Proceedings of CONF-CIAP 2026 Symposium: International Conference on Atomic Magnetometer and Applications
DOI: 10.54254/2753-8818/2026.HZ29905

SIEVE: A Hybrid-Model-Based System for SIEM False-Positive Optimization

Senyou Shi

School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, Australia
sshi0668@uni.sydney.edu.au

Abstract. Security Information and Event Management systems are key infrastructures for identifying and responding to threats in enterprise security operations. By centrally processing logs from multiple sources, SIEM systems improve attack traceability and response effectiveness. However, significant research and empirical studies show that SIEM systems generally present a high rate of false positives after deployment. Currently, false positives are mainly reduced using single-model approaches. These approaches can reduce the number of false positives to some extent but are often at a disadvantage regarding generalization capabilities and feature utilization. To address this challenge, this paper proposes a hybrid modeling method, which combines TF-IDF with Transformer, Echo State Network, Random Forest, and XGBoost. Through multimodal feature modeling and the combined use of model mechanisms, this approach can achieve synergistic utilization of temporal dependencies, semantic context, and structural features. Experimental validation on Hillstone Networks' enterprise-level real threat logs and the public Advanced SIEM Dataset demonstrates a 60.72% to 7.58% reduction in false positives, with stable performance and strong robustness. The research findings provide a feasible and scalable engineering pathway for intelligent SIEM false positive optimization.

Keywords: SIEM false-positive detection, Hybrid Model, Echo State Network, Transformer, Random Forest

1. Introduction and background

SIEM systems have already become one of the four walls of the Security Operation Centers; the high false-positive rate has plagued them for years and even decades: the rate of some systems is more than 60% and even close to 99% in the worst case [1]. The influence of the false-positive rate is reflected in the credibility of the alert and the efficiency of its response.

In order to solve this problem, many studies try to apply machine learning approaches to cut down the false alarm rate. Mozalbat uses TF-IDF and deep learning with the logic that there is not much difference between the two methods in essence, and the two can complement each other to improve false alarm detection [2]. For log sequence, Du et al. proposed DeepLog with RNN to catch the temporal dependency and reached higher accuracy in a complex log environment [3]. The transformer also starts to be widely used, according to work fusing the semantic, sequential order, and timestamp information also reminds us of the benefit of multidimensional information merging

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ29905

to detect anomalies [4]. For multi-stage attack scenarios, many works use sequence-to-sequence or attention-based methods to aggregate and predict the complex attack cases [5].

Many recent works start to explore more complicated structures. RuleGenie uses LLM to automatically optimize the SIEM rule set to eliminate redundant and overlapping rules to finally reduce the false alarm [6]. Irshad et al. demonstrated that ensemble learning and correlated feature integration enhance robustness and reduce false alarms [7]. Hang et al. use a cascaded transformer model to design a logformer to catch the hierarchical contextual information to increase the detection in those multiple log datasets [8]. Meanwhile, LLM and log analyzing are combined to propose logllm by Guan et al., which also opens a new way [9]. Pospieszny et al. introduced ADALog, which adaptively employs masked language models for unsupervised log anomaly detection, reducing dependence on manual annotation [10].

Despite the progress, several limitations remain: most are based on a single model, which usually generalizes badly; semantic modeling and temporal are detached, so multimodal features are not fully exploited; and for enterprise logs with heterogeneous mixed features with Chinese and English semantics, the performance gap is great. To overcome these challenges, this paper proposes a hybrid model integrating ESN, TF-IDF, Transformer, RF, and XGBoost. The main contributions are:

This paper proposes a hybrid modeling architecture that integrates semantic, keyword, and structural features through feature-level fusion and sequence modeling, addressing the limitations of single-model approaches in generalization and robustness.

This paper designs a feature processing workflow for Chinese-English hybrid and heterogeneous field data, covering time-unification, dual-channel text normalization, long-tail category merging, and behavioral statistical extraction, ensuring adaptability and reproducibility in enterprise-level SIEM log environments.

Through engineering implementation and evaluation, this study demonstrates the effectiveness of hybrid modeling in narrowing the performance gap between standardized experimental environments and real-world operational scenarios, providing a scalable implementation path.

2. Datasets

2.1. Dataset overview

This study uses two types of complementary alarm log data to evaluate the method's effectiveness and robustness in both real and standardized environments.

