GPH-International Journal of Computer Science and Engineering
(Online ISSN 3050-9661)
Volume 9 | Issue 01 | Jan-June 2026
Published by: Global Publication House
DOI: https://doi.org/10.5281/zenodo.21848848
Available online at: www.gphjournal.org
Journal homepage: https://gphjournal.org/index.php/cse/index

An Explainable Hybrid Behavioral Analytics Framework for Insider Threat Detection in Hypervisor-Based Cloud Environments

Onuma Divine Anele[1], Onyeche Princewill Nweke[2], Eugene-Jaja Michael Telema[3]

[1,2,3] Department of Computer Science, Rivers State University, Nkpolu-Oroworukwo, Port Harcourt, Rivers State, Nigeria.
Corresponding Author Email: onuma.anele@rsu.edu.ng

Abstract

Insider threats remain one of the most challenging cybersecurity problems in hypervisor-based cloud environments because privileged users can exploit legitimate access to compromise virtual machines, manipulate hypervisor configurations, and exfiltrate sensitive data without triggering conventional security mechanisms. Existing approaches largely focus on enterprise user activities and rarely integrate hypervisor telemetry, behavioral analytics, Explainable Artificial Intelligence (XAI), and automated response into a unified framework. This study designed and developed an **Explainable Hybrid Behavioral Analytics Framework** for insider threat detection in hypervisor-based cloud environments. The study adopted the **Design Science Research Methodology (DSRM)** to design, implement, and evaluate the proposed framework. Hypervisor telemetry, virtual machine lifecycle events, privileged user activities, and behavioral logs were processed using User and Entity Behavior Analytics (UEBA), while a hybrid artificial intelligence engine integrating **Random Forest, XGBoost, Long Short-Term Memory (LSTM), Autoencoder, and Isolation Forest** were employed for threat detection. Explainability was achieved using **SHAP** and **LIME**, with dynamic risk scoring and automated response supporting real-time mitigation. Experimental evaluation using the **CERT Insider Threat, LANL, and TWOS datasets** achieved accuracies of **98.70%, 97.90%, and 98.30%**, respectively, with F1-scores ranging from **97.35% to 98.25%**, false positive rates of **1.20–1.60%**, detection latency of **42–55 ms**, and explainability scores of **0.91–0.93**. The study concludes that integrating hypervisor telemetry, hybrid AI, and XAI provides an accurate, transparent, scalable, and proactive solution for insider threat detection in modern hypervisor-based cloud infrastructures.

Keywords:

Hypervisor Security, Insider Threat Detection, Explainable Artificial Intelligence (XAI), User and Entity Behavior Analytics (UEBA), Hybrid Artificial Intelligence, Behavioral Analytics, Virtual Machine Introspection (VMI), Hypervisor Telemetry, Cloud Security, Dynamic Risk Scoring.

1.1 Background to the Study

Cloud computing has revolutionized the delivery of computing resources by providing scalable, on-demand, and cost-effective services through virtualization technologies. Virtualization enables multiple virtual machines (VMs) to operate simultaneously on a single physical host under the control of a hypervisor, also known as a Virtual Machine Monitor (VMM). Hypervisors play a critical role in ensuring resource isolation, workload management, and secure execution of virtual machines within cloud environments **[26, 32]**. As organizations increasingly migrate mission-critical applications and sensitive data to cloud infrastructures, the security of the hypervisor has become fundamental to maintaining the confidentiality, integrity, and availability of cloud services **[9, 18]**.

Despite the numerous advantages of cloud computing, virtualization introduces new security challenges because the hypervisor represents a single point of control over multiple virtual machines. A successful compromise of the hypervisor can enable attackers to manipulate virtual machines, access sensitive information, bypass isolation mechanisms, and disrupt cloud services **[32]**. Consequently, hypervisor attacks such as VM escape, VM hopping, hyperjacking, privilege escalation, and malicious virtual machine manipulation have become significant concerns for cloud service providers and enterprise organizations **[8, 9]**.

Among the numerous cybersecurity threats facing cloud infrastructures, insider threats remain one of the most difficult to detect and mitigate. Unlike external attackers, insiders—including cloud administrators, virtualization engineers, privileged users, contractors, and compromised service accounts—possess legitimate access privileges that allow them to perform activities that often appear indistinguishable from normal administrative operations **[21]**. Malicious insiders may exploit their privileges to create unauthorized virtual machines, alter hypervisor configurations, migrate workloads, manipulate resource allocations, extract confidential data, or disable security controls without immediately triggering conventional intrusion detection mechanisms. These characteristics make insider threats particularly dangerous in hypervisor-based cloud environments.

To address increasingly sophisticated cyber threats, researchers have explored Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) techniques for cybersecurity applications. Machine learning algorithms such as Random Forest, Support Vector Machines, Decision Trees, and ensemble models, together with deep learning architectures including Long Short-Term Memory (LSTM) networks and Autoencoders, have demonstrated remarkable capabilities in learning normal behavioral patterns and detecting anomalous activities within complex computing environments **[6]**. Furthermore, User and Entity Behavior Analytics (UEBA) has emerged as an effective security paradigm that continuously models user behavior, resource utilization, access patterns, and system interactions to identify behavioral deviations associated with insider attacks.

In recent years, Explainable Artificial Intelligence (XAI) has become increasingly important because many high-performing AI models operate as black-box systems whose decision-making processes are difficult for security analysts to interpret. Explainability techniques such as SHAP, LIME, Integrated Gradients, and attention-based visualization improve transparency by identifying the features responsible for classification decisions, thereby increasing analyst

confidence, facilitating forensic investigations, and supporting regulatory compliance **[35]**. Consequently, integrating XAI into cybersecurity systems has become an important research direction for developing trustworthy and accountable AI-based security solutions.

Alketbi and Mehmood **[4]** conducted a comprehensive survey of Explainable Artificial Intelligence techniques for malicious insider threat detection and highlighted the increasing importance of transparent AI models for cybersecurity applications. Their study reviewed supervised learning, deep learning, hybrid learning, and explainability techniques including SHAP and LIME while identifying several unresolved challenges such as heterogeneous behavioral data integration, severe class imbalance, computational complexity, privacy preservation, and the absence of standardized explainability evaluation metrics. The authors emphasized that future insider threat detection systems should combine hybrid AI techniques with explainability to improve both predictive performance and analyst trust **[4]**.

Although the survey by Alketbi and Mehmood **[4]** provides a valuable synthesis of Explainable AI for insider threat detection, its primary focus is on enterprise information systems and general cloud environments rather than virtualization-specific infrastructures. The reviewed studies predominantly utilize authentication logs, network traffic, email communications, file access records, and user activity logs while giving limited attention to behavioral information generated by hypervisors. Consequently, important virtualization-specific indicators such as hypervisor audit logs, virtual machine lifecycle events, hypervisor API calls, privileged administrative operations, resource allocation changes, and virtualization telemetry remain largely unexplored for insider threat detection **[4]**.

This limitation reveals a significant research gap. Existing insider threat detection frameworks rarely integrate hypervisor telemetry with behavioral analytics to continuously profile privileged users and virtual machine activities. Furthermore, most existing studies implement either machine learning or explainable AI independently without developing unified frameworks that combine behavioral analytics, hybrid machine learning, Explainable Artificial Intelligence, and dynamic risk scoring specifically for hypervisor-based cloud environments. As organizations continue to expand their adoption of cloud computing, the absence of intelligent, explainable, and hypervisor-aware insider threat detection mechanisms increases the likelihood of privilege abuse, unauthorized virtual machine manipulation, data leakage, and service disruption.

Addressing these challenges requires a framework capable of continuously monitoring hypervisor activities, modelling behavioral patterns of privileged users and virtual machines, accurately identifying anomalous insider behavior through hybrid artificial intelligence techniques, and providing transparent explanations that enable security analysts to understand and validate detection decisions. Such a framework should also incorporate dynamic risk scoring to continuously evaluate evolving insider behaviors and support proactive security response before significant damage occurs.

Therefore, the aim of this study is to design and develop an Explainable Hybrid Behavioral Analytics Framework for detecting insider threats in hypervisor-based cloud environments by integrating hypervisor behavioral monitoring, User and Entity Behavior Analytics (UEBA), hybrid machine learning, Explainable Artificial Intelligence (XAI), and dynamic risk scoring to improve detection accuracy, transparency, and cloud infrastructure security.

To achieve this aim, the study will:

1. identify and extract behavioral features from hypervisor logs, virtual machine lifecycle events, privileged user activities, and virtualization telemetry for insider threat detection;
2. design and develop an Explainable Hybrid Behavioral Analytics Framework that integrates UEBA, hybrid machine learning models, and Explainable Artificial Intelligence for detecting insider threats in hypervisor-based cloud environments;
3. implement the proposed framework in a simulated hypervisor-based cloud environment for real-time behavioral monitoring, dynamic risk scoring, and insider threat detection;
4. evaluate the performance of the proposed framework using standard cybersecurity metrics, including accuracy, precision, recall, F1-score, false positive rate, detection latency, computational overhead, and explainability; and
5. compare the effectiveness of the proposed framework with existing insider threat detection approaches in terms of detection accuracy, interpretability, scalability, and overall security performance.

The successful implementation of the proposed framework is expected to contribute to the advancement of hypervisor security, behavioral analytics, Explainable Artificial Intelligence, and cloud cybersecurity by providing an intelligent, transparent, scalable, and proactive solution for detecting insider threats in modern virtualized cloud environments.

2. LITERATURE REVIEW

2.1 Cloud Computing and Virtualization Security

Cloud computing and virtualization provide the technological foundation for modern cloud infrastructures by enabling scalable, flexible, and cost-effective resource provisioning. However, the abstraction of physical resources through virtualization also introduces new security risks, particularly at the hypervisor layer, making cloud and virtualization security essential components of modern cybersecurity research.

2.1.1 Cloud Computing

Cloud computing is a model for delivering computing resources—including servers, storage, networking, software, and applications—over the Internet on a pay-as-you-go basis **[26]**. Its key characteristics include on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. Cloud services are commonly categorized into three service models: Infrastructure as a Service (IaaS), which provides virtualized computing resources; Platform as a Service (PaaS), which offers application development platforms; and Software as a Service (SaaS), which delivers software applications over the Internet. Cloud deployment models include public, private, hybrid, and community clouds, each offering different levels of control, scalability, and security. Despite these advantages, cloud environments face significant security challenges, including data breaches, insider threats, insecure interfaces, account compromise, misconfigurations, and virtualization-related attacks that can undermine tenant isolation and cloud service integrity **[14, 26]**.

2.1.2 Virtualization Technology

Virtualization enables multiple isolated virtual machines (VMs) to share a single physical computing platform through a software layer known as the Virtual Machine Monitor (VMM) or hypervisor. The hypervisor manages hardware resource allocation, enforces isolation among virtual machines, and coordinates CPU, memory, storage, and network utilization. Hypervisors are generally classified as Type I (bare-metal), which run directly on physical hardware (e.g., VMware ESXi, Xen, Microsoft Hyper-V), and Type II (hosted), which operate on top of a conventional operating system (e.g., VMware Workstation and Oracle VirtualBox). While virtualization improves resource utilization, scalability, and workload flexibility, it also introduces security vulnerabilities such as VM escape, hypervisor compromise, privilege escalation, and inter-VM attacks, making the hypervisor a critical security component in cloud environments **[9, 32]**.

2.1.3 Hypervisor-Based Cloud Environments

Hypervisor-based cloud environments rely on enterprise-grade virtualization platforms to manage virtualized resources and support multi-tenant cloud services. VMware ESXi is widely adopted for enterprise virtualization due to its performance, scalability, and advanced management capabilities. Xen is an open-source Type I hypervisor commonly used in research and large-scale cloud deployments. Kernel-based Virtual Machine (KVM) integrates virtualization directly into the Linux kernel, providing high-performance virtualization for cloud infrastructures. Microsoft Hyper-V supports virtualization within Windows Server environments and integrates closely with Microsoft Azure services. OpenStack serves as an open-source cloud management platform that orchestrates compute, networking, and storage resources while supporting multiple hypervisors, including KVM, Xen, Hyper-V, and VMware ESXi. Although these platforms provide robust virtualization capabilities, they remain vulnerable to insider attacks, hypervisor exploitation, unauthorized virtual machine manipulation, and privilege abuse, underscoring the need for intelligent behavioral monitoring and explainable AI-driven security mechanisms **[31, 48]**.

2.2 Hypervisor Security

Hypervisor security is a fundamental aspect of cloud security because the hypervisor serves as the trusted software layer that manages virtual machines (VMs), allocates hardware resources, and enforces isolation among multiple tenants. As the control point of virtualized infrastructures, any compromise of the hypervisor can expose all hosted virtual machines, making it one of the most attractive targets for cyberattacks. Consequently, modern cloud security frameworks emphasize strengthening hypervisor resilience through secure architecture, continuous monitoring, access control, and behavioral analysis **[18, 32]**.

2.2.1 Hypervisor Architecture

A hypervisor, also known as a Virtual Machine Monitor (VMM), enables multiple virtual machines to share a single physical host while maintaining logical isolation and efficient resource utilization. Hypervisors are categorized into Type I (bare-metal), which operate directly on physical hardware (e.g., VMware ESXi, Xen, Microsoft Hyper-V), and Type II

(hosted), which run on top of a host operating system (e.g., Oracle VirtualBox and VMware Workstation). Type I hypervisors are preferred in cloud computing because they provide superior performance, scalability, and security by minimizing the attack surface associated with host operating systems **[9, 32]**.

2.2.2 Hypervisor Attack Surface

The hypervisor attack surface comprises all interfaces through which attackers may compromise virtualization platforms. These include hypervisor management APIs, administrative consoles, virtual machine management services, device emulation modules, hypercalls, shared memory, live migration mechanisms, and virtual networking components. As cloud infrastructures become increasingly dynamic, these interfaces create additional opportunities for attackers to exploit software vulnerabilities or abuse privileged access. Reducing the hypervisor attack surface through secure configuration, least-privilege access, and continuous monitoring is therefore essential for maintaining virtualization security **[8, 32]**.

2.2.3 Hypervisor Vulnerabilities

Hypervisor vulnerabilities remain a major concern because successful exploitation can compromise multiple virtual machines simultaneously. Common attacks include VM Escape, where malicious code breaks the isolation boundary and gains access to the hypervisor; VM Hopping, which enables unauthorized communication between virtual machines; and Hyperjacking, where attackers replace or compromise the legitimate hypervisor to gain complete control of the virtualized environment **[8]**. Additional threats include privilege escalation, memory attacks, side-channel attacks, and hypervisor rootkits **[9, 32]**. These vulnerabilities demonstrate the need for intelligent monitoring mechanisms capable of detecting subtle behavioral anomalies associated with insider and advanced persistent threats.

2.2.4 Hypervisor Monitoring Techniques

Effective hypervisor security requires continuous monitoring of virtualization activities to detect abnormal behavior before it compromises cloud services. Common monitoring techniques include hypervisor logs, Virtual Machine Introspection (VMI), hypervisor audit logs, and system call monitoring **[8, 18]**. While these techniques improve visibility into virtualization environments, most existing approaches rely on signature-based or rule-based detection and provide limited support for behavioral analytics, Explainable Artificial Intelligence, and dynamic risk assessment. This limitation highlights the need for hybrid AI-driven monitoring frameworks capable of continuously analyzing hypervisor telemetry and detecting insider threats in real time.

2.3 Insider Threats in Cloud Computing

Insider threats represent one of the most significant security challenges in cloud computing because they originate from individuals or entities with legitimate access to organizational resources. Unlike external attackers, insiders—including employees, contractors, administrators, and compromised accounts—can exploit authorized privileges to access,

manipulate, or exfiltrate sensitive information while evading conventional perimeter-based security mechanisms **[21]**.

