

BackupIntegrity: AI-Enabled Protection and Recovery of Healthcare Database Backups in Cloud Environments

Answer Mushedhe¹; Onuh Matthew Ijiga²

¹College of Professional Studies, Northeastern University, Boston, Massachusetts, USA

²Department of Physics, Joseph Sarwuan Tarka University, Makurdi, Nigeria

Publication Date: 2026/10/14

Abstract: The increasing migration of healthcare information systems to cloud computing environments has intensified concerns regarding database backup integrity, ransomware attacks, unauthorized data modification, backup corruption, and operational disruptions affecting the availability and recoverability of sensitive patient records. Conventional backup protection mechanisms, including checksum-based validation, scheduled backup verification, rule-based intrusion detection, and static disaster recovery procedures, provide essential safeguards but may fail to recognize sophisticated integrity attacks, identify compromised recovery points, or optimize restoration under rapidly changing threat conditions. This paper proposes BackupIntegrity, a novel artificial intelligence-enabled framework integrating adaptive anomaly detection, cryptographic integrity verification, predictive backup reliability assessment, and risk-aware automated recovery for healthcare databases hosted in cloud environments. The proposed framework introduces the *Backup Integrity-Aware Adaptive Recovery Algorithm (BIARA)*, which combines a Temporal Transformer Network (TTN), an Isolation Forest (IF), a Gated Recurrent Unit (GRU), and a Multi-Objective Recovery Optimization (MORO) mechanism within a unified protection and restoration architecture. The Temporal Transformer captures long-range dependencies in backup activity sequences, while the Isolation Forest identifies abnormal backup access, modification, and encryption patterns. The GRU estimates backup failure probabilities and degradation trends, whereas the optimization component selects verified recovery points by jointly minimizing recovery time, potential data loss, residual compromise risk, and operational recovery costs. Cryptographic SHA-256 verification, authenticated backup manifests, immutable storage policies, and recovery-point dependency analysis provide additional safeguards against backup manipulation and contaminated restoration. The proposed system will be evaluated using a controlled cloud-based healthcare database testbed incorporating representative electronic health record transactions, incremental and full backup operations, simulated ransomware incidents, unauthorized backup modifications, storage failures, and recovery scenarios. Its performance will be compared with conventional checksum-based backup verification, rule-based security monitoring, standalone Isolation Forest, Long Short-Term Memory (LSTM), and XGBoost-based detection and recovery approaches. Experimental evaluation will examine detection accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC-ROC), false-positive rate, backup verification latency, recovery time objective (RTO), recovery point objective (RPO), data restoration success rate, and computational overhead. Comparative performance graphs, ROC curves, recovery-time distributions, reliability trajectories, and multi-objective optimization plots will be used to determine whether BIARA provides statistically significant improvements over the benchmark approaches. The anticipated contribution is a reproducible, integrity-aware cloud recovery architecture capable of strengthening ransomware resilience, identifying trustworthy restoration points, reducing recovery uncertainty, and supporting the continuity, confidentiality, and availability of critical healthcare information systems.

Keywords: Artificial Intelligence; Healthcare Databases; Backup Integrity; Cloud Security; Disaster Recovery.

How to Cite: Answer Mushedhe; Onuh Matthew Ijiga (2026) BackupIntegrity: AI-Enabled Protection and Recovery of Healthcare Database Backups in Cloud Environments. *International Journal of Innovative Science and Research Technology*, 11(10), 551-574. <https://doi.org/10.38124/ijisrt/26oct623>

I. INTRODUCTION

➤ Background to Cloud-Based Healthcare Database Management and Backup Security

Cloud-based healthcare database management has become an important technological foundation for maintaining electronic health records, diagnostic information,

medical imaging metadata, laboratory results, medication histories, and administrative transactions across interconnected healthcare facilities. Modern cloud architectures employ distributed database clusters, virtualized computing resources, replicated storage, encrypted object repositories, and application programming interfaces to support continuous clinical data availability. Ali et al. (2018)

established that cloud computing offers healthcare organizations opportunities to improve infrastructure scalability, information accessibility, and service integration, although security, privacy, and operational reliability remain significant adoption challenges. Healthcare database backup systems must consequently preserve not only stored information but also transaction ordering, referential integrity, schema consistency, encryption metadata, and recoverable database states. Full backups provide comprehensive recovery baselines, whereas incremental backups and write-ahead transaction logs enable reconstruction of more recent database conditions. However, maintaining geographically distributed backup copies does not automatically establish that those copies are authentic, internally consistent, or free from malicious modifications. Effective cloud backup management therefore requires independent verification of stored recovery artifacts, protected backup metadata, and mechanisms for identifying trustworthy restoration points.

