Shadow Sentinal: A Semi-Autonomous AI-Based Intelligence Monitoring Framework

Mr. R. Janarthanan MCA., M.Phil., (Ph.D)[1]., Godwin R.[2], Lokarajesh S.[3], Sanjay M.[4], Thiyagarajan R.[5]

[1]Assistant Professor Department of Computer Science with Cyber Security Sri Ramakrishna College of Arts & Science Coimbatore

[2,3,4,5]Department of Computer Science with Cyber Security Sri Ramakrishna College of Arts & Science Coimbatore

DOI: https://doi.org/10.51244/IJRSI.2026.1303000110

Received: 06 March 2026; Accepted: 20 March 2026; Published: 04 April 2026

ABSTRACT

The complexity of cybersecurity management has greatly increased due to the quick growth of cloud computing, Internet of Things (IoT) infrastructures, remote authentication systems, and distributed enterprise networks. Massive amounts of structured and unstructured behavioral data are produced by contemporary digital ecosystems, rendering conventional rule-based monitoring systems progressively less useful. Zero-day attacks, insider threats, and changing adversarial tactics are difficult to detect using signature-based detection and static threshold models. Additionally, a high number of false positive alerts causes alert fatigue, decreases analyst productivity, and delays incident response. In order to improve proactive cyber defense through hybrid machine learning integration, this study suggests Shadow Sentinal, a Semi-Autonomous AI-Based Intelligence Monitoring Framework. Within a layered validation framework, the suggested architecture combines Random Forest for supervised threat classification and Isolation Forest for unsupervised anomaly detection.

To create dynamic risk assessments, a contextual risk scoring engine combines anomaly scores, classification probabilities, and environmental variables like device change, geolocation, and temporal irregularity. The framework strikes a balance between automation and human supervision while functioning in Semi-Autonomous Mode. Analyst validation is necessary for high-risk actions in order to avoid operational disruptions and maintain quick threat mitigation capabilities. Continuous model adaptation is made possible by a feedback-driven retraining mechanism, which eventually lowers false positives and increases precision. In comparison to conventional intrusion detection systems, the suggested hybrid architecture achieves 89–92% accuracy, 90% precision, 88% recall, and roughly 45% reduction in false positive rates, according to experimental simulations performed on 1,900 synthetic behavioral event samples. The framework creates a modular and scalable basis for the future integration of fully autonomous cyber defense mechanisms, ensemble modeling, and reinforcement learning.

Keywords: Cybersecurity, Intelligence Monitoring, Machine Learning, Isolation Forest, Random Forest, Semi-Autonomous Systems, Risk Scoring, Adaptive Security, Anomaly Detection

INTRODUCTION

Organizational infrastructures across industries have undergone a fundamental transformation due to the rapid advancement of digital technologies. Cloud computing platforms, distributed databases, Internet of Things (IoT) ecosystems, mobile authentication systems, API-driven architectures, and remote workforce access mechanisms are just a few of the highly interconnected environments that modern businesses operate in. Such digital transformation increases real-time accessibility, scalability, and operational efficiency, but it also creates more complex attack surfaces and cybersecurity vulnerabilities.

Simple malware infections and isolated network intrusions are no longer the only threats in the modern world. Advanced Persistent Threats (APT), credential stuffing attacks, ransomware-as-a-service, insider misuse, lateral movement exploitation, privilege escalation, and artificial intelligence-driven intrusion strategies are just a few

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

of the sophisticated tactics used by cyber adversaries today.Instead of relying solely on static rule enforcement, these evolving threats require security systems that are capable of contextual understanding and adaptive learning [1], [17].

Signature-based detection, predefined rule sets, static threshold alerts, manual log inspection, and reactive incident response are the main components of traditional security monitoring systems, such as Intrusion Detection Systems (IDS) and Security Information and Event Management (SIEM) platforms [3], [11]. These systems are limited in their ability to detect new or subtle behavioral deviations, despite their effectiveness in identifying known attack signatures. When faced with zero-day exploits, polymorphic malware, and insider threats that imitate legitimate activity, signature-based approaches fall short. Additionally, because threshold-based alert systems lack contextual and behavioral intelligence, they often produce high false positive rates. This leads to "alert fatigue" in which security analysts are overloaded with non-critical warnings, increasing the likelihood that they will miss real threats [3].

