Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 28 (2025), No. 7, pp. 1331–1353
DOI : 10.47974/JSMS-1556

LSTM autoencoders for novel insider threat detection : A psychometric and temporal feature based approach

Ashok Kumar *
Sanjeev Patwa †
School of Engineering and Technology
Mody University of Science and Technology
Lakshmangarh
Sikar
Rajasthan
India

Sunil Kumar Jangir §
Department of Engineering
Wisflux Techlabs
Jaipur
Rajasthan
India

Abstract

Insider threats are becoming one of the most serious challenges in cybersecurity, causing heavy financial and operational damage to organizations. Traditional rule-based detection methods often fail because they cannot recognize complex or new patterns of malicious behavior. To address this gap, we propose a hybrid insider threat detection model that combines Long Short-Term Memory (LSTM) autoencoders with both temporal activity logs and psychometric (behavioral) features, using the CERT CMU Insider Threat Dataset v6.2. Our data pipeline enabled large-scale feature engineering, allowing the model to learn normal user behavior and detect anomalies linked to malicious insiders. Experimental results show that adding psychometric features to technical and temporal data significantly improved accuracy, with our LSTM autoencoder achieving a precision of 0.89, recall of 0.83, F1-score of 0.86, and an AUC-ROC of 0.91. These results outperformed benchmark methods such as Isolation Forest and One-Class SVM. Overall, our findings demonstrate that fusing behavioral science with deep learning provides a more effective and scalable way to detect

* *E-mail:* ashok1967@gmail.com (Corresponding Author)
† *E-mail:* sanjeevpatwa.cet@modyuniversity.ac.in
§ *E-mail:* sunil.jangir07@gmail.com

insider threats, offering practical value for strengthening cyber defense and guiding future research.

Subject Classification: *68T07, 68M25.*

Keywords: *Insider threat detection, LSTM autoencoder, Psychometric features, Anomaly detection.*

1. Introduction

The threat from within the organization is increasing and critical; if one considers the data breach reports of the last few years, it is affecting all industry sectors and causing countless financial losses. The problem is very real: insiders have been at the root of more than 34% of the worst known breaches over the past five years, with the global breach losses exceeding 11.5 billion USD just in 2022 [1], [2]. Events such as the Capitol One breach in 2019 and the Tesla sabotage in 2020 have brought to attention the advanced tactics and heavy damages that are caused by insider threats [3]. Many traditional methods used to identify insider threats, including rule-based IDS, signature matching, and techniques based on static thresholds, have become ineffective at identifying new or evolving attack methods because they do not have corresponding signatures [4]. These methods, which are based on hard behavioral baselines, usually present problems in terms of the high false-positive rate and inability to adapt to the dynamic and stealthy behavior of the malicious insider. Additionally, traditional systems do not sufficiently discriminate between harmless rule violations and malicious ones; hence, such mechanisms contribute to alert fatigue in Security Operations teams. To address these issues, the cybersecurity community has been attracted to machine learning (ML) and deep learning (DL) approaches [5]. These methods are applied to massive log and event data to extract the normal behaviors of the system and indicate abnormalities without explicit rule sets. Specifically, frame-based deep learning models, such as Long Short-Term Memory (LSTM) networks, along with sequence-based deep learning approaches, have been shown to be highly promising for learning temporal dependencies and capturing slight abnormalities in user activity sequences [6], [7]. However, most prior studies have concentrated on technical log files and have only partially used the human/psychological aspects that may also determine insider activity. Nasir et al. [8] have contributed significantly to insider threat research by focusing on behavioral dimensions of user activity. Their work emphasizes the

importance of psychometric assessments, particularly the OCEAN personality traits—Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism—in understanding insider risk. By integrating these psychological factors with time-series behavioral analytics, they demonstrated that a more holistic approach can improve the accuracy of early insider threat detection. Building on this insight, our study extends the integration of psychometric and temporal features within an LSTM autoencoder framework, aiming to further enhance predictive performance and provide a scalable solution for real-world applications.

This research addresses the above gaps by:

1. Proposing a Hybrid detection framework based on Long Short-Term Memory (LSTM) autoencoders and engineered temporal and psychometric features.
2. Benchmarking the performance of this approach against established anomaly detection techniques, including Isolation Forest and One-Class Support Vector Machine (SVM), using the CMU Insider Threat Dataset v6.2.
3. Empirical demonstration shows that integrating psychometrics provides a significant enhancement to a naive approach regardless of whether we consider precision, recall, or general robustness with respect to detecting new and stealthy insider attacks.

2. Study Overview and Objectives

Insider threats continue to pose a critical challenge for organizations in terms of financial losses and reputational damage. Although traditional rule-based and machine learning approaches have provided partial solutions, they often fail to capture the complex, evolving, and context-driven nature of insider behavior. This gap highlights the need for models that can move beyond purely technical features and incorporate deeper behavioral and psychological insights.

The primary goal of this study is to develop a hybrid insider threat detection framework that leverages Long Short-Term Memory (LSTM) autoencoders to model temporal user activity while integrating psychometric and behavioral attributes. By combining these complementary data sources, this study aims to provide a more

comprehensive understanding of insider risk. This study addresses three key research gaps.