Healthcare backup security becomes particularly complex when administrators, cloud service providers, applications, and automated orchestration services operate across different trust boundaries. Pool et al. (2024) examined failures in personal health data protection and identified weaknesses arising from interacting technological, organizational, and security management factors. In cloud environments, compromised service credentials may permit unauthorized snapshot deletion, backup retention modification, or manipulation of recovery metadata even when production databases remain operational. BackupIntegrity addresses these vulnerabilities through authenticated SHA-256 manifests, immutable storage policies, transaction-log continuity verification, and continuous behavioral monitoring. Its proposed BIARA incorporates a Temporal Transformer Network for examining sequential backup events, Isolation Forest for identifying anomalous operations, and Gated Recurrent Units for predicting backup degradation. These assessments inform recovery-point selection but cannot override failed cryptographic verification. The resulting architecture treats backup protection as an integrated security and recoverability problem, establishing the technical basis for comparing traditional validation procedures against intelligent, risk-aware healthcare database restoration.

➤ *Healthcare Data Vulnerabilities, Ransomware Threats, and Backup Integrity Challenges in Cloud Environments*

Healthcare cloud infrastructures face cybersecurity threats involving ransomware, compromised administrator credentials, unauthorized database modifications, insecure application interfaces, misconfigured access permissions, and attacks against backup management services. Their consequences may extend beyond information theft to interruption of patient registration, medication administration, laboratory reporting, and clinical decision support. Neprash et al. (2022) documented 374 ransomware attacks affecting United States healthcare delivery organizations between 2016 and 2021, with approximately 42 million patients' personal health information exposed and substantial disruption to healthcare operations. These

findings demonstrate why database backup protection must address continuity requirements alongside confidentiality and integrity. Sophisticated attackers may obtain privileged cloud credentials, alter backup policies, compromise snapshot repositories, or manipulate database records before encryption becomes apparent. Consequently, a backup may pass conventional checksum verification while containing transactions introduced during an undetected compromise. Particularly damaging scenarios include corruption of incremental backup chains, deletion of recent immutable retention candidates through administrative misconfiguration, and contamination of recovery snapshots by unauthorized changes to laboratory or prescription records. Reliable healthcare recovery therefore requires validation of both stored backup artifacts and the operational context surrounding their creation.

Early detection mechanisms offer opportunities to identify malicious behavior before attackers destroy critical recovery resources. Cen et al. (2024) reviewed ransomware detection approaches and emphasized recognizing suspicious activity during initial attack stages rather than relying exclusively on signatures or evidence of completed encryption. In BackupIntegrity, security telemetry will include abnormal backup-file entropy, unexpected snapshot deletion requests, unusual data rewriting rates, privileged access anomalies, and changes in transaction-log generation frequency. The Temporal Transformer Network is designed to identify dependencies between these events, while Isolation Forest detects deviations from normal operational profiles. GRU-based reliability forecasting complements these mechanisms by estimating future backup failure probabilities. Cryptographic verification and independently protected backup manifests provide separate evidence concerning artifact integrity. Controlled experiments will assess detection recall, false-positive rates, attack recognition latency, backup contamination, and recovery success across ransomware and administrative misuse scenarios. These evaluations will establish whether combining predictive detection with integrity-aware recovery selection improves resilience compared with conventional security monitoring and independently deployed machine learning algorithms.

