem uglašuje parametre ocenjevanja, nato pa z GResNeSt in SSA optimizacijo uči časovne vzorce za natančnejše in hitrejše zgodnje opozarjanje.

1 Introduction

As information technology and network infrastructure rapidly evolve, networks have become essential to social, financial, industrial, and transportation systems [1]. However, increasing environmental complexity has also intensified security threats [2]. Network Security Prediction (NSP) aims to detect potential attacks and abnormal behaviors in advance, supporting proactive defense and reducing operational risks and economic losses [3]. Existing NSP approaches include statistical analysis, Machine Learning (ML), and deep learning. Statistical methods perform poorly on high-dimensional, nonlinear time-series data, while ML models rely heavily on manual feature engineering and struggle with temporal dependency modeling and anomaly capture [4]. Although deep learning enables automatic feature extraction and complex pattern recognition, it incurs high training costs, is prone to overfitting, and adapts poorly to sudden anomalies, limiting its real-time and large-scale applicability [5]. Moreover, most existing methods lack effective integration of network traffic, user behavior, and heterogeneous data, restricting the extraction of key security information and resulting in suboptimal prediction accuracy and response speed [6].

Based on this, this study focuses on the integrated modeling problem of network security situation assessment and prediction, proposing the following hypotheses: (1) Can the introduction of the Channel-Spatial Dual Attention (CS-DA) more effectively capture the locality and dimensionality importance distribution of threat features in complex dynamic network environments, thereby improving the accuracy and stability of network security situation assessment? (2) Can the improved Harris Eagle Optimization (IHHO) algorithm for parameter optimization further alleviate the premature convergence problem in high-dimensional non-convex objective functions and improve model convergence efficiency and global search capability compared to traditional swarm intelligence algorithms (such as PSO)? (3) Can the introduction of quantified security situation values as prior knowledge into the temporal prediction network to construct an integrated

assessment-prediction framework achieve higher prediction accuracy and smaller generalization error compared to pure prediction models in scenarios of distribution drift and sudden attacks?

Building upon this foundation, this study proposes an end-to-end NSP method based on CS-DA enhanced situation assessment and the integration of Improved Harris Hawks Optimization and Global Context Residual Split-Attention Network (IHHO-GResNeSt). The method utilizes CS-DA and IHHO to construct a network security assessment method. Then, CS-DA is used to deeply extract and fuse key threat features, and the IHHO algorithm is combined to globally optimize the assessment model parameters. The quantified situation value of the evaluation output is used as prior knowledge and input into the GResNeSt for feature extraction. Finally, the network security situation value at future times is output through the regression prediction head, achieving end-to-end accurate prediction from temporal features to future situations. The research aims to enhance the prediction accuracy, stability, and adaptability of network security models in complex dynamic network environments, providing more reliable decision support for network security defense.

2 Related works

As computer technology and internet adoption advance, cyberspace has become a critical carrier for national infrastructure, social operations, and personal privacy. NSP provides early warning and decision support, reducing security risks and economic losses, and has thus become a key research direction in academia and industry [7]. Bangali et al. proposed a network attack and defense behavior tree model that integrates game theory to solve the problems of traditional attack tree models ignoring defense effects and poor scalability of defense tree models. By analyzing the logical relationships between different levels of attack behavior and integrating corresponding attack and defense trees, the model's resistance to attacks was improved [8]. Qi H et al. proposed a hybrid detection scheme based on quantum particle swarm optimization and extreme learning machine to address the problem of balancing speed and accuracy in network intrusion detection. By using partition gain feature selection and hidden layer node tuning strategies, they significantly improved the model training and detection speed while ensuring high detection accuracy, providing an efficient and lightweight solution for real-time network security protection [9]. Alsenani T R et al. proposed a feature selection model based on particle swarm optimization algorithm to address the problem of phishing websites threatening user login and payment security. By iteratively optimizing the key feature subset, they provided a solution for efficient and accurate phishing website identification [10]. Li et al. proposed an intelligent traffic prediction and anomaly detection method that integrates CNNs and LSTM networks to deal with the increasingly severe threats to campus network security and the insufficient prediction accuracy and detection capabilities of traditional methods. The detection sensitivity was automatically adjusted based on traffic fluctuations, enhancing prediction accuracy by simultaneously extracting local features and temporal dependencies from network traffic [11].

Furthermore, Mohy-Eddine et al. designed a K-nearest neighbour-based intrusion detection model with feature selection to address IoT vulnerabilities and improve detection accuracy and efficiency. Using Principal Component Analysis (PCA), univariate statistical tests, and Genetic Algorithms (GAs), they extracted optimal features to enhance performance [12]. Li et al. proposed an anomaly-based intrusion detection method for IoT security, combining multiple filtering feature selection methods with wrapper algorithms to reduce single-method bias and improve accuracy [13]. Li et al. developed a ML-based runtime monitoring mechanism to mitigate denial of service attacks on multi-processor on-chip IoT systems through offline training and policy deployment [14]. Mohy-Eddine et al. proposed an anomaly edge detection method based on the Euler framework to detect advanced persistent threats during lateral movement by modeling host logs as discrete time series and constructing graph neural networks with sequence encoding layers [15].

In summary, existing research has achieved rich results in the combination of NSP methods, intrusion detection models, and intelligent optimization algorithms, laying the foundation for improving prediction accuracy and detection capabilities. The study provides a structured summary of representative related works, as shown in Table 1.

Table1: Comparison of representative network security prediction methods

Reference	Core Algorithm	Dataset	Evaluation Metrics	Main Limitations
[8]	Attack–Defense Tree+Game Model	Private Campus Network	Detection Rate = 0.91 False Alarm Rate = 0.07	Relies on handcrafted rules; weak generalization; no temporal modeling
[9]	QPSO+ELM	Private Network Traffic	Accuracy = 0.89 Convergence Speed = 0.78	Sensitive to initial weights; limited feature adaptability
[10]	PSO+ANN	KDD99	Accuracy = 0.87 RMSE = 0.19	Uses outdated dataset; shallow temporal modeling
[11]	CNN-LSTM	Campus Traffic	Accuracy = 0.90 Detection Rate = 0.93	Weak robustness to dynamic attacks; no global feature weighting
[12]	KNN+PCA+GA	NSL-KDD	Accuracy = 0.88 F1-Score = 0.89	Heavy feature engineering; poor scalability
[13]	Hybrid Feature Selection+ML	IoT Traffic	Detection Rate = 0.92 False Alarm Rate = 0.08	Limited temporal dependency modeling
[14]	SVM/Random	NoC IoT Data	Accuracy = 0.86	Weak generalization; no deep feature learning

	Forest		Latency = 41 ms	
[15]	GNN+Temporal Encoder	Host Logs	AUC = 0.94 Precision = 0.91	High computational cost; complex deployment