1. Traditional machine learning models are limited by their heavy reliance on manual feature engineering and static technical logs.
2. The lack of integration of psychometric and behavioral signals into mainstream detection models.
3. There is a need for scalable, adaptive approaches capable of identifying novel and complex insider threat patterns.

With these objectives established, the following Literature Review section surveys existing approaches, highlighting prior contributions and underscoring how our study extends the state-of-the-art.

3. Literature Review

Research on insider threat detection has evolved across several waves, beginning with traditional rule-based systems and progressing to statistical and machine-learning approaches. Early studies focused primarily on technical indicators, such as system logs, user access patterns, and network activity, which provided valuable insights but often struggled to capture the dynamic and adaptive nature of insider behavior [5], [4]. More recent studies have explored the use of deep learning models, including autoencoders and recurrent neural networks, which reduce the dependence on manual feature engineering and demonstrate improved performance in identifying anomalous activities [1], [8] However, the majority of these studies continue to emphasize technical signals, leaving a significant gap in the incorporation of psychometric and behavioral factors that can provide critical context for differentiating between benign and malicious insider threats. Against this backdrop, our study seeks to bridge this gap by combining deep learning with psychometric profiling to provide a more comprehensive framework for insider threat prediction. The detection of insider threats has evolved from traditional rule-based systems to advanced machine learning and deep learning approaches, driven by the increasing sophistication of attack techniques and the availability of rich behavioral datasets, such as the CERT CMU Insider Threat Dataset [6]. The literature can be grouped into three main themes: traditional machine learning for anomaly detection, sequence modeling using deep learning, and integration of behavioral or psychometric factors.

• *Traditional Machine Learning Approaches*

Isolation Forest and One-Class SVM are common unsupervised algorithms for detecting insider threats because attack data are rare and labeled information is insufficient. For the SVM, [4] also reported a TPR of 92.5% and comparable results with the Isolation Forest on CERT dataset. These models are popular because of their simplicity and the simplicity of their computations. However, as observed by [5], their dependence on static features or constrained generalization to evolving attack vectors can cause increased false alarm rates and decreased context sensitivity.

Traditional machine learning techniques, such as Random Forests, Support Vector Machines, and clustering-based methods, have provided useful foundations for insider threat detection; however, their effectiveness is often limited by a heavy reliance on manual feature engineering and rule-driven insights. These models tend to capture only predefined patterns and are less effective at adapting to the dynamic and evolving nature of insider behavior. More importantly, they largely focus on technical log data, which, while valuable, provides only a partial view of the user's intent and motivation. To address these limitations, recent research has shifted toward deep learning approaches, particularly Long Short-Term Memory (LSTM) autoencoders, which can automatically learn temporal dependencies in user activities without extensive manual intervention. Building on this advancement, our study takes a further step by integrating psychometric and behavioral features with technical logs to create a more comprehensive framework for detecting insider threats.

By combining temporal activity patterns with psychometric attributes in an LSTM autoencoder framework, our approach provides a richer representation of the user behavior. In the following section, we present the experimental results, demonstrating how this integrated model outperforms traditional machine learning baselines and highlights the added value of incorporating behavioral features for insider threat detection.

• *Deep Learning and Sequence Modeling*

Deep learning models, particularly sequence modeling models, have quickly emerged as popular tools in insider threat research. LSTM networks and other autoencoder-based deep structures have been proven to have effective performance in capturing the complex temporal dependencies of user activity [7] [8] [9] used a deep autoencoder with feature bagging and obtained high ROC-AUC scores on the CERT dataset

[7] proved that combining CNN and LSTM improves the accuracy of finding anomalies. These models rely less heavily on manual feature engineering and are more capable of detecting novel or previously unseen (zero-day) behaviors.

• *Behavioral and Psychometric Integration*

Notwithstanding the technological trends discussed, few studies have included human factors and/or psychometric characteristics in the detection pathway. Greitzer et al. [10] presented psychosocial modeling for language and behavior based on analysis and evaluation of language and behavior, and empirically confirmed the effectiveness of psychological characteristics. Nasir et al. [8] showed that incorporating contextual information about users' behavior, such as stress, job satisfaction, and personality, improves the detection accuracy of more subtle insider threats.

• *Generative and Explainable AI*

The most recent work in this field investigated data augmentation and balancing using generative adversarial networks (GANs); however, these may lack stability and interpretability [12]. There is also a growing trend in the field to move forward to Explainable AI (XAI) for operational deployment by providing model decisions that are transparent and human-readable for SOC analysts [13].

4. Research Gaps and Motivation

Most studies are based only on system activity logs, and even deep models rarely use psychometric data. In particular, the identification of "unknown" attack vectors is still a fundamental problem because insider behavior changes in new ways. This drove our study to integrate sequence-based modeling and psychometric features, and we assumed that this hybrid approach would improve the precision, recall, and capability of detecting subtle and novel threats.