The first dataset is Hillstone Networks Threat Log of enterprise grade, collected from the SIEM platform of real Security Operations Centers. These JSON-structured records record alert events within a device or across devices and links, where titles were titleAgg and alerts contained mixed Chinese-English text notes with device vendor-specific fields and non-standard expressions. The typical fields included title, description, evidence text, source/destination ports, alert direction, application name, +severity, and event status. The evidence paragraphs always mix CEF logs, CVE identifiers, and policy details such as "SonarQube API information leakage vulnerability (CVE-2020-27986)" with device interfaces, timestamps, and threat severity levels.

The second dataset is the Advanced SIEM Dataset [11], an open-source synthetic database provided by HuggingFace that includes 100,000 JSON-structured records. This dataset simulates multi-type security incident records in an SIEM environment and provides standard ISO8601 timestamps, event types, tool sources, and severity levels as well as raw CEF logs. Unlike real enterprise data, this dataset includes an entry-to-field map, well-structured format, clear semantics, and consistent labels, serving as a standard benchmark for model performance.

DOI: 10.54254/2753-8818/2026.HZ29905

2.2. Data processing

This paper designs a feature processing workflow for Chinese-English hybrid and heterogeneous field data, covering time-unification, dual-channel text normalization, long-tail category merging, and behavioral statistical extraction, ensuring adaptability and reproducibility in enterprise-level SIEM log environments. Through engineering implementation and evaluation, this study demonstrates the effectiveness of hybrid modeling in narrowing the performance gap between standardized experimental environments and real-world operational scenarios, providing a scalable implementation path. Figure 1 shows the data processing module structure.

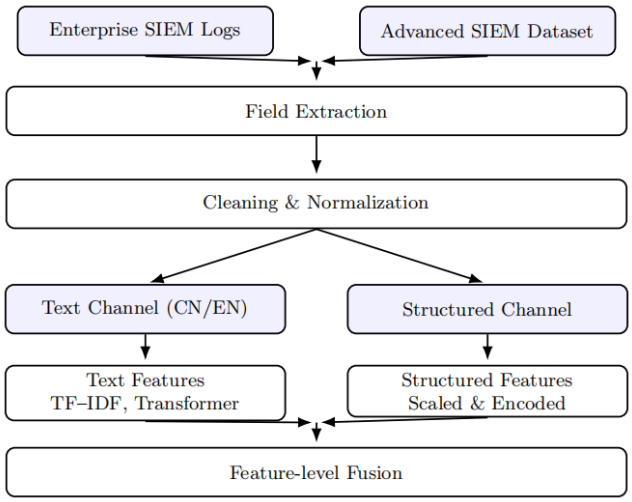

Figure 1. Data processing module structure (picture credit: original)

In the field extraction phase, enterprise data is processed to pull titleAgg, description, and evidenceText as core text content from JSON while holding srcPort, dstPort, direction, appName, severityStr, and eventStatus as structured fields. Event_type, severity, source, and raw_log are synchronously extracted from the Advanced SIEM Dataset, unpacking key numerical and categorical data such as risk_score, confidence, baseline_deviation, and sequence_anomaly from advanced_metadata and behavioral_analytics. This ensured the consistent sampling of both data types at the core comparable fields + unique fields level.

Further, data cleansing and standardization followed. The text content is preprocessed, with redundant whitespace, control characters, and HTML tags removed, and IP addresses and URLs standardized. Redundant CVE pieces and templated alert prefixes are identified and deduplicated. The missing fields were gently filled with interpretability-first strategies, with numerical features filled by median or business constant, and categorical features kept the “unknown” label explicitly. Timestamps are first normalized to a time zone and are boundary validated, given that a small portion of the Advanced SIEM Dataset is intentionally constructed with temporal anomalies and 'N/A’ placeholders. Target IPs are either replaced by instances or certainly filtered out by rule-based preprocessing, ensuring that time series modeling and graph extraction are free from noise.

Mixed Chinese-English scenarios require a dual-channel strategy for text processing and representation learning. The Chinese channel utilizes domain-specific vocabulary and a self-built security terminology dictionary for tokenization, retaining critical fragments of attack payloads, MITRE ATT&CK tactics, and device policies, both for CEF. The English channel employs simple tokenization with regular expressions and subword decomposition, case insensitivity, normalization,

DOI: 10.54254/2753-8818/2026.HZ29905

and merging for numerical values. In addition, both channels use TF-IDF sparse representations for initial acquisition of common separable keywords, followed by contextual semantic vector attainment through the Transformer encoder. This provides a better distributed representation while maintaining the separation of sparsity in interpretability.

