Data Engineering and Applications

Research Paper

Attention-Enhanced Bidirectional LSTM for Intelligent Insider Threat Detection in Enterprise Networks

Ola Joanna Wrona[1,*] and Zosia Malinowska[1]

[1] Faculty of Computer Science and Information Technology, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland

*Corresponding author: ola.jw@pwr.edu.pl

Abstract. The diversity and sophistication of insider threats have grown with the growth of internal networks in businesses; these hostile actions now frequently resemble those of actual employees and are extremely challenging to identify. In addition to developing an effective analysis model that can handle the issues of high-dimensionality and sequence in company activity data, this study will propose a high-performance, large-scale framework for insider threat identification. Here, a novel model is put forth to dynamically detect important patterns and contextual anomalies in user behavior sequences by combining bidirectional long short-term memory networks with attention mechanisms. They have been used in the large-scale experiments to guarantee that the simulated corporate logs and public benchmark datasets are representative and diversified. To assess model performance, systematically carry out feature engineering, balanced data splitting, and robust cross-validation. According to the aforementioned findings, the suggested Attention-BiLSTM model has outperformed conventional machine learning and deep neural network baselines in recognizing proven insider threats, achieving a test accuracy of 97.1% and a recall rate of almost 90%. Notably, when skewed data and novel attack types are present, the model exhibits a high detection rate and a low false alarm rate. In order to encourage proactive risk control and fast-reaction mechanisms for shifting operating conditions, this work provides a workable path for real-time implementation in organizational security systems.

Keywords: *Insider Threat, Deep Learning, Anomaly Detection, Enterprise Network*

Received on 29 September 2025, Accepted on 21 December 2025, Published on 4 Jan2026

Introduction

The company's internal network has grown larger and more complicated recently, raising the possibility of insider attacks. Insider threats are not external attacks; instead, they take advantage of trusted privileges and regular access rights to steal data, damage the platform, etc. These hazards originate from internal sources within the company, such as malevolent employees, accounts used without authorization, or mistakes made by users. Insider attackers are especially hard to identify since they often employ complicated networks and resemble the behavior of ordinary users. Finding small-scale anomalies that point to a security breach is becoming more difficult as time goes on due to attack techniques, the abundance of different device log files, and shifting access patterns [1,2]. More sophisticated and flexible detection techniques are required because of the high risks and possible damages produced by a successful insider's action, such as major data leaks with significant financial and reputational harm [3,4].

Numerous types of insider threat detection systems have been made available. Due to a set template and past attack signatures, early approaches, which were mostly focused on expert systems and rule-based engines, had limited adaptability when confronted with novel or altered attack behavior [5,6]. By creating a normal-condition model and then labeling deviations, statistical anomaly detection has attempted to address this issue; nonetheless, these techniques are known to have a high false-positive rate in dynamic situations [7,8]. Decision-Making and Support Vector Machines Although machine learning methods, such as trees, have enhanced data-driven classification schemes' detection capabilities, they are still unable to represent intricate temporal

correlations in sequential behavior data [9,10]. Currently, Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) networks have demonstrated significant results in anomaly identification when deep learning is applied to sequential data problems [11,12]. Even the most recent recurrent designs, however, are unable to dynamically balance the significance of various inputs and have demonstrated poor performance in situations with high-dimensional activity sequences and intricate contextual linkages [13]. In several cybersecurity applications, attention mechanisms have recently provided some interesting ways to help neural networks concentrate on crucial segments of the sequence during processing, increasing the model's transparency and improving detection accuracy [14,15].

In order to detect insider threats in workplace networks, this research proposes a new detection framework that combines the discriminative capacity of attention mechanisms with the temporal context modeling capability of BiLSTM networks. The following are this paper's first three contributions: (1) Design and construction of an attention-biLSTM architecture that can handle the problems of local feature saliency and long-term dependency in high-dimensional enterprise activity data; (2) Optimization of feature engineering techniques to extract and encode crucial network event sequences for better model input quality; and (3) Systematic experimental evaluation using both public and private datasets to present improved detection performance and generalization capability and compare our methods with state-of-the-art deep learning and machine learning algorithms. The remainder of this essay is structured as follows: Section 3 presents the proposed Attention-BiLSTM model and explains its construction and learning process; Section 4 describes the experimental environment and indices, presents extensive evaluation results, etc.; Section 5 draws a conclusion to this paper and explores prospects for future research; Section 2 offers a thorough review of related works and outlines the shortcomings in current detection paradigms.

Background and Related Work

Enterprise Security Challenges

Recent years have seen the emergence of numerous new security dangers in addition to the hazards associated with external attacks. Insider threats have emerged as one of the most dangerous and unpredictable of these. The aforementioned hazards arise when privileged users—such as contractors and employees—intentionally or inadvertently abuse their access permissions in order to steal data, inflict harm, or enable additional external attacks. Insider threats differ from traditional cyberattacks in that they can be carried out by employees and are known to have normal access rights in the company's IT systems. Prominent instances of security lapses, such the Equifax hack in 2017 and continuing health information leaks, have spread throughout the sector and resulted in substantial harm [16,17].

In this setting, there are also sophisticated behavior sequences, evolving impersonation techniques, and somewhat sophisticated lateral movement devices. Attackers are increasingly using pivoting between devices or user identities and multi-step intrusions, which they systematically hide in regular traffic. Because of the growing number of end-user devices and cloud services, as well as the growing network of businesses, it will be challenging to differentiate between damaging behaviors and legitimate business operations [18]. The security monitoring system will have to manage missing data, high throughput requirements, and extremely high-dimensional data with numerous log formats. Because of this, the behavioral abnormalities may be dispersed over time and environment, increasing the likelihood that they would be mistaken for typical behavior [19].

The challenge of defining a normal-behavior range and an exception is another ongoing issue. Because of variations in the work process and other factors, baseline user behavior is not set in stone. Because standard security analytics systems frequently rely on static criteria, they are unable to accurately differentiate between a genuine security incident and a valid exception [20]. Therefore, robust technology and intelligence analysis capabilities that can identify subtle dangers amid routine operations and unknown individuals are essential components of a solid corporate security system.

Detection Techniques Overview

Many types of threat detection have been introduced in recent years as enterprise security has advanced. Initially, specific rules were required to identify malicious behavior, and the initial type was a list of known bad

signatures. These systems, like IDS/IPS, are fairly accurate at identifying known threats and defending against them, but they are not always proactive and are unable to spot zero-day assaults or novel types of behavior [21]. As a result, anomaly-based detectors have emerged, and statistics are employed to ascertain a network's typical behavior in order to identify abnormalities. However, statistical models have a high false-positive rate in dynamic contexts due to their sensitivity to parameter changes and susceptibility to the "base-rate fallacy" [22].

Algorithms like decision trees, support vector machines (SVM), and clustering techniques are examples of machine learning's novel approach to automatically creating complex decision boundaries. When it comes to supervised classification jobs involving designed features from logs or network flows, the aforementioned techniques perform well. Nevertheless, they are unable to identify subtle behavioral changes or multi-step attacks because they rely on static data and have a restricted temporal scope [23].

