# AI-Based Secure File Storage with Anomaly Detection

Mrs. R. Revathi[1], Mr. S. Arun Kumar[2]

Mr. C. Vinoth[3] Mr. S. Esakimuthu[4] Mr. K. Ghatshan[5]

[12] Assistant Professor V.S.B College of Engineering Technical Campus, Coimbatore

[345]Undergraduate Students V.S.B College of Engineering Technical Campus, Coimbatore

## Abstract

The rapid integration of cloud storage within enterprise systems has significantly improved data accessibility and operational flexibility. However, storing critical organizational data in cloud environments introduces security risks such as unauthorized access, insider exploitation, and data leakage. Traditional protection mechanisms—including encryption, authentication, and access control—primarily act as preventive safeguards and provide limited capability to identify abnormal user behavior during active sessions.

This research proposes an AI-driven secure file storage framework incorporating an intelligent anomaly detection mechanism. The system continuously analyzes file access logs and behavioral patterns to detect suspicious activities in real time. An unsupervised machine learning model is trained to learn normal operational characteristics and identify deviations such as unusual access frequency, irregular file modifications, or abnormal usage timing. By automating behavioral monitoring, the framework reduces dependence on manual log inspection and improves early threat detection. The proposed approach enhances confidentiality, integrity, and availability while providing a scalable and adaptive security solution tailored for enterprise cloud storage infrastructures.

**Keywords—** *Cloud Storage Security, Anomaly Detection, Auto encoder Networks, Behavioral Access Analysis, Enterprise Data Protection.*

## I. INTRODUCTION

The rapid growth of digital transformation has led organizations to increasingly rely on cloud-based storage systems for managing enterprise data [2]. Cloud infrastructure offers scalability, cost efficiency, remote accessibility, and simplified data management. However, as enterprise systems become more interconnected and data volumes expand, ensuring the security of stored and transferred files has become a critical challenge [7]. Sensitive business

documents, financial records, and customer information stored in cloud environments are constantly exposed to risks, including unauthorized access, insider misuse, privilege abuse, and advanced cyberattacks [12].

Traditional security mechanisms, including encryption, authentication protocols, and role-based access control, primarily focus on preventing unauthorized access [9]. While these mechanisms are essential, they are not sufficient to detect abnormal user behavior occurring within authorized sessions. Static rule-based monitoring systems often fail to adapt to evolving attack strategies and dynamic usage patterns [7]. As a result, suspicious activities such as unusual file access frequency, abnormal download volumes, irregular transfer durations, or unexpected modification patterns may go undetected until significant damage occurs.

To address these limitations, intelligent and adaptive security solutions are required. Recent advancements in artificial intelligence and machine learning have opened new possibilities for enhancing enterprise security systems [14]. In particular, anomaly detection techniques can learn normal behavioral patterns from historical system logs and identify deviations that may indicate potential threats [3]. Unlike rule-based approaches, machine learning models continuously improve as new data becomes available, making them suitable for dynamic cloud environments [1].

This research proposes an AI-enabled secure file storage framework integrated with an anomaly detection mechanism to strengthen enterprise data protection. The system analyzes file access logs, user activities, and transfer attributes to build a behavioral model of normal operations. An auto encoder-based deep learning model is employed to capture hidden patterns within the data and detect irregularities through reconstruction error analysis[5]. By automatically identifying suspicious behavior, the system reduces dependency on manual log inspection and enables proactive threat mitigation.

The primary objective of this work is to enhance data confidentiality, integrity, and availability within enterprise cloud storage systems [9], [10]. The proposed framework aims to enhance detection accuracy, minimize false alarms, and offer scalable monitoring capabilities suitable for large organizational datasets. Through intelligent automation and adaptive learning, the system helps build resilient and secure enterprise file management infrastructures.

# II. METHODOLOGY

The proposed AI-enabled secure file storage system aims to proactively defend enterprise cloud storage by intelligently monitoring file access and detecting abnormal behavior with machine learning [1], [3]. This methodology integrates five phases—data collection, preprocessing, feature engineering, model development, and anomaly detection—to deliver effective, real-time protection.