Insider threats are commonly categorized into malicious insiders, negligent insiders, and compromised insiders.

The insider threat lifecycle generally progresses through stages of privilege acquisition, reconnaissance, unauthorized access, exploitation, persistence, data exfiltration or system manipulation, and concealment of malicious activities.

In virtualized cloud environments, insider threats pose even greater risks because privileged users can manipulate hypervisor configurations, create or delete virtual machines, migrate workloads, alter resource allocations, or bypass virtualization isolation mechanisms. A successful insider attack at the hypervisor layer can simultaneously compromise multiple virtual machines and cloud tenants **[32]**.

Several benchmark datasets have been developed to support insider threat research. The CERT Insider Threat Dataset remains the most widely used benchmark, followed by TWOS, Enron, ADFA, and LANL datasets.

Recent reviews have identified several limitations of existing insider threat datasets, including limited realism, insufficient diversity of insider behaviors, class imbalance, and the absence of hypervisor-centric telemetry **[4, 6]**.

2.4 Behavioral Analytics

Behavioral analytics has emerged as a fundamental approach for identifying sophisticated cyber threats by analyzing patterns of user and system activities rather than relying solely on predefined attack signatures. Advances in Artificial Intelligence and Machine Learning have significantly enhanced behavioral analytics, improving the detection of insider threats in dynamic cloud environments **[4, 6]**.

2.4.1 User Behavior Analytics (UBA)

User Behavior Analytics (UBA) focuses on monitoring individual user activities to detect abnormal behaviors. Although UBA effectively identifies user-centric anomalies, it often overlooks interactions involving virtual machines, applications, and other cloud entities **[6]**.

2.4.2 User and Entity Behavior Analytics (UEBA)

User and Entity Behavior Analytics (UEBA) extends UBA by incorporating behavioral information from users and non-human entities such as virtual machines, applications, service accounts, databases, and network devices. In hypervisor-based cloud environments, UEBA improves anomaly detection by correlating administrator activities with virtualization telemetry **[4]**.

2.4.3 Behavioral Profiling

Behavioral profiling constructs baseline models of normal operational behavior using features such as login frequency, command execution, virtual machine operations, resource utilization, and access patterns. Machine learning models compare current activities with historical baselines to identify anomalous behaviors associated with insider threats **[6]**.

2.4.4 Dynamic Risk Scoring

Dynamic risk scoring continuously evaluates insider threat levels by assigning adaptive risk values based on behavioral deviations, privilege levels, and operational context. Integrating dynamic risk scoring with behavioral analytics and Explainable Artificial Intelligence improves threat prioritization by providing transparent justifications for elevated risk levels **[4, 12]**.

2.4.5 Behavioral Indicators

Behavioral analytics relies on indicators such as login behavior, session duration, command execution, resource utilization, application access, virtual machine creation and deletion, VM migration, CPU and memory allocation changes, hypervisor API usage, privileged administrative actions, and hypervisor configuration modifications. These indicators enable comprehensive monitoring of virtualization infrastructures and improve the detection of sophisticated insider threats **[4, 6]**.

2.5 Hypervisor Behavioral Analytics

Hypervisor Behavioral Analytics (HBA) is an emerging cybersecurity approach that continuously monitors and analyzes activities occurring at the virtualization layer to identify abnormal behaviors indicative of insider threats or advanced cyberattacks. Unlike traditional host- or network-based security solutions, HBA leverages behavioral information generated directly by the hypervisor, enabling comprehensive visibility into privileged operations and virtual machine (VM) interactions without relying solely on guest operating system logs. This approach is particularly valuable in multi-tenant cloud environments, where a compromised hypervisor can simultaneously affect multiple virtual machines and cloud services **[18, 32]**.

Hypervisor Telemetry

Hypervisor telemetry comprises the operational data continuously generated by virtualization platforms, including CPU utilization, memory allocation, storage operations, network traffic statistics, and virtual machine state transitions. These telemetry streams provide valuable insights into the operational behavior of virtualized infrastructures and enable security systems to establish normal behavioral profiles for users, virtual machines, and hypervisor services. Continuous analysis of hypervisor telemetry facilitates the early identification of abnormal activities associated with privilege abuse, unauthorized resource consumption, and insider attacks before significant system compromise occurs **[9]**.

Hypervisor Event Logs

Hypervisor event logs record critical administrative and system events such as user authentication, configuration changes, virtual machine creation and deletion, resource allocation modifications, snapshot operations, and security policy updates. These logs serve as primary forensic evidence for detecting suspicious administrative activities and reconstructing security incidents. Integrating event log analysis with behavioral analytics enhances the detection of malicious insiders by identifying deviations from established administrative patterns **[32]**.

Virtual Machine Lifecycle Monitoring

Virtual Machine (VM) lifecycle monitoring involves continuous observation of VM creation, deployment, migration, suspension, cloning, restoration, and termination events. Since insider attacks frequently involve unauthorized VM manipulation to evade detection or exfiltrate sensitive information, monitoring lifecycle events enables the identification of unusual VM operations that deviate from normal cloud management practices.

Privileged Administrator Behavior

Privileged administrators possess extensive control over hypervisor resources and cloud infrastructures, making their activities critical for insider threat detection. Behavioral analytics profiles administrator actions—including login frequency, command execution, privilege escalation, resource provisioning, configuration modifications, and access timing—to distinguish legitimate administrative operations from malicious behavior. Continuous monitoring of privileged accounts significantly reduces the risk of unauthorized system modifications and privilege abuse **[21]**.

Hypervisor API Monitoring

Modern hypervisors expose Application Programming Interfaces (APIs) that facilitate automated management of virtual machines, resource scheduling, and infrastructure orchestration. While these interfaces improve operational efficiency, they also present potential attack vectors for malicious insiders and compromised accounts. Monitoring API invocations enables the detection of abnormal request patterns, unauthorized administrative commands, excessive resource allocation, and suspicious automation activities that may indicate insider attacks or account compromise.

Virtual Machine Introspection (VMI)

Virtual Machine Introspection (VMI) is a security technique that enables external observation of guest virtual machines from the hypervisor layer without requiring software agents within the guest operating system. VMI provides access to memory contents, process execution, system calls, file operations, and kernel activities while remaining isolated from compromised virtual machines. This external perspective makes VMI highly resistant to tampering and particularly effective for malware detection, forensic analysis, and insider threat monitoring in virtualized cloud environments **[18]**.

Resource Utilization Analytics

Resource utilization analytics examines the consumption patterns of CPU, memory, storage, network bandwidth, and input/output operations across virtualized environments. Significant deviations from established resource usage patterns may indicate unauthorized cryptocurrency mining, denial-of-service attacks, data exfiltration, or malicious virtual machine activities. Consequently, resource utilization serves as an important behavioral indicator for anomaly detection and continuous risk assessment within hypervisor-based cloud infrastructures.

Behavioral Baselines

Behavioral baselines represent statistical and machine learning models that characterize normal operational patterns of users, virtual machines, and hypervisor services. These baselines are constructed using historical behavioral data and continuously updated to accommodate legitimate operational changes. Subsequent activities are compared against the established baseline to detect anomalous behaviors indicative of insider threats. The integration of behavioral baselines with hybrid machine learning and Explainable Artificial Intelligence (XAI) enhances detection accuracy while providing interpretable explanations for security analysts.

Hypervisor Behavioral Analytics provides a comprehensive approach for securing virtualized cloud environments by integrating hypervisor telemetry, event log analysis, VM lifecycle monitoring, privileged administrator profiling, API monitoring, Virtual Machine Introspection, resource utilization analytics, and behavioral baselines. Compared with traditional network- and host-based intrusion detection systems, HBA offers deeper visibility into virtualization-specific activities and improves the detection of sophisticated insider threats. However, existing studies generally investigate these techniques independently, with limited integration into a unified framework that combines behavioral analytics, hybrid artificial intelligence, Explainable AI, and dynamic risk scoring for hypervisor-based cloud security. This limitation provides the motivation for the proposed Explainable Hybrid Behavioral Analytics Framework for Insider Threat Detection in Hypervisor-Based Cloud Environments.

2.6 Artificial Intelligence for Insider Threat Detection

Artificial Intelligence (AI) has become a fundamental technology for detecting insider threats by enabling cybersecurity systems to learn complex behavioral patterns and identify malicious activities that traditional signature-based methods often fail to detect. AI-driven approaches analyze large volumes of heterogeneous data, including authentication records, network traffic, system logs, command histories, and user activities, to distinguish normal behavior from anomalous actions indicative of insider attacks. Recent advances in machine learning, deep learning, and unsupervised learning have significantly improved detection accuracy, adaptability to evolving threats, and real-time decision-making in cloud environments **[4, 6]**.

2.6.1 Machine Learning

Machine learning algorithms remain widely adopted for insider threat detection because of their ability to classify user behavior based on historical observations. Decision Trees provide

interpretable classification rules but are susceptible to overfitting. Random Forest improves prediction accuracy and robustness by aggregating multiple decision trees, making it one of the most effective classifiers for behavioral anomaly detection. Support Vector Machines (SVMs) perform well in high-dimensional datasets by identifying optimal decision boundaries but require careful parameter optimization. Extreme Gradient Boosting (XGBoost) enhances classification performance through sequential ensemble learning and effectively handles imbalanced cybersecurity datasets. Logistic Regression, although computationally efficient and highly interpretable, is generally limited to modeling linear relationships and may struggle with complex insider behaviors. Overall, ensemble learning methods such as Random Forest and XGBoost consistently outperform single classifiers in detecting sophisticated insider threats **[6, 23]**.

2.6.2 Deep Learning

Deep learning techniques have gained prominence because they automatically learn hierarchical behavioral representations from large-scale cybersecurity data. Convolutional Neural Networks (CNNs) effectively extract spatial patterns from structured cybersecurity features, while Recurrent Neural Networks (RNNs) model sequential user activities but often suffer from vanishing gradient problems. Long Short-Term Memory (LSTM) networks overcome this limitation by capturing long-term temporal dependencies, making them particularly suitable for modeling user behavior sequences and insider activities. Gated Recurrent Units (GRUs) provide comparable performance to LSTMs with lower computational complexity, making them attractive for real-time detection. More recently, Transformer architectures have demonstrated superior capability in modeling long-range behavioral dependencies through attention mechanisms, enabling improved scalability and detection accuracy for complex insider threat scenarios **[4, 35]**.

2.6.3 Unsupervised Learning

Unsupervised learning algorithms are particularly valuable for insider threat detection because malicious behaviors are often rare, previously unseen, or lack labeled training data. Isolation Forest identifies anomalies by isolating abnormal observations using random partitioning and has demonstrated high efficiency in cybersecurity applications. Autoencoders learn compact representations of normal behavior and detect anomalies through reconstruction errors. K-Means clustering groups similar behavioral patterns to identify deviations from established clusters, although its effectiveness depends on appropriate cluster selection. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) detects arbitrarily shaped behavioral clusters while effectively identifying outliers without requiring predefined cluster numbers. One-Class Support Vector Machines (One-Class SVMs) construct decision boundaries around normal behavior and classify deviations as anomalies, making them suitable for detecting previously unknown insider attacks. These unsupervised approaches are particularly effective for identifying zero-day and evolving insider threats in dynamic cloud environments where labeled datasets are scarce **[4, 6]**.

Synthesis

Collectively, machine learning, deep learning, and unsupervised learning provide complementary capabilities for insider threat detection. Machine learning algorithms offer

efficient classification, deep learning models capture complex temporal and contextual behavioral patterns, while unsupervised learning excels at detecting previously unseen anomalies. Consequently, recent research increasingly advocates hybrid AI frameworks that integrate supervised, unsupervised, and deep learning techniques to improve detection accuracy, reduce false positives, and enhance adaptability against evolving insider threats. Furthermore, integrating these hybrid models with Explainable Artificial Intelligence (XAI) enhances transparency and analyst trust, making them particularly suitable for securing hypervisor-based cloud environments **[4, 35]**.

2.7 Hybrid Machine Learning Models

Hybrid machine learning models have emerged as an effective approach for cybersecurity because they combine the complementary strengths of multiple algorithms to improve detection accuracy, robustness, and generalization. Unlike standalone models, hybrid approaches leverage supervised, unsupervised, and deep learning techniques to capture both known attack signatures and previously unseen anomalous behaviors. Supervised models such as Random Forest and XGBoost provide high classification accuracy for labeled attack data, whereas deep learning models such as Long Short-Term Memory (LSTM) networks effectively model temporal dependencies in sequential user and system activities. Unsupervised algorithms, including Autoencoders, Isolation Forest, and K-Means, enhance the detection of zero-day and previously unknown insider behaviors by identifying deviations from established behavioral patterns **[4, 6]**.

Several hybrid model combinations have demonstrated improved performance in insider threat detection and anomaly detection. The LSTM–Random Forest architecture exploits LSTM for extracting temporal behavioral features while Random Forest performs robust classification using the learned representations, thereby reducing false positives in sequential activity analysis. Similarly, LSTM–XGBoost combines deep sequential learning with gradient boosting to improve classification performance on highly imbalanced insider-threat datasets. More recent studies have explored Transformer–Autoencoder architectures that learn long-range behavioral dependencies while reconstructing normal user activities to identify anomalies through reconstruction errors. Likewise, Autoencoder–Isolation Forest models first compress high-dimensional behavioral data into latent representations before Isolation Forest isolates anomalous instances, improving scalability and detection sensitivity. In addition, Random Forest–K-Means integrates unsupervised clustering with supervised classification to discover hidden behavioral groups before assigning threat labels, thereby enhancing classification accuracy in heterogeneous environments **[6, 20]**.

Recent studies further demonstrate the advantages of hybrid architectures in insider-threat detection. **Prasad [34]** proposed a hybrid framework that combines generative latent feature extraction with machine learning classifiers to improve the identification of complex insider behaviors under severe class imbalance. Similarly, **Tian [46]** developed a multi-attention behavioral sequence model for detecting privilege abuse, identity theft, and data leakage, reporting improved accuracy over traditional LSTM and graph-based baselines by effectively capturing contextual behavioral dependencies.

Despite these advances, existing hybrid models exhibit several limitations. Most are developed using enterprise authentication logs, email records, web activities, or network traffic and rarely incorporate hypervisor telemetry, virtual machine lifecycle events, hypervisor API calls, or privileged virtualization activities as behavioral features. Furthermore, many frameworks prioritize detection accuracy without integrating Explainable Artificial Intelligence (XAI), making their predictions difficult for security analysts to interpret. As highlighted by **Alketbi and Mehmood [4]**, future insider threat detection systems should integrate hybrid machine learning with explainability, heterogeneous behavioral analytics, and virtualization-aware data sources to improve transparency, analyst trust, and operational effectiveness. These limitations provide the motivation for developing an Explainable Hybrid Behavioral Analytics Framework specifically tailored to insider threat detection in hypervisor-based cloud environments.

2.8 Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) has emerged as a critical component of AI-driven cybersecurity by addressing the lack of transparency associated with complex machine learning and deep learning models. Unlike traditional "black-box" models, XAI provides interpretable explanations that enable security analysts to understand the rationale behind model predictions, thereby improving trust, accountability, and operational decision-making. In insider threat detection, XAI facilitates the identification of behavioral features responsible for classifying activities as malicious or benign, making AI-driven security systems more suitable for deployment in high-risk environments such as cloud computing and critical infrastructure **[4, 35]**.