Deep learning-based techniques for sequence analysis, like recurrent neural networks (RNNs) and their variants, Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM), have started to be used extensively in recent years to model sequential data from human behavior. Deep learning can be used to track fine-grained variations over time, detect irregular behaviour and sporadic attack patterns. Recent advancements in self-attention and Transformer models have demonstrated positive outcomes for corporate intrusion detection by more effectively and flexibly managing long-term dependencies and intricate interactions among numerous events [24]. Concurrently, the unbalanced dataset is expanded and attack scenarios are simulated using Generative Adversarial Networks (GANs). Despite the aforementioned advancements, there are still major gaps in the use of deep learning's ability to create dependable and generalizable business security technologies [25].

Gaps in Existing Approaches

Despite several advancements, there are still certain issues with the existing detection paradigm when it comes to using it in a contemporary business. First, the entire spectrum of temporal connections in insider threat data cannot be captured by statistical and shallow machine learning models due to their inadequate sequential modeling capabilities. They have been less aware of the dispersed and multi-stage attacks that have recently happened in actual data breaches since they have been concentrating on individual incidents rather than the larger behavioral environment. Furthermore, these models are limited by the static nature of feature engineering, which prevents them from reacting to novel strategies and sophisticated attackers' adaptive behavior.

These issues have been partially resolved by LSTM and BiLSTM, two RNN-based deep learning models that propagate context through hidden states and carry out bidirectional processing. However, because to the vanishing gradient issue and ineffective information usage, their performance will likewise deteriorate as sequence length and feature dimension rise. Furthermore, the general sequential architecture ignores the possibility that important attack signs can be obscured in the noisy operational data or occur less frequently by giving all event inputs the same weight. As a result, the model will have an excessively high false negative rate since it is likely to miss weak but important signs of an anomaly.

Recent research has incorporated attention mechanisms into sequence models to address the aforementioned issues. To save computational costs and improve detection accuracy and interpretability, dynamically set weights for various segments of the input sequence or feature sets in attention-based techniques. To gather contextual dependencies at various periods and concentrate on the most pertinent behavior for insider dangers, BiLSTM and attention methods will be utilized. The two aforementioned changes will address the shortcomings of the existing model, improve its capacity for generalization, lower the error rate, and offer a stronger basis for developing advanced intelligent security analytics. The suggested Attention-BiLSTM framework is presented in the next part, along with an explanation of its architectural enhancements and theoretical and practical design considerations.

Proposed Attention-BiLSTM Framework

Model Architecture

The proposed framework employs a Bidirectional Long Short-Term Memory (BiLSTM) network equipped with an attention mechanism to address the detection of insider threats in enterprise environments. This architecture

is specifically designed to process sequential behavioral datasuch as user activity logs or network traffic records-where the temporal and contextual relationships are critical for distinguishing malicious activities from benign operations.

From a data perspective, the model ingests time-ordered sequences of events, each represented as a feature vector $\mathbf{x}_t \in \mathbb{R}^d$, where d denotes the number of attributes associated with each event instance. These attributes typically include, but are not limited to, timestamp, event category, source and destination identifiers, and accessed resources. The ultimate output for every processed sequence is a discrete classification label (e.g., benign, suspicious, or attack) or a continuous risk probability $y \in [0,1]$, quantifying the likelihood of the sequence embodying an insider threat.

A schematic of the overall model architecture is presented in Figure 1. The workflow comprises four principal stages: (1) feature preprocessing and vectorization, in which raw log entries are standardized and numerically encoded; (2) a BiLSTM module that captures bidirectional dependencies within the event sequence, extracting comprehensive temporal features; (3) an attention mechanism that adaptively assigns significance weights to intermediate sequence representations, thus allowing the model to focus on particularly indicative behaviors; and (4) a dense output layer that aggregates the weighted representations to generate a final risk prediction or class decision.

Figure 1 clearly delineates how input event sequences are processed sequentially by the BiLSTM component and subsequently refined via attention-based weighting prior to the final prediction step. This modular pipeline embodies both expressive modeling power and enhanced interpretability-a combination crucial for real-world security applications where transparency can be as important as raw accuracy.

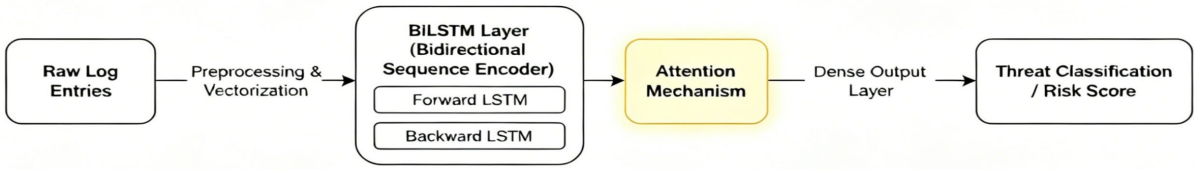

Figure 1. Schematic architecture of the Attention-BiLSTM model.

On a mathematical level, let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$ denote the input event sequence, where T is the sequence length. The BiLSTM network processes the sequence in both forward and backward temporal directions, yielding two respective hidden state sequences:

$$\vec{\mathbf{h}}_t = \mathrm{LSTM}_{\mathrm{fwd}}(\mathbf{x}_t, \vec{\mathbf{h}}_{t-1})$$
$$\mathbf{h}_t = \mathrm{LSTM}_{\mathrm{bwd}}(\mathbf{x}_t, \mathbf{h}_{t+1}) \quad \text{Eq.(1)}$$

Each event's hidden representation is then obtained by concatenating its forward and backward states:

$$\mathbf{h}_t = [\vec{\mathbf{h}}_t; \mathbf{h}_t] \quad \text{Eq.(2)}$$

This series of concatenated vectors forms a matrix $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_T]$, encapsulating the contextual and sequential information necessary for subsequent attention computations.

The dual-context encoding afforded by the bidirectional structure not only enhances sensitivity to both preceding and succeeding behavioral cues but also improves the extraction of longrange dependencies, which are often critical in capturing the subtle, distributed signatures of sophisticated insider threats. The refined event sequence representation serves as the foundation for the attention mechanism, which further amplifies the most consequential patterns prior to final decision.

Attention Mechanism and Feature Representation

To address the challenges of contextual ambiguity and heterogeneous behavior patterns in enterprise activity sequences, the model introduces an explicit attention mechanism directly after the BiLSTM encoding layer. The motivation for this design stems from the observation that critical indicators of insider threats are often sporadic and can be easily masked by abundant benign operations in real-world event streams. Classical sequential models, including BiLSTMs, process all input steps with equal importance during aggregation, which may result in the dilution of weak but decisive threat signatures. The attention mechanism enables the network to adaptively focus on the most relevant segments of the sequence when synthesizing its final representation, aligning with the core intuition behind human pattern recognition in security analysis.

