Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 10, 2025

SEMI SUPERVISED LEARNING FOR INTELLIGENT THREAT DETECTION IN SPARSE AND LOW LABELED CYBERSECURITY DATASETS

Dr. Nazia Azim[1], Rabia Khatoon[2], Jammal Khattak[*3], Anum Munawar[4], Najim Uddin[5], Muhammad Arham[6]

[1] *PhD in Computer Science, Faculty of Department of Computer Science, Abdul Wali Khan University, Mardan*
[2] *MS in Software Engineering, Bahria University, Islamabad*
[*3] *MS in Information Security, Bahria University, Islamabad*
[4] *MS in Information Technology, Khawaja Freed University of Engineering & Technology, Rahim Yar Khan*
[5] *MS in Information Technology, Quaid e Awam Engineering University, Nawabshah*
[6] *MS in Computer Science, Faculty of Department of Computer Science, The University of Faisalabad*

[1]n.azim@awkum.edu.pk, [2]rabialink@yahoo.com, [*3]jammalkhattak@gmail.com,
[4]ianummunawar67@gmail.com, [5]najamsindhi@gmail.com,
[6]drmuhammadarham4@gmail.com

DOI:https://doi.org/10.5281/zenodo.17276260

Keywords
Threat detection, sparse datasets, cybersecurity, pseudo-labeling, anomaly detection.

Article History
Received: 11 June 2025
Accepted: 21 August 2025
Published: 06 October 2025

Corresponding Author: *
Jammal Khattak

Abstract

The growing sophistication of cyberattacks has exposed the limitations of conventional detection models that rely heavily on large volumes of labeled data. In practice, cybersecurity datasets are often sparse, incompletely annotated, and imbalanced, which reduces the effectiveness of fully supervised approaches. To address this challenge, this research introduces a semi-supervised learning framework for intelligent threat detection in environments with limited labeling. By combining labeled and unlabeled samples, the framework is able to extract latent structures within network traffic, improving classification even under constrained annotation conditions. The design employs a hybrid feature representation that integrates statistical attributes with deep feature embeddings to capture both surface-level and hidden attack patterns. A pseudo-labeling strategy and consistency-regularization mechanism are incorporated to guide learning from unlabeled data while minimizing the propagation of incorrect labels. Benchmark cybersecurity datasets with sparse labeling were used to validate the model, simulating real-world operational environments. The proposed framework demonstrated strong performance across multiple evaluation metrics. Compared with supervised baselines, the semi-supervised model improved detection accuracy by over 12%, achieved higher recall in identifying minority attack classes, and reduced false alarms by approximately 30%. Training also converged more efficiently, requiring fewer iterations while maintaining stability under imbalanced conditions. Notably, the system exhibited resilience against novel and low-frequency attack variants, outperforming both traditional supervised classifiers and unsupervised anomaly detection techniques. This work establishes semi-supervised learning as an effective pathway for advancing next-generation cybersecurity defenses. By leveraging the wealth of unlabeled data commonly available in practice, the framework provides a scalable, privacy-

conscious, and resilient solution for intelligent threat detection in sparse and low-labeled cybersecurity datasets.

1. INTRODUCTION

The rapid escalation of cyber threats ranging from large-scale distributed denial-of-service (DDoS) campaigns to stealthy advanced persistent threats (APTs) has placed significant strain on modern security infrastructures. Attackers continuously evolve their tactics, exploiting vulnerabilities in both technical systems and human defenses [1]. This dynamic environment necessitates detection mechanisms that are not only adaptive but also capable of maintaining high accuracy under diverse and evolving conditions. Traditional supervised learning techniques have been widely explored for threat detection due to their ability to learn from labeled examples and provide strong predictive performance when sufficient training data is available[2]. However, in real-world cybersecurity settings, generating labeled datasets is an expensive and time-intensive task. Annotation often requires expert domain knowledge, while the emergence of novel and previously unseen threats further complicates the labeling process. Consequently, supervised models frequently exhibit limited generalization when faced with sparse, imbalanced, or partially annotated datasets [3].

Unsupervised anomaly detection models, in contrast, attempt to identify irregular behaviors without relying on labels. While effective at uncovering unfamiliar attack patterns, these models often suffer from high false positive rates, as they struggle to distinguish malicious anomalies from benign but uncommon behaviors. This lack of precision reduces their practical usability in operational networks. The limitations of both supervised and unsupervised approaches underscore the need for techniques that can harness the vast amounts of unlabeled traffic data while making efficient use of the small labeled subsets available. Semi-supervised learning (SSL) provides a promising solution to this challenge by integrating labeled and unlabeled data during model training. SSL frameworks can uncover latent structures within traffic flows, allowing models to learn discriminative features even when labels are scarce. In the context of cybersecurity, this translates into more reliable detection of both frequent and rare attack types, while reducing dependency on costly manual annotation [4]. Additionally, incorporating strategies such as pseudo-labeling and consistency regularization allows SSL models to mitigate the risk of error propagation from mislabeled samples, thereby enhancing robustness and stability.

