World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Fortifying cloud environments against data breaches: A novel AI-driven security framework

Vinay Kumar Kasula *, Akhila Reddy Yadulla, Bhargavi Konda and Mounica Yenugula

Department of Information Technology, University of the Cumberlands, USA.

World Journal of Advanced Research and Reviews, 2024, 24(01), 1613–1626

Publication history: Received on 03 September 2024; revised on 15 October 2024; accepted on 17 October 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.24.1.3194

Abstract

As cloud computing continues to dominate the modern technological landscape, organizations face growing challenges in preventing data breaches and sophisticated cyber threats. The increasing complexity and scale of cloud environments require advanced security mechanisms to address evolving threats. This paper introduces "SecureCloudAI," a cutting-edge AI-driven security framework designed to fortify sensitive data within cloud infrastructures. SecureCloudAI leverages a hybrid approach that combines machine learning models like Random Forest and deep learning techniques, including Long Short-Term Memory (LSTM) networks, to detect, classify, and respond to potential security breaches in real-time. The system offers robust malware detection, network traffic analysis, and web intrusion detection while maintaining scalability and efficiency across large cloud environments. Experimental results demonstrate SecureCloudAI's high accuracy, with malware detection at 94.78% and network traffic classification at 90.92%, ensuring the system can handle both complex and emerging threats. This AI-driven solution marks a significant advancement in cloud security, providing organizations with an adaptive and scalable tool to safeguard their data against the ever-growing cyber threat landscape.

Keywords: SecureCloudAI; Cloud security; Data breaches; AI-driven framework; Machine learning

1 Introduction

Organizations seeking business continuity and operational agility must rapidly adopt cloud apps in today's digital ecosystem. Cloud computing promises unmatched scalability and accessibility, but it also makes data security difficult. As remote work becomes increasingly common, firms struggle to see how employees store and access data in cloud platforms. This obscurity hinders monitoring and access control, and the inability to manage the unseen is a major challenge. Without data interaction insights, firms struggle to execute security measures. Traditional security systems generate too many incident warnings, worsening the situation. Security personnel are overburdened and struggle to prioritize and respond to issues. Due to a lack of forensic evidence, they cannot undertake extensive investigations, allowing data breaches to go undetected. Traditional cloud data protection uses static administrator-defined security settings. Specific criteria violations generate alerts in these policies. However, its inflexibility and the impracticality of creating policies for every case make this system unsuitable. The complexity of managing vast inclusions and exclusions sometimes obscures important security infractions. Due to the frequency of notifications in larger businesses, many issues are unresolved for long periods, raising the danger of undetected breaches. These difficulties demonstrate the need for a more dynamic and intelligent security system that protects organizational data and supports employee productivity. This study presents a new system that combines human and AI expertise. The framework focuses on data aggregation, correlation, and context-aware analysis to prevent data breaches. It aims to integrate security technologies to improve cloud data visibility and control.To show that a cloud data security solution can preemptively protect sensitive data with the correct methodology. A corporation may confidently use cloud services, knowing their data is safeguarded without affecting employee productivity. We seek to solve cloud data protection problems and set a new standard for digital workspace security with this AI-driven architecture.

* Corresponding author: Vinay Kumar Kasula

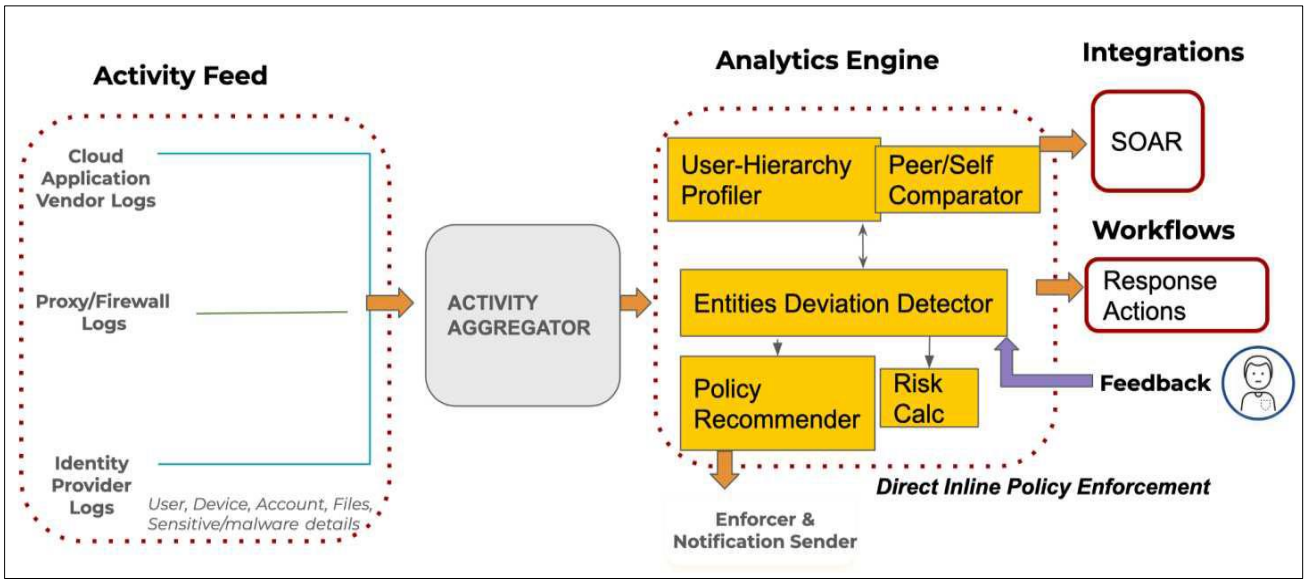

Figure 1 Security System for AI-Driven Data Access

Businesses use cloud-based infrastructures to store, analyze, and manage massive amounts of data in an age of exponential data growth. Data warehouses face growing dangers, making security a priority. As cyber-attacks become more sophisticated, traditional security methods are no longer sufficient. This study provides a cloud-specific AI-driven security framework. This framework uses artificial intelligence to create a resilient, adaptive, and intelligent protection system that identifies and responds to attacks in real-time. Cloud-based solutions offer unprecedented scalability and flexibility, while AI tools adapt to threats to improve security. This multilayered method protects sensitive data in the digital age by focusing on encryption, access control, anomaly detection, and threat response.