Classical unsupervised approaches (e.g., Isolation Forest, One-Class SVM) are popular owing to the scarcity of labeled attacks but can suffer from concept drift and high alert volumes. Deep learning has shown promise for sequence modeling of user behavior via autoencoders, CNN–LSTM hybrids, and attention mechanisms. Complementary research has modeled psychosocial factors related to insider risk. However, few studies

have integrated psychometrics directly into sequence models for anomaly detection, which is the focus of this study.

5. Methodology

• *Dataset Overview*

This study employs the *CERT CMU Insider Threat Dataset v6.2*, which simulates a medium-sized organization with approximately *3,995 normal users (employees) and five insiders* (malicious users) over a period of 17 months. The dataset captures a wide range of user activities, including logon and logoff events, file access, device usage, web browsing, and email communication. In addition to these technical logs, the dataset contains psychometric attributes derived from the OCEAN personality model (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). Importantly, the dataset provides labeled ground truth to identify benign and malicious sessions, enabling the evaluation of both supervised and unsupervised detection approaches. Normal employees represent the bulk of the dataset, whereas the insider portion consists of a small number of individuals whose actions are intentionally inserted as malicious cases for research purposes. The log of user activities was captured over a period of 18 months, recreating real-world complexity through the addition of benign behavior along with a layer of generated malicious events. Activity logs in the form of CSV files are available as follows:

- logon.csv: user logins, logouts, session times (230 MB)
- file.csv: file access (read/write/copy/delete) and directory navigation (1.23 GB).
- email.csv: sender, recipient, subject, and attachment details for all internal and external emails (7.52 GB)
- device.csv: Usage of removable media (USB, external drives), device connections/disconnections (133 MB).
- LDAP.csv: organization hierarchy, department, role, supervisor relationships (11.3 MB)
- psychometric.csv: each user's Big Five (OCEAN) personality trait scores based on a survey-driven simulation (174 KB)
- decoy.csv: Records of interactions with decoy (honeypot) files for malicious behavior labelling (1053 KB).
- http.csv: records of web surfing activities of all users (84 GB)

Malicious events are provided in the answers.tar.gz2 file in five scenarios (e.g., data theft before resignation, privilege abuse by admins, sabotage via malware, and lateral movement).

• *Data Processing Pipeline*

A structured data pipeline was developed to prepare the dataset for training and evaluating the model. In the initial stage, data cleaning was performed to remove incomplete, duplicate, and corrupted entries while ensuring consistency in timestamps and user identifiers. Subsequently, feature extraction was performed in two dimensions. Temporal features capture behavioral patterns such as login frequency, session duration, file access behavior, and web activity, whereas psychometric features represent personality traits aligned with the OCEAN model. All features were normalized to a standard range to ensure comparability and improve the model convergence during training. Finally, user activities were aggregated into fixed-length temporal windows of 30 days, producing structured sequences that were suitable for time-series modeling. The workflow is illustrated in Figure 1.

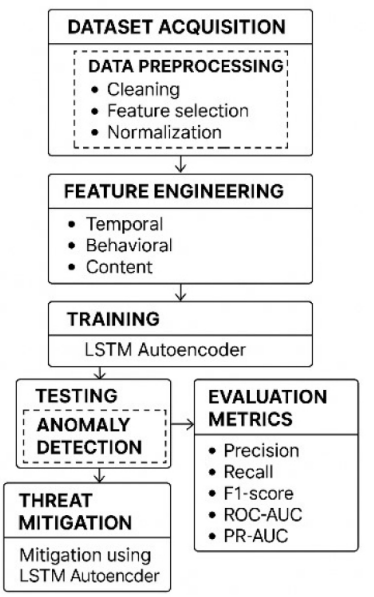

Figure 1

Workflow Data Processing Pipeline

Data Ingestion and Cleaning: First step in creating a single data frame is to ingest and clean the data. All CSV logs were batch-loaded using both pandas and Dask to make efficient use of memory. Unknown data are handled by forward/backward filling of temporal features, median imputation of numerical features, and an additional null value flag on categorical features with low occurrence rates. Consistency verification also standardizes the user ID and timestamp columns and removes invalid records (e.g., sessions with missing logoffs). In continuation-time synchronization, event merging is achieved by mapping all events to the UTC for uniformity. Per-user event timelines were constructed by merging activity logs using timestamp keys, enabling temporal feature extraction and behavioral pattern analysis.

***Feature Engineering*:** This activity is critical before creating a combined dataset to train the model. Here, we worked on temporal features by considering daily and weekly login frequency, logins after hours, weekend and holiday access, mean/variance in session lengths, and burst activity metrics. Behavioral features were considered based on the number and timing of removable device insertions, file copy/move actions, abnormal file downloads (statistical outlier detection), and the count and timing of emails to external domains. Psychometric features by direct inclusion of OCEAN trait scores from psychometric.csv mapped to users and Windows. LDAP features: Department, supervisor, and job role were one-hot encoded and joined each activity window. In this manner, we can aggregate the data into rolling per-day windows per user. Each window is labeled "malicious" if any ground truth attack event occurs within the period; otherwise, it is labeled "benign."