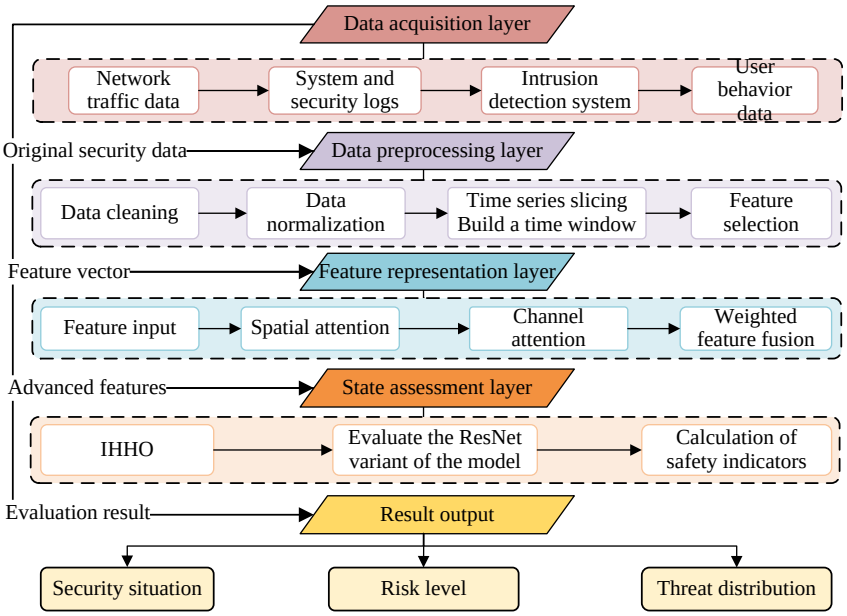


Figure 1: An NSSA system based on CS-DA and IHHO.

However, current research has shortcomings in deep modeling, feature extraction and weight allocation, and generalization ability for complex temporal features. To this end, the study proposes the end-to-end NSP via dual attention-enhanced situation assessment and IHHO-GResNeSt Integration method, which improves the ability to extract key features through CS-DA, optimizes model hyperparameters using IHHO, and constructs an evaluation-prediction integrated framework to achieve situational awareness closed-loop. Its innovation lies in the deep integration of DA and IHHO, forming a dual drive architecture of "feature optimization-parameter optimization". At the same time, by introducing historical situation values as prior knowledge for temporal prediction, it breaks through the bottleneck of insufficient modeling of dynamic threat features in traditional methods.

3 Methods and materials

3.1 NSSA method based on CS-DA and IHHO

To address the issues of uneven distribution of dimensional importance of network security traffic features and insufficient spatial focusing capability for anomaly regions, this study introduces a network security status assessment method based on channel attention and spatial attention, specifically CS-DA and IHHO. The "dual" characteristic explicitly refers to the two components: channel dimensional weight modeling and spatial location saliency modeling. Channel attention characterizes the global importance of different feature channels in threat identification, while spatial attention locates key regions of potential anomalous behavior in the feature map. The two attention modules are cascaded and applied to the input features in a "channel first, then spatial" manner, achieving a feature enhancement path of "first filtering dimensions, then locating regions." This is then combined with IHHO to optimize the evaluation model parameters, improving the accuracy and adaptability of situational awareness. The network security status assessment method is shown in Figure 1.

Figure 1 shows the network security status assessment method based on dual attention and IHHO. Its core is to achieve high-precision situational awareness through the synergistic effect of dual attention mechanism and IHHO. First, the collected network traffic and host/application logs and other heterogeneous data are preprocessed, including deduplication and missing value processing (deleting invalid records, setting abnormal fields to empty or filling them according to the median/mode), outlier pruning (such as truncating extreme values according to the 1%-99th percentile) and category label consistency. In terms of feature construction, a basic feature set is formed based on traffic quintuples and session statistics, mainly including protocol type (TCP/UDP/ICMP), source/destination port, connection duration, number of packets/bytes, average packet length and packet interval, etc.; and for typical application layer traffic (such as HTTP/DNS/FTP/SMTP), statistical features such as request frequency, response latency and error rate are added to characterize the differences in attack behavior under different protocols and traffic types. Subsequently, Min-Max normalization is performed on continuous features to eliminate the influence of dimensions, and its calculation is shown in Equation (1) [16].

$$Z_{scaled} = a + \frac{(Z - Z_{min})(b - a)}{Z_{max} - Z_{min}} \quad (1)$$

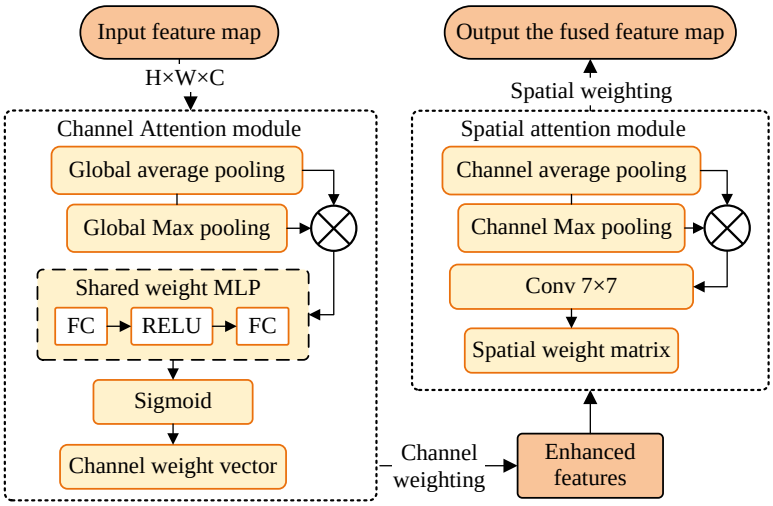


Figure 2: DA feature extraction and fusion mechanism

In Equation (1), Z represents the original feature value, Z_{min} and Z_{max} are the mini and max values in the dataset, a and b denote the upper and lower limits of the target scaling interval, and Z_{scaled} means the normalized result value.

To reduce feature redundancy and improve training efficiency, this study adopts a collaborative feature engineering process of "statistical filtering + correlation screening + dimensionality reduction compression". First, approximately constant features are eliminated based on the variance threshold; second, correlation screening is performed using the Pearson correlation coefficient. This step is used to identify the linear correlation with the target situation label and the potential collinearity between features: when $|\rho|$ is lower than the threshold, the feature is considered to have limited information contribution and is eliminated; when $|\rho|$ between two features is higher than the threshold, only the one with higher information contribution is retained to reduce collinearity. The calculation is shown in Equation (2).

$$\rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \quad (2)$$

Equation (2) adopts the Pearson correlation coefficient formula, where X and Y respectively represent two characteristic variables, cov refers to the covariance, and σ means the standard deviation.

Next, to mitigate the redundancy and noise effects of the high-dimensional feature space, PCA is performed on the feature vectors after screening, retaining principal components with a cumulative variance contribution rate of no less than 95% to obtain a compact representation. Based on this, a GA is used with the validation set accuracy as the fitness function to further refine the feature subset, thereby compressing the input dimensionality while maintaining performance. The resulting high-quality feature set is used as input to the state assessment model, highlighting key threat information through a dual attention mechanism, and global optimization of key parameters is performed by IHHO.

