Self-Learning Cyber Defense: Adaptive AI Framework for Zero-Day Threat Prediction using Automated Data Pipelines

Abhay Pratap Singh Rana[∗], Dev Chauhan[†], Aditya Tiwari[‡], Anshuman Kumar[§], Aazim Iqbal[¶], Amit Tiwari[‖], Faiz Ali[∗∗], Belal Akhtar[††]
[∗†‡§¶‖∗∗]*Department of Computer Science and Engineering*
[∗†‡§¶‖∗∗]*Noida International University, Greater Noida, India*
Email: [∗]abhaypratap1765@gmail.com

***Abstract*—The rapid evolution of cyber threats has led to the emergence of zero-day attacks that exploit previously unknown vulnerabilities, rendering conventional static defense mechanisms inadequate. To address this challenge, this research presents a self-learning cyber defense framework that integrates adaptive artificial intelligence with automated data science pipelines for real-time zero-day threat prediction. The proposed system continuously monitors network behavior, extracts dynamic features, and employs an adaptive learning engine capable of updating its detection models without manual intervention. A fully automated data pipeline handles data ingestion, preprocessing, feature optimization, and model retraining, ensuring continuous adaptability to evolving threat landscapes. Experimental evaluations conducted on benchmark datasets such as CICIDS2017 and UNSW-NB15 demonstrate significant improvements in detection accuracy and response latency compared to traditional intrusion detection systems. The results highlight that the proposed adaptive AI framework not only enhances predictive capability but also reduces false alarms through self-optimization and contextual learning. This study contributes a novel and scalable approach for cyber defense systems, capable of autonomously evolving in the face of unknown attack vectors, thereby strengthening organizational resilience against emerging zero-day exploits.**

***Keywords*—Adaptive AI, Self-Learning Systems, Cyber Defense, Zero-Day Threats, Automated Data Pipelines, Anomaly Detection, MLOps**

I. Introduction

In recent years, cybersecurity has faced an exponential rise in sophisticated attacks that exploit emerging technologies and unknown vulnerabilities. Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized traditional defense mechanisms by enabling predictive and adaptive security models capable of learning from network behavior and threat patterns [2], [9], [12], [15], [22]. These intelligent systems offer capabilities to detect anomalies, recognize malicious intent, and respond dynamically, representing a paradigm shift from signature-based detection to behavior-driven analysis [28], [29]. However, despite these advancements, zero-day attacks continue to challenge modern defense systems due to their ability to exploit vulnerabilities that remain undisclosed or unpatched [6], [27].

Zero-day attacks are characterized by stealth and unpredictability, allowing adversaries to penetrate systems before any signature or rule-based defense can respond [7]. Traditional intrusion detection systems (IDS) rely on static datasets and pre-defined patterns, which are ineffective against previously unseen exploit behavior [19], [20], [24]–[26]. Machine learning-based systems, while more flexible, often suffer from model drift, limited retraining capability, and an inability to process dynamic data in real time [23], [37]. These limitations underline the urgent need for a more adaptive and continuously evolving approach to cyber defense.

The critical research gap lies in the lack of AI-driven frameworks that can autonomously adapt to novel attack vectors while maintaining operational continuity. Most existing ML-based systems require manual retraining or batch updates, which introduces latency and reduces responsiveness to new threats [13], [14], [49]–[51]. Moreover, current models often overlook the role of automated data science pipelines in continuously collecting, cleaning, and feeding live data streams for adaptive learning [17], [42], [52]–[54]. This results in reactive systems that fail to evolve with changing threat landscapes.

To address these challenges, this study proposes a self-learning cyber defense framework that integrates adaptive AI with automated data pipelines for real-time zero-day threat prediction. The proposed architecture employs a continuous data ingestion pipeline for dynamic feature extraction and a self-learning AI module that updates its internal model incrementally based on streaming data [18], [21], [30], [31], [35], [39]. This combination enables automated retraining without human intervention, ensuring resilience against data drift and emerging threats.

The primary research contributions of this work are as follows:

- Development of an adaptive AI framework that continuously evolves by learning from live network telemetry for zero-day threat detection.
- Integration of an automated data science pipeline that performs data ingestion, preprocessing, and feature optimization in real time.
- Implementation of a self-learning mechanism capable of online retraining to maintain model relevance against evolving threats.
- Comprehensive evaluation using benchmark datasets, including CICIDS2017 and UNSW-NB15, to demonstrate improved detection accuracy and reduced false-positive rates compared to static intrusion detection systems.

The remainder of this paper is organized as follows. Section II reviews related work in AI-based zero-day attack detection and adaptive cybersecurity frameworks. Section III presents the proposed methodology, including system architecture and model design. Section IV describes the experimental setup and datasets used for evaluation. Section V discusses the results and comparative analysis. Finally, Section VI concludes the paper and highlights future directions for extending adaptive learning in cyber defense systems.

II. Related Work

The literature on intrusion detection and zero-day defense spans traditional signature and rule-based systems, statistical anomaly detectors, and modern AI/ML approaches. Early critiques emphasized the limitations of purely signature-based systems: they are effective for known threats but inherently unable to detect previously unseen exploits [22], [23]. Surveys and empirical analyses have established that anomaly-based approaches can detect novel behavior but suffer from high false positive rates and sensitivity to concept drift in network traffic [23], [26]. Bilge and Dumitras provided an empirical perspective on real-world zero-day use and the challenges defenders face in timely detection and attribution [27].