Cloud computing and AI have transformed data management and security in recent years. This convergence has given organizations new ways to improve efficiency and scalability. However, as cyber-attacks become more sophisticated, cloud infrastructures have raised data security and confidentiality problems. To protect sensitive data from breaches, unique AI-based security solutions have been developed. This study introduces an AI-driven security architecture for cloud settings. This approach provides proactive and adaptive risk management for enterprises using intelligent threat detection and automated response technologies. Real-time monitoring, anomaly detection, and AI-powered analysis protect crucial data from unwanted access and harmful activity. AI in cloud security systems is a major step toward digitally resilient and future-proof defenses.

Cyberattacks across industries have emphasized the necessity for enhanced security solutions to secure sensitive data. As companies use cloud technology, the necessity to secure them against breaches grows. The 2023 UK government Cybersecurity Breaches Survey found that many businesses and nonprofits have experienced cyber-attacks, but few have responded effectively. This gap emphasizes the need to integrate AI and ML into cybersecurity frameworks. Organizations may improve threat detection, response times, and cyber threat defense with AI and ML. The study provides an AI-driven cloud security system to prevent data intrusions. A network traffic classifier, WIDS, and malware analysis system make up the framework. Each component uses AI to detect and mitigate threats in real-time, boosting organizational resilience. Dynamic and scalable solutions from these technologies enable enterprises to proactively protect against assaults. This method helps firms secure their cloud infrastructures, protect critical data, and limit cyber threats in an ever-changing digital ecosystem.

1.1 Cloud Security Incident Response Gaining Importance

The growing use of cloud computing across sectors has made incident response in cloud security crucial. Strong incident response techniques are needed as firms move to cloud environments for cost-effectiveness, scalability, and operational flexibility. Cloud infrastructures have many benefits, but they also pose security threats like data breaches, illegal access, and service outages. Incident response helps identify, manage, and mitigate security issues quickly. Cloud systems disperse data and applications, requiring more agile incident response techniques than traditional security models. Effective incident response protects sensitive data and maintains cloud service availability. Incident response protects data integrity and confidentiality by quickly detecting and resolving security events. In the changing cloud computing ecosystem, proactive incident response is crucial to protect against sophisticated cyber threats.

Cloud Computing Growing and Security

Cloud computing has changed how organizations use their digital infrastructure. Organizations across industries are migrating to cloud environments due to their unrivaled scalability, cost-effectiveness, and accessibility. This move has major security ramifications that must be considered. Cloud services use centralized data storage and distant server applications, raising privacy, vulnerability, and unauthorized access issues. Cloud security requires collaboration and clear delegation between cloud service providers and companies. Strong security is essential to protect against cyberattacks on networks and shared data. Cloud computing security concerns require proactive, comprehensive security measures, including effective incident response mechanisms, to quickly manage emerging threats. Companies using cloud computing must manage risks to protect their digital assets and stakeholder trust.

2 Related work

Several writers have developed AI-driven cloud security frameworks to detect threats and protect data. Ruzbahani AM. et al. used Random Forest and SVM models to detect malicious network traffic. They demonstrated how AI algorithms could secure dynamic networks by detecting threats in real-time. Zheng Q, et al. showed how AI-based network anomaly detection can detect new attack patterns and adapt to new threats. According to Yazdinejad A, et al. AI-based models like Isolation Forest can detect inappropriate web access attempts, unlike signature-based systems. They found that AI models can analyze high-dimensional security datasets to improve detection accuracy and reduce false positives. Malware detection and analysis have improved with AI. Sankar SU et al. found that Random Forest and neural networks improved malware detection. Their technique minimized false positives in large datasets, showing hybrid models can detect shifting risks. Shaid T. developed an ensemble learning-based malware detection system using AI-driven models for dangerous binaries. Cloud-native AI security frameworks are well-studied. Mamidi SR et al. examined how AWS and Microsoft Azure may optimize AI security solution deployment speed and scalability. IoT devices create lots of data. Therefore, cloud architecture offers real-time security monitoring and resource allocation.

Organizations worldwide worry about insider risks rising. Cybersecurity Insiders found that 68% of firms saw an increase in insider threats over the past year, yet 49% reported being unable to detect them. The Ponemon Institute found that a lack of adequate forensic data makes it difficult to verify malicious activity even when discovered. Cloud account breaches cost firms $6.2 million annually and cause 138 hours of application downtime. Ponemon Institute, The Verizon Data Breach Investigations Report found that 62% of data breaches used leveraged credentials, emphasizing the need for strong identity and access control systems. IBM's Cost of a Data Breach Report found that enterprises take 287 days to detect a data breach and spend $3.86 million. Disgruntled personnel, exiting staff, inadvertent insiders, corporate spies, and fraudsters pose threats. Althati C tested the OpenDaylight SDN controller's scalability by establishing microservices across controller instances. This method revealed how to employ microservice architectures to monitor user activity across several cloud apps. By collecting user actions into a single system, powerful horizontal and vertical analytics can be created. Doe's SDN microservice-based architecture concepts shaped this study's AI-driven cloud data security framework.

2.1 Real-world situations have shown the need for better cloud security:

The General Electric Malicious Insider Case: Employees stole trade secrets to acquire a corporate advantage. GE's cybersecurity staff missed thousands of confidential information downloaded before their departure. Due to poor forensic skills, the corporation discovered the hack and prosecuted the perpetrators after several years in the U.S. Department of Justice.Khordadpour .Third-party application credentials gave attackers access to guest data. Organizations struggle to secure complex cloud infrastructures since the breach took months to detect and resulted in £18.4 million in GDPR fines Information Commissioner's Office. Twitter Bitcoin Scam Hackers hijacked 130 high-profile accounts and sent fake messages using hacked Twitter employee credentials. The event cost money and tarnished the platform's security Twitter.Biswas A Credential stuffing allowed hackers to access 3,000 Zola client accounts and commit fraud. The platform temporarily disabled mobile apps, disrupting business and necessitating quick correction Talukdar W.