The spatial attention module focuses on critical spatiotemporal locations of attack behaviors, while the channel attention module adaptively weights feature dimensions to enhance representation of covert threats and complex attack patterns. The SA-DA feature extraction and fusion mechanism is illustrated in Figure 2.

Figure 2 illustrates the CS-DA structure, which consists of a cascade of channel attention and spatial attention. Channel attention first adaptively weights the feature dimensions, while spatial attention then significantly enhances the key anomaly regions. Firstly, channel attention generates channel descriptors via global average and max pooling and computes weight vectors using a fully connected network to model feature-dimension importance. Subsequently, spatial attention applies average and max pooling across channels to generate spatial weight matrices that highlight potential abnormal behaviors or attack locations. These two mechanisms are cascaded into a “dimension filtering first, region localization next” feature extraction path, integrating global and local information to support subsequent NSSA and prediction. Among them, dual attention first models the dimensionality importance of the input feature map through the channel attention module, and its weight calculation process is shown in Equation (3) [17].

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (3)$$

In Equation (3), F indicates the input feature map, $AvgPool$ and $MaxPool$ represent global average pooling and global max pooling, MLP is the multi-layer

perceptron with shared weights, and σ means the Sigmoid activation function. Secondly, based on the feature map after channel attention weighting, a spatial attention module is introduced to model the spatial distribution of potential abnormal behaviors. The weight calculation process is shown in Equation (4).

$$M_s(F_c) = \sigma(f^{7\times7}([AvgPool(F_c); MaxPool(F_c)]) \quad (4)$$

In Equation (4), F_c represents the feature map weighted by channel attention, $[;]$ represents the channel dimension concatenation operation, and $f^{7\times7}$ denotes the convolutional layer with a kernel size of 7 × 7. The final output feature is obtained by fusing channel attention and spatial attention in a cascaded manner, i.e., $F' = Ms(Fc) \otimes Fc$, where $\otimes$ represents the element-wise multiplication operation. This cascaded structure achieves dual modeling of the "dimensional importance + spatial saliency" of network threat features without introducing additional temporal dimension attention.

Furthermore, to overcome the problems of premature convergence and local optimum dwell in the standard Harris Hawks Optimization (HHO) algorithm during the optimization of high-dimensional network security feature parameters, this study improves HHO from three aspects: enhanced initialization diversity, dynamic balance adjustment in search, and embedding of local escape mechanisms, thus constructing the IHHO algorithm. The improvement process of HHO is shown in Figure 3.

Figure 3 shows the IHHO process. In the initialization stage, logistic chaotic mapping is introduced to generate the initial population position, to enhance the ergodicity and uniformity of the initial solution in the high-dimensional search space, as shown in Equation (5).

$$x_k^{(0)} = \mu\, x_k^{(0-1)} \left(1 - x_k^{(0-1)}\right), \mu \in (3.8,\ 4) \quad (5)$$

In Equation (5), $x_k^{(0)}$ represents the initial position of the k eagle in generation 0, and μ is the chaos control parameter. This mapping can effectively avoid the local region clustering problem caused by standard random initialization, thereby improving the quality of the algorithm's global exploration starting point.

During the location update phase, an adaptive inertial weight is introduced to dynamically balance the global exploration and local development capabilities, and its update form is shown in Equation (6).

$$\begin{cases} \omega(t) \\ = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \cdot t / T \\ X(t+1) \\ = \omega(t) \cdot X_{traditional} + [1 - \omega(t)] \cdot X_{rabbit} \end{cases} \quad (6)$$

In Equation (6), t is the current iteration number, T is the maximum iteration number, $X_{traditional}$ is the standard HHO update position, and X_{rabbit} is the current global optimal solution. This weight is set to a larger value in the early stage of iteration to enhance the global search capability, and is gradually reduced in the later stage to enhance the local convergence accuracy, thereby alleviating the convergence oscillation phenomenon of the standard HHO in complex non-convex problems.

During the search process, when the improvement of the optimal fitness is less than the threshold for several consecutive generations, a local perturbation exit mechanism is introduced to perform random perturbation on the current individual position, as shown in Equation (7).

$$X_i' = X_i + \delta \cdot randn() \quad (7)$$

In Equation (7), δ is the perturbation amplitude factor, and $randn()$ is a standard normally distributed random variable. This mechanism can break the local optimal attraction domain during the search stagnation phase, thereby enhancing the algorithm's escape capability and steady-state convergence robustness.

The objective function for cybersecurity status assessment and prediction tasks is characterized by high dimensionality, strong nonlinearity, and multiple local extrema. Standard Hierarchical Homing (HHO) is prone to premature convergence and search stagnation in such scenarios. This study improves the coverage of the initial solution through chaotic initialization, balances the exploration and development intensity through adaptive inertial weights, and introduces a perturbation exit mechanism during the search stagnation phase. This makes the Integrated Hierarchical Homing (IHHO) more suitable for the complex non-convex problem of optimizing cybersecurity feature parameters, thereby improving the accuracy and stability of the final solution while ensuring convergence speed.