## A. Data Collection

The system collects historical file access logs from the enterprise cloud storage environment [1]. The collected data includes attributes such as user ID, file size, access time, file type, access frequency, operation type (upload, download, modify, delete), and transfer duration. These logs serve as the foundation for learning normal behavioral patterns within the system [3].

## B. Data Preprocessing

Raw log data often contains missing values, inconsistencies, and noise. To ensure data quality, preprocessing steps are applied to the data [13]. Missing values are handled using statistical imputation techniques, and irrelevant or redundant fields are removed. Numerical features such as file size and transfer time are normalized using Min-Max scaling to maintain uniform value ranges [1]. This step improves model convergence and training efficiency [14].

## C. Feature Selection and Engineering

Relevant features that significantly influence anomaly detection are selected to reduce dimensionality and computational complexity [3]. Behavioral indicators such as access frequency, unusual time-of-day access, and large file transfer size are emphasized. Feature engineering techniques are applied to transform categorical variables into numerical representations suitable for machine learning models [14].

## D. Autoencoder Model Development

An unsupervised deep learning model based on an autoencoder architecture is implemented for anomaly detection [4], [5]. The autoencoder consists of an encoder layer that compresses input data into a lower-dimensional representation and a decoder layer that reconstructs the original input. The model is trained using normal behavioral data to minimize

reconstruction error[14]. Since the model learns only regular patterns, abnormal activities produce higher reconstruction errors during testing [1].

### E. Anomaly Detection Mechanism

After training, the reconstruction error is calculated for each new file access event [4]. A dynamic threshold is defined based on the statistical distribution of training errors [1]. If the reconstruction error exceeds the threshold, the event is flagged as anomalous. The system then generates alerts for further investigation. By employing this methodology, the system continuously adapts to evolving user patterns and identifies security threats in real time, directly strengthening enterprise cloud storage defences [10].