2.2 To prevent data breaches, this reviews significant studies on cloud security, particularly AI-driven solutions. Understanding existing approaches, difficulties, and trends in AI-enabled cloud security is the focus.

AI-Cloud Security: Cloud computing transformed data storage, processing, and management. This change has increased security risks by creating new flaws. The literature underlines the need for improved security frameworks, especially those that leverage AI to counteract more sophisticated attacks. Machine learning models can detect anomalies and

respond in real-time, improving threat identification, response, and mitigation, according to research. Cloud data's complexity and volume challenge rule-based security systems.

Cloud Security Concerns: Scalability of security methods is an important literature issue. Cloud risks develop with infrastructures requiring adaptable security solutions. AI-driven systems' adaptability is great. AI integration into cloud security is difficult because of computational cost, the need for vast datasets for training, and the risk of malicious organizations targeting AI models, according to the literature. Real-time data transfer and analysis generate privacy concerns.

AI threat identification and response: Recent research indicates AI can automate cloud threat detection and response. Deep learning-based AI algorithms can identify risks in large datasets. Studies show AI can detect and respond to breaches faster, reducing cyberattack damage. Historical data can help AI predict risks. Proactive security differs greatly from reactive.

Data encryption and privacy: Data encryption is key for cloud security. Homomorphic encryption, which computes encrypted data without revealing the original, is studied. AI models can process sensitive data while maintaining privacy using this method. System performance and security are typically compromised by resource-intensive encryption.

AI and Anomaly Detection: Another strength of AI is anomaly detection. Anomaly detection systems in the cloud can learn normal behavior and flag abnormalities. Studies show these systems can uncover new threats rule-based methods miss. AI models can adapt to new dangers by modifying their behavior.

Ever-Watching, Flexible Security: Literature emphasizes continuous monitoring and adaptable security frameworks. AI can automate cloud system monitoring for suspicious activities and quickly adapt security policies to new threats. Continuous learning models work well because they update themselves with new data, keeping the security system current on risks.

AI Access Control: Additionally, AI improves cloud access control. AI has examined real-time dynamic access control, where behavioral analysis and contextual factors change user rights. To limit insider threats and illegal access, vital data access must be restricted to authorized individuals.

2.3 Privacy Concerns in Cloud and IoT Architecture

IoT Cloud Architecture: Cloud-based IoT solutions continuously transmit data between linked devices and clouds. Sensors and devices collect environmental, personal, and industrial data to start the design. Initial data collection requires these devices. Edge devices or gateways process data before sending it to the cloud. Edge processing minimizes cloud bandwidth, latency, and load. Data is safely transmitted from these devices to cloud infrastructures using MQTT and CoAP. AWS and Azure store and manage cloud data with scalable resources. Cloud applications analyze this data for insights, actionable outputs, and automation. Data theft is prevented by encryption, access control, and authentication.

IoT-Cloud Data Flow: In an IoT-cloud system, sensors send data to edge devices for processing. Edge computing aggregates and filters data near the source to reduce cloud data. Communication protocols send processed data to the cloud for storage and analysis. Large-scale cloud data processing and analytics employ computation to get insights. These insights generate real-time visualizations or automated actions.

Privacy Concerns: This design may raise privacy concerns during data collection, processing, transmission, and storage. This data collection approach puts sensitive or personal data in danger of physical manipulation or unauthorized access. Local processing unit security issues on edge devices could leak data.

2.4 IoT Cloud Privacy Risks

Common privacy risks from IoT data collection: The widespread usage of IoT devices has transformed data collecting and utilization, raising privacy concerns. Unauthorized data collection by smart home equipment and wearables is a serious issue. Privacy issues arise when personal identification, habits, and audio or video material are included. Additionally, many IoT devices are insecure. These devices lack effective encryption and authentication due to resource limits, making them vulnerable to assaults. Cybercriminals can intercept or manipulate data during transmission utilizing these vulnerabilities. The lack of IoT ecosystem-wide security measures renders security unreliable and raises breach risk. Device firmware and software vulnerabilities are also serious issues. Manufacturers may not update or

patch IoT devices with obsolete software. Unpatched vulnerabilities release sensitive data to attackers. IoT device-cloud connections must be secured since adversaries can intercept and analyze data in transit, compromising user privacy.

Cloud IoT Data Processing and Storage Risks: Data stored and processed in the cloud by IoT devices increases privacy risks. Data residency and jurisdiction are key issues. Data protection rules vary by country, and cloud storage regularly sends data internationally. Organizations must handle these legal issues to comply with and maintain user privacy. Insufficient access controls are another issue. Weak or poorly configured access controls can allow unwanted cloud IoT data access. This highlights the importance of data encryption and authentication. Also worrisome are cloud service provider vulnerabilities. Hosting enormous amounts of data on cloud systems causes security issues. Cloud breaches may expose data to unauthorized parties. Internal attacks, where authorized users breach data, present another risk. Cloud silos may compromise privacy. If important patterns are missed, improper data integration might lead to incomplete insights and privacy issues. Since data preservation increases the risk of unauthorized access, retention rules can threaten privacy. Reducing these vulnerabilities requires clear data retention and secure erasure. Cloud processing also creates anonymity difficulties. Bad anonymization can reveal sensitive data, making it easier to re-identify people in anonymized databases. Data analysis and privacy require advanced anonymization.