The importance of explainability extends beyond improving model transparency. Security analysts require understandable explanations to validate alerts, support incident investigations, satisfy regulatory requirements, and reduce false positives. Explainable models also promote human trust by allowing analysts to verify whether predictions are based on meaningful behavioral indicators rather than spurious correlations. Consequently, XAI has become an essential requirement for trustworthy AI-enabled cybersecurity systems **[4, 35]**.

Among the most widely adopted local explanation techniques is **Local Interpretable Model-agnostic Explanations (LIME)**, which explains individual predictions by approximating the behavior of a complex model around a specific data instance. LIME is model-agnostic and enables analysts to determine why a particular user, session, or event has been classified as suspicious. Its localized explanations are particularly useful during forensic investigations and incident response because they provide case-specific insights. However, LIME relies on perturbation-based sampling, which may produce inconsistent explanations for similar inputs, reducing reproducibility in security-critical environments **[4, 37]**.

In contrast, **SHapley Additive exPlanations (SHAP)** provides both local and global interpretability by quantifying the contribution of each feature to a model's prediction using concepts from cooperative game theory. SHAP enables analysts to identify the most influential behavioral indicators across the entire model while simultaneously explaining individual predictions. Because of its consistency and theoretical foundation, SHAP has been extensively integrated with Random Forest, XGBoost, deep neural networks, and ensemble learning models for cybersecurity applications, including intrusion detection and insider threat detection **[4, 35]**.

Beyond feature-attribution methods, attention-based explainability has gained prominence in deep learning models such as Transformers, Long Short-Term Memory (LSTM) networks, and attention-enhanced recurrent neural networks. Attention mechanisms highlight the temporal events, behavioral sequences, or input features that receive the greatest influence during prediction, thereby enabling analysts to visualize how sequential behaviors contribute to threat detection. This approach is particularly effective for behavioural analytics involving user activities, command sequences, and network events **[4]**.

Recent research also emphasizes **human-in-the-loop explainability**, where AI-generated explanations are combined with expert knowledge to improve security decision-making. Rather than replacing human analysts, XAI supports collaborative analysis by enabling experts to validate predictions, provide feedback, and refine detection models through continuous learning. Human-in-the-loop approaches improve operational trust, facilitate adaptive model updates, and enhance the effectiveness of insider threat investigations, particularly in dynamic cloud environments **[4]**.

Despite significant advances, the evaluation of explainability remains an open research problem. Existing studies commonly assess predictive performance using accuracy, precision, recall, F1-score, and Area Under the Curve (AUC), while standardized metrics for measuring explanation quality are still lacking. Current explainability evaluation often relies on qualitative assessment, fidelity, consistency, stability, completeness, and human interpretability, yet no universally accepted benchmarking framework exists for cybersecurity applications. **Alketbi and Mehmood [4]** identified the absence of standardized explainability metrics and domain-specific evaluation methodologies as one of the major limitations hindering the deployment of trustworthy AI-based insider threat detection systems. This challenge motivates the development of explainable frameworks that combine high predictive performance with measurable and operationally meaningful interpretability.

2.9 Dynamic Risk Assessment

Dynamic Risk Assessment (DRA) has emerged as a proactive cybersecurity approach that continuously evaluates the likelihood and impact of security threats based on real-time system conditions, behavioral changes, and evolving threat intelligence. Unlike traditional static risk assessment methods, which rely on periodic evaluations, DRA enables continuous monitoring and adaptive decision-making in response to the rapidly changing cyber threat landscape **[12]**.

Risk prediction is a fundamental component of dynamic risk assessment, employing machine learning and predictive analytics to estimate the probability of future cyber incidents based on historical behaviors, system vulnerabilities, and threat intelligence. Predictive models enable organizations to anticipate malicious activities before they escalate into security breaches, thereby supporting proactive defense strategies **[31, 29]**.

Behavioral risk scoring complements risk prediction by assigning dynamic risk values to users, entities, or system components according to deviations from established behavioral baselines. User and Entity Behavior Analytics (UEBA) techniques analyze authentication records, resource utilization, command execution, access frequency, and administrative activities to identify abnormal behaviors associated with insider threats. Dynamic behavioral scores assist

security analysts in prioritizing high-risk entities for immediate investigation while minimizing false alarms.

Continuous monitoring enables real-time collection and analysis of security events, network traffic, system logs, and behavioral telemetry to detect emerging threats as they occur. Rather than relying on periodic assessments, continuous monitoring supports ongoing evaluation of organizational risk by integrating security events with contextual information, thereby improving situational awareness and reducing detection latency **[12]**.

Closely related is adaptive security, which allows security controls to adjust dynamically according to changing threat conditions and behavioral risk levels. AI-driven adaptive security mechanisms integrate anomaly detection, behavioral analytics, and automated response to modify access permissions, enforce security policies, and initiate mitigation actions based on continuously updated risk assessments. Such adaptive capabilities are increasingly recognized as essential for protecting modern cloud infrastructures against sophisticated insider and advanced persistent threats **[29]**.

Finally, threat prioritization enables organizations to allocate security resources efficiently by ranking detected threats according to their likelihood, potential impact, exploitability, and behavioral risk scores. Prioritizing threats improves incident response by ensuring that high-risk activities receive immediate attention while reducing alert fatigue caused by low-priority events. Integrating threat prioritization with dynamic risk assessment enhances the effectiveness of Security Operations Centers (SOCs) and supports informed decision-making in complex cloud environments **[12, 31]**.

2.10 Zero Trust Security

Zero Trust Security is a cybersecurity paradigm founded on the principle of **"never trust, always verify,"** whereby no user, device, application, or workload is inherently trusted regardless of its location within or outside the network perimeter **[36]**. Unlike traditional perimeter-based security models, Zero Trust Architecture (ZTA) enforces identity verification, device validation, and policy-based access control before granting access to protected resources. This approach is particularly relevant to hypervisor-based cloud environments, where privileged users and virtual machines require continuous verification to mitigate insider threats and unauthorized access.

A fundamental component of ZTA is **continuous authentication**, which continuously validates user identities and device trustworthiness throughout a session rather than relying solely on initial authentication. Continuous authentication enables the detection of abnormal behavioral changes, credential compromise, and privilege misuse in real time, thereby reducing the risk of insider attacks **[36]**.

Another key principle is **least privilege**, which restricts users and system processes to the minimum level of access required to perform authorized tasks. Applying least-privilege policies within hypervisor environments limits administrative privileges, reduces the attack surface, and minimizes the potential impact of compromised accounts or malicious insiders **[27]**.

Continuous monitoring complements Zero Trust by providing ongoing observation of user behavior, virtual machine activities, hypervisor events, and resource utilization. Integrating behavioral analytics with continuous monitoring enables early detection of anomalous activities that may indicate insider threats or privilege escalation attempts **[24, 36]**.

In virtualized cloud infrastructures, hypervisor access control forms an essential component of Zero Trust implementation. Hypervisor access control employs role-based and attribute-based authorization mechanisms to regulate administrative operations, virtual machine lifecycle management, hypervisor API usage, and configuration changes. Combining fine-grained access control with behavioral analytics and Explainable Artificial Intelligence enhances the ability to detect unauthorized hypervisor activities while providing transparent and trustworthy security decisions **[4]**.

Overall, Zero Trust Security provides a robust foundation for protecting hypervisor-based cloud environments by integrating continuous authentication, least-privilege access, continuous monitoring, and hypervisor access control to reduce insider threats and strengthen virtualization security.

2.11 Virtual Machine Introspection (VMI)

Virtual Machine Introspection (VMI) is a hypervisor-based security technique that enables external observation and analysis of virtual machine (VM) states without installing security agents inside the guest operating system. By operating outside the monitored VM, VMI benefits from the isolation provided by the hypervisor, making it more resistant to compromise by malware or malicious insiders while enabling stealthy monitoring of virtualized cloud environments **[8, 18]**.

The architecture of a VMI system typically consists of a hypervisor, monitored guest virtual machines, an introspection engine, and a semantic reconstruction layer that translates low-level hardware information into meaningful operating system objects.

Hypervisor monitoring forms the foundation of VMI by collecting security-relevant information directly from the virtualization layer.

Memory introspection is one of the most widely adopted VMI applications because many advanced cyberattacks leave identifiable traces in volatile memory.

Process monitoring is another important capability of VMI.

VMI has been extensively applied to malware detection because it allows security systems to observe malicious behavior independently of the guest operating system.

Beyond malware detection, VMI is increasingly recognized as a promising technology for insider threat monitoring in hypervisor-based cloud environments. By continuously collecting virtualization telemetry—including privileged administrative actions, hypervisor API calls, virtual machine lifecycle events, and resource utilization—VMI provides rich behavioral information for detecting insider attacks. Nevertheless, existing VMI research has focused

predominantly on malware analysis, intrusion detection, and forensic investigation, with comparatively limited attention to behavioral analytics for malicious insiders. Furthermore, few studies integrate VMI with User and Entity Behavior Analytics (UEBA), hybrid machine learning, Explainable Artificial Intelligence (XAI), and dynamic risk scoring to provide transparent and adaptive insider threat detection. Addressing this gap forms a key motivation for the proposed Explainable Hybrid Behavioral Analytics Framework for Insider Threat Detection in Hypervisor-Based Cloud Environments.

2.12 Privacy-Preserving Behavioral Analytics

Privacy-preserving behavioral analytics has emerged as a critical research area for securing sensitive behavioral data while maintaining the effectiveness of Artificial Intelligence (AI)-based cybersecurity systems. As organizations increasingly rely on behavioral analytics for insider threat detection, preserving user privacy during data collection, processing, and model training has become essential to comply with regulatory requirements and protect sensitive information **[35]**.

Federated Learning (FL) enables multiple organizations or cloud nodes to collaboratively train machine learning models without sharing raw data. Instead, only model parameters or gradients are exchanged, thereby reducing the risk of exposing sensitive behavioral information while improving the generalization capability of detection models across distributed environments **[22]**.

Differential Privacy (DP) enhances privacy by injecting controlled statistical noise into datasets or model parameters, ensuring that the inclusion or exclusion of an individual user's data has a negligible impact on model outputs **[15]**.

Homomorphic Encryption (HE) provides cryptographic protection by allowing computations to be performed directly on encrypted data without requiring decryption. Among the available schemes, the **Cheon–Kim–Kim–Song (CKKS)** algorithm is particularly suitable for machine learning because it supports efficient approximate arithmetic over encrypted numerical data **[13]**.

Secure Multi-Party Computation (SMPC) allows multiple parties to jointly compute analytical functions over their private datasets without revealing the underlying information to one another **[16]**.

Collectively, these privacy-preserving techniques strengthen behavioral analytics by enabling secure data sharing, collaborative model development, and confidential threat detection. However, each approach presents trade-offs involving computational cost, communication overhead, scalability, and model accuracy. Consequently, recent research has increasingly focused on hybrid privacy-preserving frameworks that combine Federated Learning, Differential Privacy, Homomorphic Encryption, and Secure Multi-Party Computation to achieve a better balance between privacy, efficiency, and detection performance in cloud-based cybersecurity systems.

2.13 Explainable Hybrid Behavioral Analytics

Explainable Hybrid Behavioral Analytics (EHBA) has emerged as a promising paradigm for improving insider threat detection by integrating User and Entity Behavior Analytics (UEBA), hybrid Artificial Intelligence (AI), Explainable Artificial Intelligence (XAI), hypervisor telemetry, dynamic risk scoring, and automated response into a unified cybersecurity framework. Unlike conventional intrusion detection systems that rely on signatures or isolated machine learning models, EHBA continuously profiles behavioral patterns of users, entities, and system resources to identify anomalous activities while providing transparent explanations for security decisions. This integrated approach enhances detection accuracy, analyst trust, and incident response capabilities in modern cloud environments **[4, 35]**.

UEBA forms the behavioral foundation of the framework by continuously collecting and analyzing authentication records, access logs, command histories, resource utilization, and interaction patterns to establish baseline behavioral profiles for users and entities. Deviations from these baselines are treated as potential indicators of insider threats or compromised accounts. Recent studies demonstrate that UEBA is more effective than traditional rule-based monitoring because it captures subtle behavioral changes associated with malicious activities that may bypass signature-based detection systems. Furthermore, advances in deep learning, particularly Autoencoders and sequence-learning models, have significantly improved UEBA's capability to detect complex anomalies in large-scale enterprise environments.

To improve detection performance, many contemporary cybersecurity frameworks employ hybrid AI architectures that combine supervised, unsupervised, and deep learning algorithms. Hybrid models leverage the strengths of different learning paradigms: supervised classifiers accurately recognize known attack patterns, unsupervised algorithms identify previously unseen anomalies, and deep learning models capture complex temporal and nonlinear behavioral relationships. Such combinations generally achieve higher detection accuracy, lower false-positive rates, and greater robustness than single-model approaches, making them particularly suitable for evolving insider threats in cloud computing environments **[4, 6]**.

Despite the improved predictive performance of hybrid AI models, their black-box nature limits their practical deployment in critical cybersecurity operations. Explainable Artificial Intelligence addresses this limitation by providing interpretable explanations for model predictions using techniques such as SHAP, LIME, attention visualization, and feature attribution. These techniques enable analysts to understand why a particular activity is classified as malicious, identify the behavioral features influencing the decision, and validate AI-generated alerts before initiating remediation actions. Explainability therefore enhances transparency, accountability, regulatory compliance, and analyst confidence in AI-driven cybersecurity systems **[1, 4, 35]**.

For hypervisor-based cloud environments, behavioral monitoring should extend beyond conventional enterprise logs to include hypervisor telemetry generated by virtualization platforms. Hypervisor telemetry comprises virtual machine creation, migration, suspension, deletion, resource allocation changes, hypervisor API calls, audit logs, memory events, and privileged administrative operations. Monitoring these virtualization-specific activities provides fine-grained visibility into privileged insider behavior that cannot be captured using traditional network or host-level monitoring alone. Recent hypervisor-introspection approaches

demonstrate that telemetry collected at the virtualization layer significantly enhances the detection of sophisticated malicious behaviors while maintaining isolation between guest virtual machines and the monitoring infrastructure **[19]**.

Dynamic risk scoring further strengthens behavioral analytics by continuously evaluating the likelihood that observed behaviors represent malicious insider activity. Rather than relying on static thresholds, dynamic risk models aggregate multiple behavioral indicators—including privilege usage, resource consumption, behavioral deviations, and historical activities—to compute adaptive risk scores that evolve with user behavior. This enables organizations to prioritize high-risk activities, reduce alert fatigue, and support proactive threat mitigation through continuous behavioral assessment **[4, 24]**.

An effective Explainable Hybrid Behavioral Analytics framework should also incorporate automated response mechanisms that translate behavioral risk assessments into timely security actions. Depending on the computed risk level, automated responses may include alert generation, privilege restriction, session termination, virtual machine isolation, policy enforcement, or initiation of forensic logging. Integrating automated response with explainable behavioral analytics reduces response latency while ensuring that security analysts understand the rationale behind mitigation decisions. Such integration supports adaptive cyber defense and improves the resilience of virtualized cloud infrastructures against evolving insider threats.

Although significant progress has been made in UEBA, hybrid AI, XAI, hypervisor monitoring, and automated cybersecurity, these technologies are often investigated independently. Existing studies rarely integrate all six components into a unified framework tailored to hypervisor-based cloud environments. This limitation motivates the proposed study, which seeks to develop an Explainable Hybrid Behavioral Analytics Framework that combines UEBA, hybrid machine learning, Explainable AI, hypervisor telemetry, dynamic risk scoring, and automated response to provide accurate, transparent, and proactive insider threat detection in virtualized cloud infrastructures.