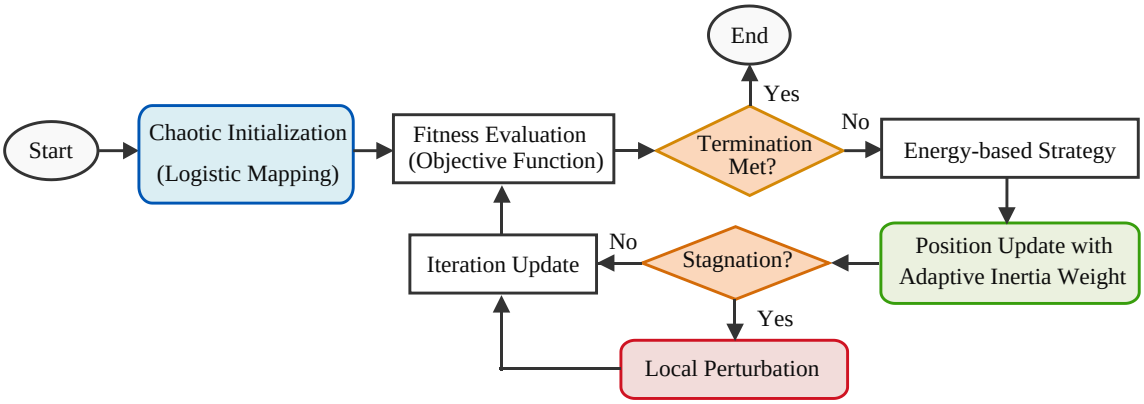


Figure 3: Framework of the IHHO

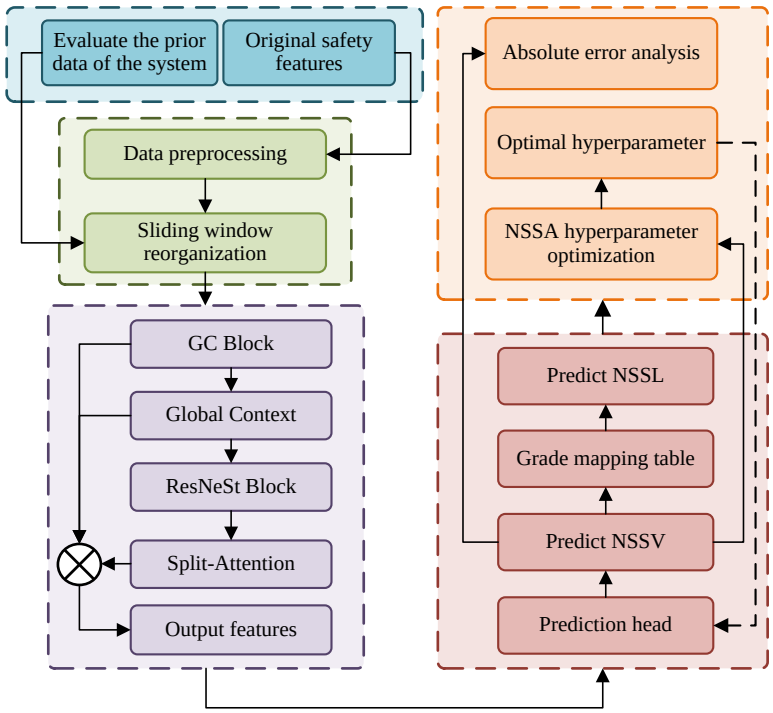


Figure 4: NSP model based on NSSA

3.2 NSP model based on NSSA

To address the limitations of existing NSP methods, including insufficient accuracy, slow convergence, and limited generalization, a new NSP model based on the NSSA method is proposed. A GResNeSt temporal feature extraction backbone integrates historical situation values as prior knowledge, while the Sparrow Search Algorithm (SSA) is used to globally optimize prediction hyperparameters. By employing a dual-output strategy that combines regression prediction with horizontal mapping, the NSS value and situation level can be jointly predicted, thereby improving the model's stability and generalization ability. The architecture of the NSP model is shown in Figure 4.

Figure 4 shows a NSP model based on NSSA. The model combines the original network traffic and log features, and inputs the improved GResNeSt backbone network for feature extraction after data preprocessing and sliding window reconstruction, as denoted in equation (8) [18].

$$\begin{cases} \tilde{x}_k = \sum_{c=1}^{C} a_{k,c}(x) \cdot x_c \\ g_i = F_i + W_{v2} \cdot ReLU(LN(W_{v1} \cdot GAP(F))) \end{cases} \tag{8}$$

Equation (8) is mainly composed of cross-feature interaction $\tilde{x}_k$ and long-range dependency modeling g_i. Among them, x_c represents the c th feature group, C refers to the total amount of feature groups, and $a_{k,c}$ denotes the attention weight. F denotes the input feature, F_i denotes its local feature, GAP represents the global contextual feature, W_{v1} and W_{v2} are two 1x1 convolutional layers, LN represents the normalization layer, and $\mathrm{Re}\,LU$ is the activation function.

In the prediction stage, the model outputs the quantified value of the safety status for the next period

through the regression head, and combines it with level mapping to complete situation prediction. At the same time, an improved SSA is introduced for global optimization of hyperparameters to balance convergence speed and prediction accuracy, as denoted in Equation (9).

$$\hat{y}_{t+1} = f(\mathrm{X}_t;\Theta) \tag{9}$$

In Equation (9), $\hat{y}_{t+1}$ represents the NSS value predicted at time $t+1$, X_t denotes the input feature sequence at time t, $f(\cdot)$ represents the entire GResNeSt prediction model, and Θ is the set of parameters to be trained for the model.

This model not only ensures that the predicted results are consistent with the actual situation, but also has higher stability and generalization ability, providing support for security warning and decision-making in complex network environments. Among them, the model feature optimization and parameter optimization strategies are denoted in Figure 5.

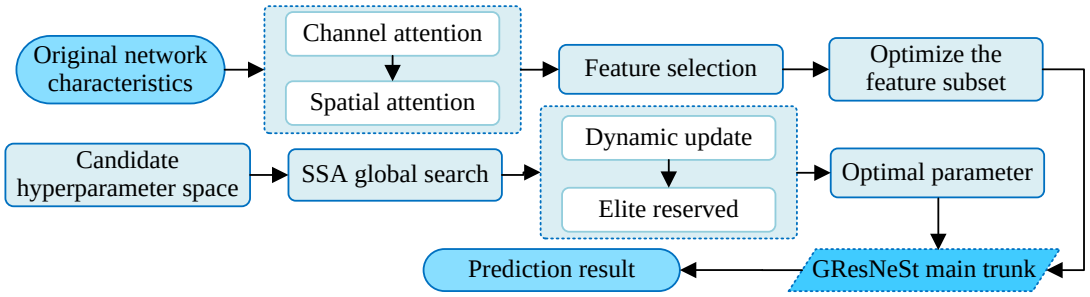


Figure 5: Feature optimization and parameter optimization strategies

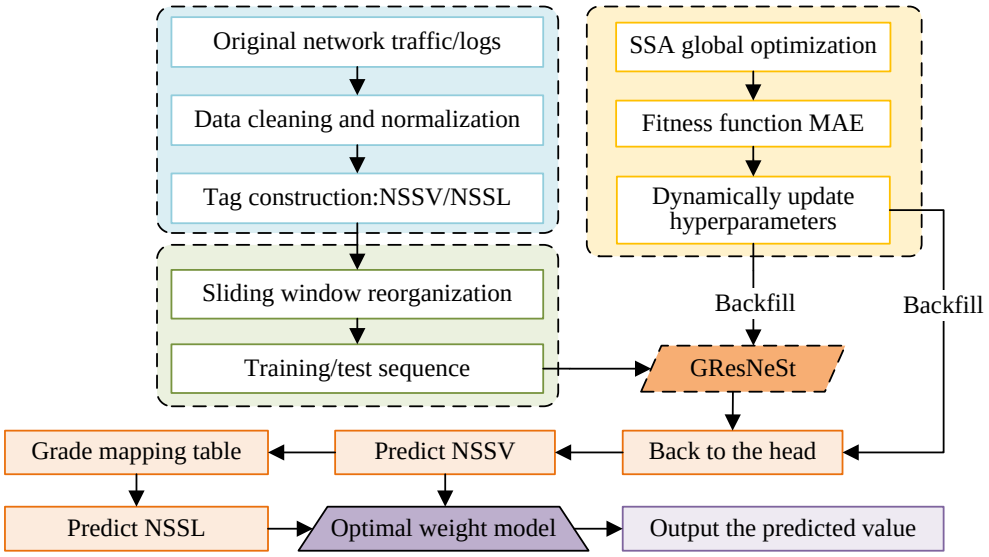


Figure 6: Training and implementation of NSP models

Figure 5 shows the feature optimization and parameter optimization strategies during the prediction phase. At the feature level, DA modeling is used to input data, channel attention is used to screen key dimensions, spatial attention is used to highlight abnormal areas, and PCA, statistical testing, and genetic operators are combined to further compress and screen features, as denoted in Equation (10) [19].

$$I_F = S_{select}\left(M_c(F)\otimes F + M_s(F_c)\otimes F_c\right) \tag{10}$$

In Equation (10), F is the original input feature matrix, M_c and M_s represent the channel and spatial attention weight matrices, $\otimes$ stands for the element wise multiplication operator, F_c denotes the channel weighted feature, S_{select} represents the feature selection function, and I_F is the final optimized feature subset output.