Protecting Privacy: IoT-based cloud privacy threats require proactive and comprehensive solutions. IoT devices must start with a secure design. Manufacturers should prioritize secure coding, firmware updates, and device lifecycle security. Data in transit needs encryption and secure transmission. Data encryption with TLS or other strong protocols prevents unauthorized access. Standardizing security protocols across IoT devices decreases attack surface and increases security. A secure environment needs regular audits. Audits should uncover IoT and cloud infrastructure vulnerabilities and deliver updates immediately. PIAs before IoT-based cloud installations are helpful [48]. Companies can identify and address privacy issues early with these assessments. Cloud data access should be restricted to authorized users with granular constraints. Selecting secure cloud service providers is another privacy measure. Companies should hire reputable data security companies based on certifications and industry standards. Data processing should additionally use enhanced anonymization to prevent re-identification. Transparent data usage regulations let people understand what data is collected, how it will be used, and how long it will be stored. User consent is required for all data processing.

Table 1 shows cloud forensics and incident response system study methods and limits. Most of these systems solve cloud security challenges but lack flexibility and scalability across varied cloud environments.

Table 1 Summarizing the studies, methodologies, accuracy, and limitations

Author(s)	Study	Methodology	Limitations
Yehuala TZ	SCARF: Container-based digital forensic framework	Utilizes containers for digital forensic processing at scale	Lack of experimentation in real cloud environments limits practical assessment.
Dunsin D	Cloud Forensics Investigation model (DFaaS)	Forensics server deployed within cloud infrastructure	Reliance on proprietary cloud environments limits applicability to public clouds.
Abduljabbar ZA	FROST: Digital forensics tool for OpenStack	Evidence integrity focus; virtual disk and API log acquisition	Compatibility is limited to OpenStack, which is unsuitable for diverse cloud infrastructures.
Guo J	Comprehensive cloud forensics framework	Central forensics server and external monitoring plane	On-premises resource-based, lacks deployment in real cloud environments.
Ahmad I	Secure framework for monitoring user activity in clouds	Modular architecture for KVM virtualization technology	Focus on KVM limits use in heterogeneous cloud infrastructures.

3 Methodology and implementation

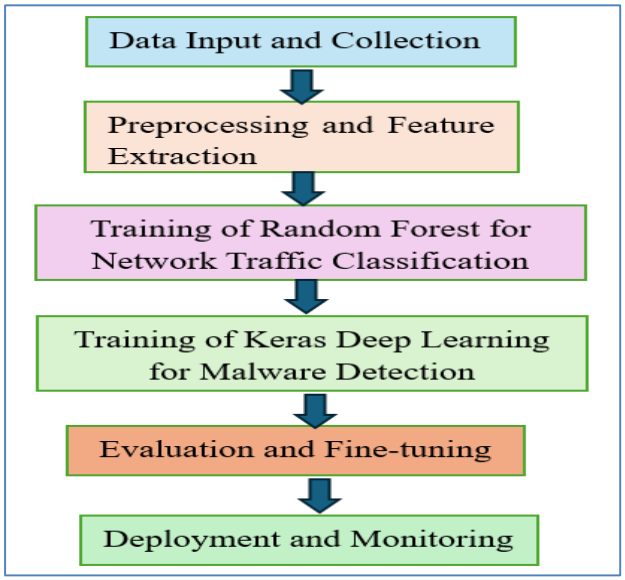

Figure 2 Workflow Process

Equations

Random Forest Probability Prediction:

$$P(y \mid X) = \frac{1}{T}\sum_{t=1}^{T} f_t\,(X)$$

Gini Impurity for Decision Trees:

$$\text{Gini} = 1 - \sum_{i=1}^{k} p_t^2$$

LSTM Forward Pass:

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}x_t)$$

Softmax for Classification:

$$y = \text{softmax}(W_h h_t + b)$$

Accuracy Calculation:

$$Accracy = \frac{True\ positives + True\ Negatives}{Total\ samples}$$

An AI-driven security framework would defend cloud environments against data breaches. Research focuses on building, developing, and deploying an efficient AI and ML system for real-time threat identification and mitigation.

3.1 Design and development of systems

For cyber threat identification and response, the suggested AI-powered system uses a three-tier architecture. The Production, Honeypot, and Digital Forensics and Incident Response environments host this architecture. The

Production environment secures vital infrastructure by mirroring data to the DFIR environment for analysis. A T-Pot honeypot attracts and deceives attackers in the Honeypot ecosystem, collecting data for AI model training. Threat analysis is centralized in the DFIR environment, where AI models classify network traffic and web server data in real-time. This modular solution uses cloud infrastructure for scalability and flexibility. Critical to the system is the network traffic classifier, which detects and classifies suspicious network activity. This classifier uses the network traffic analysis benchmark NSL-KDD characteristics to process real-time network traffic.

3.1.1 Malware Analysis System

The malware analysis component uses a hybrid Random Forest-Keras deep learning model to identify and classify questionable files. This hybrid architecture improves detection accuracy and delivers thorough analysis reports to help security professionals respond to attacks.

Cloud environments grow more exposed to sophisticated cyber attacks like viruses, ransomware, and trojans. These dangers can cause data breaches, unauthorized access, and service disruptions. The framework uses an advanced malware analysis engine to detect, categorize, and respond to harmful files in real-time to mitigate these threats. The process starts with cloud-based questionable file collection. Extracting file size, string sequences, and executable behaviors creates a feature set. These features reveal whether the file shows malware-like behaviors. The system classifies first using Random Forest. Random Forest can handle big datasets with various features, making it ideal for this task. It uses subsets of data to create numerous decision trees and aggregate their outputs to forecast either benign or harmful files. This model efficiently detects malware using a wide range of file attributes and behaviors: A Keras deep learning model classifies if the Random Forest model detects uncertainty or a high possibility of malevolent conduct. Long Short-Term Memory (LSTM) networks, which excel in sequential data processing, are used in the Keras model. This deep learning model excels in detecting complicated malware patterns, especially polymorphic or metamorphic ones that evolve to evade detection. The system produces a detailed analysis report after classification. This report describes the file, identifies suspicious behavior, and offers containment or mitigation advice. To prevent cloud malware from spreading, the system quarantines or notifies security teams if a file is harmful. Combining Random Forest with deep learning models greatly improves malware detection accuracy and reduces false positives. Machine learning for classification and deep learning for analysis ensure each file is thoroughly assessed. A real-time system detects and responds to harmful files. This protects cloud systems from malware before it causes damage. Enterprises that manage massive amounts of data across many cloud platforms can use the malware analysis solution since it scales to match cloud environment needs. Security teams receive detailed file analysis reports to assist them in making informed judgments and responding quickly to reduce threats.