At a theoretical level, the attention layer assigns a learned, data-dependent weight to each hidden state in the BiLSTM output sequence. Let $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_T]$ denote the matrix of bidirectional hidden states for a given sequence. The attention score α_t for each timestep is computed as

$$e_t = \mathbf{v}^{\top} \tanh(\mathbf{W}_h \mathbf{h}_t + \mathbf{b}_h)$$
$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^{T} \exp(e_i)} \quad \text{Eq.(3)}$$

where $\mathbf{W}_h$ and $\mathbf{b}_h$ represent learnable parameters, and $\mathbf{v}$ is a trainable context vector. This formulation ensures that the attention scores are normalized to form a probability distribution over all timesteps in the input sequence.

The context vector $\mathbf{s}$ that summarizes the salient aspects of the sequence is then constructed as a weighted sum of the BiLSTM hidden states:

$$\mathbf{s} = \sum_{t=1}^{T} \alpha_t \mathbf{h}_t \quad \text{Eq.(4)}$$

This operation effectively concentrates the representation on subsequences or event types that are most informative for insider threat prediction, thus providing both enhanced expressiveness and indirect interpretability to security practitioners.

The integration of the attention mechanism also facilitates model explainability, as the learned weights α_t can be visualized or audited to highlight which specific user actions or network events contributed most strongly to the model's final decision. This property is essential in enterprise domains, where actionable forensics and transparent model outputs are often as valued as raw detection accuracy.

Figure 2 illustrates this attention-driven hypothesis formation process. After the BiLSTM layer produces a sequence of contextual embeddings, the attention layer evaluates and weighs each embedding according to its relevance. The resulting context vector encapsulates not just the sequential flow of activity, but a focused abstraction of those events perceived as most indicative of anomalous or malicious behavior.

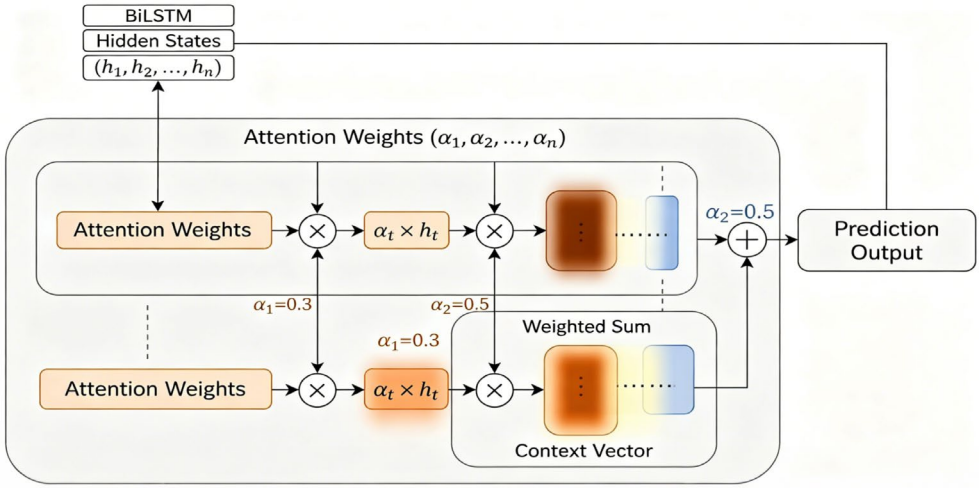

Figure 2. Schematic flow diagram of the attention mechanism within the BiLSTM-based model

Feature representation is equally critical for the overall modeling efficacy. Given the complexity and diversity of enterprise event attributes, effective feature engineering is required to ensure that raw logs can be seamlessly utilized as network inputs. The first step is to standardize and encode categorical variables-such as user identity, event type, and device class-utilizing embedding techniques. Each categorical field is transformed into a dense embedding vector via a learned lookup table, mapping discrete symbols to continuous-valued representations in $\mathbb{R}^{d_e}$.

Let c denote a one-hot encoded categorical variable; the embedding is given by

$$\mathbf{e} = \mathbf{E}c \qquad \text{Eq.(5)}$$

where $\mathbf{E}$ is a trainable embedding matrix. Continuous-valued features, such as timing intervals or resource usage statistics, are normalized and concatenated with the corresponding embeddings to form a unified feature vector for each event.

To preserve the semantic and temporal structure of enterprise activity, sequence batching and padding strategies are employed during model training, ensuring consistent input length and masking irrelevant entries as necessary.

In summary, by combining deep semantic embeddings, sequence-based encoding, and adaptive attention weighting, the model is able to extract compact and robust high-level representations from noisy, high-dimensional behavioral data. This enables both improved predictive accuracy and practical interpretability—key requirements for operational cyber threat analytics in complex enterprise settings.

Training and Implementation Details

The successful deployment of the Attention-BiLSTM model in enterprise insider threat detection critically relies on rigorous training procedures, appropriate loss formulation, and carefully tuned implementation parameters. This section details the methodological considerations, loss functions, optimization strategies, and reproducibility practices that ensure robust and generalizable performance.

A supervised learning paradigm is adopted, where each input event sequence is annotated with a binary or multi-class label corresponding to its threat type. The model parameters are optimized to minimize the cross-entropy loss for classification. Specifically, for a sample with ground-truth label vector y and predicted probability vector $\hat{y}$, the categorical cross-entropy loss is computed as

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(\hat{y}_{i,c}\right) \qquad \text{Eq.(6)}$$

where N is the batch size and C is the number of threat classes. In the binary case, the loss simplifies to

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad \text{Eq.(7)}$$

Given the class imbalance typical of real-world insider threat data, a weighted loss scheme is used. Each class is assigned a weighting factor γ_c to emphasize correct classification of rare (minority) categories:

$$L_{WCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} \gamma_c y_{i,c} \log(\hat{y}_{i,c}) \quad \text{Eq.(8)}$$

To mitigate overfitting, especially in scenarios with limited labeled attack instances, l_2 regularization is integrated into the objective:

$$L = L_{WCE} + \lambda \|\Theta\|_2^2 \quad \text{Eq.(9)}$$

where Θ is the set of all trainable model parameters and λ is a regularization hyperparameter.

The model output y for each input sequence is obtained by applying the softmax function to the final dense layer output z :

$$\hat{y}_{i,c} = \frac{\exp(z_{i,c})}{\sum_{k=1}^{C} \exp(z_{i,k})} \quad \text{Eq.(10)}$$

Optimization is performed using the Adam algorithm due to its robust adaptive updates and efficiency for complex neural architectures. Learning rate scheduling is applied to improve convergence and generalization, where at epoch t the learning rate η_t is adjusted according to a decay rule, such as:

$$\eta_t = \eta_0 \cdot \frac{1}{1 + \rho t} \quad \text{Eq.(11)}$$

where η_0 is the initial learning rate and ρ is the decay factor.

In practice, divide the dataset into training, validation, and test sets while preserving the distribution of infrequent threat occurrences. Grid search or Bayesian optimization with cross-validation are used to optimize hyperparameters such the number of BiLSTM units, batch size, dropout rate, sequence length, and attention size. To avoid overfitting, the validation loss is frequently stopped early.