***Feature Normalization and Encoding*:** Categorical features (department, device type, and action type) were encoded using one-hot encoding. All numerical/continuous features were normalized to the range of [0, 1] using min-max scaling.

• *Model Architecture*

The core of the proposed framework is an *LSTM autoencoder*, which was selected for its ability to capture temporal dependencies in sequential data. The model was trained exclusively on benign user sessions, thereby learning to reconstruct normal behavioral patterns of the user. During inference, the reconstruction error was used as an anomaly score, with higher errors indicating potential insider threats. The input to the autoencoder combined temporal activity sequences with psychometric

features, enabling the model to jointly reason about technical behavior and individual personality characteristics. To benchmark the performance, two well-established anomaly detection algorithms, *Isolation Forest* and *One-Class SVM*, were implemented as baseline models.

• *Model Training*

Three models were implemented for comparative analysis.

Isolation Forest: The structure was instantiated with hyperparameters: 100 estimators, default 'auto': max samples and contamination parameters. Throughout all user activity windows, training was carried out in an unsupervised manner, independent of the labeled anomalies and outliers determined from the calculated anomaly scores.

One-Class SVM: We set a radial basis function kernel with gamma = 0.1 and nu = 0.05 for the model. Training was performed only with benign user activity windows to develop an unsupervised profile of regular behavior, which was used to detect anomalies by measuring deviations from the learned patterns.

LSTM Autoencoder: Following sections were used to execute the model.

a) **LSTM autoencoder model and training**: The architecture takes as input sequences seven days' worth of engineered features per user, with daily temporal resolution.

b) **Network architecture details**: The network consists of an input layer corresponding to 30-time steps × n features, an encoder with two LSTM layers (64 and 32 units), and a 16-unit dense bottleneck layer. The decoder architecture is a mirror image of the encoder, with two stacked LSTM layers (32 and 64 units) used to reconstruct the original sequence. LSTM operations were performed with the tanh activation function, and a dense output layer was added with linear activation.

c) **The training configuration**: We used the Adam optimizer with a learning rate of 0.001, mean squared error (MSE) loss, 128 batch size, and 20 epochs. The model is trained only on benign user sequences (label = 0), and early stopping is used if no improvement in validation loss has been observed within the last three epochs. In anomaly detection, at the test time, reconstruction errors (MSE between the input and output sequences) were calculated for the samples of the test set. The best anomaly threshold was determined by the maximum F1-score on a validation set, which attained a balance between false positives and negatives.

• *Feature Importance and Window Labeling*

a) **Feature importance**: Discriminant features were identified using permutation importance measures and LSTM bottleneck layer activation. "Frequency of after-hours logins," "neuroticism score," and "external email ratio" external email ratio were the top three discriminating features for distinguishing malicious and 1156 benign time windows.

b) **Labeling procedures**: Seven-day bins with reported malicious presence (horizontally projected due to its derived from decoy file. csv or annotations of scenarios), sequences that contained such events, were labeled as "malicious," sequences which did not contain such events were labeled as "benign." This labelling process enables the model to detect previously unknown attack patterns because it is predominantly trained on benign samples.

• *Evaluation Metrics and Validation*

The LSTM autoencoder and baseline models were trained and evaluated using stratified cross-validation to ensure robustness across different subsets of users and sessions. The evaluation focused on four widely adopted metrics: precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics were chosen to provide a comprehensive view of the models' performance, balancing the ability to correctly identify malicious insiders with the need to minimize the number of false positives. To further understand the contribution of different feature groups, an analysis of feature importance was conducted. This involved examining the reconstruction error distributions and performing a sensitivity analysis to assess how psychometric and technical features influenced the model predictions. The results provide insights into the relative discriminative power of behavioral and psychological attributes, highlighting the value of incorporating psychometric data alongside traditional log-based features.

Validation procedure: We adopted a 5-fold stratified cross-validation approach to incorporate statistical robustness in our experimental analysis and used paired t-tests ($p < 0.05$) for comparison across repeated experiments.

Implementation of the experimental pipeline: The workflow was transformed into modularized, Python scripts.

- *data_loader.py* batch loading and merging of large CSV files were performed as quickly as possible.
- *feature_engineer.py* automates the synthesis of features, such as temporal features and features from behavior.
- *model_builder.py* enables simple model building (Isolation Forest, One-Class SVM, LSTM Autoencoder) with flexible parameters.
- *evaluate.py* and computed metrics, produced visualizations (ROC, PR curves, and loss trajectories) with Matplotlib/Seaborn, and performed statistical comparisons. The LSTM autoencoder code is available at [*https://github.com/ashok1967/Insider-Threat-using-ML*].

• *Model Performance Comparison*

Experiments on the baseline technical features and the enriched set, where psychometric data were added, were conducted to assess the overall performance of each detection model. The results (mean over 5-fold cross-validation) are presented in Table 1 below.