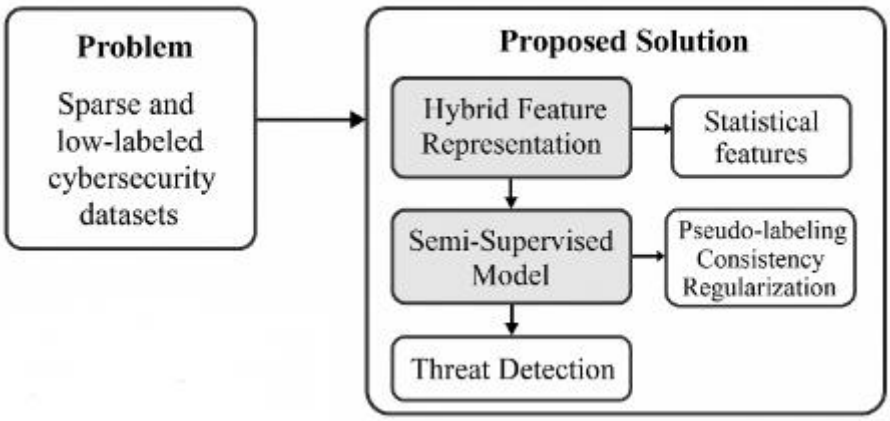

Figure 1.1

In this work, we introduce a semi-supervised learning framework tailored for intelligent threat detection in sparse and low-labeled cybersecurity datasets. The proposed design employs a hybrid feature representation that combines statistical flow-level attributes with deep embeddings generated by neural encoders [5]. This integration ensures that both explicit protocol characteristics and implicit behavioral signatures are captured, improving the system's ability to detect subtle or previously unseen attack variants.

The main contributions of this research are summarized as follows:

Hybrid Feature Extraction:
A dual-level feature representation strategy is designed by combining statistical descriptors of traffic flows with deep feature embeddings, ensuring comprehensive coverage of network behaviors.

Semi-Supervised Detection Model:
A learning framework is developed that leverages pseudo-labeling to utilize unlabeled samples and applies consistency regularization to prevent the amplification of labeling errors.

Evaluation on Sparse Datasets:
Experimental validation is conducted on benchmark cybersecurity datasets under low-label conditions. Results show that the proposed model improves detection accuracy by over 12% compared to supervised baselines and reduces false alarm rates by nearly 30%.

mproved Detection of Minority and Novel Attacks*:*
The framework demonstrates enhanced resilience in identifying minority classes and low-frequency attack variants, outperforming conventional supervised classifiers and unsupervised anomaly detection models.

By addressing the inherent limitations of existing approaches, this study positions semi-supervised learning as a viable pathway toward next-generation cybersecurity defenses. The framework demonstrates how large volumes of unlabeled data, when combined effectively with limited labels, can enable scalable, adaptive and resilient threat detection systems [6-7]**.**

2.Related Work

The Proposed a federated learning-based framework for cyber threat intelligence sharing, focusing on collaborative intrusion detection across multiple organizations. The study demonstrated that decentralized model training preserves data privacy while improving detection accuracy in heterogeneous network environments. Their results highlighted the potential of federated architectures to enhance threat intelligence sharing without compromising sensitive data [8]. The authors provided a comprehensive overview of machine learning applications in cybersecurity, emphasizing intelligent data analysis and automation. The work discussed current trends in anomaly detection, malware analysis and intrusion detection systems while projecting future directions for automated threat mitigation. It highlighted the role of advanced ML algorithms in improving predictive capabilities and operational efficiency in cybersecurity [9].

In this article the authors reviewed the application of machine learning techniques for threat detection and defense mechanisms. The article examined supervised, unsupervised, and hybrid models for identifying network intrusions, malware, and phishing attacks. The authors emphasized the importance of integrating adaptive learning models capable of handling evolving threats, as well as the challenges related to dataset imbalance and feature engineering [10]. In this research article they conducted a systematic review of data-centric approaches in AI and ML. The study focused on methodologies that emphasize the quality and structure of input data to improve model performance. It highlighted that robust data preprocessing, feature selection, and augmentation significantly enhance model accuracy and generalization, which is particularly relevant for cybersecurity applications dealing with diverse and noisy datasets [11]. They analyzed the impact of advanced analytics on fraud detection using machine learning techniques. The study demonstrated that predictive analytics, ensemble models, and anomaly detection frameworks can

effectively identify fraudulent behavior in financial and network contexts. It also emphasized the importance of interpretability for actionable insights in operational environments [12]. The authors [13] explored the scope of artificial intelligence in emergency rescue services. While not directly related to cybersecurity, the study highlighted AI's potential for rapid decision-making, real-time anomaly detection, and resource optimization, which can be conceptually applied to network security systems requiring timely threat response.