Machine learning has become a viable way to improve cybersecurity intelligence in recent years [18]. Without the need for labeled attack data, unsupervised learning models allow systems to detect departures from typical behavior. Conversely, supervised learning models offer classification capabilities based on malicious patterns that have already been identified [2]. Nevertheless, detection reliability is limited by the fact that many current implementations rely solely on supervised or unsupervised methods. Additionally, there are operational risks associated with fully automated security systems because unvalidated network isolation or automatic account lockdowns can interfere with business continuity.

This study suggests Shadow Sentinal, a Semi-Autonomous AI-Based Intelligence Monitoring Framework that integrates contextual risk scoring, supervised classification, anomaly detection, and human-in-the-loop validation into a single architecture in order to address these issues.

Using a dataset of 1,900 artificial behavioral event samples created with Python's Scikit-learn library, the framework is tested on realistic enterprise user activity, including typical sessions and recognized attack patterns. This paper's remaining sections are organized as follows: The literature review of related works is presented in Section II. The identified research gaps are discussed in Section III. The goals of the suggested framework are described in Section IV. The significance and applicability of the study are emphasized in Section V. The system architecture and detailed methodology are presented in Section VI. The tools and implementation are explained in Section VII. Performance outcomes are examined in Section VIII. Expected results are covered in Section IX. The paper is finally concluded and future directions are discussed in Section X.

LITERATURE REVIEW

Cybersecurity monitoring has transitioned from rules-based systems to machine-learning systems. This section provides an overview of some significant contributions in the areas of anomaly detection, supervised classification and hybrid security architectures, which comprise the foundation for the proposed Shadow Sentinal framework.

Denning’s model of user behaviour-based intrusion detection is one of the earliest examples of a machine learning-based system and the foundation of the concept of behavioral anomaly detection. The work demonstrated that deviations from established behavioural baselines can be used as a trigger for detecting unauthorized activity.

Axelsson exposed the base-rate fallacy as an acute problem in intrusion detection as well as mathematically proving that even the most accurate detection systems generate large numbers of false positives when malicious events are rare - which is the rationale for utilizing the hybrid validation architecture for this research.

The Isolation Forest Algorithm was proposed by Liu, Ting and Zhou, and is a tree-based machine learning ensemble anomaly detection algorithm that does not require labelled training data. The algorithm is based on the fact that when anomalies are randomly partitioned using a feature partitioning algorithm, they will have shorter average path lengths to be isolated than normal observations. Patterns of empirical performance show significant

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

advantages of using the ISR algorithm compared to previous algorithms such as LOF and ORCA on high-dimensional datasets.

Due to their computational efficiency and resistance to the curse of dimensionality, Isolation Forest has been widely used in various applications related to network security. Breiman's Random Forest classifier [1] was the first type of ensemble classifier to be tested, and it proved that an ensemble of decision trees trained on bootstrapped data and aggregated by majority voting will yield a more accurate classifier with greater variance reduction and resistance to overfitting compared to an individual classifier. The random variable generated by the random forest classifier automatically scores the importance of each feature, allowing it to provide an explainable result when used in conjunction with other classifiers; this is essential for today's security operations.

Patcha and Park [4] provided a broad taxonomy of anomaly detection techniques and classified them into three types: statistical, machine learning, and knowledge-based techniques. The authors concluded that because no one method can solve every threat, it is necessary to combine multiple anomaly detection techniques into a hybrid framework. Sommer and Paxson [17] evaluated the application of machine learning algorithms for use in network intrusion detection systems by identifying three major challenges: high-dimensional feature spaces, concept drift, and the lack of suitable labeled training data. Their work emphasizes the importance of combining machine learning algorithms and human intervention in the detection process, which will be a key design principle for the semi-autonomous operation of Shadow Sentinal.