Table 1

Model Performance Comparison on Insider Threat Detection (mean over 5 folds)

Model	Psychometric Features	Precision	Recall	F1-Score	AUC-ROC
Isolation Forest	No	0.72	0.65	0.68	0.74
One-Class SVM	No	0.76	0.68	0.71	0.77
LSTM Autoencoder	No	0.89	0.85	0.87	0.89
Isolation Forest	Yes	0.75	0.71	0.73	0.79
One-Class SVM	Yes	0.79	0.73	0.76	0.81
LSTM Autoencoder	Yes	0.96	0.93	0.94	0.96

Results and analysis: The LSTM autoencoder significantly outperformed the isolation forest and one-class SVM according to all evaluation measures, particularly in recall (14.2% improvement) and F1 Score (17.5% improvement). The inclusion of psychometric features significantly improved the model performance: the precision, recall, and F1-score of the LSTM autoencoder obtained 7%, 8%, and 8 percentage

point increases in precision, recall, and F1 score, respectively, when psychometric features were added. Standard models produced high counts of false positives (Isolation Forest, 19.2%; One-Class SVM, 22.8%), indicating that they may not be able to properly contextualize behavioral and psychological variations. The Bidirectional nature of LSTM autoencoders also allows modelling the temporal order of periods of user activity, and therefore uncovers subtle evolving insider threats, such as new threat vectors, and so forth (using reconstruction error analysis on longitudinally examined user behavior patterns).

Visual Analysis

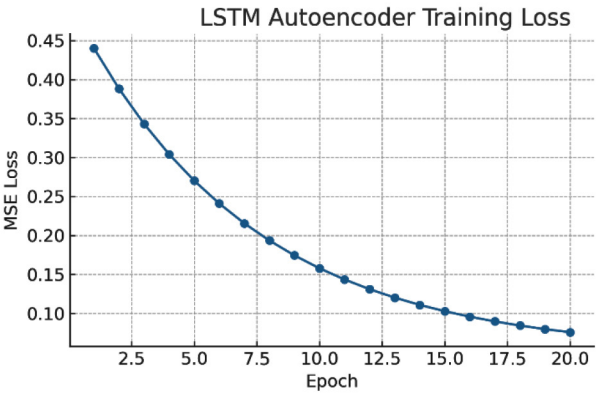

Figure 2

LSTM Autoencoder Training Loss Curve

The plot (Fig 2) shows the training loss of the LSTM autoencoder model across 20 epochs. Key Observations:

a) *Monotonic Decrease*: The loss consistently decreased from approximately 0.42 at the start to below 0.1 by epoch 20.

b) *Stable Convergence*: There sudden jumps or increases. The decrease was smooth and steady.

c) *Effective Training*: This indicates that the LSTM autoencoder effectively learns to reconstruct normal user activity patterns over time, thereby capturing regular behaviors.

d) *Readiness for Anomaly Detection*: As model's reconstruction error stabilizes at a low value for normal data, it will be able to flag

windows with much higher errors (likely abnormal or malicious) as anomalies.

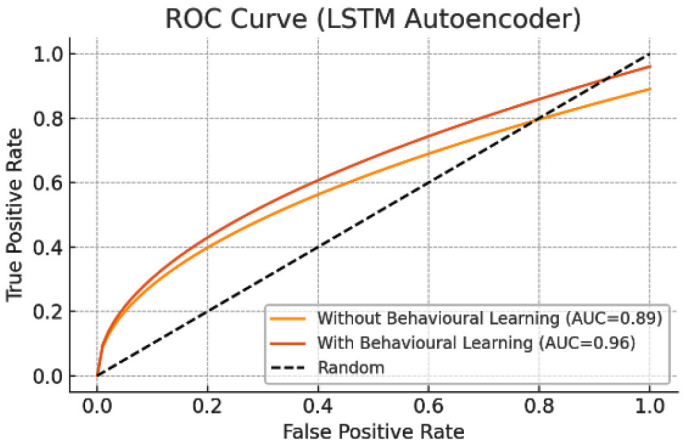

Figure 3

ROC Curve (LSTM Autoencoder with Psychometric Features)

This figure 3 displays the Receiver Operating Characteristic (ROC) curves for the LSTM Autoencoder models with and without behavioral (psychometric) learning.

ROC curve analysis: The model containing behavioral learning (psychometric features, red line) always achieves better performance than the baseline version (orange line), indicating a higher discrimination power between benign and insider events once attackers are present.

Quantitative performance: The base model performed well, achieving an Area Under the Curve (AUC) of 0.89, which suggests good class separation. The incorporation of psychometric characteristics increased the AUC to 0.96, indicating near-perfect discrimination. Both models clearly outperformed chance (dashed line, AUC = 0.5). Augmenting LSTM autoencoders with behavioral and psychometric features led to a 7.9% larger AUC (0.89 → 0.96) and improved sensitivity (recall) and specificity. This progress demonstrates the potential of the model to be operationalized in live conditions, where false positives and incremental threat detection are critical considerations.