➤ *Conventional Database Backup Protection, Verification, and Disaster Recovery Mechanisms*

Conventional healthcare database protection employs scheduled full backups, incremental snapshots, transaction-log archiving, storage replication, encryption, checksum verification, and predefined disaster recovery procedures. These safeguards provide mechanisms for restoring database services following infrastructure failures, accidental deletion, software defects, or malicious interference. Cryptographic hash functions such as SHA-256 produce fixed-length digests capable of detecting modifications when comparison values are securely generated and authenticated. However, matching hash values do not establish that source data were clinically accurate or uncompromised when the backup was created. Jin et al. (2018) investigated integrity and freshness assurance for cloud data, demonstrating the importance of verifying both stored information and its currency. Their findings are particularly relevant to healthcare systems requiring

transactionally consistent restoration of rapidly changing records. For example, a valid database snapshot may omit recently administered medication, completed laboratory investigations, or emergency treatment updates. Conventional recovery planning therefore distinguishes the recovery time objective, representing acceptable service restoration duration, from the recovery point objective, representing acceptable data-loss exposure. Both measures require careful interpretation when identifying the most appropriate restoration point.

Remote cloud auditing introduces cryptographic mechanisms for assessing outsourced data integrity without necessarily downloading complete backup repositories. Garg et al. (2020) developed a publicly verifiable cloud integrity auditing protocol using bilinear pairing-based techniques and cryptographic signatures, demonstrating opportunities to reduce selected verification costs. Nevertheless, storage auditing alone cannot establish transaction-log replay feasibility or determine whether a structurally valid backup contains maliciously modified clinical records. BackupIntegrity extends conventional verification through authenticated hash manifests, incremental backup dependency analysis, transaction-log continuity checking, and predictive security assessment. Candidate recovery points must satisfy mandatory cryptographic and database consistency conditions before entering the optimization process. BIARA subsequently evaluates eligible candidates according to expected recovery duration, potential transaction loss, predicted compromise probability, and cloud resource consumption. The comparative experimental design will measure checksum verification latency, restoration completeness, RTO compliance, RPO performance, and computational overhead under different backup frequencies and database workloads. This approach creates a basis for determining whether integrated security assessment improves practical recoverability beyond conventional backup existence and checksum validation.

➤ *Artificial Intelligence and Predictive Analytics for Intelligent Backup Security and Recovery Management*

Artificial intelligence provides computational techniques for identifying abnormal database backup operations, predicting storage-related failures, and supporting adaptive security decisions across distributed healthcare cloud infrastructures. Conventional security monitoring often depends on predefined thresholds that may fail to distinguish malicious activity from legitimate variations in clinical database workloads. Machine learning algorithms instead examine multidimensional relationships involving backup frequency, snapshot size, storage input/output rates, encryption activity, administrative access patterns, transaction-log growth, and historical verification outcomes. Pang et al. (2022) systematically examined deep anomaly detection methods and demonstrated how learned representations can support the identification of complex deviations in high-dimensional datasets. These capabilities motivate the proposed BackupIntegrity architecture, which combines unsupervised and temporal learning mechanisms to distinguish abnormal backup activities from routine database operations. Isolation Forest identifies statistically unusual

observations through randomized partitioning, whereas the Temporal Transformer Network captures contextual dependencies among successive security events using attention mechanisms. For example, an unexpected administrative login followed by retention-policy modification and rapid snapshot deletion may represent a more significant security anomaly than any individual event examined independently.

Predictive modelling further enables backup management systems to estimate reliability deterioration before a restoration attempt becomes necessary. Zamanzadeh Darban et al. (2024) examined deep learning methods for time-series anomaly detection, identifying methodological challenges associated with temporal dependence, contextual abnormalities, and changing data distributions. BackupIntegrity applies these principles through GRU-based forecasting of backup failure probabilities using historical integrity violations, incomplete transfer events, storage latency, transaction-log inconsistencies, and recovery-test outcomes. BIARA combines these reliability estimates with Transformer-derived threat representations, Isolation Forest anomaly scores, and independently authenticated integrity evidence. The resulting recovery-point ranking process is constrained so that probabilistic predictions cannot override failed cryptographic verification. Experimental comparisons will involve standalone Isolation Forest, LSTM, XGBoost, and conventional rule-based security models, evaluated under equivalent training conditions and controlled cloud attack scenarios. Performance indicators will include precision, recall, F1-score, ROC-AUC, false-positive rate, prediction calibration, and inference latency. Component-level ablation studies will establish whether model integration provides measurable performance gains and whether those improvements justify additional computational complexity.