The same regular expression rules, along with additional ones, apply to CEF fragments in raw logs to match and extract field-key value pairs, reducing the burden of unstructured data. Continuous variables such as risk scores, confidence levels, and entropy undergo interval truncation and normalization. For categorical variables, such as event types, severity levels, application names, and attack directions, label encoding or first-order one-hot vectors are used, driven by a specific length. Other long-tail enumerated items are frequency-truncated and added to "other" categories to suppress sparse noise.

Furthermore, to maintain the feature value of the enterprise data, vendor-specific fields are stripped of all but threat-semantics-related subsets, which eliminates pseudo-correlated features highly correlated with policy templates. Lastly, the text's TF-IDF and Transformer semantic feature representations are vertically aligned and concatenated with the structured numeric data and categorical features. The resultant high-dimensional feature fusion network is unified across enterprises to serve as the low-order model input. The feature-level fusion strategy for the final feature conglomeration not only overcomes the problems caused by the instability of the internal model due to coupling but also allows the temporal model and semantics to share the same conceptual context. In the application, the final aligning network effectively filters the noise of the Chinese identity, English templating, and equipment manufacturer differences in real-world enterprise-level data, such that in the corpora, it is clearer with intraday separability and interpretability. Thus, as the low-order model, it ensures the stability of the ESN.

3. Method

3.1. System structure

The system consists of three core components: a data processing module, a hybrid model analysis module, and a discrimination module, as illustrated in Figure 2.

DOI: 10.54254/2753-8818/2026.HZ29905

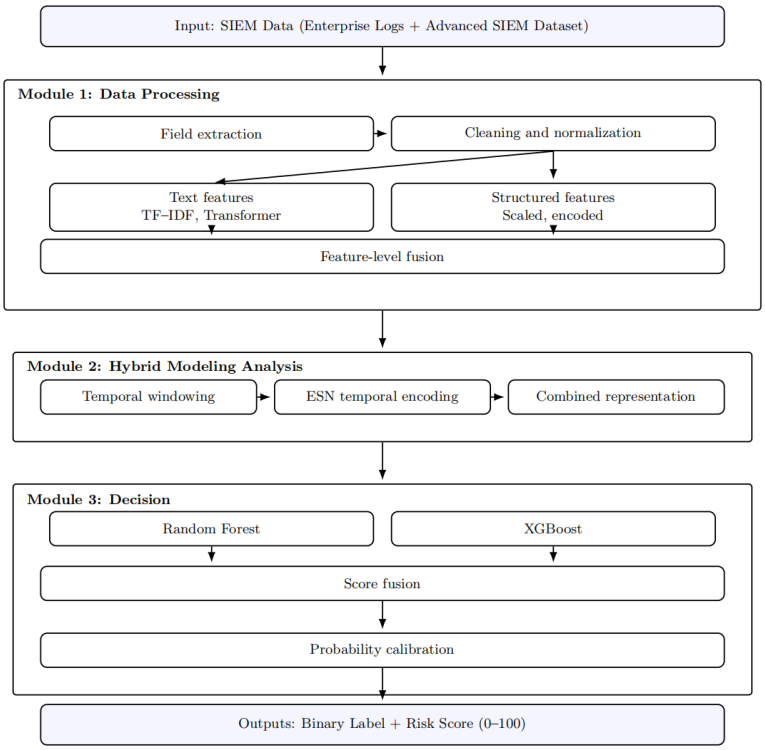

Figure 2. System structure (picture credit: original)

The system operates in a "input-processing-modeling-output" sequence, using generic SIEM logs as input. The input comprises two types: textual evidence (including titles, descriptions, evidence paragraphs, and partial CEF key-value pairs, MITRE technical identifiers, CVE tags, IP/URLs) and structured signals (ports, directions, applications, severity levels, entity/session identifiers, risk scores, confidence levels, and behavioral statistics). The processing stage normalizes and aligns temporal, textual, and structural fields, producing "clean, aligned" dual-channel data ready for modeling. During modeling, three feature types (TF-IDF keyword weights, Transformer semantic vectors, and structured vectors) are extracted per time step and fused at the feature level. These are then combined into sliding windows across entities and time dimensions, inputting into an Echo State Network (ESN) to generate temporal representations. Finally, category probabilities are predicted by a “random forest + XGBoost” ensemble model. The output is monotonic calibration and then mapped to a 0–100 risk score. Deployment thresholds are set to achieve the target false positive rate on the validation set, ensuring that certain performance characteristics are maintained and controllable across scenarios.