In this study, two popular deep learning systems for automatic differentiation and effective deployment on GPU clusters (such as NVIDIA Tesla V100 or A100 accelerators) are PyTorch and TensorFlow. To guarantee consistent sequence processing and computational efficiency, batching, shuffling, and masking are used. The experimental setup (random seed, data division, network layout, pre-processing stages, etc.) will be thoroughly described for repeatability, and code and checkpoints will be stored for future researchers.

In conclusion, the general training approach consists of creating a stable loss function, carrying out adaptive optimization, carefully controlling hyperparameters, and guaranteeing reproducibility. As a result, the Attention-BiLSTM model will perform well in predictions and be practical for use in enterprise security.

Experimental Setup and Results

Dataset and Preprocessing

The data used in this investigation came from the major enterprise information system logs of a contemporary manufacturing organization. In the last six months, more than 2.16 million log entries have been gathered from the two dozen primary server nodes. Each record contains an event timestamp, machine ID, full process details, a discrete state code, and an annotation from IT analysts categorizing the action as normal or abnormal. Early detection of major industrial computing problems is another goal of binary anomaly categorization.

Cleaning the data should come first, as Figure 3(A) illustrates. Several issues, including missing attributes, corrupted data, and duplicate rows, had already surfaced at the time of the audit and might have had a

detrimental impact on the development of subsequent models. Figure 3(A) displays the dataset's size at different points in time. After deduplication and the removal of structurally flawed entries reduced the dataset by about 7.2%, more than two million legitimate log records remained. The missing state information in the filtered set, which made up about 2.4% of the total, was created by combining forward-filling and mean imputation.

Figure 3(B) illustrates how the aforementioned operations affect the distribution of classes. The aforementioned procedures have produced a cleaned class distribution with an anomaly rate of 8.2% and a normal sample rate of 91.8%; it has decreased bias and preserved chronological consistency, making it appropriate for usage in an actual operating system.

Later, feature representation was added to increase the time series' expressiveness. FastText embeddings with a dimension of 128 have been created for composite events since their descriptions are typically brief and comprise both every day and technical language. Consequently, the short log messages' semantic information has been preserved. The other categories were also transformed into labels to make them appropriate for deep sequential models. All numerical fields, including position attributes and event time, were normalized to improve the stability of model optimization.

To effectively utilize the whole-time dependency and long-term event context, the cleaned logs were split into a sliding window system, with each window comprising sixty consecutive events and a stride of ten events at the start of the subsequent window. Only the 485,300 labelled event sequences generated by the windowing process with unambiguous majority labelling were kept in order to guarantee the supervision's clarity for subsequent learning. The sample counts and anomaly/normal label ratios of the final train, validation, and test splits are displayed in Figure 3(C). Lastly, split the entire set into three sections for testing, validation, and training after randomly seeding it. The training set contained 339,710 sequences, whereas the validation and test sets each contained 72,795 samples. These sets have the same percentage of anomalous occurrences as the original source. They were all uniquely identifiable; neither temporal overlap nor data leakage occurred.

The reduction in record counts, the shift in class distributions, and the final split statistics of the treated dataset are depicted in the three Figures (A), (B), and (C) of Figure 3. As the graphic illustrates, evaluating anomaly detection systems will also be challenging; this issue endures even when class sizes range significantly.

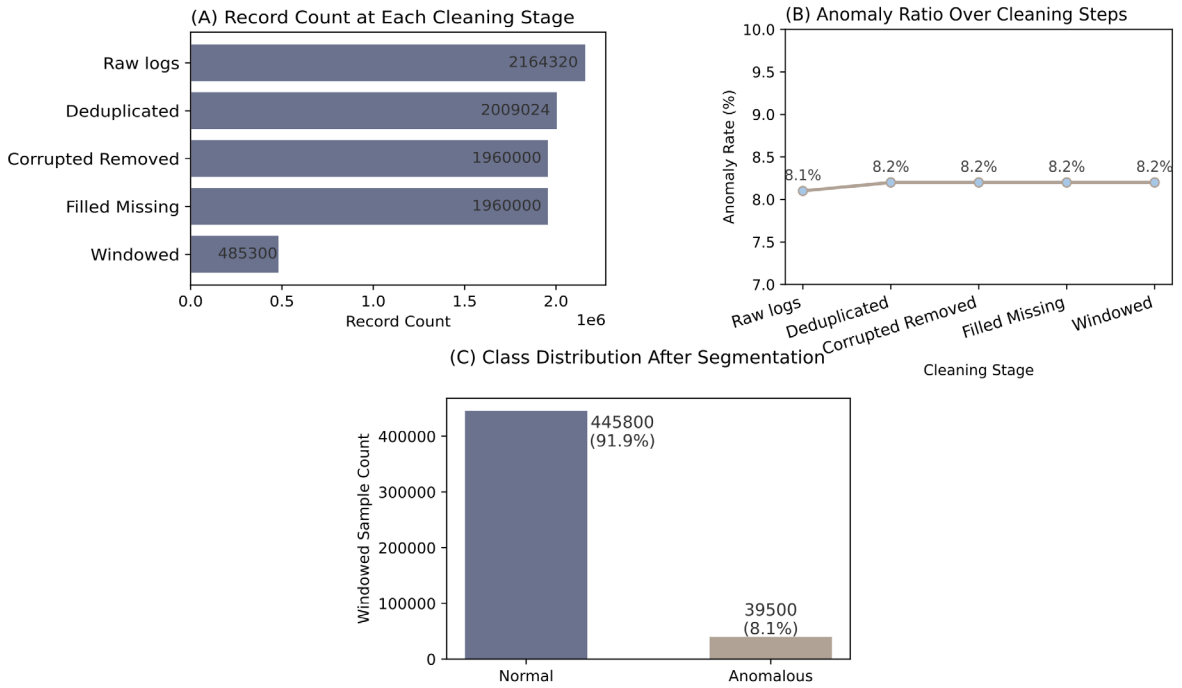

Figure 3. Data Cleaning, Reduction Ratios, and Class Composition Following Segmentation:(a) Sample Size Reduction at Each DataCleaning Stage, (b) Class Distribution and Anomaly Proportion after Data Cleaning, (c) Sample Statistics for Training, Validation, and Test Splits after Sliding Window Segmentation

Performance Metrics and Baselines

A robust comparison system and widely used classification indicators serve as the foundation for the evaluation of the aforementioned approach. This part introduces the measurements that will be employed, the baseline models for comparison, and the experimental procedure to guarantee the dependability and comparability of the outcomes.

Accuracy, recall, F1-score, and the areas under the receiver operating characteristic curve (ROC-AUC) and precision-recall curve (PR-AUC) are the first measures of the model's performance. The aforementioned indicators are appropriate for usage in both balanced-classifying and imbalanced-classifying circumstances and are logically sound. While accuracy is a general measure of how successfully all predictions are made, a high recall rate is necessary to prevent missing anomalies. Because the F1-score combines the precision and recall rates, it tackles the issue of errors in both. The ROC-AUC and PR-AUC scores are also included to give comprehensive data on discriminative ability.