➤ *Limitations of Existing Cloud Backup Integrity Verification and Automated Recovery Approaches*

Existing cloud backup integrity verification approaches exhibit limitations involving computational efficiency, dependence on trusted metadata, incomplete threat visibility, and insufficient integration with actual database recovery operations. Conventional checksum comparison can detect changes in stored artifacts but cannot independently determine whether malicious modifications occurred before reference hashes were created. Public auditing protocols provide stronger evidence concerning outsourced data possession and integrity, although their practical implementation requires consideration of communication costs, cryptographic processing, dynamic updates, and metadata management. Zafar et al. (2017) reviewed cloud integrity verification techniques and identified significant challenges associated with security assumptions, storage efficiency, auditing overhead, and changing outsourced datasets. These limitations become particularly important in healthcare systems processing large volumes of electronic health records, diagnostic images, medication transactions, and continuously generated clinical data. Full backup verification may consume considerable storage bandwidth, whereas infrequent sampling may delay recognition of corruption. Conventional restoration procedures may

additionally prioritize the newest available backup without examining whether its incremental dependencies are trustworthy or whether malicious transactions were incorporated before the snapshot was generated.

Distributed integrity verification mechanisms seek to reduce dependence on centralized auditing infrastructure and improve transparency across cloud storage environments. Han et al. (2022) examined blockchain-based integrity auditing and identified opportunities for decentralized verification alongside unresolved challenges concerning efficiency, scalability, and security management. However, an authenticated audit history does not independently establish the clinical correctness or transactional recoverability of the underlying database. Similarly, machine learning threat detectors may identify suspicious activity without determining which historical recovery point remains safe to restore. BackupIntegrity addresses this separation by integrating mandatory cryptographic checks, predicted compromise indicators, backup dependency validation, and multi-objective recovery optimization. BIARA excludes candidates failing integrity requirements before evaluating feasible restoration plans according to recovery duration, data-loss exposure, residual security risk, and resource consumption. Its proposed experimental evaluation will examine corrupted-backup rejection, recovery-point selection accuracy, verification latency, restoration success, and computational scalability. These comparisons will determine whether integrating security intelligence with restoration decisions offers benefits beyond standalone auditing protocols, reactive anomaly detection, and predetermined backup recovery schedules.

➤ *Research Problem, Knowledge Gap, and Motivation for Developing the BackupIntegrity Framework*

The fundamental research problem addressed in this study concerns the limited coordination between healthcare database backup integrity verification, predictive cyber threat detection, backup reliability assessment, and secure restoration decision-making. Existing backup technologies can preserve database copies, verify stored file digests, and support recovery after infrastructure failures, but these capabilities do not automatically establish whether a restoration point is free from malicious changes or suitable for clinical operations. Al-rimy et al. (2018) reviewed ransomware threat mechanisms, enabling factors, and countermeasures, demonstrating the importance of considering attack progression and weaknesses in conventional detection methods. However, ransomware mitigation and trustworthy database restoration represent different technical problems. A compromised healthcare database may generate cryptographically valid backups containing unauthorized modifications, while seemingly complete restoration chains may include corrupted incremental snapshots or missing transaction logs. These conditions create uncertainty that cannot be resolved solely through signature-based detection, backup availability checks, or isolated anomaly classification. The research gap therefore concerns a unified mechanism capable of evaluating backup authenticity, behavioral compromise evidence,

predicted reliability, and restoration feasibility within a consistent decision framework.