At the parameter level, the improved SSA performs global optimization on key hyperparameters such as learning rate, batch size, and regularization coefficient, using prediction error as the fitness function, balancing global search and local convergence, and avoiding falling into local optima through an elite retention mechanism, as denoted in Equation (11).

$$Fitness = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - f(Z_i;\Theta)\right| \tag{11}$$

In Equation (11), $Fitness$ represents the fitness value of the SSA, N refers to the amount of samples, y_i stands for the true NSS value of the i th sample, Z_i

means the input feature subset after feature optimization, and $f(\cdot)$ represents the GResNeSt prediction model

The synergistic effect of optimizing features and optimal parameters improves the convergence speed, generalization performance, and prediction accuracy of the prediction model. Ultimately, accurate prediction of NSS is achieved through training. The training and implementation of the prediction model is denoted in Figure 6.

Figure 6 shows the training and implementation of the NSP model. Firstly, the raw network traffic and log data are cleaned, normalized, and labeled, and combined with the previous state evaluation results, NSS values and NSS level labels are generated. Subsequently, the sliding window method is applied to restructure the data to capture short-term dynamics and long-term trend features, as denoted in Equation (12) [20].

$$\mathrm{D} = \{(X^{(i)}, y^{(i)}) | i = 1, 2, ..., M - s\} \quad (12)$$

In Equation (12), $X^{(i)}$ indicates the input feature sequence of the i th sample, s is the sliding window step size, $y^{(i)}$ means the corresponding prediction target, and M indicates the total amount of time steps.

In the model training stage, the optimized features are input into the GResNeSt backbone network, and the NSS value is output through the regression head. The level mapping table is utilized to predict the NSS level, as denoted in Equation (13).

$$\mathrm{Pr}\,edicted\ NSSL = \sum_{k=1}^{K} k \cdot \mathrm{I}\{\hat{y} \in Interval_k\} \quad (13)$$

In Equation (13), $\hat{y}$ is the predicted continuous NSS value, K means the total number of preset risk levels, $Interval_k$ is the numerical interval corresponding to the k th level, and $\mathrm{I}\{\cdot\}$ is the indicator function.

4 Results and analyses

4.1 NSSA and verification

To assess the reliability and usefulness of the end-to-end NSP via dual attention-enhanced situation assessment and IHHO-GResNeSt integration method, experimental performance verification and analysis were performed. Firstly, the experimental environment and configuration are denoted in Table 2 below.

Based on the experimental environment and parameters in Table 2, this study selected three publicly available benchmark datasets-NSL-KDD, UNSW-NB15, and CICIDS2017 to comprehensively validate the proposed method. NSL-KDD, released by the Canadian Institute for Cybersecurity, is a de-redundant version of KDDCup99, containing 41-dimensional traffic features and 5 major categories of intrusion behaviors, used to verify the model's performance on classic benchmarks. UNSW-NB15, built by the University of New South Wales, Australia, contains 49-dimensional network traffic features and 9 types of modern attack samples, reflecting a moderately complex real-world network environment. CICIDS2017, released by the Canadian Institute for Cybersecurity in 2017, covers various real-world attack types such as DoS, DDoS, Brute-Force, and Web attacks, and has traffic patterns that more closely resemble the distribution of actual network threats. Instead of using k-fold cross-validation, this study divided each dataset into training/validation/test sets based on a fixed ratio: 70% training set, 15% validation set, and 15% test set. All three datasets were independently partitioned as described above, without any cross-dataset training or joint normalization, to avoid potential data leakage risks.

Redundancy removal was applied to the original data, and Min-Max normalization parameters and class imbalance resampling ratios were fitted on the training set, then consistently applied to the validation and test sets to ensure fair and reproducible evaluation. This prevented interference across datasets and improved result comparability and stability. Each model was independently run five times under identical datasets and parameter settings, and the mean performance values were reported. The Wilcoxon signed-rank test was used to assess statistical significance between the proposed model and comparison models, with all performance differences passing the significance test ($p < 0.05$).

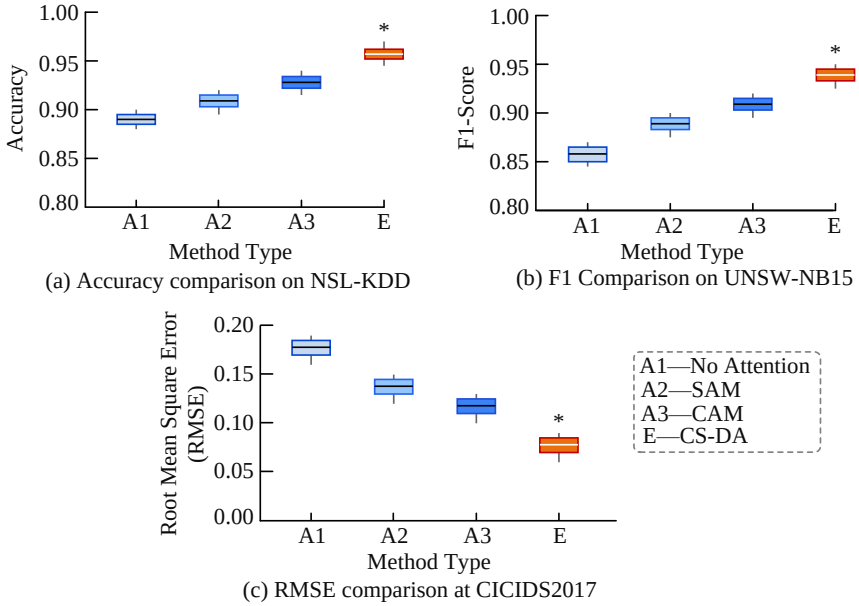
First, an ablation experiment using the dual attention mechanism was conducted to verify the contribution rates of channel/spatial/dual/no attention on different datasets. The ablation group was configured with No Attention, Channel Attention Network (CAM), and Spatial Attention Network (SAM), while the experimental group used CS-DA. The experimental results are shown in Figure 7.