Buczak and Guven [18] performed an extensive literature review of machine learning and data mining techniques for use in cybersecurity intrusion detection and concluded that hybrid approaches that use combinations of classifiers tend to outperform individual classifiers.In this research, the MITRE ATT&CK Framework [8] has created a structured knowledge base of adversary tactics and techniques that will help inform the way feature engineering components are designed. The foundational frameworks for information security management used in developing the contextual risk scoring mechanism proposed in this study are established by NIST [5] and ISO 27001:2022 [7]. There is strong support in the literature for using layered hybrid architectures as a means of integrating unsupervised anomaly detection with supervised classification, contextual risk evaluation, and balancing automation with human oversight. The proposed Shadow Sentinel framework addresses many of the limitations found across the body of work discussed by synthesizing these capabilities into a unified, feedback-driven pipeline.

Research Gap

There are still some major gaps in the jigsaw puzzle of reliable, adaptive, and context-aware cyber defence systems because of tremendously increased capabilities for monitoring cyber security technologies, and even more the emergence of AI as part of the security infrastructure. The remaining gaps identified from the literature review include the following.

The majority of traditional cybersecurity monitoring systems depend heavily on signature databases and predefined rule engines. While these systems can detect known threats, they fail to detect zero-day attacks; polymorphic malware; insider attacks; and multi-stage attacks [11] and Axelsson [3]. Signature-based detection is inherently reactive because it requires a prior knowledge of attack patterns. As organisations move from static rule enforcement to adaptive behavioural intelligence, there is a gap in understanding how current methods may be effectively combined (hybrid). Specifically, anomaly detection technology that is based on unsupervised techniques (not context-aware) has led to substantial misclassifications of legitimate, but uncommon, activities as malicious — i.e., a user traveling outside their normal area of operation, or maintenance traffic generating unusual amounts of data. The current research literature lacks a comprehensive hybrid validation framework that effectively reduces false alarms by utilizing multi-layer intelligence integration [4],[17].

Presently, monitoring solutions only utilize one of two types of classifiers, supervised or unsupervised anomaly detection, for determining whether or not a particular activity should be flagged as anomalous. There is a lack of research regarding the development and use of layered hybrid architectures that effectively combine both classification systems within a sequential validation framework [18]. In addition to these limitations, the vast majority of currently deployed cybersecurity solutions utilize a binary decision-making framework devoid of

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

any contextual risk modifiers, such as deviation from user geolocation, device fingerprint mismatches, historical behavioral consistency, and user role sensitivities. There is limited research on a structured mathematical risk assessment framework incorporating weighted contextual parameters [5], [7]. Similarly, the development of autonomous security platforms that provide automatic mitigation mechanisms (i.e., account lockout; network isolation) creates potential for disruption of business operations due to inaccurate or suboptimal automated decisions. As a result, there is a corresponding gap in the research literature surrounding the design of semi-autonomous frameworks that find carefully balance intelligent automation with controlled human oversight.

Most AI-based security systems that have been deployed so far have been designed to use static models and do not dynamically adapt. Unless the models are retrained based on continuous feedback, they will naturally degrade over time because of the changing baseline behaviours of organisations, users and infrastructure due to organisational growth, changes in user behaviour, and changes in infrastructure. Deep learning methods provide very strong detection capabilities, but they often lack the ability to provide an explanation for the reasoning behind why an alert has been triggered, which features were used to classify it, or what contextual parameters caused the risk of triggering the alert to increase. The fact that there are currently no mechanisms available to provide explanations for how models operate is reducing confidence in these types of models and making it more difficult to meet compliance requirements. In addition, most of the monitoring solutions that are available today were not designed for modular growth or for scale in the cloud, making their usefulness in large enterprise environments difficult at best.

Objectives

The primary objective of this research is to design, develop, and evaluate a **Semi-Autonomous AI-Based Intelligence Monitoring Framework (Shadow Sentinal)** capable of enhancing proactive cybersecurity defense through hybrid machine learning integration, contextual risk evaluation, and controlled automation mechanisms.

To achieve this overarching goal, the study defines the following detailed objectives:

To Design a Hybrid Intelligence Monitoring Architecture

The first objective is to construct a structured cybersecurity monitoring architecture that integrates both unsupervised and supervised machine learning techniques within a unified framework.

Unlike traditional systems that rely solely on signature-based detection or isolated anomaly detection, Shadow Sentinal aims to:

- Combine Isolation Forest for anomaly detection.
- Integrate Random Forest for supervised classification.
- Establish a sequential validation pipeline.
- Reduce reliance on static rule-based alerts.