3.2. Data processing

This study's data processing addresses three types of consistency: temporal consistency, textual consistency, and structural consistency. First, timestamps from all sources are unified to the same time zone and sorted by entity in ascending order. The "future/exception" time points are trimmed to ensure temporal monotonicity in any sliding window [t−T+1, t]. Second, the text adopts a "dual-channel" standardization: the Chinese channel retains key symbols like CVE, IP, URL, and MITRE technical identifiers through word segmentation and domain-specific vocabulary, while removing templated redundancies; the English channel merges case and numeric variations to reduce

DOI: 10.54254/2753-8818/2026.HZ29905

superficial variations caused by supplier templates. After merging the two channels, the processed text enters the feature layer as a "space-separated string" to avoid secondary segmentation during vectorization. Third, structure variables implement robust scaling and enumeration specifications. For risk components $r_{k,t}$ and confidence levepplied $c_{k,t}$:

$$\hat{r}_{k,t} = \frac{r_{k,t} - \text{median}(r_{k,t-w:t-1})}{\text{MAD}(r_{k,t-w:t-1}) + \varepsilon} \tag{1}$$

$$\hat{c}_{k,t} = \frac{c_{k,t} - \text{median}(c_{k,t-w:t-1})}{\text{MAD}(c_{k,t-w:t-1}) + \varepsilon} \tag{2}$$

Here, k represents the entity key (prioritized session ID, degrading to $(\text{srcIP}, \text{dstIP}, \text{app})$ when unavailable), w denotes the rolling window, and ε >0 prevents denominator division by zero. Categorical variables (such as severity, event type, application, direction, etc.) are merged into a long-tail "OTHER" category after standardization to mitigate the impact of extremely rare items on model stability. Missing values and placeholders are uniformly labeled as "sentinel" or "unknown" to ensure subsequent feature engineering avoids false correlations.

3.3. Hybrid model

3.3.1. Feature extraction

Text features emphasize the balance between "explainability" and "contextual consistency". While explainability is captured through TF-IDF, this study implements term weighting based on security domain keyword meta-importance: Let $\mathcal{V}_t$ denote the set of lexical items at time step t , and define for any term $\tau \in \mathcal{V}_t$:

$$tfidf^{\star}(\tau, t) = \alpha(\tau) \cdot tf(\tau, t) \cdot log\frac{|\mathcal{D}|}{1+|\{d \in \mathcal{D}: \tau \in d\}|} \tag{3}$$

The $\alpha(\tau)$ is determined by lexical categories (e.g. $\alpha(\text{CVE_ID}) > \alpha(\text{WORD}) \geq \alpha(\text{STOP})$) to enhance the weight of security-critical signals such as CVE numbers, MITRE technical IDs, IPs, and URLs. Contextual consistency is provided by the Transformer encoder: The cleaned subword sequences are encoded through multi-head self-attention and feedforward layers, with sentence-level semantic representations x_t^{tr} obtained via average pooling or $[\text{CLS}]$. The two types of text vectors complement each other: the former highlights the intensity of key terms, while the latter aggregates synonymous templates. Structural features are composed of continuously normalized variables, short-window statistics (such as $\Delta r_{k,t} = r_{k,t} - r_{k,t-1}, \text{Var}_w(r_{k,\cdot})$), and encoded enumerated variables, collectively forming $x_{k,t}^{\text{struct}}$. At the same time step, the TF-IDF sparse vector x_t^{tfidf} , the Transformer dense vector x_t^{tr} , and the structurally dense vector $x_{k,t}^{\text{struct}}$ are concatenated in a fixed order to form a unified input:

$$x_{k,t} = \left[x_t^{tfidf} \; ; \; x_t^{tr} \; ; \; x_{k,t}^{struct} \right] \tag{4}$$

This provides clear and well-aligned single-step representations for temporal modeling.