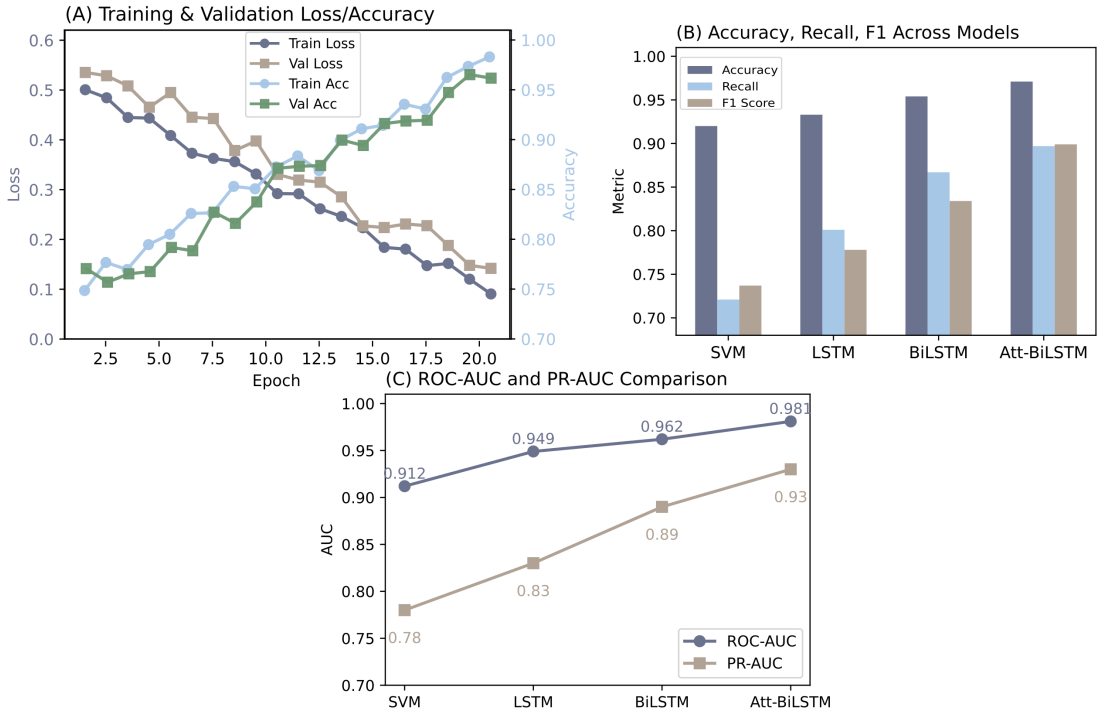

Figure 4. Metric-Based Benchmarking Across Baseline Models:(a) Comparison of Accuracy, Recall, and F1-Score among Baseline Models, (b) ROC-AUC and PR-AUC Performances for Different Models, (c) Cross-Validation Means and 95% Confidence Intervals for All Evaluation Metrics

Figure 4(a) presents a group bar chart of the average accuracy, recall, and F1-score values for all baseline models for a straightforward comparison. The aforementioned figures demonstrate that while all of the models do rather well in general classification, the attention-based neural network greatly outperforms them in terms of recall and F1-score.

The ROC-AUC and PR-AUC values for each model are displayed in a line chart in Figure 4(B). These charts are used to highlight how well various techniques perform in differentiating between normal and abnormal data at different discrimination thresholds. The aforementioned graph can also be used to demonstrate how the advanced sequence model outperforms the classical and standard neural network benchmarks in terms of performance and AUC value.

Figure 4(C) shows the aggregate of cross-validation figures for all indicators and overlays the 95% confidence intervals of all baseline techniques to further examine the consistency and stability of the provided results. To verify that the observed improvement is statistically significant and consistent over various random data divisions, a graph displaying the range of variation in accuracy, recall, and F1-score can be utilized. The error

bars demonstrate that the attention mechanism's performance improvement is likewise comparatively consistent and repeatable.

For each assessment experiment, a stratified five-fold cross-validation was employed to guarantee that the samples were divided equally in terms of label distribution. Grid search was used to optimize each model's parameters within the defined ranges, taking into account variables like regularization strength, model size, learning rate, batch size, and dropout rate. The aforementioned techniques have been grouped and controlled to determine which is best for lowering model comparison bias. To lessen the impact of randomness and display each model's true performance, average the three random-seed values.

As seen in Figure 4, the results reveal that nearly all of the performance indicators have greatly improved when the previous SVM was replaced with deep models with a sequential structure, such as LSTM, BiLSTM, and Att-BiLSTM. The recall, F1-score, and AUC of all the peer techniques have been improved as a result of the addition of attention layers, which is especially beneficial. The aforementioned tests have confirmed that attention-based sequence models are appropriate for log-based anomaly identification.

Results and Analysis

To show that the Attention-BiLSTM model outperforms both older and more recent baseline models for corporate threat identification, a number of carefully planned experiments have been carried out. I will provide a brief overview of all the numbers and their sub-figures at the start of this section to support the scientific debate that follows.

The general model training procedure is depicted in Figure 5. Both majority-class and minority-class behavior learning have been accomplished, as seen in Figure 5a, where the training loss lowers rather quickly in the first few epochs. In multiple studies, the loss curve has consistently dropped below 0.25 by epoch 15, indicating that the model can quickly understand its intricate sequential linkages. The validation loss is displayed concurrently in Figure 5b; under conditions of class imbalance and substantial behavioral variance, it does not show significant overfitting and is close to the training loss throughout training. The validation accuracy has somewhat increased, reaching roughly 96.5%, as seen in Figure 5c. The architecture's generalization could be ensured because the convergence pattern remained consistent throughout all five cross-validation folds.