As shown in Figure 7, on the NSL-KDD dataset, the proposed CS-DA method achieved an accuracy of 0.96, outperforming CAM (0.93), SAM (0.91), and No Attention (0.89). On the UNSW-NB15 dataset, CS-DA achieved an F1 score of 0.94, also higher than CAM (0.91) and SAM (0.89). On the CICIDS2017 dataset, CS-DA reduced the RMSE to 0.08, demonstrating superior situational awareness prediction accuracy compared to CAM (0.12), SAM (0.14), and No Attention (0.18). In summary, CS-DA enhances the model's ability to represent key threat information by jointly modeling feature dimension weights and anomaly spatial locations, fully demonstrating the core role of this mechanism in improving the accuracy and stability of situational awareness assessment.

Table 2: Experimental environment and key parameters

Experimental environment		Key configuration		
CPU	Intel i7-12700	Training	Optimizer	Adam
GPU	NVIDIA RTX 3080 10GB		Learning rate	1e-3
			Batch size	64
RAM	32GB		Epochs	150

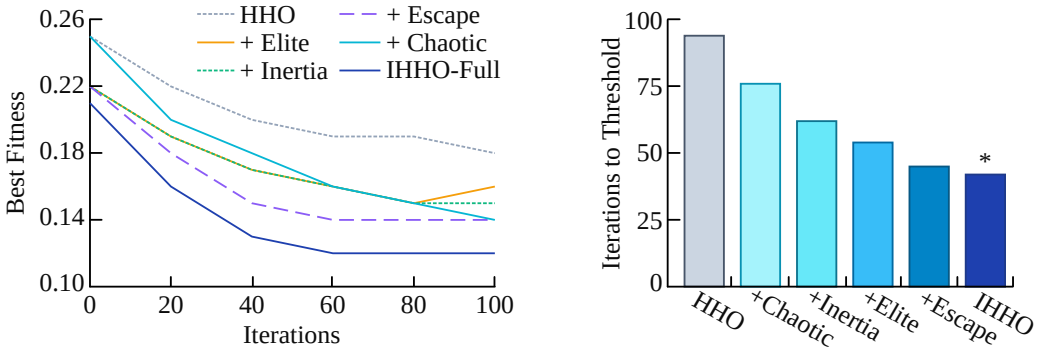
Storage	1TB SSD		Dropout rate	0.4
Operating system	Ubuntu 22.04 LTS		Embedding dimension	128
			Attention heads	4
Deep learning framework	PyTorch 1.10.2	IHHO	Population size	30
			Max iterations	100
CUDA version	CUDA 11.7		Chaotic map	Logistic
			Inertia weight	0.4 → 0.9
Python version	Python 3.10	SSA	Population size	30
			Max iterations	80



(a) Accuracy comparison on NSL-KDD (b) F1 Comparison on UNSW-NB15

(c) RMSE comparison at CICIDS2017

Figure 7: Analysis of ablation experiments using dual attention mechanisms



(a) Convergence Curves of IHHO Variants (b) Speed to Threshold (0.15)

Figure 8: IHHO improved module sub-component ablation analysis

Next, component ablation of the improved IHHO module was performed to evaluate the contribution rate of the improved IHHO. The ablation group and experimental group were configured as follows: HHO, +Chaotic init, +Inertia weight, +Elite retention and +Dynamic escape (Full IHHO). The experimental results are shown in Figure 8.

As shown in Figure 8, in terms of convergence quality, the proposed Full IHHO method achieved a final optimal fitness of 0.12, a 33.3% improvement over HHO's 0.18. The data comparison between +Chaotic (0.14) and +Escape (0.14) confirmed the crucial role of the initialization and perturbation modules in escaping local optima. Regarding convergence speed, Full IHHO reached the preset threshold of 0.15 in an average of only 42.25 iterations. Furthermore, the data for +Chaotic (76.45 iterations), +Inertia (62.18 iterations), and +Elite (54.30 iterations) showed a clear stepwise decreasing trend, quantifying the cumulative contribution of each component to the search efficiency. This demonstrates the synergistic gain effect of chaotic mapping, adaptive weights, elite preservation, and dynamic escape mechanisms, enabling IHHO to achieve faster and more accurate parameter optimization in the high-dimensional and complex network security feature space.

To quantitatively characterize the deviation between the network security situation assessment output value and the actual situation label, this study introduced Situation Score Deviation (SSD) as an evaluation index

in this experiment, which is defined as shown in Equation (14).

$$SSD = \frac{1}{N}\sum_{i=1}^{N} S_i^{pred} - S_i^{true} \quad (14)$$

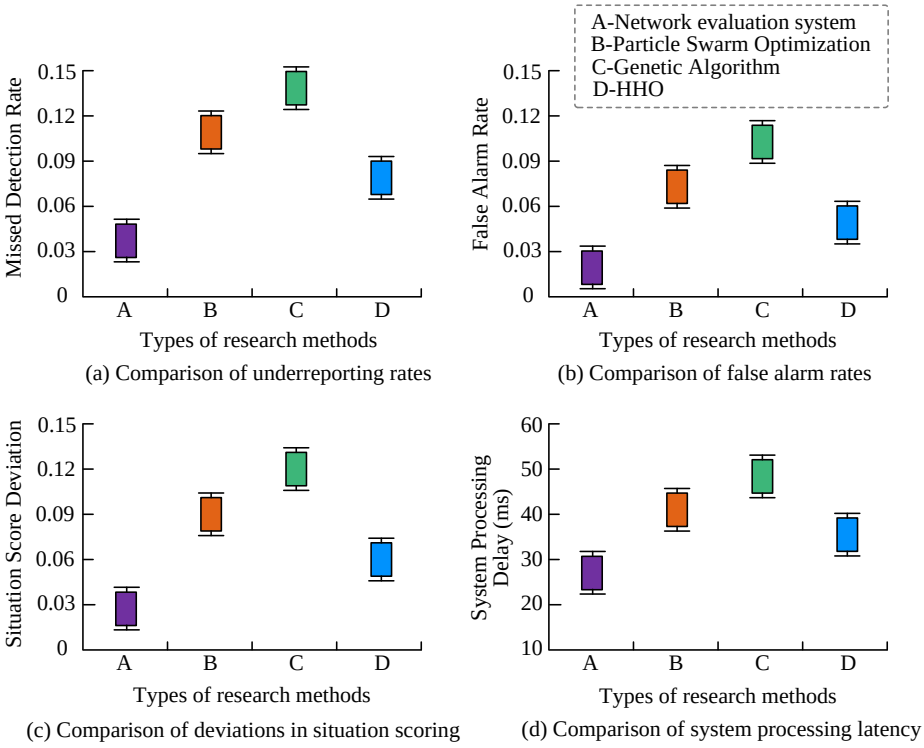
In Equation (13), S_i^{pred} represents the situation score output by the model for the i sample, S_i^{true} represents the corresponding real situation label or expert annotation score, and N is the total number of samples. This index is essentially the mean absolute error of the situation score, used to measure the accuracy of the model's assessment at the level of quantitative output of security situation. The smaller the value of SSD, the closer the situation assessment result is to the real state. The experimental results are shown in Figure 9.