Experimental research into ransomware mitigation provides an important foundation for evaluating coordinated defensive strategies. Berardi et al. (2023) developed data-flooding mechanisms intended to delay ransomware encryption and preserve opportunities for defensive intervention, reporting reduced file encryption losses across controlled attack scenarios. Their findings illustrate the value of intervention timing, although the selection of trustworthy healthcare database recovery points requires additional integrity and transactional evidence. BackupIntegrity proposes the Backup Integrity-Aware Adaptive Recovery Algorithm as a technical response to this challenge. BIARA integrates Temporal Transformer event modelling, Isolation Forest anomaly detection, GRU-based reliability prediction, authenticated SHA-256 backup manifests, and constrained multi-objective recovery optimization. The framework will be evaluated using representative healthcare database transactions, full and incremental backup operations, controlled ransomware activity, unauthorized snapshot modifications, storage failures, and restoration scenarios. Comparisons against conventional checksum validation, rule-based security monitoring, Isolation Forest, LSTM, and XGBoost will examine detection performance, verification latency, recovery duration, data-loss exposure, and restoration success. Sensitivity analysis and component-level ablation will determine the contribution of each architectural element, establishing whether the proposed system improves recovery decisions without imposing disproportionate computational or operational costs.

➤ *Research Aim, Objectives, and Research Questions*

• *Research Aim*

The primary aim of this study is to develop and experimentally evaluate BackupIntegrity, an artificial intelligence-enabled healthcare database backup protection and recovery framework that integrates cryptographic integrity verification, temporal anomaly detection, predictive reliability assessment, and multi-objective restoration optimization to improve the security, trustworthiness, and operational efficiency of cloud database recovery.

• *Research Objectives*

- ✓ To design a secure cloud healthcare database backup architecture integrating authenticated SHA-256 manifests, immutable storage, transaction-log verification, and continuous integrity monitoring.
- ✓ To develop the Backup Integrity-Aware Adaptive Recovery Algorithm using Temporal Transformer Networks, Isolation Forest, and GRU-based predictive reliability modelling.
- ✓ To formulate an integrated backup trust assessment model incorporating cryptographic validity, behavioral anomaly scores, predicted failure probabilities, and restoration dependency integrity.
- ✓ To design a multi-objective recovery optimization mechanism minimizing restoration duration, potential

clinical data loss, residual compromise risk, and cloud resource consumption.

- ✓ To implement a reproducible healthcare cloud database testbed incorporating representative clinical transactions, full and incremental backups, ransomware attacks, unauthorized modifications, and storage failure scenarios.
- ✓ To compare BIARA against checksum-based verification, rule-based detection, Isolation Forest, LSTM, and XGBoost using standardized threat detection and predictive performance metrics.
- ✓ To evaluate recovery-point selection accuracy, RTO, RPO, restoration success, verification latency, computational overhead, and scalability under varying database workloads.
- ✓ To conduct ablation studies, sensitivity analysis, statistical significance testing, and robustness assessments to establish the contributions and operational limitations of the proposed framework.

• Research Questions

- ✓ How can authenticated cryptographic verification, immutable cloud storage, and transaction-log validation be integrated to strengthen healthcare database backup integrity and recoverability?
- ✓ How effectively can BIARA combine Temporal Transformer Networks, Isolation Forest, and GRU prediction to detect malicious backup activity and anticipate reliability deterioration?
- ✓ To what extent can integrated cryptographic and predictive evidence distinguish trustworthy recovery points from corrupted, compromised, or transactionally inconsistent database backups?
- ✓ How can multi-objective restoration optimization reduce recovery duration, data-loss exposure, residual compromise risk, and computational resource requirements?
- ✓ How do ransomware severity, database transaction intensity, backup frequency, storage failures, and incremental backup dependencies influence the performance of BackupIntegrity?
- ✓ How does the proposed BIARA framework compare with conventional checksum verification, rule-based monitoring, Isolation Forest, LSTM, and XGBoost across standardized detection metrics?
- ✓ What improvements or trade-offs arise in recovery-point selection accuracy, RTO compliance, RPO performance, restoration success, verification latency, and scalability?
- ✓ Which components of BackupIntegrity contribute most significantly to threat detection, backup reliability prediction, recovery optimization, and system robustness under different failure conditions?