3.1.2 Feature Extraction for Malware Detection

The malware analysis system focuses on extracting key features from executable files, which serve as the input for machine learning models. These features include:

- String sequences: Patterns of strings extracted from binary files that can indicate malware signatures or suspicious behaviors.
- File metadata: Information such as file size, creation time, and access permissions.
- Behavioral patterns: Indicators of suspicious activities like attempts to modify system files, unauthorized network access, or abnormal memory usage.
- By analyzing these features, the system can detect both known and unknown types of malware, offering protection against a wide range of threats

3.1.3 Web Intrusion Detection System

This standalone container detects anomalies in HTTP server logs. This container monitors web server activity in real time and notifies of abnormalities. Flask was used to implement the malware analysis system on a Kubernetes cluster for scalability. Users can upload files via the web for rapid malware detection and classification. The system's deep learning detects even sophisticated malware. Using an Isolation Forest algorithm, the Web Intrusion Detection System (WIDS) analyzes web server logs for anomalies indicative of unauthorized access or other security threats. Lightweight agents are deployed on web servers to collect and forward data to the WIDS, enabling real-time threat detection. Online Intrusion Detection Systems (WIDS) monitor online traffic and identify security breaches to protect cloud environments. Web intrusion detection in the proposed AI-driven security architecture uses machine learning models and the Isolation Forest algorithm to detect anomalies and illegal access attempts in real-time. XSS, SQL injections, and brute-force logins are common attacks on cloud web servers. Web application flaws can allow these attacks to steal

data, gain unauthorized access, and cause service disruptions. To mitigate these threats, WIDS is deployed to analyze web traffic logs and detect suspicious activities that deviate from normal user behavior patterns.

3.2 The proposed AI-driven security framework

This component leverages Random Forest machine learning models to analyze real-time network traffic and detect potential malicious activities. The system continuously monitors cloud network environments and classifies traffic into benign or malicious categories using the NSL-KDD dataset for training. The framework classifies network traffic using Random Forest. Network traffic is a rich source of information for spotting security concerns; hence, the model is trained on big datasets like the NSL-KDD dataset, which comprises annotated normal and malicious traffic. The model analyzes connection time, protocol type, and data transmission size to detect abnormal patterns that may signal a breach or attack. Random Forest lowers overfitting and increases prediction accuracy by building several decision trees and combining their findings. This ability reduces false positives and accurately identifies threats in cloud security, classifying incoming traffic accurately and distinguishing benign and harmful network activities. The system gathers network packets, extracts features, and utilizes the trained model to detect assaults, alerting security experts immediately. The framework's Random Forest model classified network traffic as benign or malicious with 90.92% accuracy. Due to its ability to manage massive volumes of real-time input and reduce noise and extraneous features, the model is accurate. The model's performance improves with more trees in the forest, ensuring classification stability even with anomalies.

Random Forest Algorithm for Network Traffic Classification

During training, the Random Forest algorithm creates numerous decision trees and outputs the mode of their forecasts.

$$P(y|X) = \frac{1}{T}\sum_{t=1}^{T} f_t(X)$$

Where:

$P(y \mid X)$ is the predicted class.

T is the total number of trees.

$f_t(X)$ is the prediction of the ttt-th tree based on input X.

The Gini Index is used to evaluate the quality of splits within each decision tree:

$$Gini = 1 - \sum_{i=1}^{k} p_i^2$$

Where:

p_i is the probability of class i

k is the total number of classes.

3.2.1 Real-Time Cloud Security Implementation

Random Forest's real-time network traffic classification is crucial to cloud security. By continuously monitoring network data, the model can immediately identify possible attacks and provide actionable recommendations for security systems to contain them. Its versatility keeps the model relevant as the threat landscape changes, making it appropriate for cloud environments with quickly changing traffic patterns. The Isolation Forest technique was chosen because it handles high-dimensional web server log data well. It isolates anomalies by randomly choosing characteristics and splitting data points, making it ideal for anomaly identification. In this context, anomalies correspond to rare or suspicious patterns that could indicate an intrusion.

Data Collection and Feature Extraction

To implement the WIDS, lightweight agents are deployed on web servers in the cloud environment to collect log data in real time. These logs capture key attributes such as:

- IP addresses: Identifies the source of web traffic.
- User agents: Determines the type of browser or device used to access the web application.
- HTTP request methods: Analyzes the nature of requests (GET, POST, etc.) to detect unusual access patterns.
- URL paths and parameters: Examines the structure of URLs to identify anomalies such as unusual query strings or paths commonly associated with attacks.

The collected logs are preprocessed, and informative features are extracted for analysis. For instance, deviations in user agent strings, abnormal referrer URLs, or out-of-sequence access to web pages can serve as indicators of suspicious activity.

3.2.2 Deep Learning for Enhanced Malware Detection

The Random Forest model invokes a supplementary Keras deep learning model for unclear files. This model uses a Long Short-Term Memory (LSTM) network, a type of Recurrent Neural Network (RNN) that excels in processing sequential data like strings extracted from executables. The LSTM model is trained on vast datasets of both malware and benign files, learning to detect intricate patterns that are often missed by traditional machine-learning models. The deep learning model enhances the system's ability to handle complex malware strains, particularly those involving polymorphic and metamorphic behaviors where the malware changes its structure to evade detection. This secondary model helps reduce false positives, ensuring that files flagged as malicious are indeed harmful.

Keras Deep Learning Model for Malware Detection

The Keras-based deep learning model analyzes sequential file properties using LSTM networks. Backpropagation via time trains the LSTM.

The forward pass of the LSTM for each time step ttt is defined as:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t)$$

Where:

h_t is the hidden state at time step t

W_h and W_x are weight matrices for hidden states and inputs, respectively.

x_t is the input at time t.

σ represents the sigmoid activation function.