In Figure 9, the network state evaluation method had the lowest False Negative Rate (FNR) (0.05) and False Positive Rate (FPR) (0.03), which were reduced by 0.04 and 0.03 respectively compared to the low performing original HHO method's FNR (0.09) and FPR (0.06), proving that it has the highest detection accuracy. Its situation rating deviation (0.04) was the smallest, which was 0.03 lower than the original HHO (0.07). It comprehensively verified the effectiveness and superiority of network state assessment methods in overall performance.

4.2 Validation of NSP model

After verifying the performance of the NSSA method, the study proceeded to validate the effectiveness of the NSP model. First, the superiority of SSA parameter tuning was verified. Grid Search, Bayesian Optimization, and Manual Tuning were selected as controls. The experimental parameters and environment were the same as above. The experimental results are shown in Table 3

As shown in Table 3, in terms of prediction accuracy, the accuracy of Proposed SSA after optimization reached 0.96, which is better than Bayesian Optimization (0.94), Grid Search (0.92), and Manual Tuning (0.89). Regarding search efficiency, Proposed SSA converged in only 64 trials, far fewer than Bayesian Optimization (120 trials), reducing the total hyperparameter tuning time from 14.26 hours for Grid Search to 2.25 hours for Proposed SSA. In terms of deployment performance, the inference latency of the model optimized by Proposed SSA was reduced to 32.16 ms, and the throughput was increased to 31.10 samples/s. This demonstrates that the SSA algorithm possesses extremely high search efficiency and global optimization capability in complex hyperparameter spaces, not only improving model performance but also balancing computational cost and engineering feasibility.



(a) Comparison of underreporting rates (b) Comparison of false alarm rates

(c) Comparison of deviations in situation scoring (d) Comparison of system processing latency

Figure 9: Overall performance verification of the network status assessment system

Table 3: Performance comparison of different hyperparameter tuning strategies

Evaluation dimension	Manual Tuning	Grid Search	Bayesian Optimization	Proposed SSA
Best Accuracy	0.89±0.02	0.92±0.01	0.94±0.01	0.96±0.01*
Search Budget (Trials)	25±4	400±0	120±8	64±6*
Total Tuning Time (h)	1.84±0.35	14.26±0.50	4.12±0.22	2.25±0.15*
Inference Latency (ms)	38.25±2.10	36.42±1.85	34.18±1.42	32.16 ± 1.10*
Throughput (samples/s)	26.14±1.25	27.45±1.10	29.26±0.85	31.10±0.92*

Note: * indicates that the proposed method is significantly superior to other methods (p<0.05).

Table 4: Performance and computational cost comparison of different methods

Evaluation dimension	CNN-LSTM	XGBoost	Transformer	Proposed model
Root Mean Square Error(RMSE)	0.17±0.01	0.20±0.02	0.15±0.01	0.08±0.01*
F1-Score	0.88±0.01	0.85±0.01	0.90±0.01	0.94±0.01*
Train time / epoch (s)	2.40±0.20	1.90±0.20	3.90±0.30	3.60±0.30*
Peak GPU memory (GB)	2.10±0.10	1.80±0.10	3.60±0.20	3.40±0.20*
Inference latency (ms)	41.80±2.10	36.50±1.80	48.20±2.40	32.16±1.90*

Note: * indicates that the proposed method is significantly superior to other methods (p<0.05).

Finally, the experiment compared the overall performance and computational cost of the prediction model. The control group was the same as above. The experimental results are shown in Table 4.

As shown in Table 4, the proposed model outperformed CNN-LSTM, XGBoost, and Transformer in both RMSE (0.08) and F1-Score (0.94), indicating higher prediction accuracy in cybersecurity situation prediction tasks. Meanwhile, despite the introduction of dual attention and IHHO parameter optimization mechanisms, the proposed model maintained manageable single-round training time (3.60 s) and peak memory usage (3.40 GB), not significantly exceeding that of deep learning models like Transformer. Furthermore, its inference latency was the lowest (32.16 ms), better than the comparison methods, demonstrating good real-time performance and engineering deployability. In summary, the proposed model achieves a good balance between prediction performance and computational efficiency, making it suitable for security situation prediction tasks in complex and dynamic network environments.

5 Discussion

This study addresses the limitations of existing NSP methods in insufficient feature extraction and low parameter optimization efficiency, and proposes an end-to-end NSP method based on CS-DA enhanced situation assessment and an IHHO-GResNeSt ensemble. Compared with PSO-based optimization methods [9][10], which are sensitive to initial populations and prone to premature convergence in high-dimensional non-convex spaces, the IHHO-based model achieves higher prediction accuracy (0.96) and lower RMSE (0.08) on NSL-KDD, UNSW-NB15, and CICIDS2017 datasets. This indicates that IHHO improves global search capability and convergence stability through chaotic initialization, adaptive inertial weights, and stagnation perturbation, outperforming traditional PSO in network security feature parameter optimization.

Compared with CNN-LSTM-based temporal modeling methods [11], the proposed method exhibits superior feature enhancement and threat locality modeling. CNN-LSTM lacks explicit modeling of feature importance and abnormal region salience, making it vulnerable to noise in complex attack scenarios. By introducing CS-DA in the state evaluation stage, the proposed method adaptively enhances discriminative feature dimensions and highlights abnormal regions, providing more informative inputs for prediction and achieving higher accuracy and F1-Score than CNN-LSTM-like methods.

Furthermore, compared with Transformer-like models, the proposed method demonstrates better stability and generalization under small-sample and high-noise conditions. While Transformers excel at long-range dependency modeling, they are sensitive to data scale, class imbalance, and noise. By integrating quantified situation values as prior knowledge within the IHHO-GResNeSt framework, the proposed method jointly exploits raw traffic features and historical situation information, mitigating performance degradation under distribution drift and sudden attack scenarios.

In summary, compared to representative state-of-the-art methods such as PSO-type optimization methods, CNN-LSTM, and Transformer, the performance advantage of the proposed method does not stem from improvements to a single module, but rather from the synergistic effect of CS-DA feature enhancement, IHHO parameter optimization, and the integrated assessment-prediction structure.

6 Conclusion

Addressing insufficient feature extraction, low parameter optimization efficiency, and limited prediction stability in existing NSP methods, this study proposes an end-to-end NSP framework based on CS-DA–enhanced situation assessment and IHHO-GResNeSt integration. CS-DA and IHHO are first used to construct a network security status assessment model for deep threat feature characterization and global parameter optimization. The quantified situational values are then introduced into the GResNeSt network for time-series feature learning and prediction, with regression prediction and level mapping combined to jointly output situational values and risk levels. Experimental results show that CS-DA achieves an accuracy of 0.96 on the NSL-KDD dataset, outperforming CAM (0.93), SAM (0.91), and NO Attention (0.89), validating the complementary gains of channel weighting and spatial saliency enhancement. At the system level, the proposed method attains an FNR of 0.05, an FPR of 0.03, and an inference latency of 32.16 ms, demonstrating high assessment accuracy and real-time performance. In prediction validation, the SSA-tuned model achieves a best accuracy of 0.96 and outperforms CNN-LSTM, XGBoost, and Transformer in RMSE (0.08) and F1-Score (0.94), while maintaining the lowest inference latency (32.16 ms), indicating good engineering deployability. Overall, the proposed method achieves a favorable balance between prediction performance and computational efficiency and is suitable for security situation prediction in complex and dynamic network environments. Future work will extend

validation to real-world network scenarios to support large-scale real-time deployment.