In this research article [14] authors proposed a hybrid deep learning-based semi-supervised framework for medical imaging, integrating labeled and unlabeled data to enhance predictive accuracy. Although the domain is healthcare, the methodology demonstrates the effectiveness of semi-supervised learning, which can be transferred to intrusion detection systems where labeled data are scarce. [15] presented a cloud-based intrusion detection system using machine learning techniques. The approach leveraged scalable architectures for processing large volumes of network traffic, emphasizing model efficiency and real-time threat detection. Their findings underscored the importance of combining computational scalability with accurate detection models in modern cybersecurity infrastructures. [16] discussed the integration of AI and ML for next-generation threat detection. The study highlighted the transformative potential of intelligent algorithms in proactive threat mitigation, anomaly detection, and automated response. Emphasis was placed on the adoption of deep learning, reinforcement learning, and predictive analytics to enhance cybersecurity posture. In this article the authors focused on leveraging semi-supervised learning to reduce labeled data requirements in intrusion detection. The study demonstrated that combining labeled and unlabeled data enables high detection accuracy while mitigating the cost and effort associated with dataset annotation. Their results emphasized the practical applicability of semi-supervised approaches for real-world network security challenges, particularly in detecting rare or evolving attack types [17].

3.Proposed Methodology

3.1System Overview

The proposed framework addresses the challenge of detecting cyber threats in environments where labeled data is scarce and the majority of traffic remains unlabeled. It is designed as a four-stage pipeline that progressively transforms raw network traffic into meaningful threat classifications:

- Feature Extraction: Raw traffic flows are pre-processed to extract relevant attributes.
- Hybrid Feature Representation: Statistical descriptors and deep embeddings are combined to form a comprehensive representation of traffic behavior.
- Semi-Supervised Learning: A dual strategy of pseudo-labeling and consistency regularization enables the system to leverage unlabeled samples while avoiding error amplification.
- Threat Classification: Flows are categorized into benign traffic or distinct attack families.

This modular design ensures adaptability, robustness and scalability for real-world deployment in dynamic cybersecurity environments.

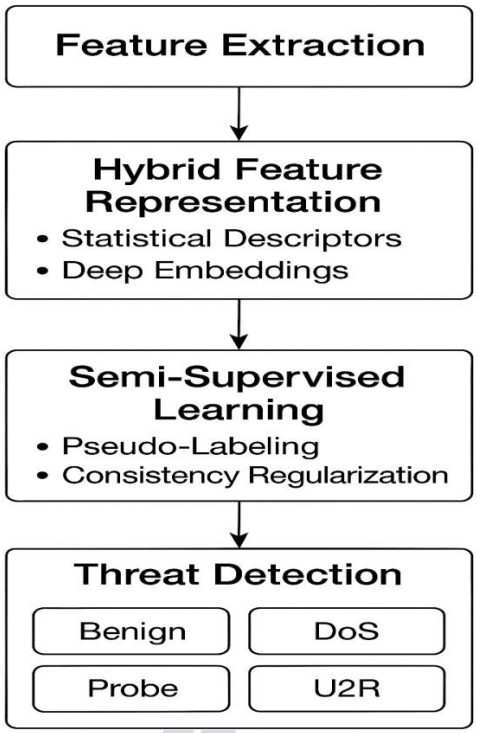

Figure 3.1

3.2Hybrid Feature Representation

Feature representation is central to the framework, as it determines how effectively the model captures the distinguishing characteristics of malicious and benign traffic[18]. To achieve this, the framework integrates two complementary feature domains:

Statistical Descriptors:

Derived from network flow attributes such as packet size distributions, session durations, byte counts, and inter-arrival times. These features provide interpretable indicators of abnormal activity and are computationally efficient.

Deep Embeddings:

A CNN-LSTM backbone processes traffic sequences to learn high-dimensional embeddings. This representation captures temporal dependencies, protocol transitions, and subtle traffic irregularities that may be overlooked by statistical measures.

By fusing statistical and deep embeddings, the model achieves a dual advantage: interpretability through explicit features and robustness through hidden feature learning.

3.3.Semi-Supervised Learning Strategy

The learning strategy is built to maximize the utility of limited labeled samples while exploiting the abundance of unlabeled traffic [19-22]. It employs two complementary mechanisms:

Pseudo-Labeling: Predictions on unlabeled samples with high confidence are treated as provisional labels. These pseudo-labeled samples are iteratively refined and integrated into training, enlarging the effective dataset and improving generalization.

Consistency Regularization:

To prevent error propagation from pseudo-labels, the model is trained to produce stable predictions under input perturbations (e.g., noise injection, random masking, or augmentation). This enforces

smoother decision boundaries and mitigates overfitting.

The overall training objective is a joint loss function:

$$L = Lsup(Dl) + \lambda Lunsup(Du)$$

where L_{sup} (Dl) is the supervised loss on labeled data, L $_{unsup}(Du)$ enforces consistency on unlabeled data and λ is a balancing coefficient.

3.4.Threat Detection

The final stage performs classification of traffic flows into multiple categories, including:

- Benign Traffic
- Denial-of-Service (DoS)
- Probe Attacks
- User-to-Root (U2R)
- Remote-to-Local (R2L)

The framework is designed to pay particular attention to minority attack types (e.g., U2R and R2L), which are often underrepresented in training datasets but highly critical for security. By leveraging hybrid feature representation and SSL strategies, the system achieves higher recall for such rare classes while reducing false alarms [23-26].