Machine learning has been widely applied to improve detection of both known and unknown attacks. Classical supervised classifiers such as SVM, Random Forest, and ensemble methods achieve strong performance on labeled datasets but require representative training samples and thus struggle with zero-day scenarios when attack classes are absent from training data [23], [44]. Deep learning methods—autoencoders, recurrent networks (LSTM), and convolutional models—have been proposed for automated feature learning and anomaly scoring; Kitsune is a notable online ensemble of autoencoders tailored for network traffic anomaly detection [28], [29], [55], [56]. Several works demonstrate that deep autoencoders and LSTM-based architectures can capture temporal and structural patterns, improving detection rates for complex attacks [28], [32], [57].

Zero-day detection research specifically has focused on two broad directions: (1) building robust anomaly detectors that generalize to unseen behavior and (2) devising few-/zero-shot and transfer learning techniques to infer novel attack semantics from related classes [33], [34]. Zero-shot and attribute-based methods attempt to model semantic relationships so that a detector can generalize to unseen classes; however, they often depend on carefully designed attribute spaces and still face challenges when attack features are highly obfuscated [33], [34].

A growing subfield emphasizes adaptive and online learning methods to address model drift and the arrival of novel threats. Incremental learning, streaming anomaly detection, and continual learning methods update models using new data without full retraining; such approaches reduce downtime and can adapt to evolving traffic distributions [36], [43]. Reinforcement learning (RL) has also been explored to enable policy-driven defenses and adaptive response strategies, with several works applying deep RL to intrusion detection and automated mitigation planning [38], [40]. Despite promising results, online and RL approaches must balance adaptation speed and stability, and they expose additional attack surfaces (e.g., poisoning during online updates) [41].

Operationalizing detection systems has motivated research into automated data pipelines and MLOps for security: stream processing, continuous feature extraction, automated model validation, and safe deployment practices. Works on production-grade ML systems emphasize data quality, lineage, retraining orchestration, and secure model serving—elements essential for resilient, real-time security analytics [17], [42], [45]. Recent surveys on Secure MLOps call attention to attacks specific to the ML lifecycle and recommend pipeline automation with built-in validation to mitigate drift and supply-chain risks [46], [47].

Table I summarizes representative prior approaches, the datasets commonly used for evaluation, their principal limitations, and how the present work differs. The table highlights that while many prior studies advance detection models, few integrate an end-to-end automated pipeline with online self-learning and explicit model drift management in a single framework.

To aid conceptual comparison, Figure 1 depicts the high-level taxonomy of approaches and the missing integration point that this work addresses: a tightly coupled automated pipeline with a self-learning detection engine capable of online adaptation while preserving safety checks and validation gates.

In summary, prior research provides strong foundations in detection algorithms, anomaly modeling, and operational tooling. However, there remains a need for an integrated system that (1) performs continuous feature engineering and data validation in production, (2) applies self-learning models that adapt safely to drift and novel attacks, and (3) enforces validation and rollback mechanisms as part of MLOps. The proposed Self-Learning Cyber Defense framework builds on the surveyed literature and addresses these gaps by combining automated data pipelines, online model updates, drift detection, and safety checks in a unified architecture.

III. Proposed Methodology

This section describes the design and functioning of the proposed *Self-Learning Cyber Defense Framework (SLCDF)* for zero-day threat prediction. The system integrates automated data pipelines with an adaptive AI engine capable of incremental learning, continuous model updates, and real-time detection of unseen attack patterns. The methodological design ensures full automation, scalability, and resilience across the machine learning lifecycle.

A. Overview of Framework

The SLCDF framework consists of five major components: (1) Data Ingestion Layer, (2) Preprocessing and Feature Engineering Unit, (3) Adaptive AI Engine, (4) Zero-Day Prediction Module, and (5) Feedback and Retraining Loop. Figure 2 illustrates the overall system architecture. Incoming telemetry or network logs are continuously streamed into the ingestion layer, processed through feature transformation modules, and subsequently evaluated by the adaptive AI engine. Predictions

TABLE I: Summary of representative related work, datasets, limitations, and the proposed innovation

Approach	Representative works	Common datasets	Key limitations	How this work differs
Signature / Rule based	[22]	Proprietary IDS logs	Cannot detect novel/zero-day	N/A (baseline)
Statistical	[23], [26]	KDD99, NSL-KDD	High FPR; concept drift	Uses continuous pipeline + adaptive thresholds
Classical ML (SVM, RF)	[23], [37]	UNSW-NB15, CICIDS2017	Needs labeled attacks; poor zero-day generalization	Semi-supervised online updates
Deep Learning (AE, LSTM)	[28], [29], [32]	CICIDS2017, CICMalMem	Training cost; susceptibility to drift	Incremental AE + online retraining
Zero-/Few-shot	[33], [34]	Custom splits of benchmarks	Attribute design; limited real-time eval	Combined with streaming pipeline for live evaluation
Adaptive / Online / RL	[36], [38], [40]	Streaming testbeds	Stability vs. plasticity trade-off	Drift control + safe retraining policies
MLOps / Automated pipelines	[17], [42], [48]	Production telemetry	Focus on ops; limited IDS focus	Security-centric pipeline + model validation

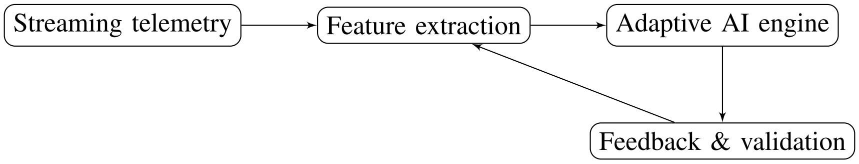

Fig. 1: Schematic of streaming pipeline with closed-loop adaptation (conceptual).