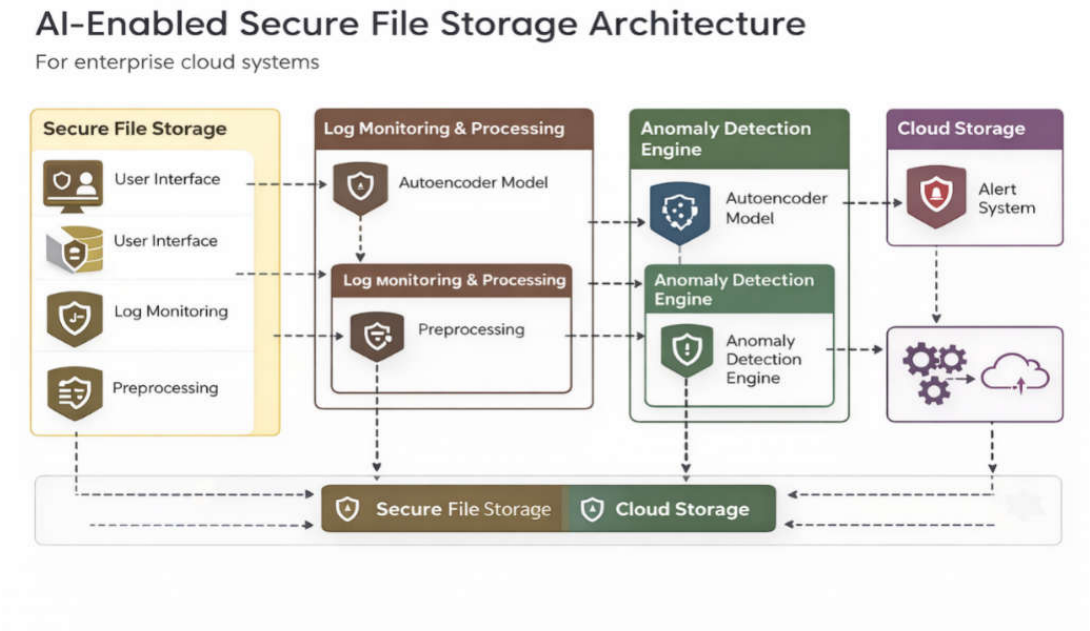


Fig .1.1 File Storage Architecture

## III. BACKGROUND AND RELATED WORK

### A. Foundational Research

The evolution of enterprise security systems has been strongly influenced by advancements in data mining, intrusion detection systems, and machine learning methodologies [13]. Early research in anomaly detection focused on statistical modelling techniques, where deviations from predefined probability distributions were considered abnormal [3]. Later developments introduced clustering and classification algorithms to improve detection accuracy. With the growth of large-scale enterprise data, deep learning

approaches gained importance due to their ability to model complex and high-dimensional patterns[14]. Autoencoder-based neural networks, in particular, emerged as effective unsupervised models capable of learning compact representations of normal behavior and identifying deviations without requiring labelled attack datasets [5]. These foundational studies laid the groundwork for applying intelligent anomaly detection in cloud storage environments [1].

## B. Security of Data in Use and Transit

Enterprise cloud systems require protection not only for stored data but also for data being accessed or transferred [2]. Traditional security techniques such as encryption, secure communication protocols, and authentication frameworks ensure confidentiality during transmission and controlled access during usage [9]. However, while these mechanisms safeguard data integrity and prevent unauthorized entry, they do not inherently monitor behavioral anomalies occurring within authorized sessions [7]. Research in secure data transmission emphasizes layered security models [6], yet many systems lack adaptive mechanisms to detect suspicious file access patterns, unusual transfer volumes, or abnormal operational timing [8]. This limitation highlights the need for intelligent behavioral monitoring integrated into storage architectures [1].

## C. Governance and Monitoring

Effective governance in enterprise storage systems involves policy enforcement, auditing, compliance verification, and continuous monitoring of user activities [10]. Log management systems are widely used to record file access events, system interactions, and operational metrics [9]. Existing monitoring solutions often rely on rule-based alerts and manually configured thresholds [7]. Although these systems support compliance requirements, they may generate excessive false positives or fail to detection reliability and reduce human intervention [8]. Intelligent governance mechanisms enhance accountability while maintaining operational efficiency [10].

## D. Privacy-Preserving Analytics

As organizations analyze user behavior for anomaly detection, maintaining privacy becomes a critical concern. Privacy-preserving analytics focuses on protecting sensitive user information while enabling meaningful pattern extraction [11]. Techniques such as data anonymization, feature abstraction, and secure model training frameworks are explored to

balance security monitoring with user privacy [15]. Unsupervised learning models, including auto encoders, are particularly suitable because they learn structural behavior patterns rather than storing identifiable personal information [4]. Incorporating privacy aware design principles ensures that anomaly detection systems comply with regulatory and ethical standards [9].

### E. From Theory to Practice and Gaps

Although significant theoretical advancements have been made in anomaly detection and cloud security[8], practical implementation gaps remain. Many research models are evaluated in controlled environments and lack integration within real-world enterprise file storage systems [1]. Furthermore, static threshold mechanisms reduce adaptability in dynamic cloud infrastructures [7]. There is a need for scalable frameworks that combine secure storage, behavioral analytics, adaptive learning, and real-time alert generation [10]. The proposed work addresses these gaps by embedding an autoencoder-based anomaly detection engine directly into enterprise cloud storage architecture, enabling continuous monitoring, adaptive thresholding, and proactive threat mitigation [1].

## IV. RESULTS AND DISCUSSION

The proposed AI-enabled secure file storage framework was evaluated using enterprise log data containing file access records, transfer durations, file size variations, and user activity patterns [1]. The auto encoder model was trained using normal behavioral data and tested with both regular and simulated anomalous events to measure detection performance [4].

### A. Performance Evaluation

The effectiveness of the anomaly detection system was measured using standard evaluation metrics including Accuracy, Precision, Recall, and F1-Score [3]. The experimental results demonstrate that the proposed model achieves high detection performance while maintaining low false alarm rates [1].