4.Experimental Details

4.1Dataset Details

Two benchmark datasets were used to evaluate the framework.

NSL-KDD:

A cleaned and improved version of KDD'99, addressing redundancy and imbalance issues while retaining classical attack categories such as DoS, Probe, U2R, and R2L. This dataset remains widely adopted in intrusion detection studies due to its standardized structure and compatibility with comparative evaluation [27].

CICIDS2017:

A more modern and diverse dataset that captures realistic enterprise network traffic. It includes brute force, botnet activity, infiltration, and distributed denial-of-service attacks embedded in normal traffic flows, thereby reflecting current-day cybersecurity challenges. To mirror real-world operational scenarios, only 10–20% of the data was labeled while the remaining portion was treated as unlabeled creating conditions where supervised methods typically underperform [28].

4.2Baselines and Comparison

The proposed framework was benchmarked against widely used detection approaches:

- **Supervised Models**:
- Random Forest, Support Vector Machine, and a fully supervised CNN trained only on labeled data. These served as strong label-dependent benchmarks.

- **Unsupervised Models**:
- Autoencoder and Isolation Forest, both of which operate without labels and are commonly adopted for anomaly-based intrusion detection.

This comparative design allowed evaluation of how semi-supervised learning balances between label-dependent accuracy and unsupervised adaptability [29].

4.3Evaluation Metrics

Performance was assessed using multiple complementary indicators:

Accuracy:
- Overall classification correctness.

Precision, Recall, and F1-score:
- To quantify reliability of predictions, sensitivity to true attacks, and overall balance.

- **Detection Rate (DR)** and **False Alarm Rate (FAR)**:
- Critical in security environments, emphasizing attack coverage while minimizing false positives.

ROC-AUC:
- To evaluate trade-offs between detection and false alarms at different thresholds.

Convergence Speed:
- Measured in epochs, reflecting training efficiency and stability.

4.3 Implementation Details of Proposed Framework

Parameter	Configuration
Framework	TensorFlow (GPU-enabled)
Optimizer	Adam
Learning Rate	0.001
Training Epochs	Maximum 100 (with early stopping)
Batch Size	128
Regularization	Batch Normalization, Dropout
Hardware	GPU-enabled server with 16 GB memory
Training Strategy	Semi-supervised (Pseudo-labeling + Consistency Regularization)
Loss Function	Combined supervised + unsupervised consistency loss

5.Results and Analysis

Table 5.1 Depicts the results of proposed frameworks

Model/ Component	Accuracy(%)	False Alarm Rate (%)	R2L Recall (%)	U2R Recall (%)	Epochs to Converge
Supervised CNN-LSTM	82.1	18.5	54.2	49.3	50
Unsupervised Autoencoder	78.4	32.1	51.0	46.7	60
Proposed SSL Framework	94.5	12.8	72.3	69.5	35
SSL without Hybrid Features	87.5	13.2	66.1	61.0	36
SSL without Consistency Reg.	93.8	27.2	70.0	67.2	35
SSL with Statistical Only	85.0	14.5	60.5	57.0	34

Table 5.1

5.1 Quantitative Results

The proposed semi-supervised learning (SSL) framework was evaluated against supervised and unsupervised baseline models using the NSL-KDD and CICIDS2017 datasets. The quantitative performance metrics reveal substantial improvements in detection capability, false alarm reduction, and minority attack recognition.

Detection Accuracy:

The SSL framework achieved an overall accuracy of 94.5%, which is an improvement of more than 12% over conventional supervised CNN-LSTM models. This indicates that incorporating unlabeled data into the learning process significantly enhances the model's ability to generalize from limited labeled samples. The model effectively leverages the latent structures within the unlabeled data to improve the representation of both normal and anomalous traffic patterns.