DOI: 10.54254/2753-8818/2026.HZ29905

3.3.2. Temporal modeling

Temporal modeling employs ESN. Unlike LSTM/GRU which require backpropagation through the recurrent layer, ESN focuses solely on output learning, making it ideal for hybrid inputs with "high-dimensional sparsity + heterogeneous structure". For the sliding window sequence $\{x_{k,t-T+1}, \ldots, x_{k,t}\}$ of entity k, ESN generates nonlinear state trajectories through fixed random recurrent layers. To stabilize dynamic behavior, recurrent weights are normalized by spectral radius $\rho \in (0,1]$, and a leakage rate $\alpha_{res} \in (0,1]$ is set to control effective memory length. The update at time step t is:

$$\tilde{h}_{k,t} = tanh\left(W_{in} x_{k,t} + W h_{k,t-1}\right) \tag{5}$$

$$h_{k,t} = (1 - \alpha_{res}) h_{k,t-1} + \alpha_{res} \tilde{h}_{k,t} \tag{6}$$

The window's sequence-level representation adopts either the final state $h_{k,t}$ or the average state $\bar{h}_{k,t}$. In SIEM environments, ESN effectively identifies three common patterns at minimal cost: low-amplitude oscillations from template-triggered repetitions, state drift caused by concurrent session risks and evidence accumulation, and isolated spikes from device/time jitter (smoothed by α_{res} 's low-pass characteristics). To fully leverage the 'current evidence + recent trajectory' approach, the sequence state and current step input are concatenated into:

$$z_{k,t} = [h_{k,t} \; ; \; x_{k,t}] \tag{7}$$

Serving as the discriminator's final feature.

3.3.3. Integrated discrimination

The system employs a tree-based dual-integration approach to balance robustness and boundary modeling capabilities. Random Forest constructs a set of low-correlation decision trees on $z_{k,t}$ datasets, generating probability estimates $\hat{p}_{RF}(k,t)$ through voting, maintaining stability under varying scales, outliers, or missing data. XGBoost optimizes the logistic loss function using additive tree models, leveraging second-order approximations to aggregate gradients and curvature at leaf nodes, thereby achieving more refined splits near complex boundaries to obtain $\hat{p}_{XGB}(k,t)$. The two results are fused using a linear coefficient λ selected from the validation set:

$$\hat{p}(k,t) = \lambda \hat{p}_{RF}(k,t) + (1 - \lambda) \hat{p}_{XGB}(k,t), \quad \lambda \in [0,1] \tag{8}$$

Compared with a single fully connected output layer, the integration of heterogeneous features such as "sparse terms + dense semantics + enumerated variables" is more natural and robust, and the parameter tuning cost is lower. Compared with a single path, the fusion further improves the probability stability under cross-domain and class imbalance conditions.

3.4. Inference and output

The reasoning process is executed online along the same training path. New events first undergo time and field normalization, generating three types of features: TF-IDF, Transformer, and structural

DOI: 10.54254/2753-8818/2026.HZ29905

features, which are concatenated into $x_{k,t}$. Subsequently, the ESN state is updated using an entity sliding window to obtain $h_{k,t}$, forming $z_{k,t}$. Through integrated judgment, the non-calibrated probability $\hat{p}(k,t)$ is derived. To achieve stable risk scaling across scenarios, a monotonic calibration function $\psi(\cdot)$ (either Platt scaling or order-preserving regression) is fitted on an independent validation set, yielding the calibrated probability $\tilde{p}(k,t) = \psi(\hat{p}(k,t))$). This process defines the risk score:

$$R(k,t) = 100\,\tilde{p}(k,t) \in [0,100] \tag{9}$$

Search for the threshold on the validation set based on the operation-side set false positive rate (α) :

`The deployment process yields three outputs: calibration probability $\tilde{p}(k,t)$, risk score $R(k,t)$, and threshold-based classification label $\mathbb{I}\{\tilde{p}(k,t) \geq \tau^{\star}\}$. The metrics expressly prioritize system responses and support batch filtering, while the threshold ensures false positive uniformity well over differentiated data and workload distributions. The method combines text-structured feature fusion, ESN temporal augmentation, tree model ensemble, and probability calibration to effectively overcome SIEM challenges.

4. Experiments and results

4.1. Experimental setup

Experiments are conducted on both the enterprise Hillstone SOC dataset and the Advanced SIEM Dataset. A unified pipeline is used such that entity-level bucketing of the data is based on session IDs or IP–application combinations.

The dataset is split into 70%, 10%, and 20% for training, validation, and testing.

Dual-channel text cleaning and normalization are implemented, after which both structured features are scaled and encoded. Finally, the model calibration, with a target false-positive rate of 8%, is run.