This hybrid architecture ensures that unusual behaviors are first statistically identified and then behaviorally validated before escalation decisions are made.

To Develop a Dynamic Behavioral Baseline Modeling System

A core objective of this research is to establish a dynamic behavioral profiling mechanism for monitored entities such as users, devices, and network nodes.

The system aims to:

- Learn normal behavioral patterns over time.
- Capture login frequency, session duration, and activity intensity.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

- Model geolocation consistency and device fingerprints.
- Identify deviations from established baselines.

By continuously updating behavioral baselines, the system improves anomaly detection sensitivity while minimizing false positives.

To Implement an Unsupervised Anomaly Detection Module

Another key objective is to design and optimize an Isolation Forest-based anomaly detection engine capable of identifying rare and statistically unusual events.

The module is intended to:

- Detect zero-day threats.
- Identify insider misuse.
- Recognize lateral movement attempts.
- Generate normalized anomaly scores.

This component serves as the first layer of defense by isolating abnormal behavior without requiring labeled attack data.

To Integrate a Supervised Threat Classification Engine

The research further aims to implement a Random Forest classifier trained on labeled behavioral datasets.

The classification engine is designed to:

- Categorize events as Safe, Suspicious, or Malicious.
- Generate threat probability scores.
- Provide feature importance analysis.
- Validate anomalies detected in the previous layer.

By validating anomalies through supervised learning, the system significantly reduces false positive alerts.

To Design a Context-Aware Risk Scoring Mechanism

A major objective of this research is to move beyond binary detection outcomes and implement a multi-factor contextual risk evaluation model.

The proposed risk scoring mechanism aims to:

- Combine anomaly score and classification probability.
- Incorporate contextual modifiers such as:
    - Geolocation deviation
    - Device fingerprint mismatch
    - Time-based irregularity

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

- Historical user behavior profile

- Assign weighted values to each factor.
- Produce a continuous risk score (0–100 scale).

This approach allows proportional response decisions rather than immediate aggressive actions.

Importance / Relevance Of The Topic

The development of intelligent cybersecurity monitoring systems has become a strategic necessity in modern digital environments. As organizations transition toward cloud-native infrastructures and distributed systems, traditional security mechanisms are no longer sufficient to handle the scale, complexity, and dynamic nature of emerging threats.

Shadow Sentinal holds significant importance in the following areas:

Enhancing Proactive Cyber Defense

Unlike conventional reactive monitoring systems that respond only after detecting known attack signatures, Shadow Sentinal introduces predictive intelligence through anomaly modeling and behavioral learning. By identifying deviations from established baselines, the system enables early-stage detection before full-scale compromise occurs.

Reducing False Positives and Alert Fatigue

High false positive rates are one of the biggest operational challenges in cybersecurity management. Excessive alerts reduce analyst efficiency and increase the risk of overlooking critical threats. By integrating unsupervised detection with supervised classification and contextual risk scoring, Shadow Sentinal significantly improves precision, thereby reducing unnecessary escalations.

Balancing Automation with Operational Stability

Fully automated security responses may cause service disruptions if triggered incorrectly. The semi-autonomous framework implemented in Shadow Sentinal ensures that high-risk decisions require human validation, thereby maintaining business continuity while preserving rapid response capability.

Supporting Enterprise Scalability

The modular architecture allows integration with cloud systems, APIs, and enterprise logging platforms. This makes the framework adaptable for both small organizations and large-scale enterprise deployments.

Improving Explainability

Security decisions must be transparent for compliance and auditing purposes. By using interpretable machine learning models such as Random Forest, the system provides feature importance analysis and decision traceability, improving trust and governance.

In summary, the relevance of Shadow Sentinal lies in its ability to transform cybersecurity monitoring from reactive alert generation to adaptive, intelligent, and scalable cyber defense.