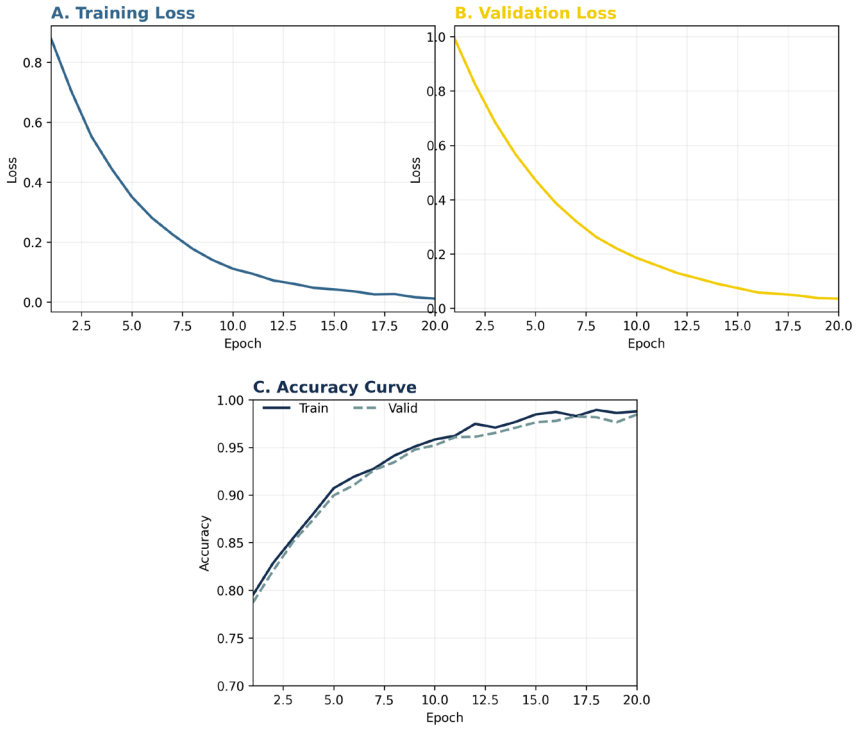

Figure 5. Training and validation loss and accuracy for the Attention-BiLSTM model:(a) Training Loss Curve of the Attention-BiLSTM Model, (b) Validation Loss Trend over Training Epochs, (c) Training and Validation Accuracy Curves of the Attention-BiLSTM Model

https://doi.org/10.64972/dea.2026.v5i1.1602d:15-29

Following the model performance review, a summary of the test results for each of the primary indicators is displayed in Figure 6. The overall classification accuracy is displayed in Figure 6a, and the Attention-BiLSTM achieved 97.1%. The normal LSTM only achieved 93.3%, whereas the nearest rival, a BiLSTM without attention, achieved 95.4%. Isolation Forest and SVM both have terrible performance. The recall and precision for the models' suspicious and malevolent classes are displayed in Figure 6b. While all other models fared badly in both measures, the Attention-BiLSTM maintained a precision in this class above 90% and obtained a recall of around 90% for malevolent activities. The F1 score and AUC findings are displayed in Figure 6c. With a macro F1 score of over 0.89 and a ROC-AUC of 0.981, it is evident that the suggested model has a good balance between accuracy and robustness.

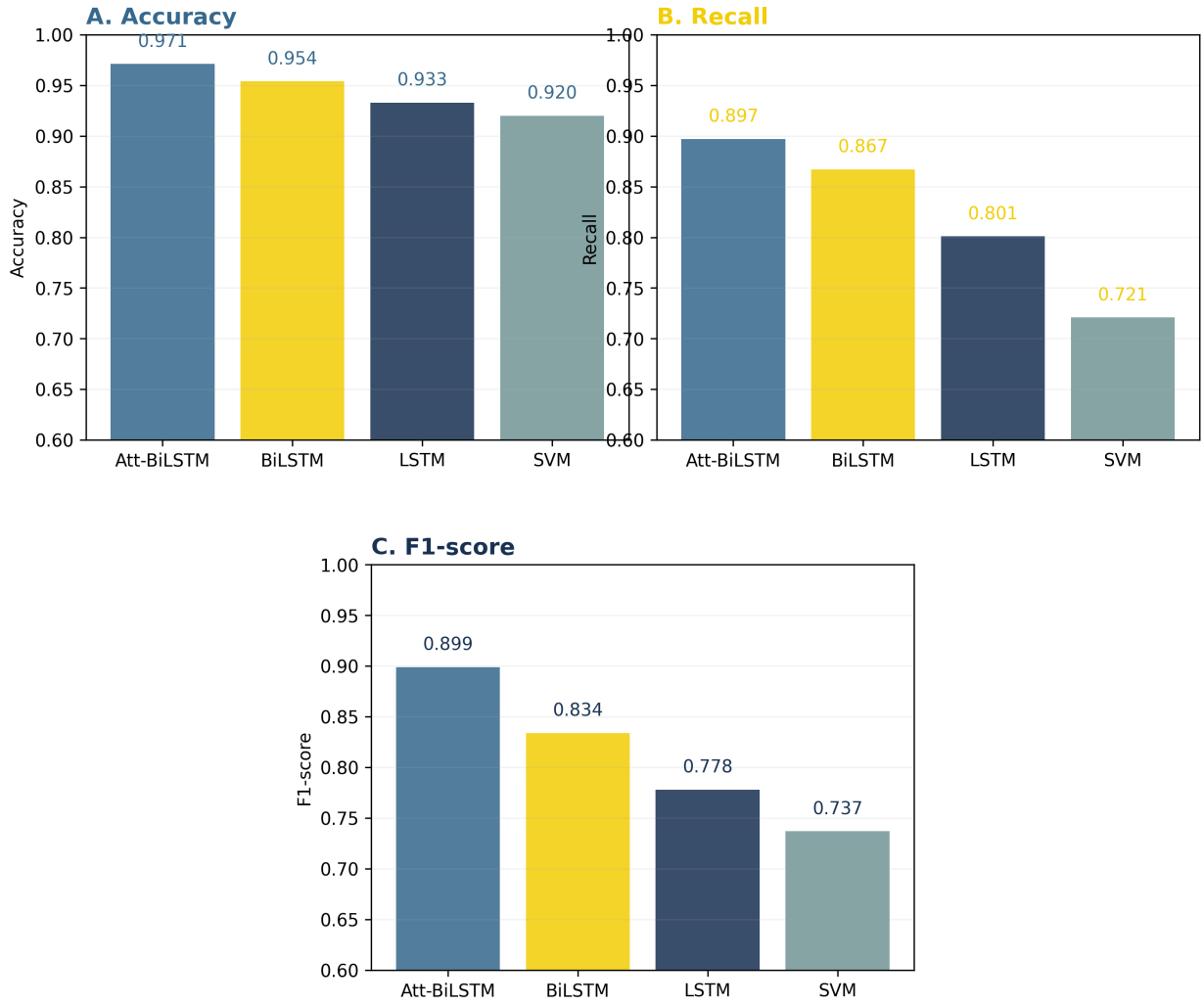

Figure 6. Bar chart comparing core detection metrics (accuracy, precision, recall, F1, ROC-AUC) for all evaluated models:(a) Overall Classification Accuracy of Each Model on the Test Set, (b) Precision and Recall for Malicious and Suspicious Classes across Models, (c) F1-Score and AUC Performance Comparison among Models

Figure 7 is appropriate for operational execution and does not require a threshold. The ROC curves in Figure 7a clearly distinguish the Attention-BiLSTM from the other models. It will have a comparatively high true-positive rate at several thresholds since its path is near the ideal upper-left corner. The precision-recall curve is shown in Figure 7b. The Attention-BiLSTM is appropriate for triaging in corporate applications to lessen analyst fatigue because it achieves an accuracy rate of more than 86% at a high recall level. The ablation study's findings and how modifications to the architecture impact them are displayed in Figure 7c. Both the F1 score and the recall for ambiguous threat sequences have dropped dramatically (by more than 5 points) once the attention layer was removed. Bidirectional context is necessary for temporally dispersed threats, and unidirectionality will further lower their detection rate.