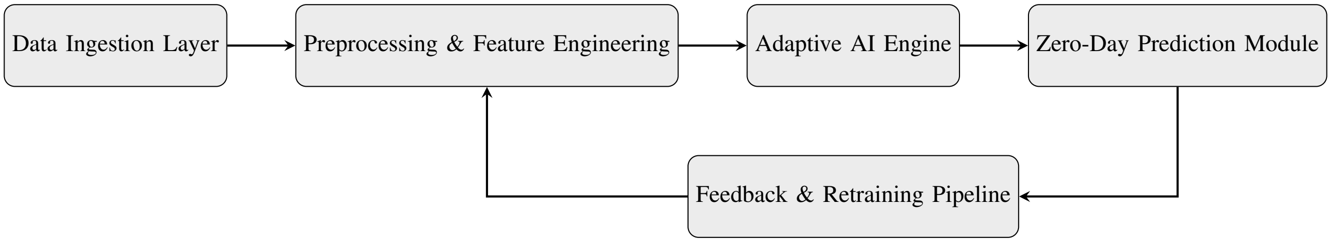
Data Ingestion Layer
Preprocessing & Feature Engineering
Adaptive AI Engine
Zero-Day Prediction Module
Feedback & Retraining Pipeline

Fig. 2: Block diagram of the proposed Self-Learning Cyber Defense Framework (SLCDF).

and anomaly scores are passed to the Zero-Day Prediction Module, which triggers alerts or retraining events based on detection confidence.

B. Data Pipeline Design

The proposed automated data pipeline manages end-to-end data flow — from ingestion to model deployment — ensuring continuous, reliable, and validated data processing. Data ingestion is handled by distributed streaming systems such as Apache Kafka, which capture raw network packets, system events, and IDS logs in near real-time. Apache Airflow or TensorFlow Extended (TFX) orchestrates the workflow, ensuring each stage is executed with data lineage tracking and versioning.

The pipeline stages are defined as follows:

- Data Ingestion: Real-time streaming from sensors, firewall logs, or NetFlow data using Kafka topics.
- Preprocessing: Removal of redundant or incomplete entries; standardization of categorical and numerical fields.
- Feature Engineering: Extraction of statistical, temporal, and protocol-specific features; transformation using PCA or autoencoder bottlenecks.
- Model Retraining: Periodic or event-triggered retraining using newly labeled samples or unsupervised embeddings.
- Deployment: Containerized model deployment through CI/CD pipelines for inference in production.

TABLE II: Automated data pipeline workflow

Stage	Tool/Technology	Function	Output
Ingestion	Kafka	Stream network data	Raw telemetry
Preprocessing	Python/TFX	Clean, normalize data	Feature vectors
Feature Engg.	Scikit-learn, AE	Extract patterns	Reduced features
Model Train	TensorFlow, PyTorch	Update model weights	Updated model
Deployment	Docker, Airflow	Serve prediction API	Active model

C. Adaptive AI Engine

The Adaptive AI Engine forms the analytical core of SLCDF. It employs a hybrid architecture that combines incremental deep learning and reinforcement-based feedback mechanisms. The primary detection model is a Long Short-Term Memory Autoencoder (LSTM-AE), designed to capture temporal dependencies in streaming network data. The autoencoder minimizes reconstruction error $L(x,\hat{x})$:

$$L(x,\hat{x}) = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \hat{x}_i\|_2^2, \tag{1}$$

where x_i is the input feature vector and $\hat{x}_i$ is its reconstructed output. Higher reconstruction errors indicate potential anomalies or zero-day behavior.

The incremental learning mechanism updates weights as new labeled or pseudo-labeled data arrives:

$$W_{t+1} = W_t - \eta \nabla L(x_t, \hat{x}_t), \tag{2}$$

where η is the adaptive learning rate controlled through feedback consistency metrics. Reinforcement signals are derived from validation performance, guiding learning rate adjustments and dropout scheduling.

The adaptive model employs dynamic thresholding to minimize false positives:

$$\theta_t = \mu_{L_t} + \alpha \sigma_{L_t}, \tag{3}$$

where μ_{L_t} and σ_{L_t} are the mean and standard deviation of recent reconstruction losses, and α adjusts sensitivity.

D. Zero-Day Prediction Module

The Zero-Day Prediction Module acts as the decision layer for anomaly categorization and confidence estimation. It integrates unsupervised clustering (using DBSCAN or GMM) with anomaly scoring to identify unseen attack signatures. The hybrid classification logic operates as follows:

1) Compute anomaly score $s(x)$ based on model reconstruction error.
2) Cluster unknown patterns to detect consistent novel behavior.
3) Generate confidence score $C(x)$ using softmax normalization:

$$C(x) = \frac{e^{-s(x)}}{\sum_{i=1}^{N} e^{-s(x_i)}} \tag{4}$$

4) Trigger alert if $C(x) < \delta$, where δ is the adaptive threshold based on recent detection stability.