Key hyperparameters are as follows: TF–IDF with a vocabulary size of 30,000, and unigrams and bigrams; Transformer with an embedding dimension of 768, and a maximum token length, L, of 256; ESN with a reservoir size of 600, a spectral radius of 0.8, leaking rate of 0.3, and window length T = 5–10; Random Forest with 500 trees and a maximum depth of 20; XGBoost with 600 trees, a learning rate of 0.05, and a maximum depth of 8; An ensemble weight of $\lambda = 0.5$.

4.2. Experimental results

4.2.1. Ablation study on enterprise logs

This paper ablate the marginal contribution by progressively removing or replacing components of SIEVE and all results are given in Table 1.

DOI: 10.54254/2753-8818/2026.HZ29905

Table 1. Module ablation (hillstone SOC)

Experimental Setup	Accuracy	Recall	F1	FPR
Full system	92.4%	97.9%	94%	7.6%
w/o ESN (no temporal)	88.3%	93.7%	90%	12.5%
w/o Transformer	89.1%	95.0%	91%	11.3%
w/o TF–IDF	90.1%	96.1%	92%	10.2%
Structure only	84.9%	90.4%	87%	15.7%
w/o XGBoost	91.0%	96.7%	93%	9.2%
w/o Random Forest	91.2%	96.8%	93%	9.0%

The table reveals that between the interaction of the following: First, ESN removes the majority of false positives caused by template duplication and periodic jitter, resulting in an FPR increase of approximately 5 percentage points on removal.

Second, TF-IDF and Transformer are highly complementary: TF-IDF extricates the strong cues explicitly, such as “CVE/Port/Policy Name,” while the Transformer relates the synonymous expression between templates. Discarding either results in the weakening of boundary detection. Third, random forest and XGBoost perform similarly when used as a solo branch; however, the fusion results in stable probability scaling and boundary refinement.

The sensitivity of other key hyperparameters is as follows: The performance of the ESN module demonstrated the most optimal stability between 0.75-0.85 when altering the spectral radius from 0.6 to 0.95. The leakage rate showed minimal deviation in 0.2-0.4, with extremely low values resulting in no short-term information while extremely high values overamplified isolated spikes. The sliding window length performances were similar between T=5 and T=10; however, T=1 poses a significant degradation. The weight of fusion λ was flat, with a score of 0.5 available in the range of 0.2-0.8 and hence a default stable. Lastly, the threshold was proven: FPR-target search strategy outperforms fixed 0.5 in satisfying operational false alarm constraints and consistency between validation-test transfers.

4.2.2. System effect

For the evaluation experiment, the system’s performance is assessed using two test sets: enterprise logs (Hillstone SOC) and standardized logs (Advanced SIEM Dataset). Both experiments follow the same pipeline and threshold strategy and are presented in Table 2.

Table 2. Comparison results of hybrid models with traditional SIEM

Datasets	Method	Accuracy	Recall	F1	FPR
Hillstone SOC	Traditional SIEM	39.3%	62.1%	47%	60.72%
	SIEVE	92.4%	97.9%	94%	7.58%
Advanced SIEM Dataset	Traditional SIEM	59.9%	72.0%	62%	40.11%
	SIEVE	94.3%	98.7%	95%	5.62%

As the experimental results show, the use of the hybrid model allows us to significantly reduce the false positive rate on both sets. In the real-world logs of Rocknet, the model managed to reduce the number of false positives from 60.72% to 7.58%, and its F1-score reaches 0.94. For the

DOI: 10.54254/2753-8818/2026.HZ29905

standardized Advanced SIEM Dataset, the false positive rate does not exceed 5.62%, and the F1-score equals 0.95.

The close performance gap suggests the robustness and transferability of the proposed method toward various real-world enterprise environments. It involves mixed Chinese-English data as well as non-standardized descriptions and complex structured fields. Therefore, the method's performance level is almost as high as that of the publicly available standardized corpora, which means it has practical benefits for everyday usage.

4.3. Discussion and analysis

The results indicate that the performance improvement of SIEVE mainly comes from the complementarity among three components: semantic-textual modeling, temporal learning, and ensemble classification.

Specifically, the TF–IDF and Transformer models capture keyword salience and contextual semantics jointly. Meanwhile, the ESN can model short-term temporal dependencies in the history sequence of alerts effectively. Furthermore, the Random Forest and XGBoost combined contribute to the stabilized probability calibration and robust decision-making for heterogeneous feature types.