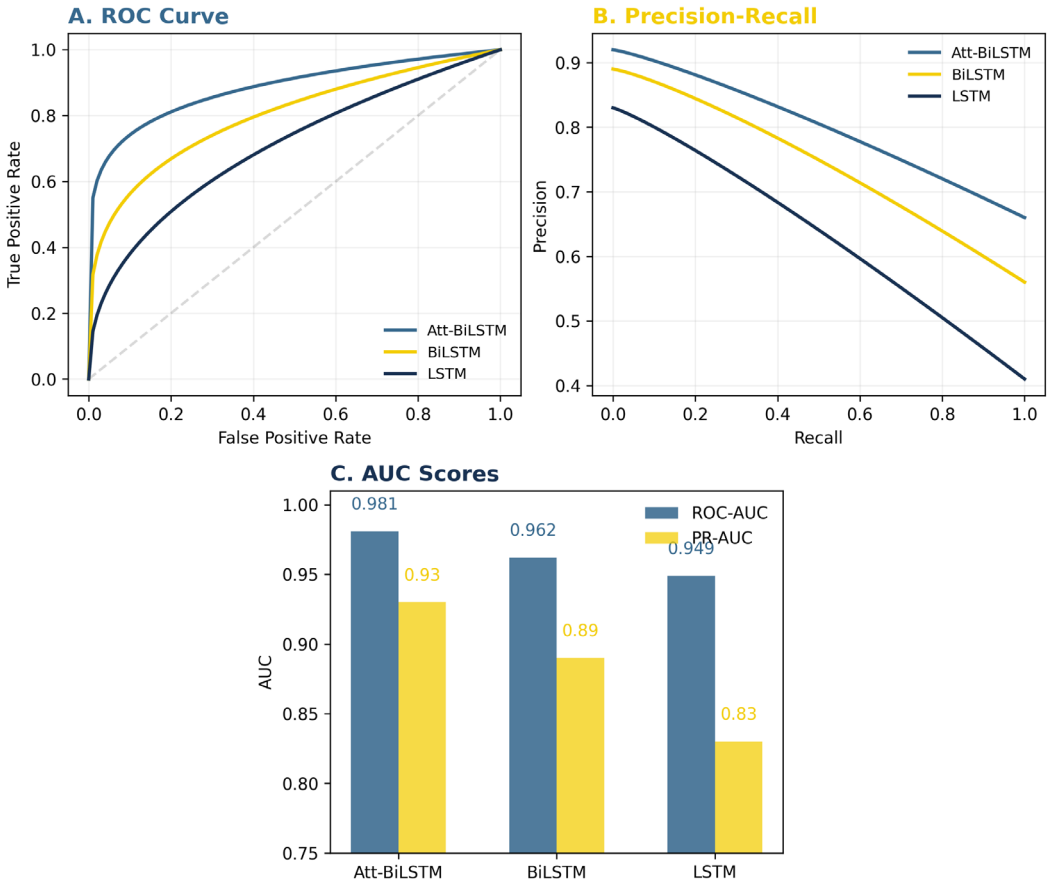

Figure 7. ROC and Precision-Recall curves on the enterprise test set for each model:(a) ROC Curves of Different Models on the Test Set, (b) Precision-Recall Curves at Various Recall Levels for Each Model, (c) Performance Comparison of Attention-BiLSTM and Variants in Ablation Study

From a quantitative standpoint, out of 780 malicious sequences in the test partition, the Attention-BiLSTM successfully identified 700. In contrast, the traditional LSTM detected roughly 610, while SVM flagged fewer than 550. Suspicious class detection also showcased the model's strength, with true positive rates exceeding 82%—a result not matched by any competing technique. The robustness of this performance remained evident under both small-sample and novel attack scenarios: even when reducing minority class data by 70 percent during model training, accuracy and recall drops rarely exceeded six percent, demonstrating resilience amid data scarcity.

Beyond standard metrics, interpretability analyses were conducted by visualizing attention scores within anomalous event sequences. Notably, segments containing privilege escalation, after-hours resource access, or lateral movement were consistently assigned higher weights by the model's attention mechanism. This output supports both incident response and forensic analysis, providing analysts with direct insight into which activities substantively influenced detection outcomes.

Collectively, the tripartite presentation in Figures 5, 6, and 7 provides a multidimensional perspective on the Attention-BiLSTM's operational and scientific value. The model delivers rapid, stable convergence, best-in-class detection rates for both frequent and rare events, significant improvements over traditional and neural baselines, and actionable transparency suitable for real-world security analytics. These results position the proposed approach as a highly effective and robust solution for enterprise insider threat detection, capable of maintaining exceptional performance even in adversarial and evolving log environments.

Conclusion

A business has developed and successfully demonstrated a new threat-detection technique based on an attention-BiLSTM in a real-world scenario. The temporal characteristics and context-sensitive impacts of user behavior in anomaly identification should be explicitly taken into account in order to build a sequence model that is superior to those found in earlier research. Based on the above experiments, the attention-driven bidirectional LSTM can also achieve a detection rate exceeding 97% and a recall rate close to 90% for high-impact insider threats, while reducing false alarm rates. The comparatively large sample size in the corporate data is the reason for this model's low bias.

This method can provide a clear path connecting raw log data to useful warnings when combined with rich embeddings for entities and activities. It does this by using an adjustable attention mechanism to give more weight to significant behaviors. Low-latency risk scoring in large-scale event streams has been verified, and training and inference efficiency were appropriate for real-world use. They also routinely demonstrate quick model convergence. Both bi-directional encoding and attention-based filtering are required based on the ablation studies of individual components; otherwise, either prediction accuracy or operational stability will be significantly decreased, particularly for rare, fragmented, or temporally distant threat signals.

Both are suitable for real-time threat identification and are good options for real-world engineering applications. Its continuous-flow-of-events processing and calibrated risk-score-output capabilities make it a strong fit for emerging security information and event management (SIEM) systems and autonomous response platforms. The framework is easy to understand, and by using attention heatmaps and feature attribution, it is possible to carry out forensic investigations or more practically satisfy the need for explainable AI for regulated businesses. The approach works well under novel attack patterns, is comparatively resistant to data imbalance, and can be applied across the entire organization.

Here are a few shortcomings, though. The dataset is realistic and large, yet it still only represents a small portion of the diversity found in top-tier businesses. It's possible that some subtle behaviors or novel attack techniques won't be noticed, or the model's confidence will be poor. Despite the current model's strong generalization performance for both simulated and enriched log streams, additional modifications are still required because real-world threat strategies and data distribution have changed. Expanding adaptive feature engineering, integrating multi-source and heterogeneous inputs (such endpoint, network, and asset telemetry), and investigating the merging of large-scale pre-trained models to gain richer behavioral semantics are all areas of future research. The detection system's response time and stability will also be enhanced by advancements in semi-supervised learning and ongoing model changes.