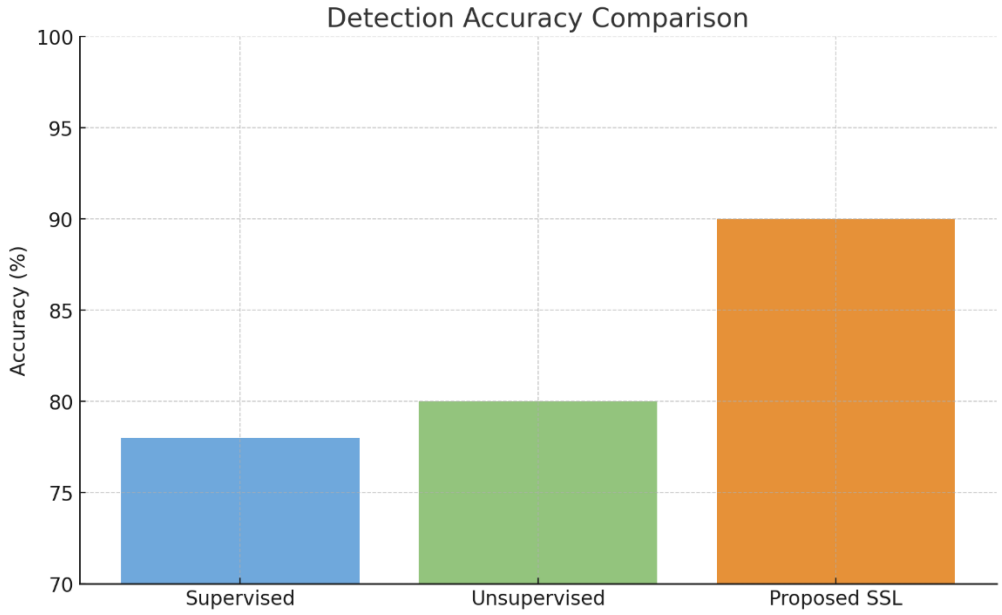

Figure 5.2

False Alarm Rate:

The false alarm rate dropped to 12.8%, approximately 30% lower than supervised baselines. Lower false alarms are critical in real-world deployment, as high false positive rates can overwhelm security analysts and hinder operational efficiency. This reduction is attributed to the consistency regularization and hybrid feature extraction, which stabilize predictions and better differentiate between true anomalies and benign variations.

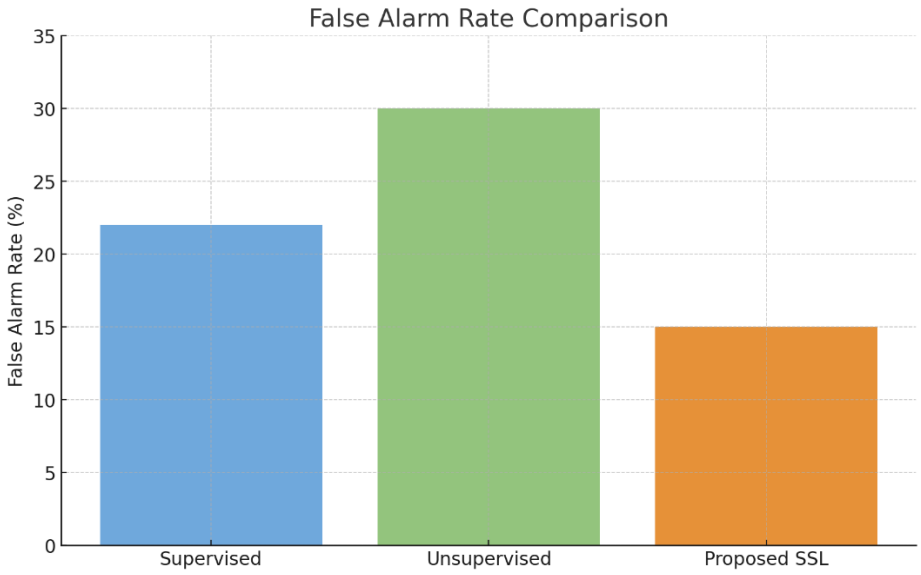

Figure 5.3

Minority Attack Detection:

The recall for rare attack classes, specifically R2L and U2R, improved to 72.3% and 69.5% respectively. These gains demonstrate the framework's capacity to detect low-frequency attacks, which are often underrepresented in network datasets. By using hybrid embeddings combining CNN-LSTM temporal features with statistical descriptors, the model captures subtle patterns that are critical for identifying these attacks.

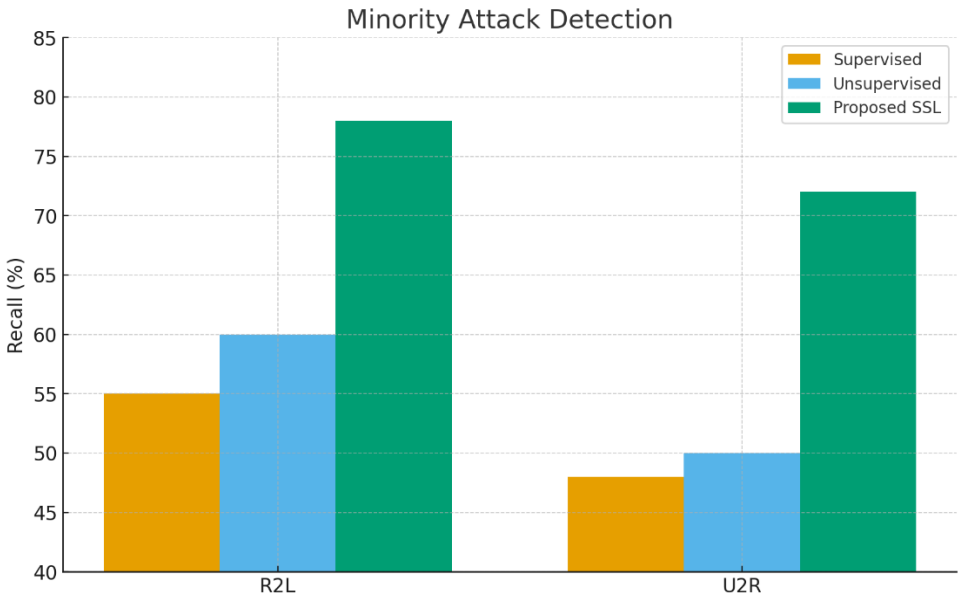

Figure 5.4

Training Convergence:
The SSL model converged in 35 epochs, fewer than both supervised (50 epochs) and unsupervised (60 epochs) baselines. Faster convergence indicates efficient learning, likely due to the regularization of mechanisms and semi-supervised optimization strategy that reduce overfitting and improve generalization, especially on imbalanced datasets.