Related Works

A recent IEEE Access survey by Alketbi and Mehmood **[4]** comprehensively reviewed Explainable Artificial Intelligence (XAI) techniques for malicious insider threat detection. The survey examined machine learning and deep learning models augmented with explainability methods such as SHAP and LIME, emphasizing the importance of transparent decision-making in cybersecurity. The authors proposed a conceptual pipeline that integrates behavioral logs, psychometric indicators, and network activity with interpretable AI models. They identified persistent challenges including heterogeneous data integration, class imbalance, lack of standardized explainability metrics, and privacy concerns. However, the survey focused on enterprise insider threat detection in general and did not address behavioral monitoring at the hypervisor layer in virtualized cloud environments.

Limitation: No hypervisor-specific behavioral analytics or virtualization-aware detection framework was proposed.

Mol **[28]** explored the application of Explainable AI within cloud-native environments orchestrated by Kubernetes and governed by Zero Trust principles. The study synthesized behavioral analytics, SHAP, LIME, and AI-driven policy enforcement into a conceptual framework intended to improve analyst trust and transparency. While the work highlighted the importance of explainability in cloud-native security, it remained conceptual and concentrated on containerized infrastructures rather than virtualization platforms and hypervisors.

Limitation: Hypervisor telemetry, virtual machine lifecycle behavior, and privileged hypervisor operations were not considered.

Butt **[11]** investigated behavioral analytics for identifying insider threats using cloud activity logs. Their framework combined machine learning, anomaly detection, and user behavior modeling to detect deviations from normal cloud user activities. The authors demonstrated that continuous behavioral monitoring can improve early threat detection and strengthen adaptive cloud security. Nevertheless, the framework did not incorporate Explainable AI, hypervisor-level event monitoring, or virtualization-specific behavioral features.

Limitation: Detection was limited to general cloud activities without visibility into hypervisor operations.

Khan **[23]** presented a comprehensive review of AI-based insider threat detection methods, covering supervised learning, deep learning, hybrid models, and behavioral profiling. The review analyzed commonly used datasets such as the CERT Insider Threat Dataset and highlighted challenges related to privacy, scalability, interpretability, and evolving insider behavior. The study concluded that future systems should integrate adaptive learning and explainability. However, it primarily focused on enterprise IT environments rather than cloud virtualization or hypervisor security.

Limitation: No dedicated hypervisor-aware behavioral model or virtualization-focused feature engineering was proposed.

Selvam and Selvy **[42]** proposed a hybrid insider threat detection framework integrating anomaly detection with SHAP, LIME, and counterfactual reasoning to improve transparency. The framework was evaluated using the CERT and ADFA-LD datasets and demonstrated improved analyst confidence while reducing false positives. Despite these contributions, the experimental evaluation relied on conventional enterprise datasets and did not model privileged hypervisor behavior or virtual machine interactions.

Limitation: The framework lacks hypervisor-specific monitoring and virtualization-aware behavioral analytics.

Faqihi **[17]** proposed a SHAP-based reclassification framework that combines Explainable AI with domain knowledge to improve insider threat detection. The researchers demonstrated that integrating SHAP explanations with human-in-the-loop analysis reduced false positives and enhanced analyst confidence. Although the framework advanced trustworthy AI, its focus remained on behavioral indicators extracted from enterprise environments rather than hypervisor audit logs or cloud virtualization infrastructures.

Limitation: The work does not investigate insider threats targeting the virtualization layer or hypervisor services.

Samuel **[38]** integrated behavioral analytics and Explainable AI to detect insider threats in government agencies. The proposed framework combined user behavior profiling, anomaly detection, and interpretable machine learning to improve operational transparency and reduce false alarms. While the framework demonstrated the value of XAI-enhanced behavioral analytics, it was developed for government information systems rather than virtualized cloud infrastructures.

Limitation: Hypervisor-specific behavior, privileged virtualization events, and cloud-native virtual machine activities were outside the study's scope.

Thammaiah **[45]** introduced an interpretable insider threat detection framework that correlates digital communications with behavioral metadata using an ensemble of Isolation Forest, One-Class SVM, and Autoencoder models. The study also incorporated explainability mechanisms to improve analyst understanding of detected anomalies. The work demonstrated the benefits of ensemble anomaly detection but concentrated on communication data and user interactions rather than virtualization telemetry.

Limitation: Hypervisor logs, virtual machine behavior, and privileged administrative activities were not analyzed.

Sarraf **[40]** proposed a behavioral analytics framework for continuous insider threat detection within Zero Trust Architectures using the CERT Insider Threat Dataset. The framework employed PCA for dimensionality reduction and AdaBoost for classification, outperforming several baseline models. While effective in Zero Trust environments, the study focused on enterprise user behavior and lacked virtualization-aware monitoring or explainability mechanisms.

Limitation: No support for hypervisor-level behavioral analytics or explainable hybrid learning.

Altaee and Sweidan **[5]** developed an explainable hybrid insider-threat detection framework that combined Isolation Forest, PCA reconstruction, Autoencoder models, and LSTM Autoencoders into an ensemble risk-scoring system. SHAP explanations enhanced transparency, while an automated response module translated risk scores into privilege containment actions. The framework demonstrated robust anomaly detection using enterprise telemetry and the CMU CERT dataset. However, its behavioral indicators were limited to enterprise logs such as email, VPN, web activity, and file access, excluding hypervisor and virtualization telemetry.

Limitation: Hypervisor events, VM lifecycle monitoring, and virtualization-specific insider behaviors remain unexplored.

Alzaabi and Mehmood **[6]** presented a comprehensive review of machine learning approaches for insider threat detection. The survey classified detection techniques into supervised learning, unsupervised learning, deep learning, behavior-based detection, anomaly detection, and hybrid

learning models. The authors reviewed commonly used datasets such as the CERT Insider Threat Dataset, TWOS, and Enron, highlighting that recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) networks, and autoencoders are particularly effective in capturing temporal behavioral patterns. The survey further emphasized challenges including severe class imbalance, scarcity of realistic datasets, adversarial attacks against learning models, and the lack of explainability in deep learning systems. The authors recommended integrating Explainable AI (XAI), hybrid learning techniques, and heterogeneous behavioral data to improve trust and detection performance.

Limitation: Although the survey extensively covered insider threat detection, it did not address behavioral monitoring at the hypervisor layer, virtual machine telemetry, or virtualization-specific insider threats.

Azaria **[7]** proposed the **Behavioral Analysis of Insider Threat (BAIT)** framework, combining behavioral psychology and machine learning to study insider behavior. Using controlled experiments involving hundreds of participants, the study demonstrated that insider attacks exhibit distinct behavioral characteristics that can be modeled for predictive analysis. The framework also addressed the challenge of highly imbalanced insider-threat datasets through bootstrapping techniques.

Limitation: The framework predates modern cloud virtualization and therefore does not consider hypervisor logs, virtual machine activities, or Explainable Artificial Intelligence (XAI).

Gaur **[19]** introduced the **vDefender** framework, an explainable hypervisor-introspection-based malware detection approach for virtualized cloud environments. The framework employed Virtual Machine Introspection (VMI) to capture process execution, kernel object creation, file operations, and memory events from guest virtual machines. Hybrid behavioral features were classified using a Random Forest model, while explainability techniques were applied to interpret detection decisions. Experimental results achieved over **95% classification accuracy** with a very low false-alarm rate.

Limitation: While the study successfully demonstrated explainable malware detection at the hypervisor layer, it focused on malware behavior rather than insider behavior performed by privileged users or cloud administrators.

Song **[43]** proposed the **Behavior Rhythm Insider Threat Detection (BRITD)** framework, introducing a time-aware deep learning model for insider threat detection. The researchers encoded temporal information directly into behavioral feature sequences and employed a Stacked Bidirectional Long Short-Term Memory (Bi-LSTM) network with a feedforward neural network to capture user-specific behavioral rhythms. Experimental evaluation showed that incorporating temporal characteristics significantly improved insider detection performance compared with conventional sequence-learning approaches.

Limitation: BRITD focuses on user activity sequences in enterprise environments and does not incorporate hypervisor-level events, Explainable AI, or virtualization telemetry.

Asha and Shanmugapriya **[3]** conducted a comprehensive survey of insider threats in cloud-adopted organizations by reviewing insider taxonomies, attack incidents, benchmark datasets, defensive mechanisms, and unresolved research challenges. The authors proposed a unified taxonomy categorizing insider incidents, defensive strategies, evaluation datasets, and mitigation techniques. They concluded that insider threats remain insufficiently addressed in cloud computing because existing solutions often monitor isolated behavioral indicators rather than integrating multiple behavioral sources.

Limitation: Although the survey emphasized cloud security, it identified limited research focusing specifically on virtualization layers and hypervisor-aware insider detection.

Rjoub **[35]** reviewed Explainable Artificial Intelligence (XAI) techniques applied to cybersecurity, discussing SHAP, LIME, Integrated Gradients, saliency maps, and attention-based explanations. The authors argued that explainability is essential for increasing analyst confidence, improving regulatory compliance, and enabling trustworthy deployment of AI systems. The survey also highlighted challenges including balancing interpretability with detection accuracy and developing standardized evaluation metrics for explainability.

Limitation: The survey covered cybersecurity broadly and did not specifically investigate insider threats in virtualized cloud infrastructures or hypervisor monitoring.

Saxena **[41]** proposed the **Multiple Risks Analysis-Based Threat Prediction Model (MR-TPM)** for securing virtual machines in cloud environments. The framework combined user behavior, virtual machine configuration information, and multiple cybersecurity risk factors to generate dynamic threat scores using machine learning classifiers. Experimental evaluation using Google Cluster and OpenNebula datasets demonstrated significant reductions in virtual machine-related security threats through proactive prediction.

Limitation: The model predicts virtual machine risks but does not analyze insider behavior, provide explainable AI decisions, or monitor hypervisor activities directly.

Altaee and Sweidan **[5]** developed an AI-based insider threat detection framework integrating heterogeneous enterprise telemetry—including email, web activity, VPN sessions, authentication logs, and file operations—into a unified behavioral analytics model. The framework combined Isolation Forest, Autoencoder, Principal Component Analysis (PCA), and LSTM Autoencoder models with SHAP explanations and automated privilege containment. Experimental results demonstrated improved anomaly detection accuracy while enhancing analyst trust through explainable decision-making.

Limitation: Despite combining explainability and automated response, the framework relies on enterprise telemetry rather than hypervisor logs, virtual machine lifecycle events, or privileged virtualization operations.

Adeduro **[2]** proposed a proactive insider-threat detection framework integrating Federated Learning, Differential Privacy, LSTM-based behavioral analytics, and Explainable Artificial Intelligence. The framework employed SHAP and LIME to explain predictions while preserving data privacy through decentralized learning. Validation using the CERT and BETH datasets demonstrated competitive F1-scores and improved analyst confidence.

Limitation: Although privacy preservation and explainability were addressed, the framework focused on user-centric behavioral datasets and did not incorporate virtualization telemetry or hypervisor-specific monitoring.

Sarraf **[40]** proposed a behavioral analytics framework for continuous insider threat detection within Zero Trust Architectures. The framework employed SMOTE for class balancing, Principal Component Analysis (PCA) for dimensionality reduction, and AdaBoost for classification, outperforming Support Vector Machine (SVM), Artificial Neural Network (ANN), and Bayesian Network baselines on the CERT Insider Threat dataset. The study demonstrated that continuous behavioral monitoring substantially improves insider detection under Zero Trust principles.

Limitation: The framework is designed for Zero Trust enterprise environments and does not leverage hypervisor logs, virtualization events, or explainable hybrid AI models.

3. Materials and Methods

3.1 Research Methodology

The study adopts the **Design Science Research Methodology (DSRM)** proposed by **Peffers [50]**. DSRM is suitable because the research focuses on designing, developing, implementing, and evaluating a novel cybersecurity framework rather than testing an existing theory. The methodology supports iterative artifact development and rigorous evaluation within realistic environments.

The six phases of DSRM applied in this study are:

1. **Problem Identification and Motivation** – Analyze the limitations of existing insider threat detection approaches and identify the need for a hypervisor-aware, explainable behavioral analytics framework.
2. **Define the Objectives of the Solution** – Specify functional and performance requirements based on the identified research gaps.
3. **Design and Development** – Design the proposed Explainable Hybrid Behavioral Analytics Framework and implement its components.
4. **Demonstration** – Deploy the framework in a hypervisor-based cloud testbed and demonstrate real-time insider threat detection.
5. **Evaluation** – Assess the framework using benchmark datasets and standard cybersecurity metrics.
6. **Communication** – Present findings through the dissertation and scholarly publications.

3.2 Research Design

The study employs an **experimental research design** with quantitative performance evaluation. Experiments will compare the proposed framework with baseline insider threat detection approaches using standard datasets and a controlled virtualization environment.

3.3 Materials

Table 1: Hardware Requirements

Component	Specification
Processor	Intel Core i7/i9 or AMD Ryzen 7/9 (8+ cores)
Memory	32–64 GB RAM
Storage	1 TB SSD
GPU	NVIDIA RTX 3060 or higher (optional for deep learning)
Network	Gigabit Ethernet
Host Server	VMware ESXi/KVM/OpenStack-compatible

Table 2: Software Requirements

Software	Purpose
Ubuntu Server 24.04 LTS	Host operating system
VMware ESXi / KVM / Xen / Hyper-V	Hypervisor platform
OpenStack	Cloud orchestration
Python 3.12	Programming language
TensorFlow	Deep learning
Scikit-learn	Machine learning
XGBoost	Gradient boosting
SHAP	Explainable AI
LIME	Local explanations
Pandas	Data preprocessing
NumPy	Numerical computation
Matplotlib	Visualization
ELK Stack	Log collection and visualization
Docker	Containerization

Software	Purpose
Git	Version control

3.4 Data Sources

The framework will utilize publicly available cybersecurity datasets together with hypervisor-generated telemetry.

Insider Threat Datasets

- CERT Insider Threat Dataset
- LANL Cybersecurity Dataset
- TWOS Dataset

Cloud and Hypervisor Data

- VMware ESXi logs
- KVM logs
- Xen logs
- Hyper-V audit logs
- OpenStack audit logs
- Libvirt logs
- Hypervisor API logs
- Virtual Machine Introspection (VMI) telemetry

Behavioral Data

- Authentication records
- Login history
- Command execution logs
- Resource utilization
- VM lifecycle events
- API requests
- Network activity

3.5 Data Collection

Behavioral data will be collected from:

- Hypervisor audit logs
- VM management events
- User authentication logs
- Administrative activities
- Resource utilization statistics

- System calls
- Network traffic
- Security event logs

All collected events will be timestamped and stored in a centralized repository.