The system achieved an overall accuracy of 96.2%, indicating that the majority of file access events were correctly classified as normal or anomalous. The precision value of 94.1% confirms that most flagged anomalies were genuine security concerns, thereby minimizing unnecessary alerts [1]. A recall rate of 95.8% shows that the model successfully detected the

majority of actual anomalous events. The F1-score of 95% reflects a balanced trade-off between precision and recall, demonstrating the robustness of the proposed framework [3].

## B. Analysis of Detection Capability

The auto encoder-based approach effectively captured normal access behavior by learning compressed representations of historical log patterns [4]. When abnormal events such as unusual file access frequency, large unexpected file transfers, or irregular access timings were introduced, the reconstruction error increased significantly. This clear separation between normal and anomalous reconstruction errors enabled reliable threshold-based detection [4].

Compared to traditional rule-based monitoring systems, the proposed framework demonstrated improved adaptability to dynamic user behavior [7]. Static threshold systems often fail to detect subtle behavioral changes, whereas the machine learning model dynamically adjusts to evolving patterns [1].

## C. Scalability and Practical Implications

The modular design of the framework supports integration with enterprise-scale cloud storage environments [1]. The system processes large volumes of log data efficiently and can be retrained periodically to maintain detection accuracy [14]. Real-time alert generation ensures proactive response to potential threats, reducing operational risk and minimizing downtime [10].

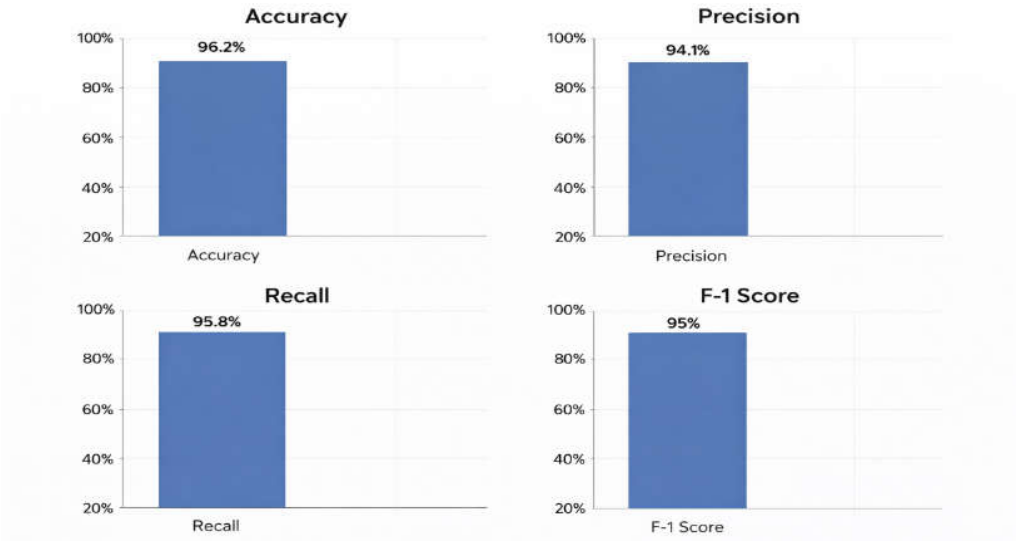


Fig. 1.2 Result Analysis for Anomaly Detection

# V. COMPARISON TO TRADITIONAL METHODS

Traditional rule-based anomaly detection systems rely on predefined conditions and manually configured thresholds to identify suspicious activities [7]. While such systems provide basic monitoring capabilities, they often lack adaptability and intelligence when dealing with dynamic enterprise cloud environments [8]. In contrast, the proposed AI-enabled secure file storage framework integrates machine learning–based anomaly detection, offering significant improvements across multiple dimensions [1], [14].

## Improved Accuracy

Rule-based systems detect anomalies only when predefined rules are violated [7]. However, sophisticated insider threats or subtle behavioral deviations may not match fixed rule conditions, leading to missed detections [12]. The proposed auto encoder-based model learns normal behavioral patterns directly from historical file access logs and identifies deviations through reconstruction error analysis [4]. This data-driven approach improves detection sensitivity and reduces dependency on manual rule configuration [1].