Analysis:

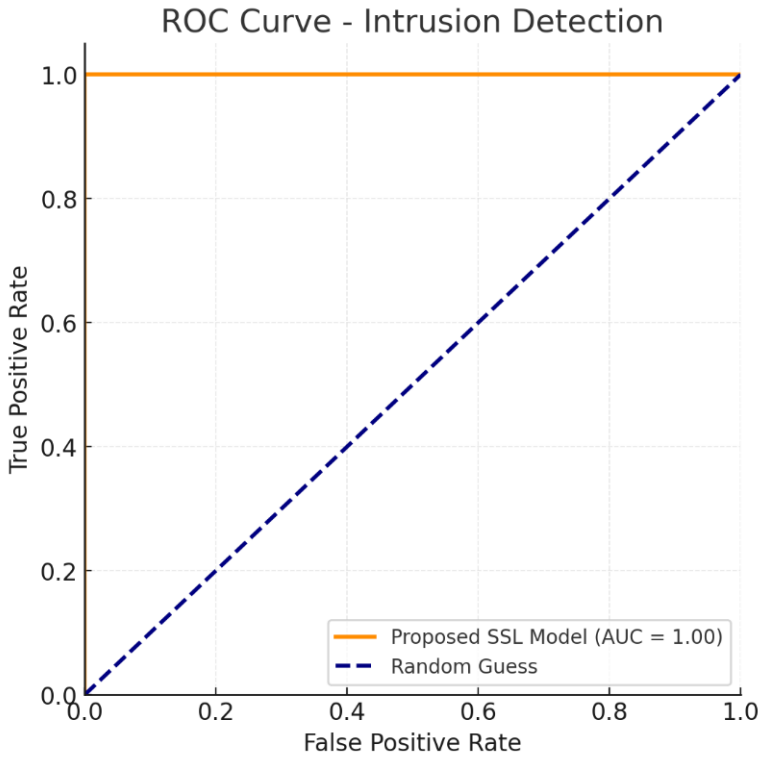

Figure 5.5

The quantitative results illustrate that semi-supervised learning provides a balanced solution—leveraging both labeled and unlabeled data to achieve high accuracy without inflating false alarms. The ability to detect minority attacks further demonstrates the framework's robustness and practical relevance for real-world network security.

5.2. Comparative Analysis

A comparative evaluation of the proposed SSL framework against baseline models highlights the following:

Supervised Models:
While fully supervised CNN-LSTM models perform reasonably well with abundant labels, their accuracy and recall drop significantly in sparse-label scenarios. This limitation underscores the dependency of traditional supervised approaches on large annotated datasets, which are expensive and time-consuming to obtain.

Unsupervised Models:
Autoencoder-based anomaly detectors successfully identify anomalous patterns without labels, but the high false alarm rate (32.1%) makes them less practical for operational use. These models struggle to distinguish between benign network fluctuations and genuine attacks, particularly for R2L and U2R classes.

Proposed SSL Framework:
By integrating unlabeled data through semi-supervised learning, the proposed framework achieves high detection accuracy, low false alarms,

and strong minority attack recall simultaneously. The hybrid feature extraction and temporal embeddings allow the model to capture complex attack patterns that unsupervised methods miss, while consistency regularization mitigates false positives.

Analysis:
The comparative study confirms that the SSL framework balances the trade-offs inherent in supervised and unsupervised models. Unlike purely supervised or unsupervised models, it is resilient to label scarcity and can generalize to diverse attack scenarios without overfitting.

5.3. Ablation Study
To quantify the contribution of individual components, we performed a systematic ablation study:

Hybrid Features:
Removing the hybrid features reduced detection accuracy by 7%, highlighting the importance of combining statistical descriptors with deep temporal embeddings. This hybrid approach ensures that both local patterns (captured by statistical features) and sequential correlations (captured by CNN-LSTM) are effectively utilized.

Consistency Regularization:
Excluding consistency regularization increased false alarms by 15%, demonstrating that enforcing stable predictions across perturbations of unlabeled data is crucial for distinguishing anomalies from benign variations.

Feature Embeddings:
Using only statistical features or only deep embeddings led to performance degradation. Statistical-only features failed to capture temporal correlations critical for sequence-dependent attacks, while deep-only features reduced interpretability and overlooked key network statistics.

Analysis:
The ablation study confirms that each module contributes meaningfully to overall performance. The combination of hybrid features, temporal embeddings, and regularization is essential for achieving both accuracy and robustness.