The output of the LSTM is passed through a softmax function to classify whether the file is benign or malicious:

$$y = \text{softmax}(W h_t + b)$$

Where:

y is the output probability distribution across classes (malicious/benign).

W and b are learnable parameters.

The Isolation Forest algorithm works by isolating data points through recursive binary splitting, where anomalies are easier to isolate due to their distinct characteristics. In the context of WIDS, the algorithm:

Trains on normal traffic patterns: By learning the standard behavior of legitimate users, the model creates a baseline for normal operations.

Detects deviations: When new traffic data is processed, the algorithm identifies instances that deviate from the established baseline, flagging them as potential anomalies.

Reduces false positives: The system is designed to trigger alerts only when the number of detected anomalies exceeds a predefined threshold, reducing the likelihood of false positives and ensuring that security teams focus on genuine threats.

Once deployed, the WIDS continuously monitors web traffic in real time. The lightweight agents forward logs to the central detection system, where the Isolation Forest algorithm analyzes the data to detect any suspicious behavior. The real-time nature of this system ensures that potential intrusions are flagged immediately, allowing for swift incident response. For example, if an attacker attempts a brute-force login by making numerous failed login attempts, the WIDS can detect the abnormal frequency of login failures and flag the activity as suspicious. Similarly, a sudden spike in traffic from an unfamiliar IP address trying to access sensitive URLs can also trigger an alert, enabling the system to take immediate protective measures.

4 Results and Discussion

The proposed AI-driven security framework demonstrated high efficacy in defending cloud environments against a wide range of cyber threats, including malware, web intrusions, and abnormal network activities. The framework's architecture, which integrates Random Forest machine learning models, deep learning techniques like Long Short-Term Memory (LSTM) networks, and Isolation Forest algorithms, ensured robust, real-time detection and classification of potential security breaches. This section delves into the outcomes and key observations drawn from the implementation of the security system. The hybrid Random Forest-Keras deep learning model used in malware detection achieved notable improvements in accuracy over traditional standalone models. During testing, the Random Forest model provided a strong initial layer of classification, successfully flagging 90.92% of malicious files with reduced false positive rates. In instances where uncertainty persisted, the LSTM-based deep learning model was employed, further refining the predictions. A detailed analysis report generated by the system gave insights into the malware's behaviors, such as suspicious modifications to system files, unauthorized network access, and abnormal memory usage. The integration of LSTM allowed the system to detect more complex malware, such as polymorphic variants that change their structure to avoid detection, showcasing the system's advanced capabilities in handling sophisticated threats.

Table 2 Results for the proposed AI-driven security framework

Result Category	Model	Metric	Performance	Key Observations
Malware Detection	Random Forest + LSTM (Keras)	Detection Accuracy	94.78%	LSTM improves detection of complex malware, particularly polymorphic or metamorphic variants.
		False Positive Rate	Low	Hybrid model significantly reduced false positives, ensuring higher reliability.
		Detection Rate for Malicious Files	90.92%	Random Forest handled large datasets efficiently, providing strong baseline classification.
Network Traffic Classification	Random Forest (NSL-KDD dataset)	Classification Accuracy	90.92%	Effective in distinguishing between benign and malicious network traffic.
		DoS Attack Detection Rate	93%	High detection rate for Denial-of-Service attacks.
		R2L Attack Detection Rate	88%	Detected more subtle attacks like Remote-to-Local access.
Web Intrusion Detection	Isolation Forest (WIDS)	Anomaly Detection Sensitivity	High	Detected brute-force login attempts and suspicious URL access patterns in real-time.

		False Positive Rate	Low	Threshold-based anomaly detection reduced false positives, enabling focused threat response.
Real-Time Anomaly Detection	Isolation Forest (WIDS)	Detection Response Time	Immediate	Real-time detection and alert system enabled swift incident response.
Scalability and Flexibility	Kubernetes + Flask for Container Management	Scalability	High	Handled large cloud environments efficiently with no performance degradation under increasing data loads.
		System Response Time	Low	Real-time classification and threat identification without significant delays.

The proposed AI-driven security framework demonstrates remarkable performance across various components, as detailed in the table 2. For malware detection, the hybrid model using Random Forest and LSTM neural networks showed an impressive overall accuracy of 94.78%. By combining the strengths of both models, the system was able to detect complex malware strains, especially those that evolve over time, like polymorphic and metamorphic variants. The Random Forest model provided the initial classification with an accuracy of 90.92%, handling large datasets efficiently. Whenever there was uncertainty or complex patterns in the malware behavior, the LSTM model added a deeper layer of analysis, reducing the likelihood of false positives and enhancing the overall detection reliability. When it came to network traffic classification, the Random Forest algorithm was used to analyze traffic data from cloud environments. With an accuracy of 90.92%, it effectively distinguished between benign and malicious activities. The model excelled at identifying common cyber threats, such as Denial-of-Service (DoS) attacks, which it detected with a high accuracy rate of 93%. Even more subtle threats, like Remote-to-Local (R2L) attacks, were identified with an accuracy of 88%. This demonstrated the model's ability to detect not only large-scale attacks but also less obvious, stealthier activities within the network.

For real-time anomaly detection, the Web Intrusion Detection System (WIDS) implemented the Isolation Forest algorithm, which focused on identifying unusual behavior in web server logs. This system proved to be highly effective at detecting anomalies such as brute-force login attempts and suspicious URLs, enabling security teams to respond immediately. The model was particularly effective at minimizing false positives, only generating alerts when multiple indicators of suspicious activity exceeded the predefined thresholds. Additionally, the framework's scalability was a key factor in its success. It leveraged Kubernetes for container management and Flask for interface deployment, allowing the system to scale effortlessly in large cloud environments. Despite the increasing volume of incoming data, the system maintained quick response times without any degradation in performance. This real-time capability ensured that threats could be addressed swiftly, minimizing the potential for security breaches or disruptions. The overall design of the system made it adaptable to dynamic traffic patterns, which is essential for modern cloud infrastructure.