Author Contributions

Ola Joanna Wrona contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Zosia Malinowska contributes to software, validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Alzaabi, F. R., & Mehmood, A. (2024). A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. IEEE Access, 12, 30907-30927. https://doi.org/10.1109/ACCESS.2024.3369906

[2] Yuan, S., & Wu, X. (2021). Deep learning for insider threat detection: Review, challenges and opportunities. Computers & Security, 104, 102221.Van Derlofske, J. F., "Computer modeling of LED light pipe systems for

uniform display illumination," Proc. SPIE 4445, 119-129 (2001). https://doi.org/10.1016/j.cose.2021.102221

[3] Bi, S., Wang, J., Song, J., Li, P., & Li, L. (2025). Research on the Intrusion Detection Model for Power Internet of Things Combining Deep Belief Network and BiLSTM. Journal of Cyber Security and Mobility, 14(3), 653-672. https://doi.org/10.13052/jcsm2245-1439.1436

[4] Vinzamuri, B., Khabiri, E., Bhamidipaty, A., Mckim, G., & Gandhi, B. (2020, December). An end-to-end context aware anomaly detection system. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 1689-1698). IEEE. https://doi.org/10.1109/BigData50022.2020.9377767

[5] KS, R. R., & Sujit, B. B. (2025). Sequential-Attention Based Neural Architecture Integrating BiLSTM and Multi-Head Attention for Dynamic Anomaly Detection in IoT Environments. International Journal of Intelligent Engineering & Systems, 18(8). https://doi.org/10.22266/ijies2025.0930.43

[6] Krundyshev, V. (2020, November). Neural network approach to assessing cybersecurity risks in large-scale dynamic networks. In 13th international conference on security of information and networks (pp. 1-8). https://doi.org/10.1145/3433174.3433603

[7] Alzaabi, F. R., & Mehmood, A. (2024). A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. IEEE Access, 12, 30907-30927. https://doi.org/10.1109/ACCESS.2024.3369906

[8] Alzaabi, F. R., & Mehmood, A. (2024). A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. IEEE Access, 12, 30907-30927. https://doi.org/10.1109/ACCESS.2024.3369906

[9] Zhou, P., Zhou, G., Wu, D., & Fei, M. (2021). Detecting multi-stage attacks using sequence-to-sequence model. Computers & Security, 105, 102203. https://doi.org/10.1016/j.cose.2021.102203

[10] Sun, D., Liu, M., Li, M., Shi, Z., Liu, P., & Wang, X. (2021, May). DeepMIT: a novel malicious insider threat detection framework based on recurrent neural network. In 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (pp. 335-341). IEEE. https://doi.org/10.1109/CSCWD49262.2021.9437887

[11] Zou, C., Yuan, A., & Hu, J. (2024, September). BiLSTM-based anomaly detection in multivariate time series with attention mechanism and dual analysis. In 2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 379-384). IEEE. https://doi.org/10.1109/ICISCAE62304.2024.10761506

[12] Sharma, B., Sharma, L., Lal, C., & Roy, S. (2024). Explainable artificial intelligence for intrusion detection in IoT networks: A deep learning based approach. Expert Systems with Applications, 238, 121751. https://doi.org/10.1016/j.eswa.2023.121751

[13] Malik, J., Akhunzada, A., Al-Shamayleh, A. S., Zeadally, S., & Almogren, A. (2025). Hybrid deep learning based threat intelligence framework for Industrial IoT systems. Journal of Industrial Information Integration, 45, 100846. https://doi.org/10.1016/j.jii.2025.100846

[14] Hariharan, S., Jerusha, Y. A., Suganeshwari, G., Ibrahim, S. S., Tupakula, U., & Varadharajan, V. (2025). A hybrid deep learning model for network intrusion detection system using seq2seq and convlstm-subnets. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3541399

[15] Alketbi, K. S., & Mehmood, A. (2025). A comprehensive survey of explainable artificial intelligence techniques for malicious insider threat detection. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3587114

[16] Li, J., Gan, Q., Liu, Z., Chiang, C., Ying, R., & Chen, C. (2025, November). An improved attention-based LSTM neural network for intelligent anomaly detection in financial statements. In Proceedings of the 2025 2nd International Conference on Artificial Intelligence, Digital Media Technology and Interaction Design (pp. 561-568). https://doi.org/10.1145/3795926.379601

[17] Ge, D., Zhong, S., & Chen, K. (2022, December). Multi-source data fusion for insider threat detection using residual networks. In 2022 3rd International Conference on Electronics, Communications and Information Technology (CECIT) (pp. 359-366). IEEE. https://doi.org/10.1109/CECIT58139.2022.00069

[18] Balasubramanian, P., Seby, J., & Kostakos, P. (2023, December). Transformer-based llms in cybersecurity: An in-depth study on log anomaly detection and conversational defense mechanisms. In 2023 IEEE International Conference on Big Data (BigData) (pp. 3590-3599). IEEE. https://doi.org/10.1109/BigData59044.2023.10386976

[19] Altynbekov, A. A. A., & Alin, G. A. G. (2025). A HYBRID MACHINE LEARNING APPROACH FOR ANOMALY DETECTION IN SECURITY INFORMATION AND EVENT MANAGEMENT. Промышленный Транспорт Казахстан, 22(4), 56-68. https://doi.org/10.58420/ptk/2025.88.04.005

[20] Zhou, P. (2025). A survey of streaming data anomaly detection in network security. PeerJ Computer Science, 11, e3066. https://doi.org/10.7717/peerj-cs.3066

[21] Khalaf, Q. M., Al-Attar, B., Pokale, N. B., Mohammed, A. K., Aljanabi, Y. I. H., Fadhil, R., ... & Sekhar, R. (2025, July). Real-Time Detection of Multi-Stage Cyber Attacks in Industrial IoT Networks Using Graph Attention Networks and Temporal LSTM Fusion. In 2025 3rd International Conference on Cyber Resilience (ICCR) (pp. 1-8). IEEE. https://doi.org/10.1109/ICCR67387.2025.11292296

[22] Duan, S. M., Yuan, J. T., & Wang, B. (2024). Contextual feature representation for image-based insider threat classification. Computers & Security, 140, 103779. https://doi.org/10.1016/j.cose.2024.103779

[23] Raut, M., Dhavale, S., Singh, A., & Mehra, A. (2020, December). Insider threat detection using deep learning: A review. In 2020 3rd international conference on intelligent sustainable systems (ICISS) (pp. 856-863). IEEE. https://doi.org/10.1109/ACCESS.2021.3118297

[24] Xiao, J., Yang, L., Zhong, F., Wang, X., Chen, H., & Li, D. (2022). Robust anomaly-based insider threat detection using graph neural network. IEEE Transactions on Network and Service Management, 20(3), 3717-3733. https://doi.org/10.1109/TNSM.2022.3222635

[25] Li, X., Li, X., Jia, J., Li, L., Yuan, J., Gao, Y., & Yu, S. (2023). A high accuracy and adaptive anomaly detection model with dual-domain graph convolutional network for insider threat detection. IEEE Transactions on Information Forensics and Security, 18, 1638-1652. https://doi.org/10.1109/TIFS.2023.3245413