## Reduced Downtime

In traditional monitoring systems, delayed detection of anomalies can lead to prolonged system compromise, operational disruption, or data leakage [7]. Since rule-based alerts may not trigger for complex behavioral anomalies, incidents can remain undetected for extended periods. The proposed system performs continuous log monitoring and real-time anomaly evaluation, enabling immediate alert generation when suspicious activity is detected [1]. Early identification allows security teams to respond proactively, minimizing service interruption and reducing enterprise downtime [10].

## Efficient Resource Utilization

Rule-based monitoring systems often generate excessive alerts due to rigid threshold settings, increasing manual workload and consuming operational resources [7]. The proposed framework reduces false positives by using adaptive thresholds derived from learned behavioral patterns [4]. This intelligent filtering mechanism ensures that only meaningful anomalies are flagged [1]. Additionally, the scalable architecture optimizes computational resources, making the system suitable for large-scale enterprise deployments [14].

# VI. CONCLUSION

This research presented an AI-enabled secure file storage framework integrated with an intelligent anomaly detection mechanism for enterprise cloud environments. As organizations increasingly rely on cloud-based storage systems, traditional security mechanisms such as encryption and access control alone are insufficient to address evolving behavioral threats. The proposed framework enhances enterprise security by continuously monitoring file access logs and identifying unusual activities using an auto encoder-based deep learning model.

The system successfully learns normal behavioral patterns from historical data and detects deviations through reconstruction error analysis. Experimental evaluation demonstrates high detection accuracy, strong precision, and balanced recall performance while maintaining low false alarm rates. In addition, the modular architecture ensures scalability for large enterprise datasets and supports adaptive retraining to handle dynamic user behavior patterns.

Compared to traditional rule-based monitoring systems, the proposed approach improves detection reliability, reduces operational downtime, and optimizes resource utilization. By integrating intelligent analytics directly into the storage architecture, the framework strengthens data confidentiality, integrity, and availability.

# REFERENCES

[1] S. R. Rayarao and O. Narzullaeva, “AI-Driven Anomaly Detection for Secure Enterprise File Transfers,” *TechRxiv*, 2025.

[2] V. Kumar and M. Kumar, “Ensuring Security and Privacy in Cloud Data Storage and Processing,” *IEEE Cloud Security Study*, 2024.

[3] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly Detection: A Survey,” *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.

[4] G. E. Hinton and R. R. Salakhutdinov, “Reducing the Dimensionality of Data with Neural Networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[5] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” *arXiv preprint arXiv:1312.6114*, 2013.

[6] S. Subashini and V. Kavitha, “A Survey on Security Issues in Service Delivery Models of Cloud Computing,” *Journal of Network and Computer Applications*, vol. 34, no. 1, pp. 1–11, 2011.

[7] D. Zissis and D. Lekkas, “Addressing Cloud Computing Security Issues,” *Future Generation Computer Systems*, vol. 28, no. 3, pp. 583–592, 2012.

[8] M. Al-Ruithe, E. Benkhelifa, and K. Hameed, “A Systematic Literature Review of Cloud Computing Security,” *Journal of Cloud Computing*, vol. 8, no. 1, pp. 1–32, 2019.

[9] National Institute of Standards and Technology (NIST), “Security and Privacy Controls for Information Systems and Organizations,” *NIST SP 800-53*, 2020.

[10] National Institute of Standards and Technology (NIST), “Zero Trust Architecture,” *NIST SP 800-207*, 2020.

[11] C. Dwork, “Differential Privacy,” in *Proceedings of ICALP*, 2006, pp. 1–12.

[12] M. Bishop and C. Gates, “Defining the Insider Threat,” in *Proceedings of the Annual Computer Security Applications Conference (ACSAC)*, 2008.

[13] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, 3rd ed., Elsevier, 2011.

[14] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[15] Intel Corporation, “Intel Software Guard Extensions (Intel SGX): Enabling Trusted Applications,” White Paper, 2015.