References

[1] Sofiane Lagraa, Martin Husák, Hamida Seba, Satyanarayana Vuppala, Radu State, and Moussa Ouedraogo. A review on graph-based approaches for network security monitoring and botnet detection. International Journal of Information Security, 23(1):119-140, 2024. https://doi.org/10.1007/s10207-023-00742-7

[2] Akoh Atadoga, Enoch Oluwademilade Sodiya, Uchenna Joseph Umoga, and Olukunle Oladipupo Amoo. A comprehensive review of machine learning's role in enhancing network security and threat detection. World Journal of Advanced Research and Reviews, 21(2):877-886, 2024. https://doi.org/10.30574/wjarr.2024.21.2.0501

[3] Li Yang, Mirna El Rajab, Abdallah Shami, and Sami Muhaidat. Enabling automl for zero-touch network security: Use-case driven analysis. IEEE Transactions on Network and Service Management, 21(3):3555-3582, 2024. https://doi.org/10.1109/TNSM.2024.3376631

[4] Deepika Saxena, Ishu Gupta, Rishabh Gupta, Ashutosh Kumar Singh, and Xiaoqing Wen. An AI-driven VM threat prediction model for multi-risks analysis-based cloud cybersecurity. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 53(11):6815-6827, 2023. https://doi.org/10.1109/TSMC.2023.3288081

[5] Anuradha Kanade, Chitra Sabapathy Ranganthan, Jyothi Babu A, Ramachandran G, Ashok Kumar Kusuma, Manav Anand, and Lokeswar Reddy DV. Analysis of wireless network security in internet of things and its applications. Indian Journal of Engineering, 21(55):1-12, 2024. https://doi.org/10.54905/disssi.v21i55.e1ije1675

[6] Fahad H. Alshammari. Design of capability maturity model integration with cybersecurity risk severity complex prediction using bayesian-based machine learning models. Service Oriented Computing and Applications, 17(1): 59-72, 2023. https://doi.org/10.1007/s11761-022-00354-4

[7] Md Abu Sayed, Badruddowza, Sonjoy Kumar Dey, Md Shohail Uddin Sarker, Abdullah Al Mamun, Norun Nabi, Fuad Mahmud, Md Khorshed Alam, and Md Tarek Hasan. Comparative analysis of machine learning algorithms for predicting cybersecurity attack success: A performance evaluation. The American Journal of Engineering and Technology, 6(9):81-91, 2024. https://doi.org/10.37547/tajet/Volume06Issue09-10

[8] Harun Bangali, Paul Rodrigues, V. Pandimurugan, S. Rajasoundaran, S. V. N. Santhosh Kumar, M. Selvi, and A. Kannan. Prediction of middle box-based attacks in Internet of Healthcare Things using ranking subsets and convolutional neural network. Wireless Networks, 30(3):1493-1511, 2024. https://doi.org/10.1007/s11276-023-03603-2

[9] Han Qi, Xinyu Liu, Abdullah Gani & Changqing Gong. Quantum particle swarm optimized extreme learning machine for intrusion detection. The Journal of Supercomputing, 80(10):14622-14644, 2024. https://doi.org/10.1007/s11227-024-06022-y

[10] Theyab R. Alsenani, Safial Islam Ayon, Sayeda Mayesha Yousuf, Fahad Bin Kamal Anik, and Mohammad Ehsan Shahmi Chowdhury. Intelligent feature selection model based on particle swarm optimization to detect phishing websites. Multimedia Tools and Applications, 82(29):44943-44975, 2023. https://doi.org/10.1007/s11042-023-15399-6

[11] Junyi Li, Yongdong Wu, Yang Li, Ziwen Zhang, Hassan Fouad, and Torki Altameem. A network security prediction method based on attack defense tree. Journal of Nanoelectronics and Optoelectronics, 18(3):357-366, 2023. https://doi.org/10.1166/jno.2023.3398

[12] Liu Zhang, and Yanyu Liu. Network security prediction and situational assessment using neural network-based method. Journal of Cyber Security and Mobility, 12(4):547-568, 2023. https://doi.org/10.13052/jcsm2245-1439.1245

[13] Yueying Li, and Feng Wu. Improved population intelligence algorithm and BP neural network for network security posture prediction. International Journal of Distributed Sensor Networks, 2023(1):9970205-9970206, 2023. https://doi.org/10.1155/2023/9970205

[14] Jun Li, Noel B. Linsangan, and Huiguo Dong. Campus network traffic prediction and anomaly detection based on deep learning. International Journal of Emerging Technologies and Advanced Applications, 1(7):8-13, 2024. https://doi.org/10.62677/IJETAA.2407123

[15] Mouaad Mohy-eddine, Azidine Guezzaz, Said Benkirane, and Mourade Azrour. An efficient network intrusion detection model for IoT security using K-NN classifier and feature selection. Multimedia Tools and Applications, 82(15):23615-23633, 2023. https://doi.org/10.1007/s11042-023-14795-2

[16] M. P. S. Bhatia, and Saurabh Raj Sangwan. Soft computing for anomaly detection and prediction to mitigate IoT-based real-time abuse. Personal and Ubiquitous Computing, 28(1):123-133, 2024. https://doi.org/10.1007/s00779-021-01567-8

[17] Mohammed Sadoon Hathal, Basma Mohammed Saeed, Dina A. Abdulqader, and Firas Mahmood Mustafa. Attack and anomaly prediction in networks-on-chip of multiprocessor system-on-chip-based IoT utilizing machine learning approaches. Service Oriented Computing and Applications, 18(3):209-223, 2024. https://doi.org/10.1007/s11761-024-00393-z

[18] Isaiah J. King, and H. Howie Huang. Euler: Detecting network lateral movement via scalable temporal link prediction. ACM Transactions on Privacy and Security, 26(3):1-36, 2023. https://doi.org/10.1145/3588771

[19] Wei Liang, Yuhui Li, Jianlong Xu, Zheng Qin, Dafang Zhang, and Kuan-Ching Li. Qos prediction and adversarial attack protection for distributed services under dlaas. IEEE Transactions on Computers, 73(3):669-682, 2023. https://doi.org/10.1109/TC.2021.3077738

[20] Yubao Wu. Construction and application of internet of things network security situation prediction model based on BiLSTM algorithm. Journal of Cyber Security and Mobility, 13(5):843-861, 2024. https://doi.org/10.13052/jcsm2245-1439.1352