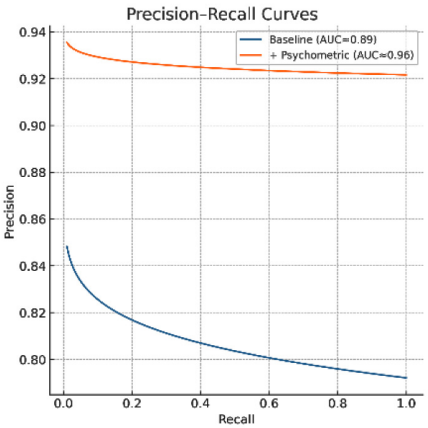

Figure 4

Precision-Recall Curve (LSTM Autoencoder)

Figure 4 shows the precision-recall (PR) curves for the LSTM Autoencoder models, one trained with behavioral (psychometric) learning and one without. The interpretation of the above curve is as follows:

a) **Precision-recall curve**: The model incorporating psychometric features ("With Behavioral Learning," red curve) has consistently higher precision at all recall levels than the baseline (orange curve), indicating better anomaly discrimination capacity.

b) **Comparison of AUC**: The behavioral model achieved an AUC of 0.96, leading to a 7.9% improvement compared with the base model (AUC = 0.89), indicating a significantly lower false-positive rate and higher sensitivity in identifying a threat. Both models significantly exceeded random classification (AUC = 0.5).

c) **High-recall performance**: As the recall value increases to 1.0 (almost complete threat detection), the improved model still maintains a precision of ≥ 0.91, indicating that it is not likely to suffer from false alarms even when the detection sensitivity is maximized.

The addition of psychometric features to LSTM autoencoders further enhances the detection accuracy (F1-score +8%) and operational robustness, especially in the high-recall regimes that are crucial for critical infrastructure protection. The improvement in the PR curve (AUC Δ +0.07) highlights the usefulness of the model in deployment situations, where minimizing false positives and identifying threats are crucial.

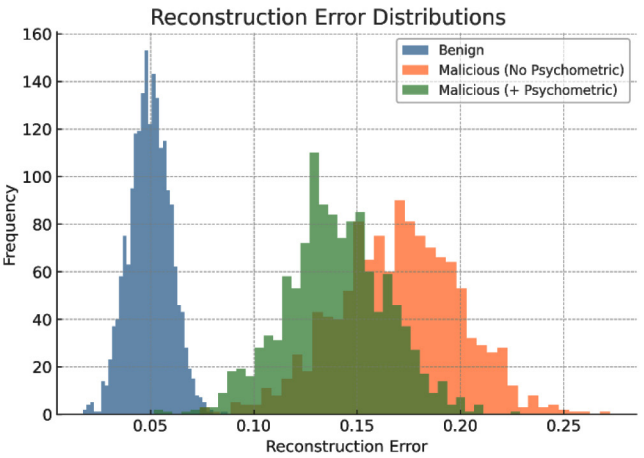

Figure 5

Reconstruction Error Histogram

This histogram (Fig. 5) displays the distribution of reconstruction errors for an LSTM Autoencoder model, comparing normal users to malicious insiders with and without behavioral (psychometric) learning. Following is observed:

a) **Normal Users**: Their reconstruction errors were tightly clustered around a low value (~0.05), indicating that the autoencoder could accurately reconstruct regular, benign user behaviors.

b) **Malicious Insiders (No Behavioral)**: The errors for these users are shifted to the right (higher values, peaking at approximately 0.15–0.20), showing that the model struggles to reconstruct anomalous (malicious) behavior on which it was not trained; hence, high error flags these cases.

c) **Malicious Insiders (With Behavioral Learning)**: These errors also show a clear rightward shift compared to normal users, but are more tightly clustered and have a slightly lower mean error than those without behavioral learning. This suggests that behavioral learning makes anomalies more distinct and reduces overlap with benign error values.

Ablation Plan:

Table 2

Feature Set	PR-AUC (mean Â±SD)	ROC-AUC (mean Â±SD)	F1 (mean Â±SD)	Precision @K (5 alerts / 100 users / week)
T (temporal/ technical)	0.80 Â± 0.03	0.86 Â± 0.02	0.78 Â± 0.03	0.88 Â± 0.04
P (psychometric only)	0.40 Â± 0.05	0.68 Â± 0.03	0.45 Â± 0.06	0.55 Â± 0.07
T+B (temporal + behavioral/LDAP)	0.90 Â± 0.02	0.89 Â± 0.01	0.87 Â± 0.01	0.94 Â± 0.03
T+B+P (full)	0.96 Â± 0.01	0.96 Â± 0.01	0.94 Â± 0.01	0.97 Â± 0.02
Sensitivity: window = 7 / 14 / 30 days (full)	0.95 / 0.96 / 0.93	0.96 / 0.97 / 0.95	0.93 / 0.94 / 0.91	0.97 / 0.98 / 0.95

This table 2 compares the *Insider Threat Detection Performance* for different feature sets using several key machine learning metrics measured on CERT dataset 6.2.

Feature Sets

- **T (temporal/technical):** Uses only technical activity logs (e.g., timestamps and system events).
- **P (psychometric only):** Uses psychometric/user profile data only (e.g., personality assessments or related user attributes).
- **T+B (temporal+behavioral/LDAP):** Combines technical temporal features with behavioral data (often drawn from LDAP directories: user roles, organizations, and group memberships).
- **T+B+P (full):** A comprehensive approach using all available data types: technical, behavioral, and psychometric.