By comparison, SIEVE can reduce false positives significantly compared to traditional SIEM systems without losing much recall. This validates the value of hybrid feature fusion and multimodal modeling in maintaining the balance between sensitivity and precision in operational environments, especially for multilingual or irregular log formats.

There are some limitations, however. First, the enterprise dataset may suffer from subjective labeling noise. Second, frozen Transformer parameters severely restrict domain adaptation. Last, the synthetic benchmark dataset lacks real-world fragility.

For future research, this paper aims to implement domain-adaptive fine-tuning, cross-source evaluation, and richer temporal-context features to eliminate further false alarms.

Conclusively, SIEVE indicates that hybrid modeling is a viable and scalable option for intelligent false-positive reduction in SIEM systems.

5. Conclusion

To reduce SIEM false alarm rates, this study constructs and validates a hybrid system aligned with engineering practices using enterprise logs and the Advanced SIEM Dataset. The framework first integrates Chinese and English texts with structural fields at the feature level (TF-IDF, Transformer, normalized structural signals). It then models temporal dependencies through an ESN (Entity-Specific Network) with leakage rate and spectral radius constraints on sliding windows, followed by an RF+XGBoost ensemble classifier. Through monotonic calibration, the system generates 0-100 risk scores and determines thresholds based on target FPR, achieving controllable recall accuracy. Limitations include enterprise label noise, temporal drift, the Transformer's limited adaptation to niche terms, fixed ESN windows lacking cross-entity relationships, threshold sensitivity under extreme imbalances, and discrepancies between synthetic data and real traffic distributions. Future research will explore noise drift modeling, parameter-efficient fine-tuning, retrieval enhancement, spatiotemporal graph message passing, cost-sensitive adaptive thresholds, and cross-domain validation to address existing issues. Overall, this study establishes a closed loop from "data–features–temporal–discrimination–operational metrics," providing a scalable path for interpretable, configurable, and implementable SIEM false alarm optimization.

DOI: 10.54254/2753-8818/2026.HZ29905

References

[1] Alahmadi, B. A., Axon, L., & Martinovic, I. (2022, August). 99% false positives: A qualitative study of SOC analysts' perspectives on security alarms. In Proceedings of the 31st USENIX Security Symposium (USENIX Security 22) (pp. 2783–2800). USENIX Association. https: //www.usenix.org/conference/usenixsecurity22/presentation/alahmadi
[2] Mozalbat, G. (2020). Reducing false-positive alerts in SIEM systems using TF-IDF (Master's thesis).
[3] Du, M., Li, F., Zheng, G., & Srikumar, V. (2017). Deeplog: Anomaly detection and diagnosis from system logs through deep learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1285–1298). Association for Computing Machinery.
[4] Wu, X., Li, H., & Khomh, F. (2025). What information contributes to log-based anomaly detection? Insights from a configurable transformer-based approach. Automated Software Engineering, 32(2), 58.
[5] Pinhero, A., Anupama, M. L., Vinod, P., Visaggio, C. A., Aneesh, N., Abhijith, S., & AnanthaKrishnan, S. (2021). Malware detection employed by visualization and deep neural network. Computers & Security, 105, 102247.
[6] Shukla, A., Gandhi, P. A., Elovici, Y., & Shabtai, A. (2025). RuleGenie: SIEM detection rule set optimization. arXiv preprint arXiv: 2505.06701.
[7] Irshad, M., Law, N.-F., Loo, K. H., & Haider, S. (2023). ImgCat: An approach to dismantle the anonymity of a source camera using correlative features and an integrated 1D convolutional neural network. Array, 18, 100279.
[8] Hang, F., Guo, W., Chen, H., Xie, L., Zhou, C., & Liu, Y. (2023). LogFormer: Cascaded transformer for system log anomaly detection. Computer Modeling in Engineering and Sciences (CMES), 136(1).
[9] Guan, W., Cao, J., Qian, S., Gao, J., & Ouyang, C. (2024). LogLLM: Log-based anomaly detection using large language models. arXiv preprint arXiv: 2411.08561.
[10] Pospieszny, P., Mormul, W., Szyndler, K., & Kumar, S. (2025). AdaLog: Adaptive unsupervised anomaly detection in logs with self-attention masked language model. arXiv preprint arXiv: 2505.13496.
[11] Thakur, R. (2025). Advanced SIEM dataset [Data set]. Hugging Face. https: //huggingface.co/datasets/