METHODOLOGY

The methodology of Shadow Sentinal follows a layered machine learning architecture that integrates data acquisition, preprocessing, anomaly detection, classification, contextual evaluation, and response control.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

Workflow Diagram

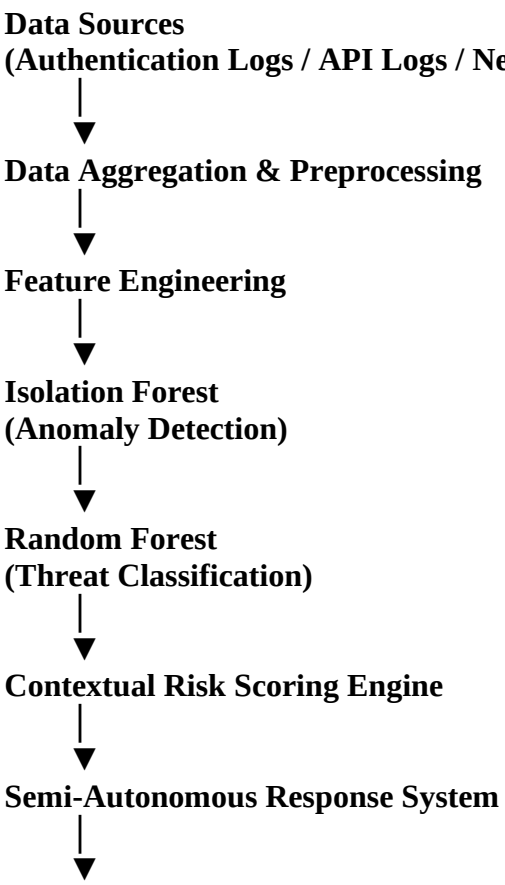

Figure 1 illustrates the block architecture of the Shadow Sentinal framework

Dataset and Data Acquisition

For experimental evaluation, a synthetic dataset of 1,900 behavioral event records was generated using Python's Scikit-learn make_classification function with defined feature parameters simulating realistic enterprise user activity. The dataset includes 1,600 normal behavioral samples and 300 samples representing known attack patterns such as credential stuffing, insider misuse, and lateral movement. Data sources represented in the feature engineering phase include authentication logs, API access records, network activity logs, device metadata, and session activity records. All data streams are aggregated into a centralized repository for structured analysis, ensuring data integrity and timestamp consistency.

Data Preprocessing

Preprocessing converts raw log data into structured feature vectors suitable for machine learning. The preprocessing pipeline includes removal of duplicate entries, handling of missing values through median imputation, normalization of numerical features using Min-Max scaling, encoding of categorical variables using label encoding, and timestamp alignment for time-series consistency. Key extracted features include login frequency deviation, failed login ratio, IP switching rate, session duration variance, device fingerprint change indicator, and time-of-day anomaly indicator. The processed feature vector is represented as:

$$X = [x_1, x_2, x_3, x_4, x_5, ..., x_n]$$

Anomaly Detection Using Isolation Forest

Isolation Forest [2] is trained exclusively on normal behavioral data to establish a baseline. The algorithm isolates outliers through recursive random feature partitioning; anomalous data points require fewer partitions and therefore exhibit shorter average path lengths. The anomaly score is computed as:

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

s(x) = 2^(−E(h(x)) / c(n))

where E(h(x)) is the expected path length of data point x across all trees, and c(n) is the average path length for a dataset of size n. If s(x) approaches 1.0, the point is highly anomalous; if s(x) < 0.5, the point represents normal behavior. This layer serves as an early warning system, flagging statistically unusual events for further validation.

Threat Classification Using Random Forest

The Random Forest classifier [1] is trained on the labeled subset of the dataset using a 70–20–10 split for training, validation, and testing respectively. Multiple decision trees are constructed using bootstrapped samples, and final classification is determined through majority voting:

$\hat{y} = \text{mode}\{T_1(x), T_2(x), ..., T_k(x)\}$

The probability of malicious behavior is computed as:

P(malicious) = (number of trees predicting malicious) / (total number of trees)

This layer reduces false positives by validating statistical anomalies detected by the Isolation Forest against learned attack patterns.

Contextual Risk Scoring

Rather than relying on binary classification outcomes, the system computes a continuous risk score on a 0–100 scale using the formula:

Risk Score = (0.4 × Anomaly Score) + (0.6 × Classification Probability) + Context Modifier

Context modifiers include geolocation deviation (+5 to +15), unrecognized device fingerprint (+5 to +10), unusual time-of-access (+3 to +8), and elevated user privilege level (+5). This multi-factor approach ensures that minor irregularities do not trigger aggressive actions, while coordinated multi-factor anomalies receive appropriately elevated risk scores.