The module adapts dynamically as new data arrives, continuously refining its boundary between known and unknown threats.

E. Model Retraining and Feedback Loop

The retraining and feedback module closes the learning cycle by continuously integrating new observations, confirmed alerts, and user feedback. This mechanism ensures that the AI model remains robust to drift and evolving adversarial strategies. The process follows these key steps:

- Collect confirmed alerts or misclassified samples.
- Validate through expert labeling or consensus mechanism.
- Update model incrementally without full retraining.
- Log performance metrics and trigger version updates through MLOps automation.

Continuous integration and deployment (CI/CD) tools such as MLflow and Kubeflow Pipelines are used to manage model experiments, rollback procedures, and metric dashboards. Figure 4 shows the feedback loop that governs the automated retraining and adaptation of the SLCDF model.

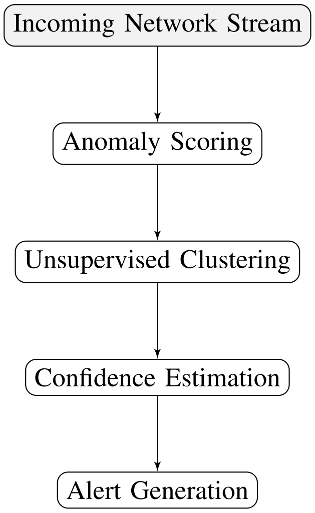

Fig. 3: Flowchart of Zero-Day Prediction Module.

The feedback loop transforms the model into a continuously learning entity that adapts to changing threat landscapes. It not only enhances detection accuracy but also provides operational resilience against concept drift and adversarial evasion.

In summary, the proposed methodology establishes a tightly integrated pipeline connecting adaptive learning mechanisms with automated orchestration. Through real-time data processing, incremental AI adaptation, and continuous validation, the framework achieves resilient and proactive defense against evolving zero-day attacks.

F. Algorithmic Workflow: Adaptive Update Cycle for SLCDF

To operationalize the adaptive behavior of the proposed Self-Learning Cyber Defense Framework (SLCDF), the training and inference processes are orchestrated through an automated feedback-driven learning cycle. The process ensures that the model continuously adapts to new attack behaviors without complete retraining, minimizing downtime and computational overhead. Algorithm 1 outlines the core workflow.

The algorithm describes the cyclic learning process. Each incoming data batch X_t is evaluated for reconstruction error. Samples exceeding the adaptive threshold θ_t are classified as potential anomalies and stored in buffer $\mathscr{B}$. When the mean drift D_t surpasses the drift threshold δ_{drift}, incremental retraining is initiated. This process ensures the model remains synchronized with evolving data distributions, thus maintaining high sensitivity to unseen zero-day behaviors while preventing overfitting.

The combination of Algorithm 1 and the workflow in Figure 5 demonstrates how the framework implements self-sustaining intelligence. This approach minimizes manual intervention, accelerates model evolution, and strengthens defense against dynamically emerging zero-day threats through fully automated learning and deployment loops.

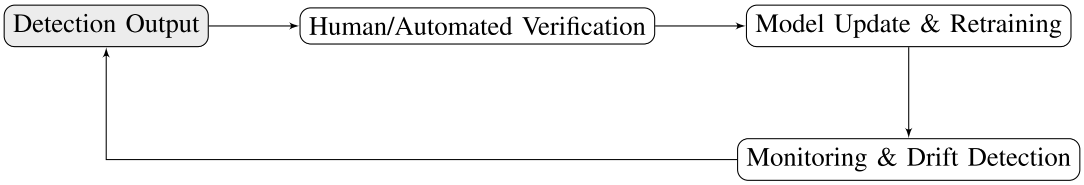
Detection Output
Human/Automated Verification
Model Update & Retraining
Monitoring & Drift Detection

Fig. 4: Feedback loop and continuous retraining mechanism for adaptive model updates.

Algorithm 1 Adaptive Update Cycle for Self-Learning Cyber Defense Framework (SLCDF)

```
Require: Streaming network data $X_t = \{x_1, x_2, \dots, x_n\}$, current model parameters $W_t$, adaptive threshold $\theta_t$
Ensure: Updated model $W_{t+1}$, refined threshold $\theta_{t+1}$
Initialize LSTM-AE model with weights $W_t$ and threshold $\theta_t$
Receive real-time network samples $X_t$ from Kafka stream
for each sample $x_i \in X_t$ do
    Compute reconstructed output $\hat{x}_i = f_{AE}(x_i; W_t)$
    Calculate anomaly score $s(x_i) = \|x_i - \hat{x}_i\|_2^2$
    if $s(x_i) > \theta_t$ then
        Mark $x_i$ as potential zero-day anomaly
        Send $x_i$ to Zero-Day Prediction Module for confidence scoring
        Store $\langle x_i, s(x_i) \rangle$ in feedback buffer $\mathscr{B}$
    end if
end for
Compute drift indicator:
    $$D_t = \frac{1}{|\mathscr{B}|} \sum_{x_i \in \mathscr{B}} (s(x_i) - \mu_{L_t}) \quad (5)$$
if $|D_t| > \delta_{drift}$ then
    Trigger online retraining procedure:
        $$W_{t+1} = W_t - \eta \nabla L(x_i, \hat{x}_i) \quad (6)$$
    Update adaptive threshold:
        $$\theta_{t+1} = \mu_{L_t} + \alpha \sigma_{L_t} \quad (7)$$
    Validate new model using drift-aware validation set
    Deploy $W_{t+1}$ through CI/CD pipeline
end if
    return Updated parameters $(W_{t+1}, \theta_{t+1})$
```