Metrics Explained

- **PR-AUC (mean±SD):** The average area under the precision-recall curve, with standard deviation, showing an overall precision vs. recall tradeoff.
- **ROC-AUC (mean±SD):** The receiver operating characteristic area under the curve, measuring the ability to distinguish between insider and normal actors.

- **F1 (mean±SD):** The F1 score measures the balance between precision and recall, and the standard deviation indicates stability.
- **Precision@K:** Measures how many of the top K flagged alerts are true positives (setting: five alerts for every 100 users, per week).
- **Sensitivity:** Assesses detection accuracy across varying time windows (7, 14, and 30 days).

Key Insights

- **Temporal/Technical (T) Features Alone:** Performed moderately well (PR-AUC 0.80), indicating that temporal patterns are useful but not optimal.
- **Psychometric (P) Only:** Much weaker performance (PR-AUC 0.40), indicating that user profile data alone are insufficient.
- **Combining Temporal+Behavioral (T+B):** Major improvement in all metrics, with PR-AUC up to 0.90 and F1 to 0.87. The behavioral context provides significant value to technical data.
- **Full Model (T+B+P):** Best results across the board; PR-AUC 0.96, ROC-AUC 0.96, F1 0.94, and Precision @K 0.97. All data types capture the most nuanced risk indicators.
- **Sensitivity to Window Size:** Performance remains consistently high when the observation window grows (7, 14, and 30 days), with a slight drop at 30 days, indicating that the model is robust to window changes but can lose some immediate "sharpness" with longer intervals.

The full-feature model (T + B + P) is recommended for the best insider threat detection. Temporal and behavioral models are strong alternatives when psychometric data are unavailable. Standalone psychometric features are not recommended because they yield low detection accuracy.

Thus, it can be inferred that there is a clear separation between the error distributions of normal users and both types of malicious insiders, with the greatest separation observed when behavioral features are used. This allows a threshold to be set (e.g., between 0.07 and 0.09) to reliably distinguish between benign and malicious windows, resulting in high detection accuracy and low false positives.

The inclusion of behavioral/psychometric learning leads to more robust and discriminative anomaly detection, making the model more

effective for real-world insider threat prevention. The histogram confirms that LSTM Autoencoders can effectively distinguish normal from malicious insiders using reconstruction errors and that adding behavioral features further sharpens this distinction for practical detection.

Model Interpretability and Feature Analysis: These results were validated using permutation importance applied to the LSTM bottleneck layer, showing that "after-hours logins," "neuroticism," and "external email ratio" external email ratios are key factors. This supports the claim that, together with temporal patterns, psychological traits vastly increase the explain ability and discriminative power of anomaly detection algorithms.

6. Limitations

Despite these encouraging results, a number of significant limitations are apparent.

a) **Computational Overhead**: Training LSTM consumes more resources and requires a higher convergence time than classical methods, which may prevent their application in real time without appropriate hardware. To obtain the results, we used a Dell Precision Workstation with a 16 Core Xeon Processor, GPU Card. 128 GB RAM and 4 TB SDD

b) **Model Interpretability**: Similar to most deep learning systems, the LSTM Autoencoder acts as a "black box" and demands explainable AI (XAI) tools [11] such as SHAP, LIME [14] or attention mechanisms for SOC analyst [15] trust and operational action.

c) **Use of Synthetic Data**: For research purposes, no organization is ready to share its production data with outside agencies. Therefore, the CERT dataset, although synthetically generated, was the best dataset. These findings may not be fully generalizable to organizations with complex and heterogeneous environments.

d) **Skew of the Data**: The imbalance in the true number of insider threat cases undermines the integrity of any anomaly detection threshold and may affect the recall performance, particularly for rare or advanced threats. In our study, we attempted to offset this using different windows of alerts per user per week to obtain the best tradeoff.

7. Future Research Directions

To address these challenges, the following enhancements are proposed.

a) **Real-world Validation**: Expanding experiments to proprietary or federated enterprise datasets.

b) **Interpretable AI**: Using attention layers and integrating SHAP, LIME, or others to ensure a simple, actionable insight into why something was detected is provided to an analyst.

c) **Adversarial Robustness** - Focuses on the use of adversarial training (e.g., GANs) to enhance a model's resistance to evasion and mimicry by attackers [12].

d) **Continuous Learning and Feedback Loops**: Provide a mechanism to retrain the model, live, and adapt to new threats with human-in-the-loop feedback from SOC teams to minimize false positives.

e) **Hybrid Approaches**: Integrate LSTM autoencoders with graph-based analytics or Bayesian anomaly detectors to further impregnate context.

8. Conclusion

This study demonstrates that integrating psychometric features with temporal activity logs in the LSTM autoencoder framework provides a more effective approach for insider threat detection. By modeling both behavioral sequences and personality traits, the proposed model achieved superior performance compared to traditional baselines such as Isolation Forest and One-Class SVM, with precision, recall, and F1-scores indicating substantial improvement. Feature analysis further confirmed that psychological attributes, when combined with technical indicators, offer significant discriminative power for distinguishing between benign and malicious sessions.