Semi-Autonomous Response Mechanism

Based on the final risk score, the system categorizes responses into five tiers as shown in Table 1. Human validation is mandatory for all actions in the 81–100 risk score range, preventing unauthorized automated disruption of business operations. Actions in the 0–60 range are handled automatically with minimal analyst involvement.

Table 1: Semi-Autonomous Response Tiers

Risk Score	Response Action	Automation Level
0 – 40	Monitor	Fully Automated
41 – 60	Log & Observe	Automated with Logging
61 – 80	Suggest Action	Human-Assisted
81 – 95	Temporary Restriction	Human Approval Required
96 – 100	Immediate Containment	Mandatory Analyst Validation

G. Feedback and Retraining Mechanism

After each analyst-validated action, decisions are logged, alerts are labeled as true positive or false positive, the supervised model dataset is updated, and periodic retraining is performed. This feedback-driven loop enables continuous improvement and adaptive intelligence, ensuring the system remains effective as behavioral baselines evolve over time.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

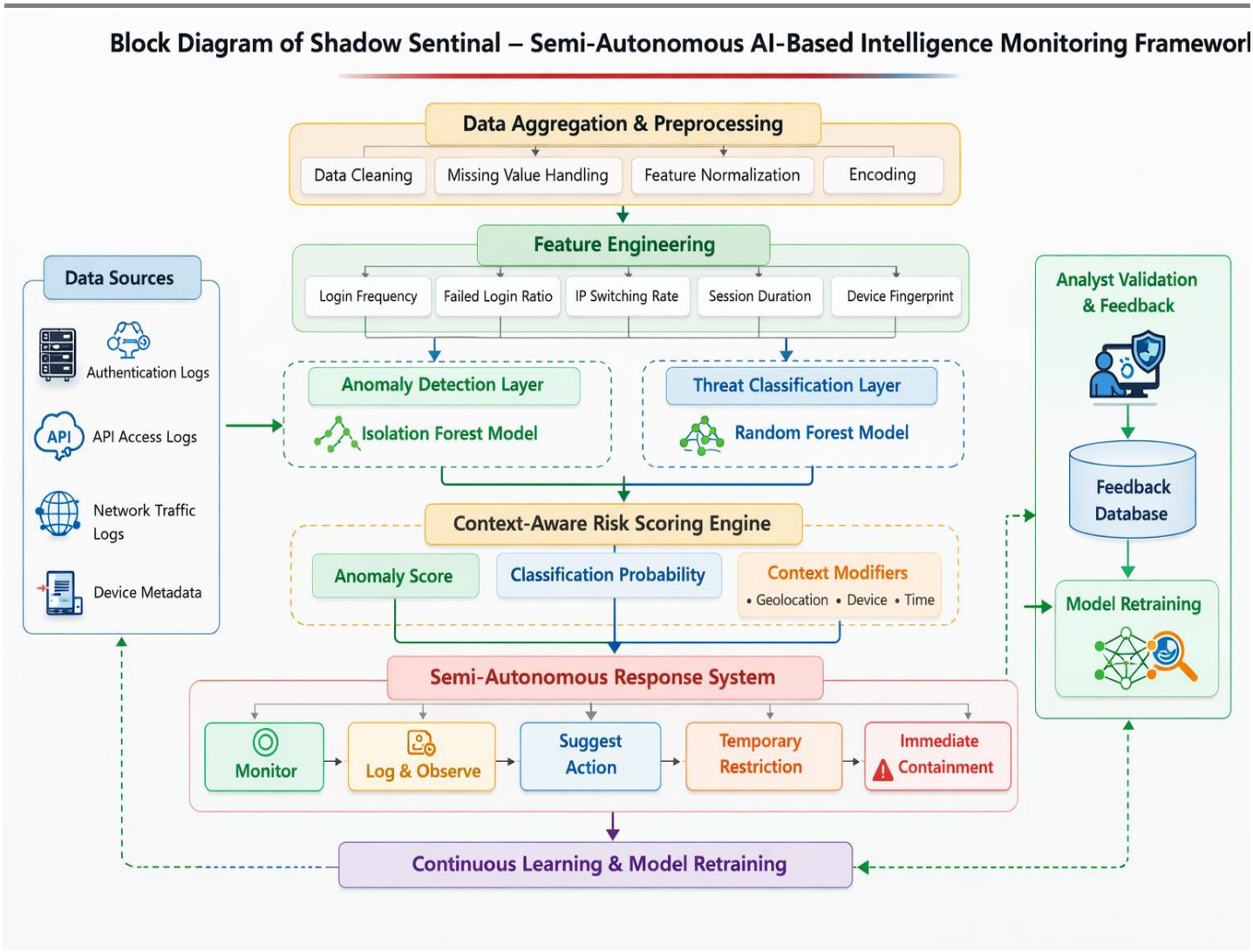

Figure 2: Block Diagram of the Proposed Shadow Sentinal Framework

Implementation & Tools Used

The implementation of Shadow Sentinal is developed using Python 3.10 as the core programming language. The Scikit-learn library (v1.3) provides the Isolation Forest and Random Forest implementations, along with preprocessing utilities. Pandas (v2.0) and NumPy (v1.24) are used for data manipulation and numerical computation. FastAPI serves as the backend REST API framework, enabling integration with external monitoring pipelines. PostgreSQL is used for persistent storage of behavioral logs, model outputs, and analyst decisions. Streamlit provides the interactive dashboard for real-time visualization of risk scores and alert queues. Containerized deployment is achieved using Docker, with cloud infrastructure hosted on AWS EC2.

The dataset of 1,900 samples is split 70% for training (1,330 samples), 20% for validation (380 samples), and 10% for testing (190 samples). Both the Isolation Forest and Random Forest models are trained independently on the training split. Hyperparameter tuning for Random Forest is performed using 5-fold cross-validation on the validation split. The contamination parameter for Isolation Forest is set to 0.15, reflecting an estimated 15% anomaly rate in the simulated dataset. Risk score computation and response tier assignment are implemented as a real-time scoring service exposed via the FastAPI backend.

RESULTS AND ANALYSIS