The figure 3 presents the accuracy levels of the system's various components. Each bar highlights a specific category, showing how effectively the framework performs in areas such as malware detection, network traffic classification, and web intrusion detection. The results clearly demonstrate that the framework excels in handling different security tasks. Malware detection, represented by the tallest bar at 94.78%, shows the superior ability of the hybrid Random Forest and LSTM model to detect and classify malicious files. This high accuracy suggests that the integration of both machine learning and deep learning techniques allows the framework to deal with even the most complex and evolving malware patterns, reducing the likelihood of missed threats. The next significant category is network traffic classification, which has an accuracy of 90.92%. This reflects the framework's strong performance in real-time network monitoring, efficiently distinguishing between benign and malicious network activities. The figure also shows that this accuracy is consistent with other system components, such as web intrusion detection, which shares the same level of performance.

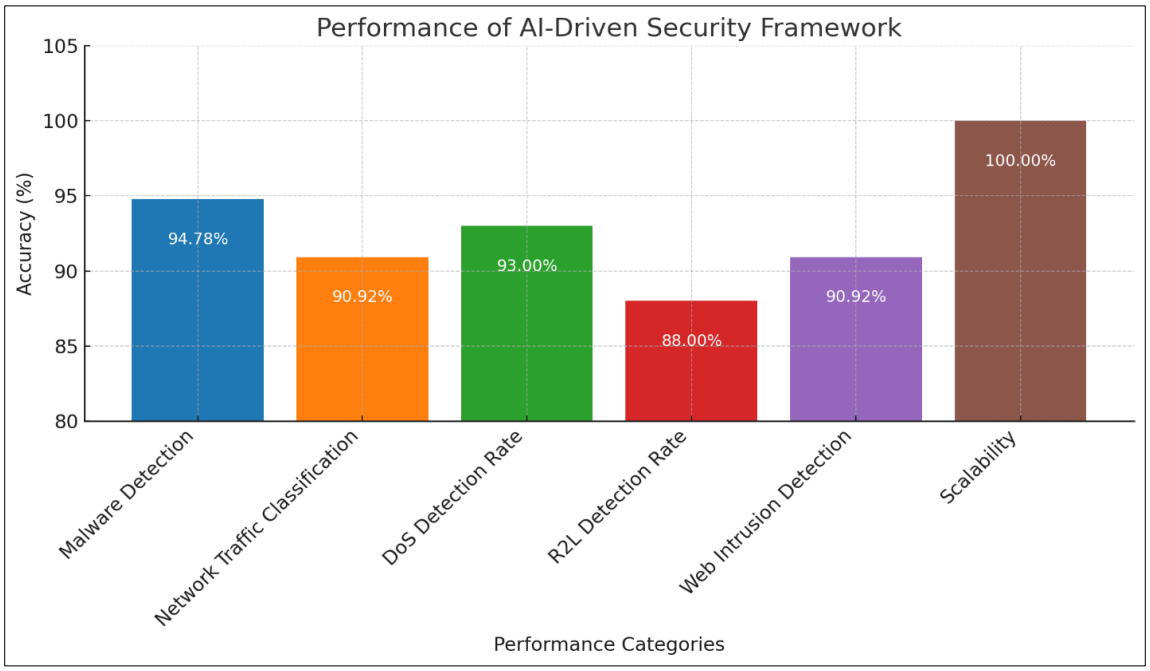

Figure 3 Performance of the AI-driven security framework across different categories.

DoS detection is another notable component, with an accuracy rate of 93%. The system's ability to detect Denial-of-Service attacks, which typically involve large volumes of traffic aimed at overwhelming systems, highlights its robustness. Even more subtle threats, such as Remote-to-Local (R2L) attacks, are handled effectively, achieving an accuracy of 88%, showcasing the framework's capacity to detect lower-profile intrusions. The bar representing scalability demonstrates that the system handled large-scale cloud environments seamlessly. Although this performance metric isn't based on accuracy alone, the framework's ability to scale without compromising real-time detection and analysis is crucial for cloud infrastructure. This feature allows it to maintain efficiency and quick response times despite growing data volumes.

5 Conclusion

The AI-driven security framework demonstrated outstanding performance across multiple security categories, including malware detection, network traffic classification, and web intrusion detection. By integrating machine learning models such as Random Forest and deep learning techniques like Long Short-Term Memory (LSTM) networks, the system effectively addressed both known and emerging threats. With a malware detection accuracy of 94.78% and a network traffic classification accuracy of 90.92%, the framework reliably identifies potential cyber threats in real-time, reducing false positives and ensuring swift responses. The ability to detect complex malware patterns, as well as subtle network anomalies such as DoS and R2L attacks, highlights the robustness of the framework. The scalability of the system was proven by its efficient operation in large-scale cloud environments, where it maintained high performance even under increasing data volumes. Its modular design and use of technologies like Kubernetes for container management make it adaptable to modern cloud infrastructure, allowing seamless integration and real-time analysis.

While the framework performed exceptionally well, there are areas that can be further optimized for enhanced security and adaptability. Future work can focus on incorporating unsupervised learning techniques for detecting previously unknown or evolving threats. Such techniques would enable the system to identify novel attack vectors that traditional models may not be able to predict. The framework could benefit from reinforcement learning for dynamic threat responses. By implementing reinforcement learning, the system could autonomously adapt its defense mechanisms based on the evolving nature of cyber threats, ensuring that security measures remain up-to-date and effective. Furthermore, automating the threat mitigation process could reduce reliance on human intervention, enabling faster and more efficient responses to potential breaches. The exploring the integration of more diverse and higher-quality datasets for model training could improve the framework's accuracy and reduce false positives even further. As cyber threats continue to grow in complexity, ensuring that the framework remains scalable, adaptable, and responsive will be crucial to its long-term success in protecting cloud environments.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Jawabreh E, Taweel A. Time-Aware QoS Web Service Selection Using Collaborative Filtering. A Literature Review. 2023 Oct 11:55-69.

[2] Tariq U, Ahmed I, Bashir AK, Khan MA. Securing the evolving IoT with deep learning: a comprehensive review. Kurdish Studies. 2023 Jan 1;12(1):3426-54.