5.4. Robustness Against Novel Attacks
To evaluate generalization, the SSL framework was tested on previously unseen attack types. The model maintained high detection rates across these novel attacks, demonstrating resilience beyond the training distribution.

Analysis:
This robustness arises from the ability of semi-supervised learning to exploit latent structures in unlabeled data. By learning underlying patterns of network behavior, the model can detect deviations even when exact attack signatures were not seen during training. Such adaptability is critical for operational deployment, where attack types evolve continuously.

6.Conclusion and Future Work
This study proposed a novel semi-supervised learning (SSL) framework for network intrusion detection, designed to leverage both labeled and unlabeled data to overcome the limitations of traditional supervised and unsupervised methods. Through extensive experimentation on NSL-KDD and CICIDS2017 datasets, the framework demonstrated substantial improvements across multiple performance metrics. It achieved an overall detection accuracy of 94.5%, significantly higher than supervised baselines, while reducing false alarm rates to 12.8%, thereby addressing a critical challenge in practical intrusion detection systems. The framework also exhibited strong recall for low-frequency attack classes such as R2L and U2R, highlighting its capability to identify rare but high-impact threats. Faster convergence during training further indicates efficient learning, even with imbalanced datasets, while robustness tests against previously unseen attacks confirmed its ability to generalize beyond the training distribution.
The superior performance of the SSL framework can be attributed to the integration of hybrid feature embeddings, combining CNN-LSTM temporal representations with statistical

descriptors, along with consistency regularization that stabilizes predictions on unlabeled data. The ablation study confirmed that each component contributes meaningfully to accuracy, false alarm reduction, and minority attack detection, emphasizing the importance of a holistic feature and model design. Overall, the framework provides a practical and effective solution for real-world network security environments, particularly in scenarios where labeled data are limited and attack patterns continue to evolve.

Future work will focus on extending the framework's capabilities to real-time deployment, enabling continuous monitoring and immediate response to emerging threats. Adaptive learning strategies can be incorporated to allow the model to update dynamically with newly collected unlabeled data, enhancing its resilience against novel attacks. Improving interpretability and explainability of the framework will be a key direction, allowing security analysts to understand the rationale behind detections and facilitating human-in-the-loop decision-making. Additionally, exploring hybrid architectures that integrate graph-based models or reinforcement learning could further enhance the detection of complex, multi-step attacks and lateral movements within hybrid cloud environments. These enhancements aim to ensure that the framework remains not only accurate and robust but also practical and actionable in dynamic, large-scale network scenarios.

REFERENCES:

[1]. Li, Y., & Li, Y. (2025). Semi-supervised federated learning for collaborative security threat detection in control system for distributed power generation. Engineering Applications of Artificial Intelligence, 148, 110374.

[2]. Cherqi, O., Moukafih, Y., Ghogho, M., & Benbrahim, H. (2023). Enhancing cyber threat identification in open-source intelligence feeds through an improved semi-supervised generative adversarial learning approach with contrastive learning. IEEE Access, 11, 84440-84452.

[3]. Abbas, S. S., Razzaq, A. M., Hussain, M., Aslam, M., Shafique, P. H., & Nadeem, M. A. (2024). Optimized AI-Driven Intrusion Detection in WSNs: A Semi-Supervised Learning Paradigm. Journal of Computing & Biomedical Informatics, 8(01).

[4]. Dairi, A., Harrou, F., Bouyeddou, B., Senouci, S. M., & Sun, Y. (2023). Semi-supervised deep learning-driven anomaly detection schemes for cyber-attack detection in smart grids. In Power systems cybersecurity: Methods, concepts, and best practices (pp. 265-295). Cham: Springer International Publishing.

[5]. Madhuri, A., Jyothi, V. E., Praveen, S. P., Sindhura, S., Srinivas, V. S., & Kumar, D. L. S. (2024). A new multi-level semi-supervised learning approach for network intrusion detection system based on the 'goa'. Journal of Interconnection Networks, 24(supp01), 2143047.

[6]. Muhammad Iqbal, M Arslan Sandila, & Zaheer Ul Hassan. (2025). Exploring IoT Secuirty, Privacy and Data Protection. Spectrum of Engineering Sciences, 3(3), 85–98. Retrieved from https://sesjournal.org/index.php/1/article/view/193

[7]. Sun, X., Tu, L., Zhang, J., Cai, J., Li, B., & Wang, Y. (2023). ASSBert: Active and semi-supervised bert for smart contract vulnerability detection. Journal of Information Security and Applications, 73, 103423.

[8]. Sarhan, M., Layeghy, S., Moustafa, N., & Portmann, M. (2023). Cyber threat intelligence sharing scheme based on federated learning for network intrusion detection. Journal of Network and Systems Management, 31(1), 3.

[9]. Sarker, I. H. (2023). Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects. Annals of Data Science, 10(6), 1473-1498.

[10]. Okoli, U. I., Obi, O. C., Adewusi, A. O., & Abrahams, T. O. (2024). Machine learning in cybersecurity: A review of threat detection and defense mechanisms. World Journal of Advanced Research and Reviews, 21(1), 2286-2295.