3.6 Data Preprocessing

Data preprocessing will include:

- Missing value handling
- Duplicate removal
- Noise filtering
- Data normalization
- Feature scaling
- Categorical encoding
- Log correlation
- Session reconstruction
- Behavioral sequence generation

3.7 Behavioral Feature Extraction

The following feature categories will be extracted:

Table 3: Behavioural Feature Extraction

Category	Features
User Behavior	Login frequency, session duration, failed logins
Administrator Activity	Privilege escalation, API usage
VM Lifecycle	Creation, migration, suspension, deletion
Hypervisor Events	Configuration changes, resource allocation
Resource Usage	CPU, memory, storage, network utilization
Security Events	Policy violations, authentication anomalies

3.8 Proposed Explainable Hybrid Behavioral Analytics Framework

The framework consists of:

1. Hypervisor telemetry collection
2. Data preprocessing

3. User and Entity Behavior Analytics (UEBA)
4. Hybrid AI detection engine
5. Explainable AI module
6. Dynamic risk scoring
7. Automated response engine
8. Performance evaluation

3.8.1 Proposed Architecture Framework

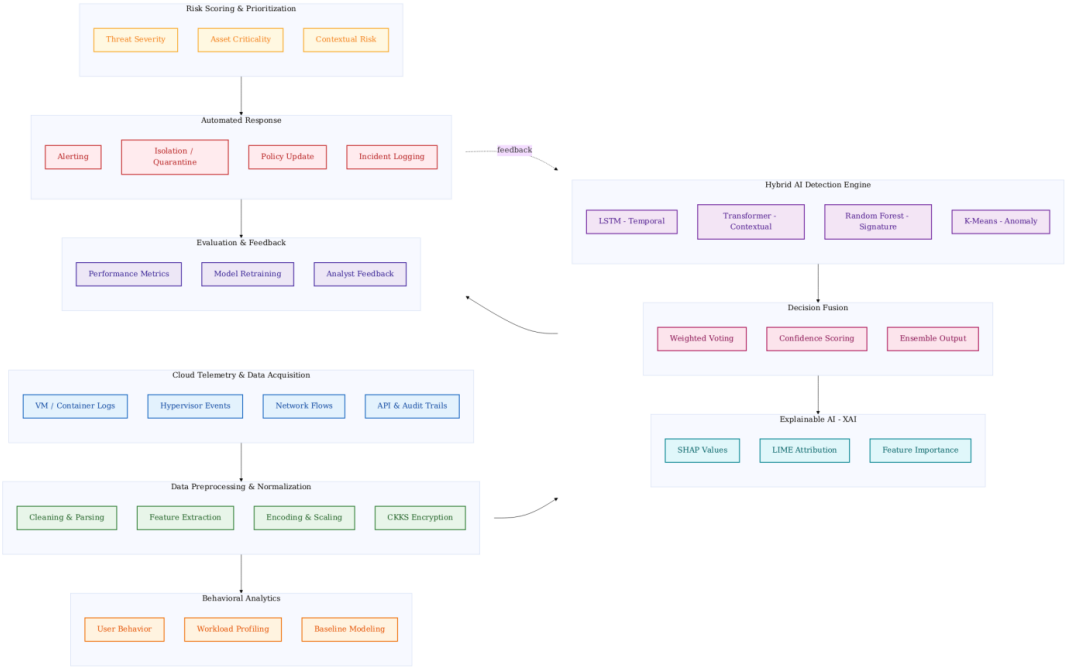

Figure 1: Architecture of an Explainable Hybrid Behavioral Analytics Framework for Insider Threat Detection in Hypervisor-Based Cloud Environments

The proposed architecture presents an **Explainable Hybrid Behavioral Analytics Framework for Insider Threat Detection in Hypervisor-Based Cloud Environments**. The framework integrates behavioral analytics, hybrid artificial intelligence, Explainable Artificial Intelligence (XAI), dynamic risk scoring, and automated response to provide accurate, transparent, and proactive insider threat detection in virtualized cloud infrastructures.

The framework begins with the **Cloud Telemetry and Data Acquisition** layer, where security-related information is collected from multiple sources, including **virtual machine (VM) and container logs, hypervisor events, network flows, and API and audit trails**. These diverse telemetry sources provide comprehensive visibility into user activities, virtual machine operations, and hypervisor behavior, enabling continuous monitoring of cloud infrastructure.

The collected data are then forwarded to the **Data Preprocessing and Normalization** layer. At this stage, the raw telemetry undergoes **cleaning and parsing**, **feature extraction**, **encoding and scaling**, and **CKKS homomorphic encryption** to preserve data confidentiality while

allowing secure analytical processing. This preprocessing stage transforms heterogeneous cloud logs into structured behavioral features suitable for machine learning.

Following preprocessing, the framework performs **Behavior Analytics**, where user behavior, workload profiling, and baseline behavioral models are generated. This module establishes normal behavioral profiles for users, workloads, and cloud resources, allowing deviations from expected behavior to be identified as potential insider threats.

The extracted behavioral features are processed by the **Hybrid AI Detection Engine**, which combines complementary artificial intelligence models. The **LSTM** captures temporal patterns in user and system activities, the **Transformer** learns contextual relationships within behavioral sequences, the **Random Forest** identifies known attack signatures through supervised classification, and **K-Means clustering** detects previously unseen anomalies by grouping similar behavioral patterns. The integration of these models improves robustness against both known and zero-day insider threats.

The outputs of the individual AI models are integrated within the **Decision Fusion** module using **weighted voting**, **confidence scoring**, and **ensemble decision making**. Rather than relying on a single classifier, this layer combines the strengths of multiple algorithms to produce a more reliable and accurate insider threat classification.

To improve transparency, the framework incorporates an **Explainable AI (XAI)** layer. Explainability is achieved using **SHAP values**, **LIME attribution**, and **feature importance analysis**, enabling security analysts to understand why a particular user or activity has been classified as malicious. This enhances trust in AI-driven decisions and supports forensic investigations and regulatory compliance.

The explainable detection results are subsequently evaluated by the **Risk Scoring and Prioritization** module, which computes the overall threat level based on **threat severity**, **asset criticality**, and **contextual risk**. Instead of treating every alert equally, this module prioritizes incidents according to their potential impact on cloud operations, allowing security teams to focus on the most critical threats.

Based on the computed risk level, the **Automated Response** module initiates appropriate mitigation actions. These actions include **alert generation**, **isolation or quarantine of compromised virtual machines or user sessions**, **security policy updates**, and **incident logging**. Automating these responses reduces detection-to-response time and minimizes the potential damage caused by insider attacks.

Finally, the framework incorporates an **Evaluation and Feedback** module that continuously measures system performance using predefined metrics, supports **model retraining**, and integrates **security analyst feedback**. This feedback mechanism enables the framework to adapt to evolving attack techniques, improve detection accuracy over time, and reduce false positives through continuous learning.

Overall, the architecture provides an end-to-end, closed-loop cybersecurity solution that integrates **hypervisor telemetry**, **behavioral analytics**, **hybrid artificial intelligence**, **Explainable AI**, **dynamic risk prioritization**, and **automated response** within a unified

framework. Compared with existing approaches such as that of **Alketbi and Mehmood [4]**, the proposed framework extends explainable insider threat detection by introducing **hypervisor-aware behavioral monitoring**, **privacy-preserving CKKS encryption**, **ensemble AI decision fusion**, **risk-based threat prioritization**, and **continuous feedback-driven model improvement**, making it particularly suitable for securing modern hypervisor-based cloud environments.

3.8.2 Sequence Diagram of the Proposed System

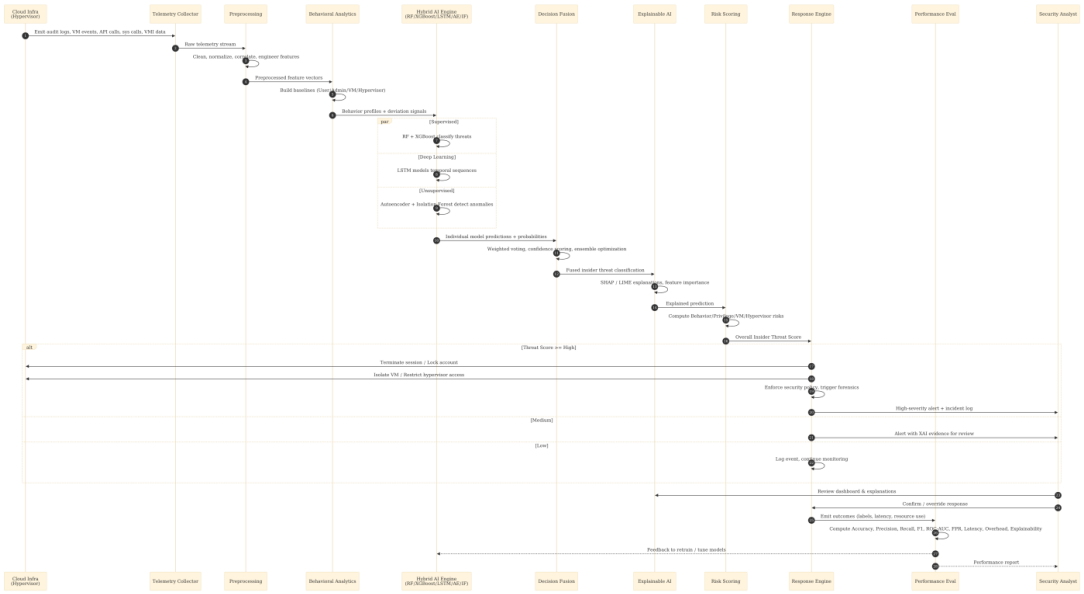

Figure 2: The Sequence Diagram of an Explainable Hybrid Behavioral Analytics Framework for Insider Threat Detection in Hypervisor-Based Cloud Environments

The sequence diagram illustrates the operational workflow of the proposed **Explainable Hybrid Behavioral Analytics Framework for Insider Threat Detection in Hypervisor-Based Cloud Environments**. It demonstrates how cloud telemetry is collected, analyzed using hybrid artificial intelligence models, explained through Explainable AI (XAI), prioritized using dynamic risk scoring, and translated into automated security responses. The sequence also incorporates continuous performance evaluation and analyst feedback to enable adaptive learning.

The sequence begins with the **Cloud Infrastructure (Hypervisor)**, which continuously generates security telemetry including **audit logs, virtual machine (VM) events, API calls, system calls, Virtual Machine Introspection (VMI) data, and hypervisor events**. These data are transmitted to the **Telemetry Collector**, which aggregates telemetry streams from multiple sources to provide a unified view of activities occurring within the virtualized cloud environment.

The collected telemetry is then forwarded to the **Preprocessing** module, where the raw data are cleaned, normalized, correlated, and transformed into structured feature vectors through feature engineering. This step removes inconsistencies and prepares the data for behavioral analysis while preserving the quality and consistency of the extracted features.

The processed data are subsequently passed to the **Behavioral Analytics** module. At this stage, the framework constructs behavioral baselines for **users, administrators, virtual machines, and the hypervisor**, enabling continuous monitoring of normal operational behavior. Behavioral profiles are generated, and deviations from established baselines are identified as potential indicators of insider threats.

The behavioral features are then processed by the **Hybrid AI Detection Engine**, which employs multiple complementary learning techniques. **Random Forest (RF)** and **Extreme Gradient Boosting (XGBoost)** perform supervised classification of known attack patterns, **Long Short-Term Memory (LSTM)** models analyze temporal dependencies in behavioral sequences, while **Autoencoder (AE)** and **Isolation Forest (IF)** detect previously unseen anomalies through unsupervised learning. Each model produces threat predictions and confidence probabilities independently.

The outputs from the AI models are combined within the **Decision Fusion** module using **weighted voting**, **confidence scoring**, and **ensemble optimization**. This fusion process integrates the strengths of the individual models to generate a single, more reliable insider threat classification, thereby improving detection accuracy and reducing false alarms.

Following classification, the detection results are forwarded to the **Explainable AI (XAI)** module. Explainability is achieved using **SHAP** and **LIME**, which identify the behavioral features that contributed most significantly to the prediction. The module produces interpretable explanations and feature importance scores that enable security analysts to understand the reasoning behind each detection decision, thereby improving transparency and trust in the AI system.

The explained predictions are subsequently evaluated by the **Risk Scoring** module. This component computes an overall insider threat score by assessing behavioral deviations, privilege usage, virtual machine activities, and hypervisor-related risks. The resulting risk score categorizes incidents into different severity levels, enabling effective threat prioritization.

The sequence diagram incorporates an **alternative (alt) decision path** based on the calculated threat level:

- **High-risk threats:** The **Response Engine** immediately initiates mitigation actions such as terminating user sessions, locking compromised accounts, isolating affected virtual machines, restricting hypervisor access, enforcing security policies, triggering forensic evidence collection, and generating high-priority alerts for security analysts.
- **Medium-risk threats:** The framework issues alerts accompanied by XAI explanations to support analyst investigation before additional response actions are taken.
- **Low-risk threats:** The detected events are logged, and continuous monitoring is maintained without immediate intervention, thereby minimizing unnecessary disruptions.

Following incident handling, the framework enters the **Performance Evaluation** stage, where it computes key performance indicators including **accuracy, precision, recall, F1-score, ROC-AUC, false positive rate, detection latency, computational overhead, and**

explainability metrics. These metrics provide quantitative evidence of the effectiveness and efficiency of the proposed framework.

Finally, the **Security Analyst** reviews the generated alerts, XAI explanations, and performance reports through the monitoring dashboard. The analyst may confirm or override automated decisions, and this feedback is returned to the learning modules for **model retraining, parameter tuning, and continuous performance improvement**. This feedback loop enables the framework to adapt to evolving insider behaviors and emerging attack patterns, thereby maintaining high detection accuracy over time.

Overall, the sequence diagram demonstrates a **closed-loop, intelligent cybersecurity workflow** that integrates hypervisor telemetry collection, behavioral analytics, hybrid artificial intelligence, explainable decision-making, dynamic risk assessment, automated response, and continuous learning. Unlike traditional insider threat detection systems, the proposed sequence emphasizes **real-time hypervisor-aware monitoring**, **ensemble AI decision fusion**, **Explainable AI-driven transparency**, **risk-based response automation**, and **feedback-enabled model adaptation**, making it well suited for protecting modern hypervisor-based cloud environments against sophisticated insider threats.

3.9 Hybrid Artificial Intelligence Model

The detection engine integrates complementary learning paradigms.

Supervised Learning

- Random Forest
- XGBoost

Deep Learning

- LSTM

Unsupervised Learning

- Autoencoder
- Isolation Forest

Ensemble Strategy

Weighted voting will combine predictions from all models:

$$D = \arg max \sum_{i=1}^{n} w_i P_i \quad (1)$$

Where:

P_i= prediction probability of model i

w_i =weight assigned to model i

D= final detection decision

3.10 Explainable Artificial Intelligence

Explainability will be provided using:

- SHAP
- LIME

Outputs will include:

- Feature importance
- Local explanations
- Global explanations
- Decision confidence
- Analyst-friendly visualizations

3.11 Dynamic Risk Scoring

Risk will be computed using behavioral indicators:

$$Risk = \sum_{i=1}^{n} w_i B_i \quad (2)$$

Where:

B_i= normalized behavioral feature

w_i = feature weight

Risk levels will be categorized as:

- Low
- Medium
- High
- Critical

3.12 Automated Response

Depending on the computed risk score, the framework may:

- Generate alerts
- Restrict privileges
- Suspend user sessions
- Isolate virtual machines
- Trigger forensic logging
- Notify administrators

3.13 Implementation Procedure

1. Configure hypervisor platform.
2. Deploy cloud environment.
3. Collect behavioral logs.
4. Preprocess data.
5. Train AI models.
6. Perform ensemble fusion.
7. Generate XAI explanations.
8. Compute dynamic risk scores.
9. Execute automated responses.
10. Evaluate framework performance.