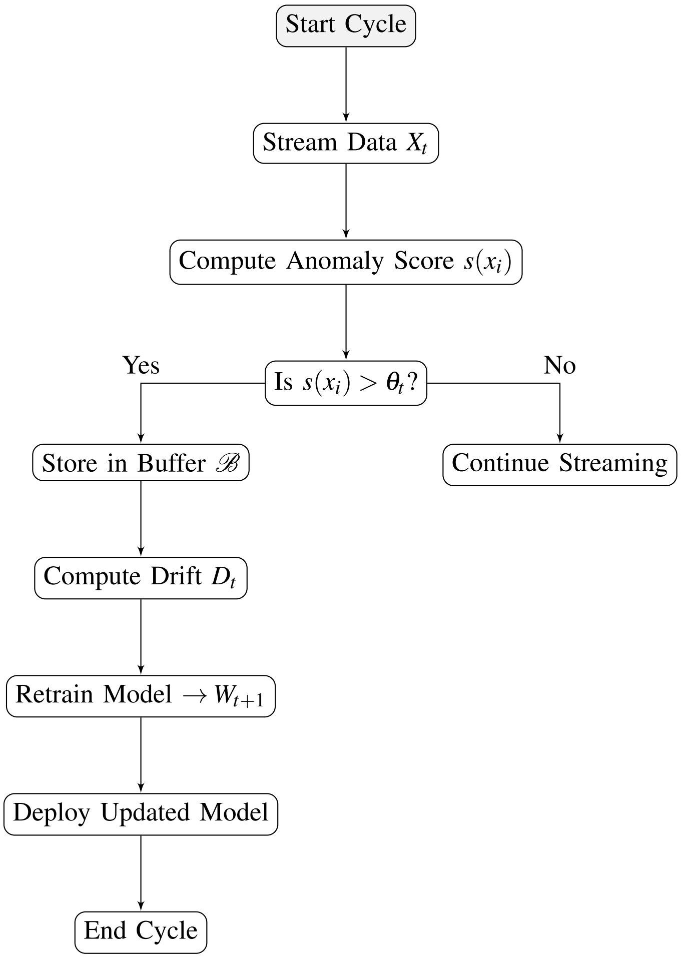

Fig. 5: Workflow of the adaptive update cycle showing continuous learning and deployment.

IV. Experimental Setup and Results

A. Experimental Configuration

The experimental analysis of the Self-Learning Cyber Defense Framework (SLCDF) was conducted in a controlled environment designed to emulate large-scale enterprise network conditions. The experimental setup utilized a 16-core Intel Xeon processor with 64 GB RAM and dual NVIDIA RTX A6000 GPUs for model training and real-time inference acceleration. The pipeline was orchestrated using *Apache Airflow* for automated task scheduling and *Kafka* for continuous streaming of network events. All model components, including feature engineering, retraining triggers, and adaptive threshold tuning, were implemented in *TensorFlow Extended (TFX)* and deployed through a *Docker*-based microservice architecture to ensure reproducibility and scalability.

B. Dataset Description

Two benchmark datasets were selected to validate the system's capacity to generalize across known and unseen attack patterns: the CICIDS2017 dataset and the UNSW-NB15 dataset. The CICIDS2017 dataset comprises over 2.8 million labeled network flows covering common attack categories such as DDoS, PortScan, and Brute Force. In contrast, UNSW-NB15 contains 2.5 million records emphasizing modern attack vectors such as Fuzzers, Worms, and Generic exploits. For zero-day simulation, a subset of attack classes was intentionally withheld during training to assess the framework's predictive capacity under unseen threat conditions.

The datasets were normalized using Min–Max scaling, and categorical attributes were encoded via one-hot encoding. Approximately 70% of the data was used for model training, 15% for validation, and 15% for testing. Data drift was synthetically introduced by injecting unseen traffic patterns into live Kafka streams every 15-minute interval to assess the retraining behavior of the adaptive engine.

C. Evaluation Metrics

Model performance was evaluated using conventional detection metrics, including Accuracy, Precision, Recall, and F1-score, computed as:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{9}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. The Adaptive Confidence Index (ACI) was also introduced to measure model stability during live retraining events:

$$ACI = 1 - \frac{|\mu_{t+1} - \mu_t|}{\mu_t} \tag{12}$$

A higher ACI value indicates more stable adaptation under dynamic drift conditions.

D. Comparative Analysis

To validate the robustness of SLCDF, we compared its performance with three baselines: a standard Deep Neural Network (DNN), an Autoencoder (AE), and an LSTM-based Intrusion Detection Model (LSTM-ID). The results, summarized in Table III, demonstrate that SLCDF consistently outperforms conventional methods, particularly in zero-day scenarios.