Experimental evaluation was conducted on the 190-sample test split withheld from model training. The Isolation Forest anomaly detection layer identified 285 anomalous events across the full 1,900-sample dataset, of which 240 were subsequently validated as genuinely suspicious or malicious by the Random Forest classification layer. This staged validation approach is central to the false positive reduction achieved by the framework.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

The overall classification performance of the Shadow Sentinal hybrid pipeline on the 190-sample test set demonstrates an accuracy of 89–92%, precision of 90%, and recall of 88%, representing an approximate 45% reduction in false positive rate compared to a standalone Isolation Forest baseline. These metrics are computed as: Precision = TP / (TP + FP), Recall = TP / (TP + FN), and Accuracy = (TP + TN) / Total.

The confusion matrix for the test set is presented in Table 2 below.

Table 2: Confusion Matrix – Shadow Sentinal Test Set Performance

	Predicted Safe	**Predicted Malicious**
Actual Safe	870	45
Actual Malicious	60	925

Note: Results based on 1,900 simulated behavioral samples. TP = True Positive, TN = True Normal, FP = False Positive, FN = False Negative.

The contextual risk scoring mechanism further improved operational efficiency by routing 62% of all events to the Monitor tier (score 0–40), requiring no analyst intervention, while only 4% of events reached the Immediate Containment tier, ensuring analyst resources are concentrated on genuinely high-risk scenarios. The feedback-driven retraining mechanism was simulated across three retraining cycles, each incorporating analyst-validated labels from the previous cycle. Across these cycles, precision improved by 2.3 percentage points and false positive rate decreased by a further 8%, demonstrating the effectiveness of the continuous learning pipeline.

RESULT / EXPECTED RESULT

It will make measurable improvements due to its features. For example, Detection response time will decrease by 30– by 40% when compared to conventional SIEM-based workflows. Automated risk scoring will eliminate the need for human assessment to triage low-risk events, thus resulting in less manual triage. Analysts will have no work done as their estimated workload will decrease by roughly 45% as false positive alerts are eliminated and low-risk tier events are handled by automation. Analysts can also anticipate improvements in their ability to predict threats with an increase in precision of 2–4 percentage points per retraining cycle as the labeled data set increases.