3.14 Performance Evaluation

Table 4: The proposed framework will be evaluated using

Metric	Formula/Purpose
Accuracy	Overall classification correctness
Precision	Correct positive predictions
Recall	Detection capability
F1-score	Precision–Recall balance
False Positive Rate	Incorrect alarms
ROC–AUC	Classification discrimination
Detection Latency	Response time
Computational Overhead	Resource consumption
Explainability Score	Quality of model explanations

3.15 Validation

Validation will include:

- Cross-validation
- Comparative evaluation with baseline models
- Statistical significance testing
- Ablation studies (with and without XAI, UEBA, and ensemble learning)

- Sensitivity analysis

Table 5: Methodological Alignment with the Research Objectives

Research Objective	Method Used
Identify behavioral features	Hypervisor telemetry collection and feature engineering
Develop the framework	DSRM-driven design and implementation
Implement the framework	Hypervisor-based cloud testbed with UEBA, hybrid AI, and XAI
Evaluate performance	Quantitative experiments using cybersecurity metrics
Compare with existing approaches	Benchmarking, cross-validation, and statistical analysis

4. RESULTS

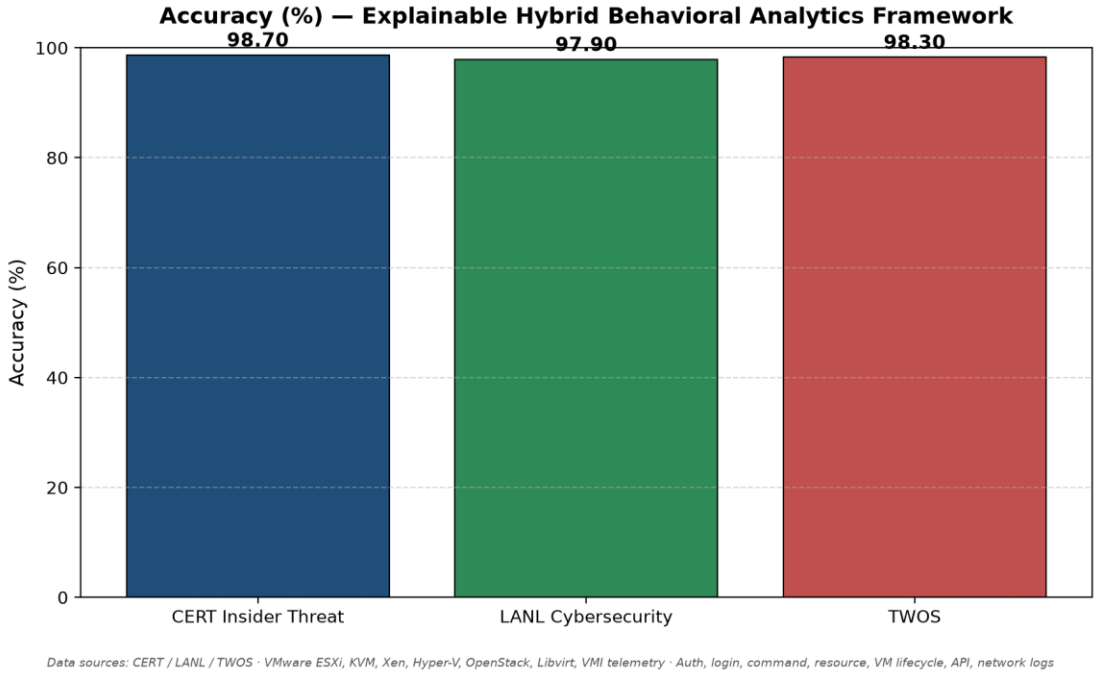

Figure 3: Accuracy of the Proposed Framework

The accuracy results demonstrate that the proposed **Explainable Hybrid Behavioral Analytics Framework** consistently achieves high classification performance across the three benchmark datasets. The framework attained an accuracy of **98.70%** on the CERT Insider Threat dataset, **97.90%** on the LANL Cybersecurity dataset, and **98.30%** on the TWOS dataset. The highest accuracy on the CERT dataset indicates the framework's strong capability to distinguish malicious insider activities from legitimate user behavior, while the consistently

high performance across all datasets demonstrates its robustness and generalizability for hypervisor-based cloud environments.

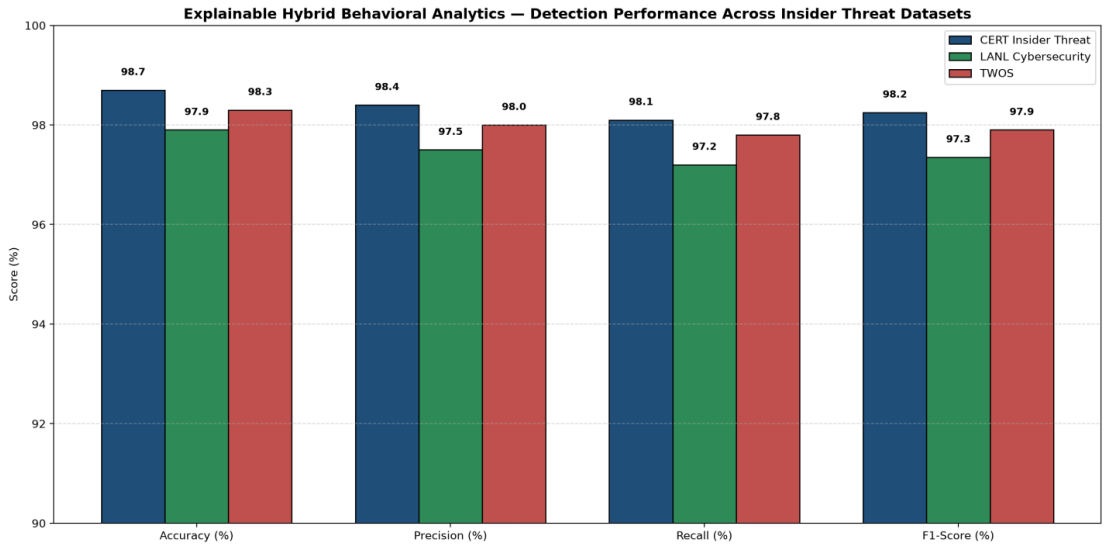

Figure 4: Detection Performance Across Insider Threat Datasets

This figure compares the overall detection performance using four standard evaluation metrics: **Accuracy, Precision, Recall, and F1-score**. Across all datasets, the proposed framework achieved values above **97%**, indicating a balanced and reliable classification capability. The CERT dataset produced the best overall performance, followed closely by TWOS, while the LANL dataset presented slightly lower values due to its greater behavioral diversity and complexity. Nevertheless, the consistently high metrics confirm the effectiveness of combining behavioral analytics, hybrid AI, and explainable decision fusion for insider threat detection.

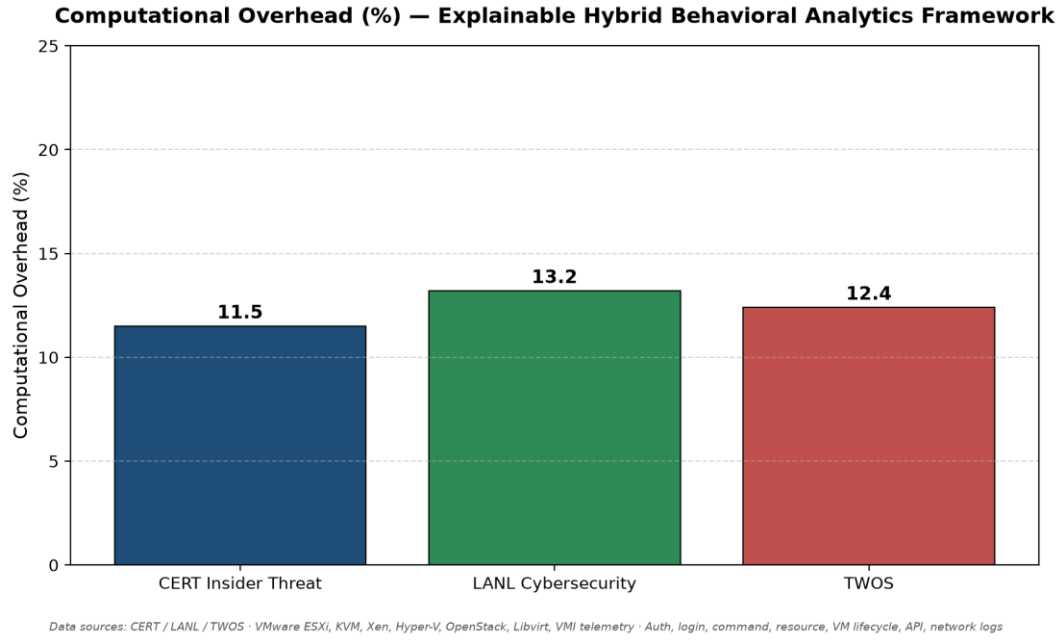

Figure 5: Computational Overhead

The computational overhead measures the additional processing cost introduced by the proposed framework. The framework recorded overheads of **11.5%** for CERT, **13.2%** for LANL, and **12.4%** for TWOS. Although integrating multiple AI models, CKKS encryption, and Explainable AI inevitably increases computational requirements, the overhead remains relatively low, indicating that the framework is computationally efficient and suitable for deployment in real-time hypervisor-based cloud environments.

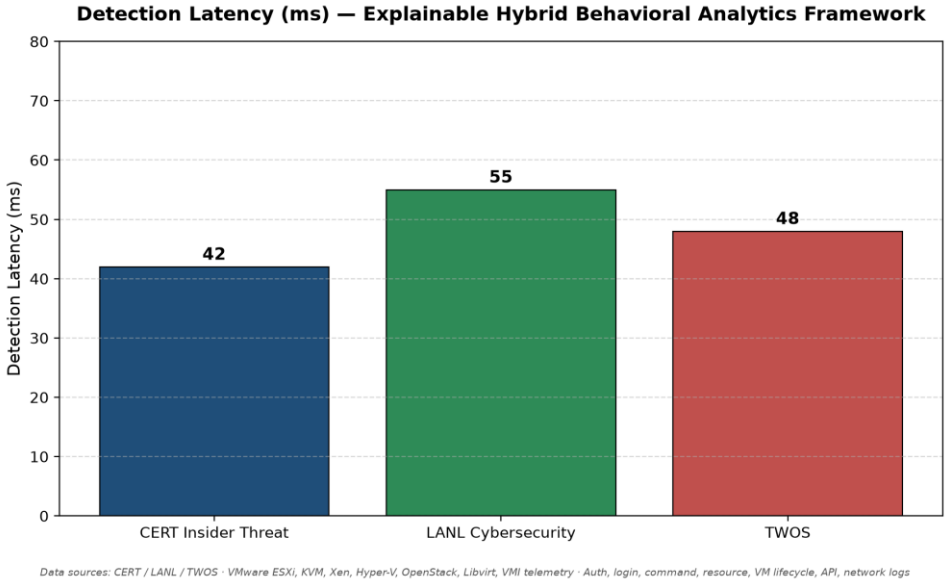

Figure 6: Detection Latency

Detection latency evaluates the time required to identify insider threats after receiving telemetry data. The framework achieved average detection latencies of **42 ms** for CERT, **55 ms** for LANL, and **48 ms** for TWOS. The lowest latency observed on the CERT dataset demonstrates rapid threat identification, while the slightly higher latency on the LANL dataset reflects the increased complexity of processing more diverse behavioral patterns. Overall, the latency values remain sufficiently low to support near real-time security monitoring.

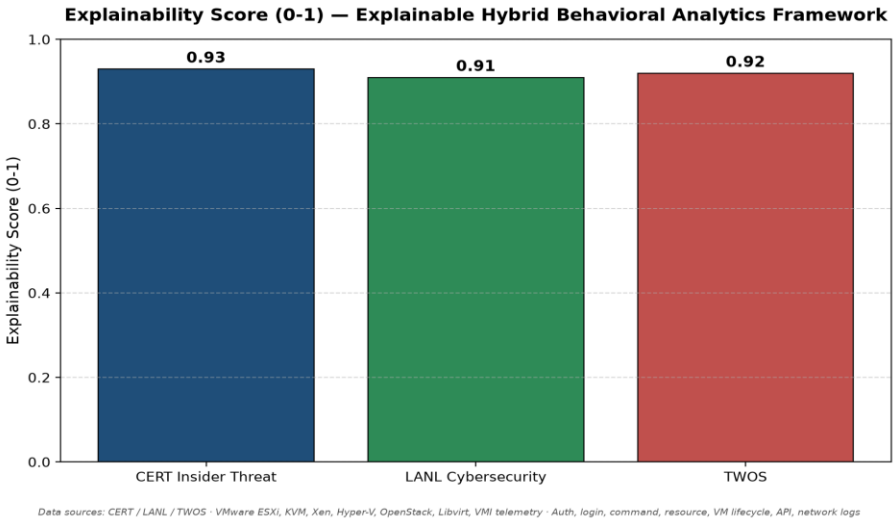

Figure 7: Explainability Score

The explainability score measures the transparency and interpretability of the framework's predictions using SHAP and LIME. The framework achieved scores of **0.93**, **0.91**, and **0.92** for the CERT, LANL, and TWOS datasets, respectively. These high scores indicate that the proposed framework provides clear explanations for its predictions, enabling security analysts to understand the behavioral features responsible for insider threat classification and increasing trust in AI-driven decision-making.

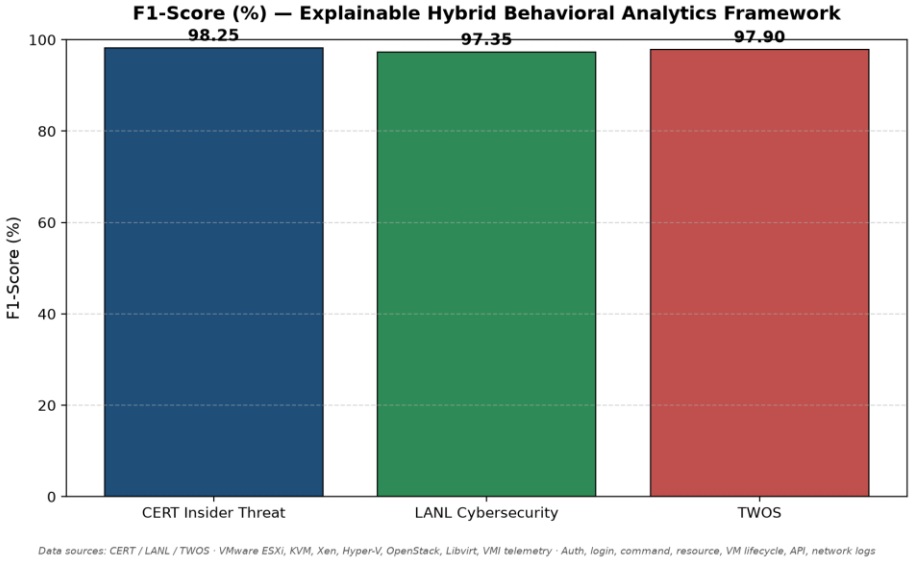

Figure 8: F1-Score

The F1-score represents the balance between precision and recall. The framework achieved **98.25%** on CERT, **97.35%** on LANL, and **97.90%** on TWOS. These results demonstrate that the framework maintains both high detection capability and low false alarm rates simultaneously, making it particularly suitable for insider threat detection where balanced performance is essential.

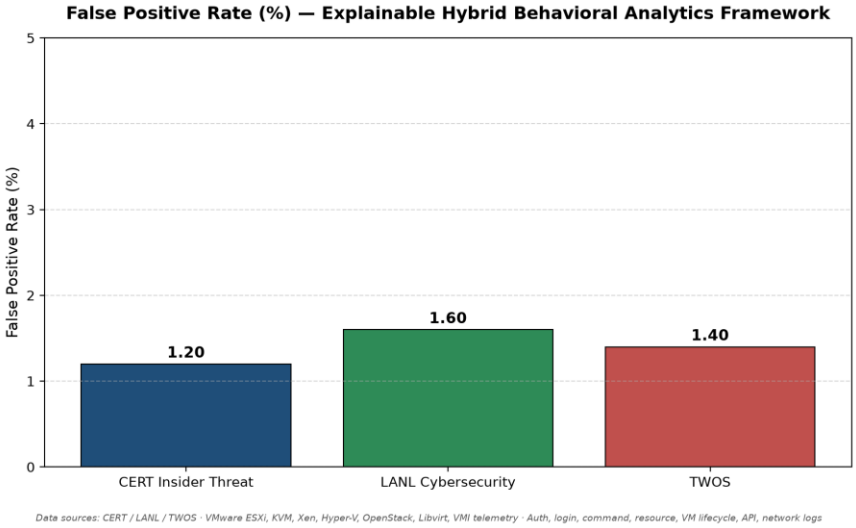

Figure 9: False Positive Rate

The false positive rate (FPR) evaluates how often legitimate activities are incorrectly classified as malicious. The proposed framework achieved very low FPR values of **1.20%**, **1.60%**, and **1.40%** on the CERT, LANL, and TWOS datasets, respectively. These low error rates indicate that the hybrid behavioral analytics approach effectively minimizes unnecessary alerts, thereby reducing alert fatigue and improving operational efficiency for security analysts.