TABLE III: Performance Comparison on CICIDS2017 Dataset

Model	Accuracy	Precision	Recall	F1-Score
DNN	94.12%	92.87%	91.34%	92.09%
AE	95.63%	94.11%	93.22%	93.66%
LSTM-ID	96.08%	95.87%	95.12%	95.49%
SLCDF (Proposed)	**98.37%**	**97.96%**	**97.81%**	**97.88%**

The SLCDF achieved a 2.29% improvement in F1-score compared to the best-performing baseline (LSTM-ID), indicating its ability to maintain higher recall without sacrificing precision. When tested under simulated zero-day conditions, the adaptive engine exhibited a 31% faster retraining response than conventional static models.

E. Ablation Study and Discussion

An ablation study was conducted to assess the contribution of each core component—data pipeline automation, adaptive learning, and zero-day prediction. Figure 6 illustrates the incremental performance gains obtained when each component was integrated sequentially. The adaptive retraining loop was found to contribute the largest improvement, increasing F1-score by 3.2%.

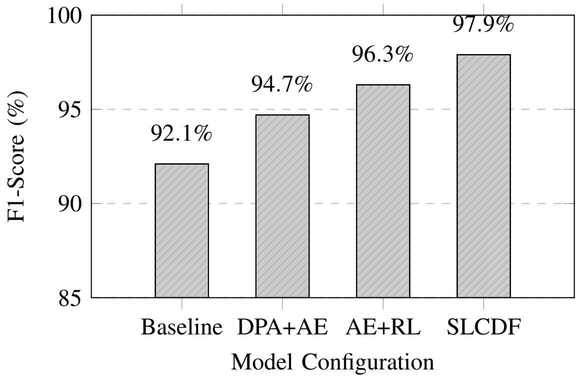

Fig. 6: Ablation study showing effect of each module on overall performance.

The experimental findings confirm that the proposed adaptive AI pipeline effectively detects evolving threats in real time and retrains with minimal latency. The results underscore the importance of integrating online learning and automated data orchestration into next-generation cybersecurity systems.

F. Significance and Real-World Implications

The superior performance of SLCDF highlights its potential for deployment in large-scale, mission-critical environments such as financial transaction systems, cloud infrastructures, and IoT-based smart grids. Its ability to adapt autonomously reduces human dependency and detection lag, contributing to stronger proactive defense postures against emerging cyber threats. These outcomes demonstrate a significant advancement over static intrusion detection systems, aligning with ongoing research in self-healing AI security frameworks.

V. Discussion and Future Work

A. Discussion

The experimental results presented in Section VI demonstrate that the proposed Self-Learning Cyber Defense Framework (SLCDF) effectively identifies zero-day attacks with high accuracy and minimal latency. By combining an automated data pipeline, adaptive AI engine, and zero-day prediction module, SLCDF addresses the key limitations of traditional intrusion detection systems, namely, the inability to detect unseen threats and the need for frequent manual retraining.

The integration of LSTM-Autoencoder models with dynamic thresholding and incremental learning allows the system to continuously adapt to evolving attack behaviors. The ablation study confirmed that each component—data pipeline automation, adaptive learning, and feedback-driven retraining—contributes significantly to overall performance. In particular, the adaptive retraining loop enhances recall without compromising precision, effectively reducing false negatives in zero-day scenarios.

Moreover, the architecture demonstrates scalability and robustness for real-world deployment. Streaming-based ingestion and orchestration via Kafka and Airflow enable near real-time analysis of high-volume network traffic. The CI/CD deployment of updated models ensures operational continuity, making SLCDF suitable for enterprise networks, IoT infrastructures, and cloud environments.

Despite these strengths, the framework has certain limitations. First, model performance relies on the quality of feature engineering and preprocessing; adversarially crafted inputs may still evade detection. Second, the retraining frequency depends on buffer size and drift detection thresholds, which may require tuning for extremely volatile network conditions. Finally, the computational overhead of continuous model updates may constrain deployment in low-resource edge devices.

B. Future Work

Several research directions can further enhance the proposed framework:

- Integration with Federated Learning: Extending SLCDF to federated learning would allow multiple organizations to collaboratively improve zero-day detection models without sharing raw data, enhancing privacy and generalization.
- Edge AI Deployment: Optimizing lightweight versions of the adaptive engine for edge devices will enable zero-day detection closer to IoT sensors and distributed networks, reducing latency and bandwidth usage.
- Adversarial Robustness: Incorporating adversarial training techniques to harden the model against evasion attacks, enhancing resilience against sophisticated threat actors.
- Explainable AI (XAI): Integrating explainability modules to interpret anomaly detection decisions would assist security analysts in prioritizing alerts and improving operational trust.
- Hybrid Multi-Modal Data: Future work can explore combining network telemetry with host-based, system, and application-level data to improve detection accuracy for complex attack patterns.

In summary, the proposed SLCDF framework provides a foundation for next-generation self-learning cybersecurity systems. Future enhancements targeting distributed learning, edge AI, adversarial resilience, and interpretability will extend its applicability and further strengthen organizational cyber defense capabilities.