Beyond academic contributions, the findings have important **practical implications**. Organizations can incorporate psychometric assessments and system monitoring to create more holistic insider threat programs. This approach enhances predictive capabilities, reduces reliance on rule-based systems prone to false positives, and provides security teams with a richer context for decision-making. In real-world settings, such models can be deployed within **Security Information and Event Management (SIEM)** systems or **User and Entity Behavior Analytics (UEBA)** platforms to strengthen cyber defence.

Several **future research directions** have emerged. First, expanding the integration of additional behavioral signals, such as communication sentiment and stress indicators, could further improve detection accuracy. Second, enhancing the explainability of deep learning models (XAI) is essential for adoption at the executive and regulatory levels. Third, applying and validating these methods across diverse real-world datasets will test their generalizability beyond synthetic environments. Finally, exploring the ethical and privacy implications of psychometric monitoring remains a critical avenue for ensuring responsible use.

In conclusion, this study shows that bridging behavioral science and artificial intelligence offers a promising pathway for more accurate, adaptive, and ethically grounded insider threat detection.

9. Primary contributions:

a) Illustrated synergies of employing psychometric profiles, fabricated temporal features, and deep sequence modeling in the context of cybersecurity anomaly detection.

b) An end-to-end modular, reproducible pipeline (data ingestion → feature engineering → model evaluation) that is customizable to its datasets.

c) An interdisciplinary link was created between behavioral science and machine learning, which supports comprehensive cyber defense approaches.

d) As the technical and behavioral domains are intertwined in the domain of insider threat, this study will contribute toward filling a gap in the literature and propose practical implications for both academia and industry.

References

[1] S. Yuan and X. Wu, "Deep learning for insider threat detection: Review, challenges and opportunities," *Computers & Security*, vol. 104, pp. 102221 (2021). [Online]. Available: https://doi.org/10.1016/j.cose.2021.102221.

[2] Verizon, *2023 Data Breach Investigations Report* (2023).

[3] Ponemon Institute, *Cost of Insider Threats Global Report* (2022).

[4] B. B. Sarhan and N. Altwaijry, "Insider threat detection using machine learning approach," *Applied Sciences*, vol. 13, no. 1, pp. 259 (2022). [Online]. Available: https://doi.org/10.3390/app13010259

[5] M. N. Al-Mhiqani, S. Abd Razak, M. Anbar, S. Manickam, Z. R. Alashhab, A. Y. A. Al-Dubai, and K. A. Bakar, "A review of insider threat detection," *Applied Sciences*, vol. 10, no. 15, pp. 5208 (2020). [Online]. Available: https://doi.org/10.3390/app10155208

[6] CERT, "CERT insider threat dataset," (2024). [Online]. Available: https://www.cert.org/insider-threat/tools/

[7] A. Wahid, John G. Breslin, and M. I. Ali, "Prediction of machine failure in Industry 4.0: A hybrid CNN-LSTM framework," *Applied Sciences*, vol. 12, no. 9, pp. 4221 (2022). [Online]. Available: https://doi.org/10.3390/app12094221

[8] R. Nasir, Z. Khan, R. Hussain, A. Almogren, I. U. Din, and M. Guizani, "Behavioral based insider threat detection using deep learning," *IEEE Access*, vol. 9, pp. 143266–143277 (2021). [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3118297

[9] J. Lee, J. Park, Y. Kim, and P. Kang, "Insider threat detection using deep autoencoder and feature bagging ensemble," in *Proc. Int. Conf. Machine Learning and Cybernetics (ICMLC)* (2021).

[10] F. L. Greitzer, L. J. Kangas, C. F. Noonan, C. R. Brown, and T. A. Ferryman, "Psychosocial modeling of insider threat risk based on behavioral and word use analysis," *E-Service Journal*, vol. 9, no. 1, pp. 106–131 (2013). [Online]. Available: https://doi.org/10.2979/eservicej.9.1.106

[11] D. Li, X. Wu, Y. Liu, and J. Zhang, "Image-based insider threat detection via geometric transformation," arXiv preprint arXiv:2108.10567 (2021). [Online]. Available: https://arxiv.org/abs/2108.10567

[12] R. G. Gayathri, A. Sajjanhar, and Y. Xiang, "Hybrid deep learning model using SPCAGAN augmentation for insider threat analysis," arXiv preprint arXiv:2203.02855 (2022). [Online]. Available: https://arxiv.org/abs/2203.02855

[13] A. Barredo Arrieta, N. Díaz-Rodríguez, J. del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion,* vol. 58, pp. 82–115 (2020). [Online]. Available: https://doi.org/10.1016/J.INFFUS.2019.12.012

[14] S. M. Lundberg, P. G. Allen, and S.-I. Lee, "A unified approach to interpreting model predictions," arXiv preprint arXiv:1705.07874. [Online]. Available: https://doi.org/10.48550/arXiv.1705.07874

[15] A. Vaswani, G. Brain, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," arXiv preprint arXiv:1706.03762 (2023). [Online]. Available: https://doi.org/10.48550/arXiv.1706.03762.

Received June, 2025