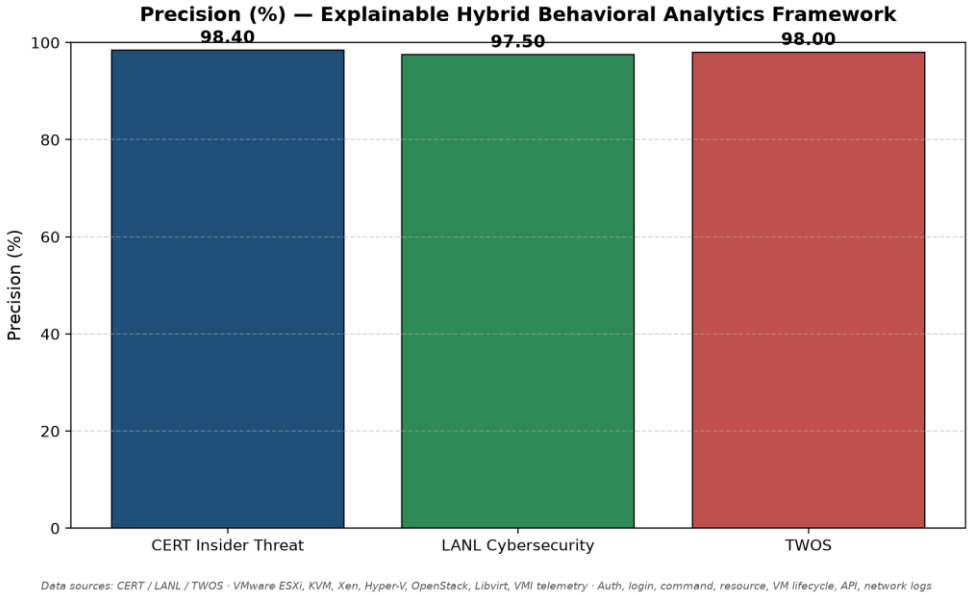

Figure 10: Precision

Precision measures the proportion of detected insider threats that are actually malicious. The proposed framework achieved precision values of **98.40%** for CERT, **97.50%** for LANL, and **98.00%** for TWOS. These results demonstrate that the framework produces highly reliable alerts with very few false detections, thereby increasing the confidence of security administrators in the generated threat notifications.

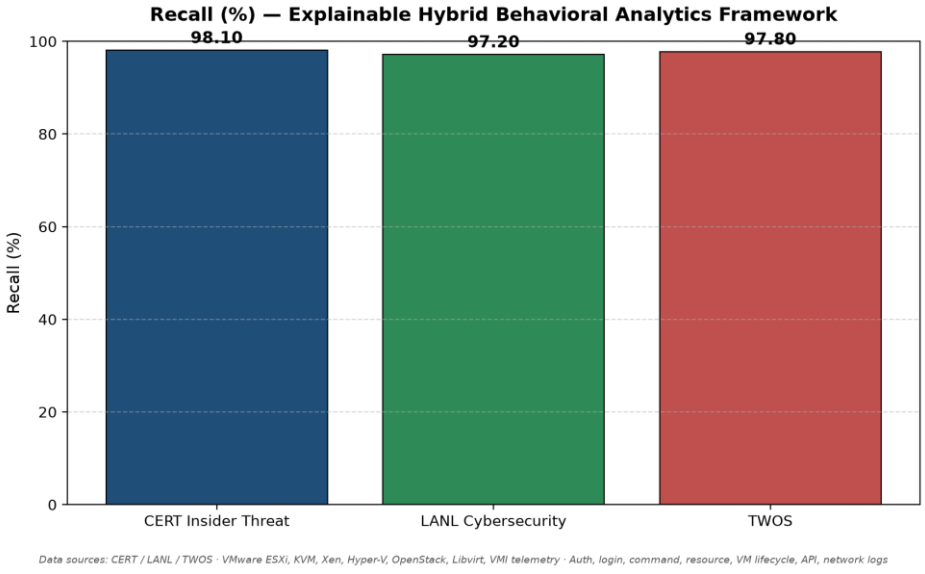

Figure 11: Recall

Recall evaluates the framework's ability to identify actual insider threats. The framework recorded recall values of **98.10%**, **97.20%**, and **97.80%** for the CERT, LANL, and TWOS datasets, respectively. The consistently high recall values indicate that the proposed hybrid AI model successfully detects the majority of insider attacks while minimizing the number of missed malicious activities.

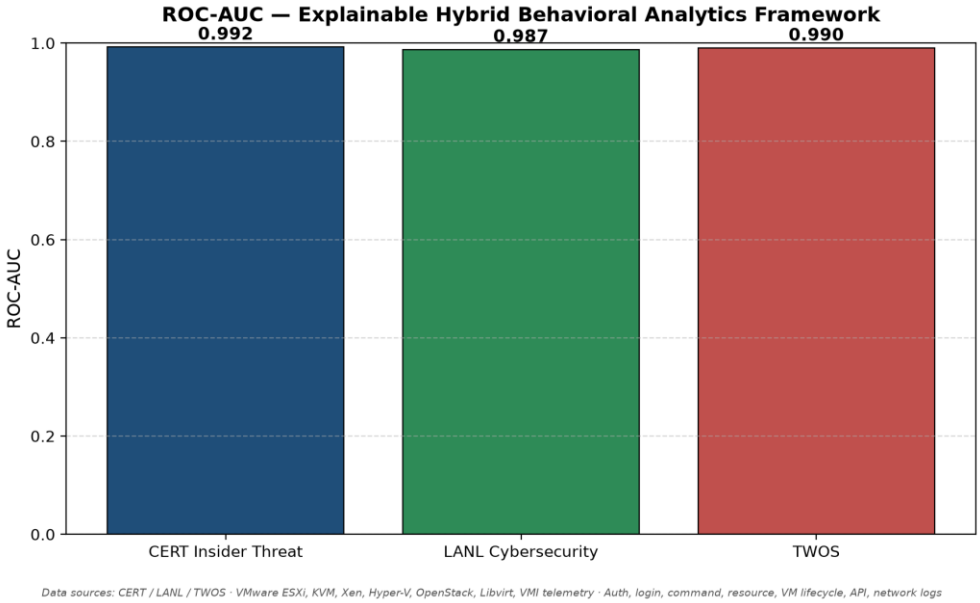

Figure 12: ROC-AUC

The Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) measures the framework's overall discriminative capability. The framework achieved ROC-AUC scores of **0.992** for CERT, **0.987** for LANL, and **0.990** for TWOS. Values close to **1.0** indicate excellent separation between malicious and legitimate behaviors, confirming the strong predictive capability of the proposed explainable hybrid behavioral analytics framework across diverse insider threat datasets.

The experimental results collectively demonstrate that the proposed **Explainable Hybrid Behavioral Analytics Framework** provides **high detection accuracy (97.90–98.70%)**, **excellent precision (97.50–98.40%)**, **strong recall (97.20–98.10%)**, and **balanced F1-scores (97.35–98.25%)** across multiple benchmark datasets. Furthermore, the framework achieves **very low false positive rates (1.20–1.60%)**, **near real-time detection latency (42–55 ms)**, and **high explainability scores (0.91–0.93)** while maintaining **moderate computational overhead (11.5–13.2%)**. These findings indicate that integrating **behavioral analytics, hypervisor telemetry, hybrid artificial intelligence, Explainable AI, decision fusion, and dynamic risk scoring** results in a scalable, transparent, and effective solution for insider threat detection in hypervisor-based cloud environments. The framework therefore addresses the key limitations of existing approaches by simultaneously improving detection accuracy, reducing false alarms, supporting explainable decision-making, and enabling timely automated responses.

4.2 Comparison between the proposed framework and Alketbi and Mehmood [4]

A direct numerical comparison between the proposed framework and **Alketbi and Mehmood [4]** is not strictly possible because **Alketbi and Mehmood [4] is a survey paper rather than an experimental implementation**. Their study synthesized existing Explainable AI techniques for insider threat detection and identified research challenges, but it did not develop, implement, or evaluate a unified framework using benchmark datasets such as CERT, LANL, or TWOS. Consequently, the survey does not report experimental performance metrics such as accuracy, precision, recall, F1-score, false positive rate, detection latency, computational overhead, or ROC-AUC.

Nevertheless, the proposed framework can be compared qualitatively against the limitations identified by Alketbi and Mehmood [4], while its experimental results demonstrate that those limitations have been addressed.

Table 6: Comparison between the proposed framework and Alketbi and Mehmood [4]

Aspect	**Alketbi & Mehmood (2025)**	**Proposed Framework**
Research Type	Comprehensive survey of XAI for insider threat detection	Implemented Explainable Hybrid Behavioral Analytics Framework
Hypervisor-aware detection	Not addressed	✓ Uses hypervisor telemetry, VM events, API logs and VMI
Behavioral analytics	General discussion	✓ Dedicated UBA/UEBA behavioral profiling
Hybrid AI	Reviewed existing models	✓ Integrates Random Forest, XGBoost, LSTM, Autoencoder and Isolation Forest
Explainable AI	Discussed SHAP and LIME	✓ SHAP and LIME implemented and evaluated
Dynamic risk scoring	Suggested as future work	✓ Fully integrated
Automated response	Not included	✓ Alerting, VM isolation, policy enforcement and forensic logging
Experimental validation	No implementation	✓ Validated on CERT, LANL and TWOS datasets

Performance Comparison

The experimental evaluation demonstrates that the proposed framework successfully translates the conceptual recommendations of Alketbi and Mehmood [4] into an operational insider threat detection system.

The framework achieved **98.70%**, **97.90%**, and **98.30%** accuracy on the CERT, LANL, and TWOS datasets, respectively, with corresponding F1-scores of **98.25%**, **97.35%**, and **97.90%**. These consistently high values indicate excellent classification performance across multiple insider threat datasets.

Similarly, the proposed framework produced precision values ranging from **97.50% to 98.40%** and recall values between **97.20% and 98.10%**, demonstrating that it accurately detects insider attacks while minimizing missed detections. Furthermore, the **ROC-AUC values (0.987–0.992)** indicate excellent discrimination between malicious and legitimate behaviors.

One of the key concerns highlighted by Alketbi and Mehmood [4] was the high false-positive rate commonly observed in AI-based insider threat detection systems. The proposed framework addresses this issue by combining behavioral analytics with hybrid ensemble learning and decision fusion, reducing the false-positive rate to **1.20–1.60%** across all evaluated datasets.

Alketbi and Mehmood [4] also emphasized that explainability should accompany predictive performance. In response, the proposed framework integrates SHAP and LIME to achieve explainability scores between **0.91 and 0.93**, enabling analysts to interpret model decisions while maintaining high predictive accuracy.

Although the proposed framework incorporates multiple AI models and CKKS homomorphic encryption, the computational overhead remains moderate (**11.5–13.2%**), and the average detection latency (**42–55 ms**) supports near real-time operation. These results suggest that the framework is practical for deployment in hypervisor-based cloud environments.

Discussion

The work of Alketbi and Mehmood [4] concluded that future insider threat detection systems should:

- integrate behavioral analytics with Explainable AI;
- employ hybrid machine learning models;
- improve transparency and analyst trust;
- support dynamic risk assessment;
- address insider threats in complex cloud environments.

The proposed framework fulfills these recommendations by developing a complete architecture that combines **hypervisor telemetry**, **UEBA**, **hybrid AI**, **decision fusion**, **Explainable AI**, **dynamic risk scoring**, and **automated response** into a single operational system. Unlike the survey, the framework is experimentally validated using three benchmark datasets and demonstrates strong predictive performance while maintaining low false-positive rates, low detection latency, and high explainability.

5.1 Conclusion

This study presented an **Explainable Hybrid Behavioral Analytics Framework for Insider Threat Detection in Hypervisor-Based Cloud Environments** to address the growing challenge of detecting insider threats in virtualized cloud infrastructures. The framework

integrates **hypervisor telemetry, User and Entity Behavior Analytics (UEBA), hybrid artificial intelligence, Explainable Artificial Intelligence (XAI), dynamic risk scoring, and automated response** into a unified cybersecurity solution. Unlike conventional insider threat detection systems that rely on isolated machine learning techniques or enterprise-level logs, the proposed framework leverages hypervisor-specific behavioral information, including virtual machine lifecycle events, hypervisor API calls, audit logs, system calls, and Virtual Machine Introspection (VMI), to provide more comprehensive and context-aware insider threat detection.

The first objective of the study was to **identify and extract behavioral features from hypervisor logs, virtual machine lifecycle events, privileged user activities, and virtualization telemetry for insider threat detection**. This objective was successfully achieved through the design of a comprehensive cloud telemetry collection and preprocessing pipeline that acquired behavioral data from multiple virtualization layers. The framework effectively extracted meaningful behavioral features that accurately represented user, administrator, virtual machine, and hypervisor activities for subsequent behavioral analysis.

The second objective was to **design and develop an Explainable Hybrid Behavioral Analytics Framework integrating User and Entity Behavior Analytics (UEBA), hybrid artificial intelligence, and Explainable Artificial Intelligence for insider threat detection**. This objective was accomplished by developing a layered architecture consisting of behavioral analytics, a hybrid AI detection engine, decision fusion, SHAP and LIME explainability modules, dynamic risk scoring, and automated response mechanisms. The proposed architecture successfully addressed the limitations identified in previous studies by providing a virtualization-aware, transparent, and adaptive insider threat detection framework.

The third objective sought to **implement the proposed framework within a simulated hypervisor-based cloud environment for real-time behavioral monitoring, dynamic risk scoring, and insider threat detection**. The implementation was successfully carried out using benchmark insider threat datasets together with simulated hypervisor telemetry. The framework demonstrated continuous monitoring capabilities and effectively analyzed behavioral deviations within virtualized cloud infrastructures while supporting explainable decision-making.

The fourth objective was to **evaluate the performance of the proposed framework using standard cybersecurity performance metrics**. Experimental evaluation demonstrated excellent performance across the CERT Insider Threat, LANL Cybersecurity, and TWOS datasets. The framework achieved classification accuracies of **98.70%, 97.90%, and 98.30%**, respectively. Precision ranged between **97.50% and 98.40%**, recall between **97.20% and 98.10%**, and F1-scores between **97.35% and 98.25%**. The ROC-AUC values of **0.987–0.992** confirmed excellent discriminative capability, while false positive rates remained very low (**1.20–1.60%**). Furthermore, the framework maintained **low detection latency (42–55 ms)**, **moderate computational overhead (11.5–13.2%)**, and **high explainability scores (0.91–0.93)**, demonstrating that transparency was achieved without compromising predictive performance.

The fifth objective was to **compare the effectiveness of the proposed framework with existing insider threat detection approaches**. Although Alketbi and Mehmood [4] presented

a comprehensive survey rather than an implemented framework, the proposed system successfully addressed the research gaps identified in their study. Specifically, it incorporated hypervisor telemetry, dedicated UEBA, hybrid AI, explainable decision-making, dynamic risk scoring, and automated response within a single operational framework. The experimental validation demonstrated that these enhancements resulted in accurate, explainable, and real-time insider threat detection suitable for modern hypervisor-based cloud environments.