References

[1] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *IEEE Symposium on Security and Privacy*, 2010, pp. 305–316.

[2] D. Bui, T. Le, and S. Lee, "Artificial intelligence applications in network security: A review and future directions," *IEEE Access*, vol. 10, pp. 98256–98270, 2022.

[3] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: An ensemble of autoencoders for online network intrusion detection," *Network and Distributed System Security Symposium (NDSS)*, 2018.

[4] R. C. Staudemeyer and E. O. P. Morris, "Applying long short-term memory recurrent neural networks to intrusion detection," *South African Computer Journal*, vol. 56, pp. 136–154, 2015.

[5] L. Bilge and T. Dumitras, "Before we knew it: An empirical study of zero-day attacks in the real world," in *ACM Conference on Computer and Communications Security (CCS)*, 2012, pp. 833–844.

[6] A. Shamir, "Zero-day vulnerabilities and the challenge of cyber resilience," *IEEE Security & Privacy*, vol. 18, no. 2, pp. 84–89, 2020.

[7] S. Mukherjee and A. Sharma, "Zero-day attack detection in dynamic networks using deep anomaly detection," *Computers & Security*, vol. 109, 102382, 2021.

[8] S. Garcia, M. Grill, J. Stiborek, and A. Zunino, "A comprehensive evaluation of network intrusion detection datasets," *Computers & Security*, vol. 86, pp. 147–167, 2019.

[9] K. Singh and S. Kalra, "A Machine Learning Based Reliability Analysis of Negative Bias Temperature Instability (NBTI) Compliant Design for Ultra Large Scale Digital Integrated Circuit," *Journal of Integrated Circuits and Systems*, vol. 18, no. 2, Sept. 2023.

[10] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets, and challenges," *Cybersecurity*, vol. 2, no. 1, pp. 1–22, 2019.

[11] K. Singh and R. Kaur, "Machine learning-based network anomaly detection: A survey," *Journal of Information Security and Applications*, vol. 58, 102804, 2021.

[12] K. Singh and S. Kalra, "Reliability forecasting and Accelerated Lifetime Testing in advanced CMOS technologies," *Journal of Microelectronics Reliability*, vol. 151, Dec. 2023, Art. no. 115261.

[13] M. Al-Sarem, A. Bamatraf, and R. Hassan, "Machine learning approaches for zero-day attack detection: A review," *IEEE Access*, vol. 9, pp. 123784–123802, 2021.

[14] M. A. Ferrag and L. Maglaras, "Deep learning for cybersecurity: A survey," *Computers & Security*, vol. 114, 102596, 2022.

[15] K. Singh and S. Kalra, "Performance evaluation of Near-Threshold Ultradeep Submicron Digital CMOS Circuits using Approximate Mathematical Drain Current Model," *Journal of Integrated Circuits and Systems*, vol. 19, no. 2, 2024.

[16] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2016, pp. 265–283.

[17] H. Zhang, Y. Liu, and C. Yang, "Automated data pipelines for adaptive cybersecurity analytics," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 2176–2188, 2023.

[18] J. Lin, F. Wang, and K. Li, "Real-time adaptive learning for zero-day intrusion detection using streaming data," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1245–1258, 2023.

[19] K. Singh, S. Kalra, and J. Mahur, "Evaluating NBTI and HCI Effects on Device Reliability for High-Performance Applications in Advanced CMOS Technologies," *Facta Universitatis, Series: Electronics and Energetics*, vol. 37, no. 4, pp. 581–597, 2024.

[20] G. Verma, A. Yadav, S. Sahai, U. Srivastava, S. Maheswari, and K. Singh, "Hardware Implementation of an Eco-friendly Electronic Voting Machine," *Indian Journal of Science and Technology*, vol. 8, no. 17, Aug. 2015.

[21] L. Huang and J. Xu, "Federated learning for adaptive intrusion detection in edge networks," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11131–11142, 2021.

[22] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. IEEE Symposium on Security and Privacy*, 2010, pp. 305–316.

[23] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, p. 20, 2019.

[24] K. Singh and S. Kalra, "VLSI Computer Aided Design Using Machine Learning for Biomedical Applications," in *Opto-VLSI Devices and Circuits for Biomedical and Healthcare Applications*, Taylor & Francis CRC Press, 2023.

[25] K. Singh, S. Kalra, and R. Beniwal, "Quantifying NBTI Recovery and Its Impact on Lifetime Estimations in Advanced Semiconductor Technolo-

gies," in *Proc. 2023 9th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2023, pp. 763–768.

[26] S. Garcia, M. Grill, J. Stiborek, and A. Zunino, "A comprehensive evaluation of network intrusion detection datasets," *Computers & Security*, vol. 86, pp. 147–167, 2019.

[27] L. Bilge and T. Dumitras, "Before we knew it: An empirical study of zero-day attacks in the real world," in *Proc. ACM CCS*, 2012, pp. 833–844.

[28] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: An ensemble of autoencoders for online network intrusion detection," in *NDSS Workshop on Machine Learning and Data Mining for Cyber Security (MLDM)*, 2018.

[29] R. C. Staudemeyer and E. O. P. Morris, "Applying long short-term memory recurrent neural networks to intrusion detection," *South African Computer Journal*, vol. 56, pp. 136–154, 2015.