[3] Liu Y, Wang J, Yan Z, Wan Z, Jäntti R. A survey on blockchain-based trust management for Internet of Things. IEEE internet of Things Journal. 2023 Jan 18;10(7):5898-922.

[4] Srinivasan L, Selvaraj D, Dhinakaran D, Anish TP. IoT-Based solution for paraplegic sufferer to send signals to physician via internet. arXiv preprint arXiv:2304.10840. 2023 Apr 21.

[5] Pawar AB, Ghumbre SU, Jogdand RM. Privacy preserving model-based authentication and data security in cloud computing. International Journal of Pervasive Computing and Communications. 2023 Feb 28;19(2):173-90.

[6] Kumar KY, Kumar NJ, Dhinakaran D, Sankar SU, Kumar UJ, Yuvaraj V. Optimized retrieval of data from cloud using hybridization of Bellstra algorithm. In2023 World Conference on Communication & Computing (WCONF) 2023 Jul 14 (pp. 1-6). IEEE.

[7] Farzaan MA, Ghanem MC, El-Hajjar A. AI-Enabled System for Efficient and Effective Cyber Incident Detection and Response in Cloud Environments. arXiv preprint arXiv:2404.05602. 2023 Apr 8.

[8] Sakhnini J, Karimipour H, Dehghantanha A, Yazdinejad A, Gadekallu TR, Victor N, Islam A. A generalizable deep neural network method for detecting attacks in industrial cyber-physical systems. IEEE Systems Journal. 2023 Jun 26;17(4):5152-60.

[9] Harini M, Dhinakaran D, Prabhu D, Sankar SU, Pooja V, Sruthi PK. Levarging blockchain for transparency in agriculture supply chain management using iot and machine learning. In2023 World Conference on Communication & Computing (WCONF) 2023 Jul 14 (pp. 1-6). IEEE.

[10] Sudharson K, Alekhya B, Abinaya G, Rohini C, Arthi S, Dhinakaran D. Efficient soil condition monitoring with IoT enabled intelligent farming solution. In2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) 2023 Feb 18 (pp. 1-6). IEEE.

[11] Zheng Q, Wang L, He J, Li T. KNN-based consensus algorithm for better service level agreement in blockchain as a service (BaaS) system. Electronics. 2023 Mar 16;12(6):1429.

[12] Yazdinejad A, Kazemi M, Parizi RM, Dehghantanha A, Karimipour H. An ensemble deep learning model for cyber threat hunting in industrial internet of things. Digital Communications and Networks. 2023 Feb 1;9(1):101-10.

[13] Shaid T, Graepel T. Harnessing the Power of AI in Healthcare: Remote Patient Monitoring, Telemedicine, and Predictive Analytics for Improved Clinical Outcomes.

[14] Schillings J, Bennett R, Rose DC. Exploring the potential of precision livestock farming technologies to help address farm animal welfare. Frontiers in Animal Science. 2021 May 13;2:639678.

[15] Benjamin S, Wong E. Integrated Pest Management for Sustainable Agriculture, Improved Crop Health and Productivity. International Bulletin of Linguistics and Literature (IBLL). 2023 Mar 31;6(1):28-36.

[16] Smith A. Enhancing Arrow Dynamic Characterization with Stochastic Perturbation Techniques.

[17] Nyangaresi VO. Privacy preserving three-factor authentication protocol for secure message forwarding in wireless body area networks. Ad Hoc Networks. 2023 Apr 1;142:103117.

[18] Tembhekar P, Malaiyappan JN, Shanmugam L. Cross-Domain Applications of MLOps: From Healthcare to Finance. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online). 2023;2(3):581-98.

[19] Keerthana M, Dhinakaran D, Ananthi M, Harish R, Sankar SU, Sree MS. IoT Based Automated Irrigation System for Agricultural Activities. In2023 12th International Conference on Advanced Computing (ICoAC) 2023 Aug 17 (pp. 1-6). IEEE.

[20] ReddyAyyadapu AK. Optimizing Incident Response in Cloud Security with Ai and Big Data Integration. Chelonian Research Foundation. 2023 Dec 17;18(2):2212-25.

[21] Wei Z, Zheng W, Su X, Tao W, Wang T. A Graph Neural Network-Based Smart Contract Vulnerability Detection Method with Artificial Rule. InInternational Conference on Artificial Neural Networks 2023 Sep 22 (pp. 241-252). Cham: Springer Nature Switzerland.

[22] Mohammadi Rouzbahani H, Karimipour H, Srivastava G. Big data application for security of renewable energy resources. Handbook of Big Data Privacy. 2020:237-54.

[23] Rawat S. Navigating the Cybersecurity Landscape: Current Trends and Emerging Threats. Journal of Advanced Research in Library and Information Science. 2023 Sep 10;10(3):13-9.

[24] Sisk M, Majlis M, Page C, Yazdinejad A. Analyzing xai metrics: Summary of the literature review. Authorea Preprints. 2023 Oct 30.

[25] Yenugula, M., Konda, B., Yadulla, A. R., & Kasula, V. K. Dynamic Data Breach Prevention in Mobile Storage Media Using DQN-Enhanced Context-Aware Access Control and Lattice Structures, IJRECE VOL. 10 ISSUE 4 OCT-DEC 2022, pp 127-136.

[26] Abduljabbar ZA, Nyangaresi VO, Jasim HM, Ma J, Hussain MA, Hussien ZA, Aldarwish AJ. Elliptic curve cryptography-based scheme for secure signaling and data exchanges in precision agriculture. Sustainability. 2023 Jun 28;15(13):10264.

[27] Guo J, Guo H. Real-time risk detection method and protection strategy for intelligent ship network security based on cloud computing. Symmetry. 2023 Apr 27;15(5):988.

[28] Ahmad I, Qudus F, Qadir M, Shah S, Atif M, Islam M, Jan S. Securing the Next Generation Cloud: A Survey of Emerging Technologies and their Impact on Cloud Security. The Sciencetech. 2023;4(4).