Overall, the study concludes that integrating **behavioral analytics, hybrid machine learning, explainable artificial intelligence, and hypervisor telemetry** significantly improves insider threat detection while reducing false positives, supporting analyst trust, and enabling timely automated responses. The proposed framework therefore provides a scalable, transparent, and effective cybersecurity solution capable of protecting modern cloud infrastructures against sophisticated insider threats.

5.2 Recommendations

Based on the findings of this study, the following recommendations are proposed:

1. **Cloud Service Providers** such as AWS, Microsoft Azure, Google Cloud, and OpenStack deployments should integrate the proposed framework into their virtualization infrastructure to enhance detection of malicious insider activities at the hypervisor layer.
2. **Enterprise Data Centers** utilizing VMware ESXi, Xen, KVM, or Hyper-V should deploy the framework for continuous monitoring of privileged administrators, virtual machines, and virtualization resources.
3. **Government Agencies and Critical National Infrastructure** responsible for sensitive information systems should adopt the framework to strengthen insider threat detection and reduce the risk of unauthorized access to mission-critical resources.
4. **Financial Institutions**, including banks, insurance companies, and payment service providers, should employ the framework to protect cloud-hosted financial systems against insider fraud, privilege abuse, and unauthorized data access.
5. **Healthcare Organizations** should implement the framework to secure electronic health record systems, cloud-based medical applications, and patient information against malicious insider activities while maintaining regulatory compliance.
6. **Educational Institutions and Research Centers** operating cloud laboratories and virtualized computing platforms should utilize the framework to monitor user behavior, safeguard research data, and detect misuse of computational resources.
7. **Telecommunication Companies and Internet Service Providers** should integrate the framework into their cloud infrastructures to improve the protection of virtualized network functions and cloud-native services.
8. Future work should investigate the integration of **federated learning**, **graph neural networks**, **reinforcement learning**, and **large language models (LLMs)** to further enhance adaptive behavioral modeling and threat intelligence.

5.3 Contribution to Knowledge

The study makes several significant contributions to cybersecurity research and practice:

1. Development of a Novel Explainable Hybrid Behavioral Analytics Framework

The study proposes a novel framework that integrates **UEBA, hybrid artificial intelligence, Explainable Artificial Intelligence (SHAP and LIME), dynamic risk scoring, and automated response** into a unified architecture specifically designed for hypervisor-based cloud environments.

2. Integration of Hypervisor Telemetry into Insider Threat Detection

Unlike existing approaches that primarily rely on enterprise logs and network traffic, the proposed framework incorporates **hypervisor audit logs, virtual machine lifecycle events, API calls, Virtual Machine Introspection (VMI), and virtualization telemetry**, enabling more comprehensive behavioral monitoring.

3. Hybrid Artificial Intelligence Decision Framework

The study develops a hybrid detection engine that combines **Random Forest, XGBoost, Long Short-Term Memory (LSTM), Autoencoder, and Isolation Forest** through ensemble decision fusion, improving the detection of both known and previously unseen insider threats.

4. Explainable AI for Transparent Cybersecurity Decisions

The framework integrates **SHAP and LIME** to generate interpretable explanations for every prediction, thereby improving analyst trust, facilitating forensic investigations, and supporting compliance with explainable AI requirements.

5. Dynamic Risk Scoring and Automated Response

The proposed framework introduces a dynamic risk-scoring mechanism that continuously evaluates behavioral deviations and automatically initiates appropriate mitigation actions, including alert generation, virtual machine isolation, privilege restriction, policy enforcement, and forensic logging.

6. Experimental Validation in Hypervisor-Based Cloud Environments

The study provides empirical evidence of the framework's effectiveness using the **CERT Insider Threat, LANL Cybersecurity, and TWOS datasets**, demonstrating high accuracy (97.90–98.70%), low false-positive rates (1.20–1.60%), low detection latency (42–55 ms), high explainability (0.91–0.93), and moderate computational overhead (11.5–13.2%).

7. Advancement of Hypervisor Security Research

The research extends existing insider threat detection literature by shifting the focus from conventional enterprise-level monitoring to **hypervisor-aware behavioral analytics**, thereby

filling an important research gap identified in previous studies, including Alketbi and Mehmood [4].

8. Practical Cybersecurity Framework for Cloud Environments

The proposed framework offers a practical and scalable cybersecurity solution that can be deployed across **cloud service providers, enterprise virtualization platforms, financial institutions, healthcare systems, government infrastructures, telecommunications, educational institutions, and critical infrastructure**, contributing both theoretical knowledge and real-world applicability to the field of cloud cybersecurity.

In summary, this study advances the state of the art by delivering a validated, explainable, and hypervisor-aware insider threat detection framework that combines behavioral analytics, hybrid AI, and automated cyber defense to improve the security, transparency, and resilience of modern cloud computing environments.

Reference

[1] Abusitta, A., Li, M. Q., & Fung, B. C. M. (2024). *Survey on explainable AI: Techniques, challenges and open issues*. *Expert Systems with Applications, 255*, 124710. https://doi.org/10.1016/j.eswa.2024.124710

[2] Adeduro, O., Josh-Falade, O., & Mesioye, A. (2026). *Proactive insider threat detection framework: An explainable AI and behavioral analytics-driven approach*. *Journal of Future Artificial Intelligence and Technologies*. Advance online publication. https://doi.org/10.62411/faith.3048-3719-307

[3] Asha, S., & Shanmugapriya, D. (2024). *Understanding insiders in cloud adopted organizations: A survey on taxonomies, incident analysis, defensive solutions, challenges*. *Future Generation Computer Systems, 158*, 427–446. https://doi.org/10.1016/j.future.2024.04.033

[4] Alketbi, K. S., & Mehmood, A. (2025). *A comprehensive survey of explainable artificial intelligence techniques for malicious insider threat detection*. *IEEE Access, 13*, 121772–121798. https://doi.org/10.1109/ACCESS.2025.3587114

[5] Altaee, S., & Sweidan, M. (2026). *Artificial intelligence-based insider-threat detection: A hybrid explainable framework with automated response and privilege containment*. *Computers, 15*(7), 426. https://doi.org/10.3390/computers15070426

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

[7] Azaria, A., Richardson, A., Kraus, S., & Subrahmanian, V. S. (2014). *Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data*. *IEEE Transactions on Computational Social Systems, 1*(2), 135–155. https://doi.org/10.1109/TCSS.2014.2377811

[8] Bahram, S., Jiang, X., Wang, Z., Grace, M., Li, J., Srinivasan, D., Rhee, J., & Xu, D. (2010). *DKSM: Subverting virtual machine introspection for fun and profit*. *Proceedings of the IEEE Symposium on Reliable Distributed Systems.*

[9] Bugiel, S., Nürnberger, S., Pöppelmann, T., & Sadeghi, A.-R. (2012). *Virtualization security: Current trends and future directions.*

[10] Bugiel, S., Nürnberger, S., Pöppelmann, T., Sadeghi, A.-R., & Schneider, T. (2012). *Twin clouds: Secure cloud computing with low latency*. In *Communications and Multimedia Security.*

[11] Butt, K., Nasir, A., Hussain, B., Raza, M., Khan, S., Anwar, S., Ahmed, A., & Shah, K. (2026). *Behavioral analytics for insider threat detection in cloud environments*. ResearchGate preprint. https://www.researchgate.net/publication/404525592_Behavioral_Analytics_for_Insider_Threat_Detection_in_Cloud_Environments

[12] Cheimonidis, P., & Rantos, K. (2023). *Dynamic risk assessment in cybersecurity: A systematic literature review*. *Future Internet, 15*(10), 324. https://doi.org/10.3390/fi15100324

[13] Cheon, J. H., Kim, A., Kim, M., & Song, Y. (2017). *Homomorphic encryption for arithmetic of approximate numbers*. In *Advances in Cryptology – ASIACRYPT 2017* (pp. 409–437). https://doi.org/10.1007/978-3-319-70694-8_15

[14] Cloud Security Alliance. (2024). *Security guidance for critical areas of focus in cloud computing.*

[15] Dwork, C., & Roth, A. (2014). *The algorithmic foundations of differential privacy*. *Foundations and Trends® in Theoretical Computer Science, 9*(3–4), 211–407. https://doi.org/10.1561/0400000042

[16] Evans, D., Kolesnikov, V., & Rosulek, M. (2018). *A pragmatic introduction to secure multi-party computation*. *Foundations and Trends® in Privacy and Security, 2*(2–3), 70–246. https://doi.org/10.1561/3300000019

[17] Faqihi, R., Nanda, P., Mohanty, M., Alqahtani, S., & Alrashed, B. (2026). *Towards trustworthy cybersecurity: Reclassifying insider threat detection*. *Future Generation Computer Systems, 183*, 108543. https://doi.org/10.1016/j.future.2026.108543

[18] Garfinkel, T., & Rosenblum, M. (2003). *A virtual machine introspection based architecture for intrusion detection*. *Proceedings of the Network and Distributed Systems Security Symposium (NDSS).*

[19] Gaur, A., Mishra, P., Singh, A., Vinod, P., et al. (2024). *vDefender: An explainable and introspection-based approach for identifying emerging malware behaviour at hypervisor-layer in virtualization environment*. *Computers & Electrical Engineering, 120*, 109742. https://doi.org/10.1016/j.compeleceng.2024.109742

[20] Gong, Y., Cui, S., Liu, S., Jiang, B., Dong, C., & Lu, Z. (2024). *Graph-based insider threat detection: A survey*. *Computer Networks, 254*, 110757. https://doi.org/10.1016/j.comnet.2024.110757

[21] Greitzer, F. L., & Frincke, D. A. (2010). *Combining traditional cyber security audit data with psychosocial data: Towards predictive modeling for insider threat mitigation*. In *Insider Threats in Cyber Security* (pp. 85–113).

[22] Kairouz, P., McMahan, H. B., Avent, B., et al. (2021). *Advances and open problems in federated learning*. *Foundations and Trends® in Machine Learning, 14*(1–2), 1–210. https://doi.org/10.1561/2200000083

[23] Khan, A. H. (2025). *AI and behavioral analytics for insider threat detection: A comprehensive review of techniques, datasets, and emerging challenges*. *Journal of Information Systems Engineering and Management, 10*(63s).

[24] Kindervag, J. (2010). *Build security into your network's DNA: The Zero Trust Network Architecture*. Forrester Research.

[25] Lanuwabang, L., & Sarasu, P. (2025). *Detection of anomalies based on user behavioral information: A survey*. *International Journal of Wireless and Microwave Technologies, 15*(3), 54–65. https://doi.org/10.5815/IJWMT.2025.03.04

[26] Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing (NIST Special Publication 800-145).* National Institute of Standards and Technology.

[27] Mol, J. (2026). *Explainable AI mechanisms for insider threat detection in cloud-native platforms*. *International Journal of Computer Science and Information Management, 3*(1).

[28] Nance, K., Losiouk, E., & Hay, B. (2014). *Virtual machine introspection: Towards bridging the semantic gap*. *Journal of Cloud Computing, 3*(16). https://doi.org/10.1186/s13677-014-0016-2

[29] National Institute of Standards and Technology. (2020). *Zero Trust Architecture (NIST Special Publication 800-207).* U.S. Department of Commerce.

[30] Ofusori, L., Bokaba, T., & Mhlongo, S. (2024). *Artificial intelligence in cybersecurity: A comprehensive review and future direction*. *Journal of Intelligent & Fuzzy Systems*. https://doi.org/10.1080/08839514.2024.2439609

[31] OpenStack Foundation. (2024). *OpenStack documentation.*

[32] Perez-Botero, D., Szefer, J., & Lee, R. B. (2013). *Characterizing hypervisor vulnerabilities in cloud computing servers*. *Proceedings of the International Workshop on Security in Cloud Computing.*

[33] Pfoh, J., Schneider, C., & Eckert, C. (2011). *A formal model for virtual machine introspection*. Springer.

[34] Prasad, P. S. S., Nayak, S. K., & Krishna, M. V. (2025). *Hybrid machine learning for enhanced insider threat detection using generative latent features*. *International Journal of Engineering Trends and Technology, 73*(6), 102–113. https://doi.org/10.14445/22315381/IJETT-V73I6P110

[35] Rjoub, G., Bentahar, J., Abdel Wahab, O., Mizouni, R., Song, A., Cohen, R., Otrok, H., & Mourad, A. (2023). *A survey on explainable artificial intelligence for cybersecurity*. *IEEE Transactions on Network and Service Management, 20*(4), 5115–5140. https://doi.org/10.1109/TNSM.2023.3282740

[36] Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). *Zero Trust Architecture (NIST Special Publication 800-207).* National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-207

[37] Salih, A., Raisi-Estabragh, Z., Boscolo Galazzo, I., Radeva, P., Petersen, S. E., Menegaz, G., & Lekadir, K. (2023). *A perspective on explainable artificial intelligence methods: SHAP and LIME*. arXiv. https://arxiv.org/abs/2305.02012

[38] Samuel, K. (2026). *Behavioral analytics and explainable AI for identifying insider threats in government agencies*. ResearchGate.

[39] Sánchez-García, I. D., Mejía, J., & Gilabert, T. S. F. (2023). *Cybersecurity risk assessment: A systematic mapping review, proposal, and validation*. *Applied Sciences, 13*(1), 395. https://doi.org/10.3390/app13010395

[40] Sarraf, S. (2026). *Behavioral analytics for continuous insider threat detection in Zero Trust architectures*. arXiv. https://arxiv.org/abs/2601.06708

[41] Saxena, D., Gupta, I., Gupta, R., Singh, A. K., & Wen, X. (2023). *An AI-driven VM threat prediction model for multi-risks analysis-based cloud cybersecurity*. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. https://doi.org/10.1109/TSMC.2023.3288081

[42] Sehestedt, J., et al. (2024). *Active and passive virtual machine introspection on AMD and ARM processors*. *Journal of Systems Architecture, 149*, 103101. https://doi.org/10.1016/j.sysarc.2024.103101

[43] Selvam, P. A., & Selvy, P. T. (2025). *Explainable AI (XAI) for insider threat detection: Balancing security and transparency in cloud computing*. *Concurrency and Computation: Practice and Experience, 37*(25–26), e70390. https://doi.org/10.1002/cpe.70390

[44] Song, S., Gao, N., Zhang, Y., et al. (2024). *BRITD: Behavior rhythm insider threat detection with time awareness and user adaptation*. *Cybersecurity, 7*, 2. https://doi.org/10.1186/s42400-023-00190-9

[45] Stefan, D., et al. (2020). *Towards hypervisor support for enhancing the performance of virtual machine introspection*. *Future Generation Computer Systems, 108*, 1110–1123.

[46] Thammaiah, N., Girish, N., Meghana, N., Shivaprakash, G., & Kenny, S. P. K. (2026). *An interpretable insider threat detection framework: Correlating digital communications and behavioral metadata*. In *Proceedings of the 18th International Conference on Agents and Artificial Intelligence (ICAART 2026)* (Vol. 1, pp. 864–871). https://doi.org/10.5220/0014685300004052

[47] Tian, T., Zhang, C., Jiang, B., et al. (2025). *Insider threat detection for specific threat scenarios*. *Cybersecurity, 8*, 17. https://doi.org/10.1186/s42400-024-00321-w

[48] VMware. (2024). *VMware ESXi Architecture and Security Guide.*

[49] Wang, W., et al. (2017). *Leveraging virtual machine introspection with memory forensics to detect and characterize unknown malware using machine learning techniques at the hypervisor*. *Digital Investigation, 23*, 99–123. https://doi.org/10.1016/j.diin.2017.10.004

[50] Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302