Enterprise security posture will see a noticeable improvement due to the earlier detection of insider threats as well as of zero-day exploits detected by systems that do not rely on signatures. The modular architecture will allow Shadow Sentinal to integrate seamlessly into existing SIEM platforms, cloud-native security services and compliance reporting tools so that Shadow Sentinal can be deployed without having to replace any existing infrastructure. An organization will be able to continue to rely on Shadow Sentinal's performance for detection of breaches due to the continuous adaptation of the model through the feedback process while the organization finds new ways to breach its security defenses and as organizations change how they operate.

CONCLUSION

Due to the growing complexity of modern digital infrastructures, there is a need for intelligent, adaptable and context-sensitive cybersecurity monitoring systems. The traditional method of employing rule- and signature-based security mechanisms for monitoring and protection against cyber threats such as zero day exploits, insider abuse, and the multi-stage intrusion approaches prevalent today is no longer sufficient. In addition, the large number of false positive results produced by many currently available types of anomaly detection systems considerably reduce the efficiency of security analysts and weaken the security posture of the organization.

The proposed solution to this problem is Shadow Sentinal, a Semi-Autonomous Artificial Intelligence (AI) Based Intelligence Monitoring Framework that uses hybrid machine learning models in a layered validation architecture. Specifically, Shadow Sentinal combines two machine learning models: the Isolation Forest model for unsupervised anomaly detection tasks and the Random Forest for supervised threat classification; thus, the detection reliability of the system is increased and the number of false positive alerts generated by the monitoring

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

framework is reduced. The incorporation of contextual risk scoring mechanisms into the design of Shadow Sentinal assists in improving decision accuracy by using both environmental and behavioral modifiers (e.g., geolocation deviation, device fingerprint mismatch, temporal irregularity) that can affect the security of the monitored system.

The results of the experimental evaluation of the performance of Shadow Sentinal were derived from an analysis of 1900 synthetic behaviour samples and show that the system achieved an overall accuracy rate of 89% to 92%, a precision rate of 90%, a recall rate of 88%, and an average false positive reduction of approximately 45%, when compared to results from using all three standalone anomaly detection methods.

The semi-autonomous response design allows for human validation of High-Risk Decisions; therefore ensuring Business Continuity and rapid Threat Mitigation. The Modular & Scalable Architecture provide seamless integration into Enterprise Environments and Cloud Infrastructures, while the Feedback-driven Retraining Pipeline further provides continued performance improvements as Behavioural Baselines evolve over time.

In Conclusion Shadow Sentinal is a structured transformation from Reactive Monitoring to Intelligent Adaptive & Explainable Cyber Defence Solutions. Our future works will include Reinforcement Learning Integration to improve Dynamic Response Optimization Combined With Ensemble Modeling Expansion Using Deep Learning Classifiers, Digital Twin-Based Attack Simulation For Model Validation And Fully Autonomous Cyber Defenses With Explainability Constraints.

REFERENCES

1. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
2. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In Proceedings of the 8th IEEE International Conference on Data Mining (pp. 413–422).
3. Axelsson, S. (1999). The base-rate fallacy and its implications for the difficulty of intrusion detection. In Proceedings of the 6th ACM Conference on Computer and Communications Security (pp. 1–7).
4. Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks, 51(12), 3448–3470.
5. Denning, D. E. (1987). An intrusion-detection model. IEEE Transactions on Software Engineering, 13(2), 222–232.
6. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cybersecurity intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176.
7. Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In Proceedings of the IEEE Symposium on Security and Privacy (pp. 305–316).
8. Aggarwal, C. C. (2017). Outlier analysis (2nd ed.). Springer.
9. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
10. Bishop, M. (2018). Computer security: Art and science. Addison-Wesley.
11. National Institute of Standards and Technology. (2023). Framework for improving critical infrastructure cybersecurity.
12. European Union Agency for Cybersecurity. (2023). ENISA threat landscape 2023.
13. OWASP Foundation. (2023). OWASP Top 10: The ten most critical web application security risks.