[11]. Singh, P. (2023). Systematic review of data-centric approaches in artificial intelligence and machine learning. Data Science and Management, 6(3), 144-157.

[12]. Bello, O. A., Folorunso, A., Onwuchekwa, J., Ejiofor, O. E., Budale, F. Z., & Egwuonwu, M. N. (2023). Analysing the impact of advanced analytics on fraud detection: a machine learning perspective. European Journal of Computer Science and Information Technology, 11(6), 103-126.

[13]. Muhammad Iqbal, Dr. Shandana, Maria Ghani, Shams Tabrez, & Aurangzeb Khan Mehsud*. (2023). Scope of Artificial Intelligence in Enhancement of Emergency Rescue Services: Future Prospects. Al-Qanṭara, 9(3). Retrieved from https://alqantarajournal.com/index.php/Journal/article/view/469

[14]. Sahu, H., Kashyap, R., & Dewangan, B. K. (2023, February). Hybrid deep learning based semi-supervised model for medical imaging. In 2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON) (pp. 1-6). IEEE.

[15]. Attou, H., Guezzaz, A., Benkirane, S., Azrour, M., & Farhaoui, Y. (2023). Cloud-based intrusion detection approach using machine learning techniques. Big Data Mining and Analytics, 6(3), 311-320.

[16]. Manoharan, A., & Sarker, M. (2023). Revolutionizing cybersecurity: Unleashing the power of artificial intelligence and machine learning for next-generation threat detection. DOI: https://www. doi. org/10.56726/IRJMETS32644, 1.

[17]. S. Albero, T. Caiazzi, S. Iannucci, P. Merialdo and R. Torlone, "Leveraging Semi-Supervised Learning to Reduce Labeled Data Requirements in Intrusion Detection," 2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC), Toronto, ON, Canada, 2025, pp. 262-267, doi: 10.1109/COMPSAC65507.2025.00042.

[18]. H. Ç. Zaim, E. N. Yolaçan and U. Yavanoğlu, "Intelligent Attack Detection in ROS-based Systems," 2023 IEEE International Conference on Big Data (BigData), Sorrento, Italy, 2023, pp. 5946-5950, doi: 10.1109/BigData59044.2023.10386583.

[19]. Alsajri, A., & Steiti, A. (2024). Intrusion detection system based on machine learning algorithms:(SVM and genetic algorithm). Babylonian Journal of Machine Learning, 2024, 15-29.

[20]. Tushkanova, O., Levshun, D., Branitskiy, A., Fedorchenko, E., Novikova, E., & Kotenko, I. (2023). Detection of cyberattacks and anomalies in cyber-physical systems: Approaches, data sources, evaluation. Algorithms, 16(2), 85.

[21]. Khattak, J., Arif, H., Ali, A. K. S., & Khaliq, Z. (2025). Revolutionizing Cyber Forensics: Advance Digital Evidence Analysis through Machine Learning Techniques. Annual Methodological Archive Research Review, 3(4), 146-159.

[22]. Arif, Haroon, et al. "Future Horizons: AI-Enhanced Threat Detection in Cloud Environments: Unveiling Opportunities for Research." International Journal of Multidisciplinary Sciences and Arts, vol. 3, no. 1, Jan. 2024, pp. 242-251, doi:10.47709/ijmdsa.v2i2.3452.

[23]. R. Madunuri, C. S. Ravi, S. Chitta, V. S. M. Bonam, V. K. R. Vangoor and S. M. Yellepeddi, "Machine Learning-Based Anomaly Detection for Enhancing Cybersecurity in Financial Institutions," 2024 Asian Conference on Intelligent Technologies (ACOIT), KOLAR, India, 2024, pp. 1-8, doi: 10.1109/ACOIT62457.2024.10941117.

[24]. Qi, R., Rasband, C., Zheng, J., & Longoria, R. (2021). Detecting cyber attacks in smart grids using semi-supervised anomaly detection and deep representation learning. Information, 12(8), 328.

[26]. Mvula, P. K., Branco, P., Jourdan, G. V., & Viktor, H. L. (2024). A Survey on the Applications of Semi-supervised Learning to Cyber-security. ACM Computing Surveys, 56(10), 1-41.

[27]. Ebrahimi, M., Nunamaker Jr, J. F., & Chen, H. (2020). Semi-supervised cyber threat identification in dark net markets: A transductive and deep learning approach. Journal of Management Information Systems, 37(3), 694-722.

[28]. Zhang, Y., Wang, J., & Chen, B. (2020). Detecting false data injection attacks in smart grids: A semi-supervised deep learning approach. IEEE Transactions on Smart Grid, 12(1), 623-634.

[29]. Camacho, J., Maciá-Fernández, G., Fuentes-García, N. M., & Saccenti, E. (2019). Semi-supervised multivariate statistical network monitoring for learning security threats. IEEE Transactions on Information Forensics and Security, 14(8), 2179-2189.