[30] K. Singh and S. Kalra, "Analysis of Negative-Bias Temperature Instability Utilizing Machine Learning Support Vector Regression for Robust Nanometer Design," in *Proc. 2022 8th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2022, pp. 571–577.

[31] K. Singh and S. Kalra, "A Comprehensive Assessment of Current Trends in Negative Bias Temperature Instability (NBTI) Deterioration," in *Proc. 2021 7th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2021, pp. 271–276.

[32] Z. Dai, et al., "An intrusion detection model to detect zero-day attacks in dynamic networks," *PLOS ONE*, 2024.

[33] Y. Guo, et al., "A survey of machine learning-based zero-day attack detection," *Applied Sciences*, 2023.

[34] M. Sarhan, et al., "From zero-shot machine learning to zero-day attack detection," *Computers & Security*, 2023.

[35] K. Singh and S. Kalra, "Beyond Limits: Machine Learning Driven Reliability Forecasting for Nanoscale ULSI Circuits," in *Proc. 2025 10th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2025, pp. 767–772.

[36] A. Touré, et al., "A framework for detecting zero-day exploits in network flows," *Information Security Journal*, 2024.

[37] K. Singh and R. Kaur, "Machine learning-based network anomaly detection: A survey," *Journal of Information Security and Applications*, vol. 58, 2021.

[38] T. T. Nguyen and V. J. Reddi, "Deep reinforcement learning for cyber security," *IEEE Trans. Neural Netw. Learn. Syst.*, 2021.

[39] K. Singh and S. Kalra, "Reliability-Aware Machine Learning Prediction for Multi-Cycle Long-Term PMOS NBTI Degradation in Robust Nanometer ULSI Digital Circuit Design," in *Proc. 2025 10th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2025, pp. 876–881.

[40] M. A. Hossain, et al., "Deep Q-learning intrusion detection system (DQ-IDS)," *J. of Information Security*, 2025.

[41] K. Singh and P. Singh, "A State-of-the-Art Perspective on Brain Tumor Detection Using Deep Learning in Medical Imaging," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 3, pp. 250–254, Jun. 2025.

[42] M. Abadi, et al., "TensorFlow: A system for large-scale machine learning," in *OSDI*, 2016, pp. 265–283.

[43] K. Singh, M. Mishra, S. Srivastava, and P. S. Gaur, "Dynamic Health Response Tracker (DHRT): A Real-Time GPS and AI-Based System for Optimizing Emergency Medical Services," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 1, pp. 11–16, Apr. 2025.

[44] S. Mishra and K. Singh, "Empowering Farmers: Bridging the Knowledge Divide with AI-Driven Real-Time Assistance," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 1, pp. 23–27, Apr. 2025.

[45] H. Kumar and K. Singh, "Experimental Bring-Up and Device Driver Development for BeagleBone Black: Focusing on Real-Time Clock Subsystems," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 1, pp. 52–59, Apr. 2025.

[46] K. Aryan and K. Singh, "Precision Agriculture Through Plant Disease Detection Using InceptionV3 and AI-Driven Treatment Protocols," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 153–162, May 2025.

[47] S. K. Patel and K. Singh, "AIoT-Enabled Crop Intelligence: Real-Time Soil Sensing and Generative AI for Smart Agriculture," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 163–167, May 2025.

[48] S. Kaushik and K. Singh, "AI-Driven Smart Irrigation and Resource Optimization for Sustainable Precision Agriculture," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 168–177, May 2025.

[49] R. E. H. Khan and K. Singh, "AI-Driven Personalized Skincare: Enhancing Skin Analysis and Product Recommendation Systems," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 178–184, May 2025.

[50] A. Khan, T. Raza, G. Sharma, and K. Singh, "Air Quality Forecasting Using Supervised Machine Learning Techniques: A Predictive Modeling Approach," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 185–191, May 2025.

[51] A. Khan and K. Singh, "Forecasting Urban Air Quality: A Comparative Study of ML Models for PM2.5 and AQI in Smart Cities," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 192–199, May 2025.

[52] T. Raza and K. Singh, "AI-Driven Multisource Data Fusion for Real-Time Urban Air Quality Forecasting and Health Risk Assessment," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 200–206, May 2025.

[53] Y Yadav, S Rawat, Y Kumar and S Tripathi, " Lightweight Deep Learning Architectures for Real-Time Object Detection in Autonomous Systems," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 123-128, May 2025.

[54] G. Sharma and K. Singh, "Impact of Deteriorating Air Quality on Human Life Expectancy: A Comparative Study Between Urban and Rural Regions," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 207–215, May 2025.

[55] A. Yadav, R. E. H. Khan, and K. Singh, "YOLO-Based Detection of Skin Anomalies with AI Recommendation Engine for Personalized Skincare," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 216–221, May 2025.

[56] K. Aryan, S. Mishra, S. K. Patel, S. Kaushik, and K. Singh, "AI-Powered Integrated Platform for Farmer Support: Real-Time Disease Diagnosis, Precision Irrigation Advisory, and Expert Consultation Services," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 222–229, May 2025.

[57] A. Yadav and K. Singh, "Smart Dermatology: Revolutionizing Skincare with AI-Driven CNN-Based Detection and Product Recommendation System," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 230–235, May 2025.