➤ Research Contributions, Technical Significance, and Organization of the Paper

This study proposes BackupIntegrity as an integrated artificial intelligence-enabled framework for strengthening healthcare database backup protection, integrity verification, and adaptive recovery within cloud computing environments. Its principal methodological contribution is the proposed BIARA, which combines a Temporal Transformer Network

for sequential security event analysis, Isolation Forest for unsupervised anomaly identification, GRU-based forecasting for backup reliability prediction, authenticated SHA-256 manifests for cryptographic integrity assurance, and constrained multi-objective optimization for trustworthy recovery-point selection. Unlike approaches that assess backup correctness, attack detection, and restoration planning independently, BackupIntegrity establishes a coordinated decision architecture in which deterministic integrity requirements govern candidate eligibility while probabilistic threat and reliability assessments inform recovery prioritization. Its recovery optimization model considers transaction-log continuity, incremental backup dependencies, expected restoration duration, potential clinical data loss, residual compromise risk, and cloud resource costs. The study also contributes a reproducible experimental evaluation methodology incorporating representative healthcare database workloads, controlled ransomware events, unauthorized backup modifications, storage interruptions, and recovery tests. Comparative performance assessment against checksum-based verification, rule-based monitoring, Isolation Forest, LSTM, and XGBoost will examine detection accuracy, precision, recall, F1-score, ROC-AUC, false-positive rates, verification latency, RTO, RPO, and restoration success. Statistical testing, sensitivity analysis, and component-level ablation will be used to determine whether the integrated approach produces measurable performance improvements and identify the conditions under which such improvements occur. The study's technical significance lies in treating reliable healthcare database restoration as a combined cryptographic, predictive, and operational optimization problem rather than a conventional backup retrieval process. Following this introduction, Section 2 examines existing healthcare database protection architectures, cryptographic auditing mechanisms, ransomware detection techniques, temporal deep learning algorithms, and intelligent recovery optimization approaches. Section 3 presents the proposed system architecture, mathematical formulation, algorithmic procedures, experimental configuration, and evaluation methodology. Section 4 presents the comparative results, integrity verification performance, recovery optimization outcomes, scalability assessments, and statistical interpretation. Section 5 addresses the principal research findings, scientific contributions, operational limitations, implementation recommendations, and future research directions.

II. LITERATURE REVIEW

➤ Healthcare Database Backup Architectures, Cloud Storage Security, and Data Integrity Assurance Mechanisms

Healthcare database backup architectures increasingly employ distributed cloud storage, transactional replication, encrypted object repositories, and geographically separated recovery facilities to preserve electronic health records against operational failures and cyberattacks. Negro-Calduch et al. (2021) identified scalability, interoperability, security, and privacy as important challenges in electronic health record modernization as represented in figure 1. Similarly, Oh et al. (2021) established that contemporary healthcare

information systems require coordinated security controls across medical data repositories, communication networks, and cloud infrastructure. These requirements extend to backup architectures, where full and incremental database copies must preserve transaction consistency, encryption metadata, snapshot dependencies, and restoration accessibility. Frank (2026a) examined healthcare quality measurement and performance monitoring, providing a clinical information-management context for preserving reliable datasets used in quality assessments. Frank (2026b) further explored predictive analytics in medical inventory management, demonstrating another healthcare application dependent on accurate operational records. Within BackupIntegrity, these information requirements motivate cryptographically authenticated backup manifests, transaction-log continuity verification, role-based access controls, and immutable cloud storage policies that protect clinical data from unauthorized alteration.

A secure implementation should separate production credentials from recovery administration, encrypt backup

objects using authenticated encryption, and independently protect the cryptographic keys used to authenticate recovery metadata. BackupIntegrity additionally proposes SHA-256 integrity verification, scheduled restoration testing, and dependency-aware validation of incremental backup chains. Ashieyi-Ahorgah (2025) examined alternative financing mechanisms for innovation; although unrelated to database security, the study provides a limited perspective on financing technology deployment. Technical evaluation must instead establish whether security controls preserve recoverability under attack. The proposed architecture therefore combines cryptographic evidence with behavioral anomaly scores and predictive reliability estimates generated by BIARA. Its experimental assessment will examine backup completeness, verification latency, transaction consistency, storage overhead, and restoration success across increasing healthcare database workloads. Recovery candidates failing mandatory integrity conditions will be excluded regardless of their predicted reliability scores.

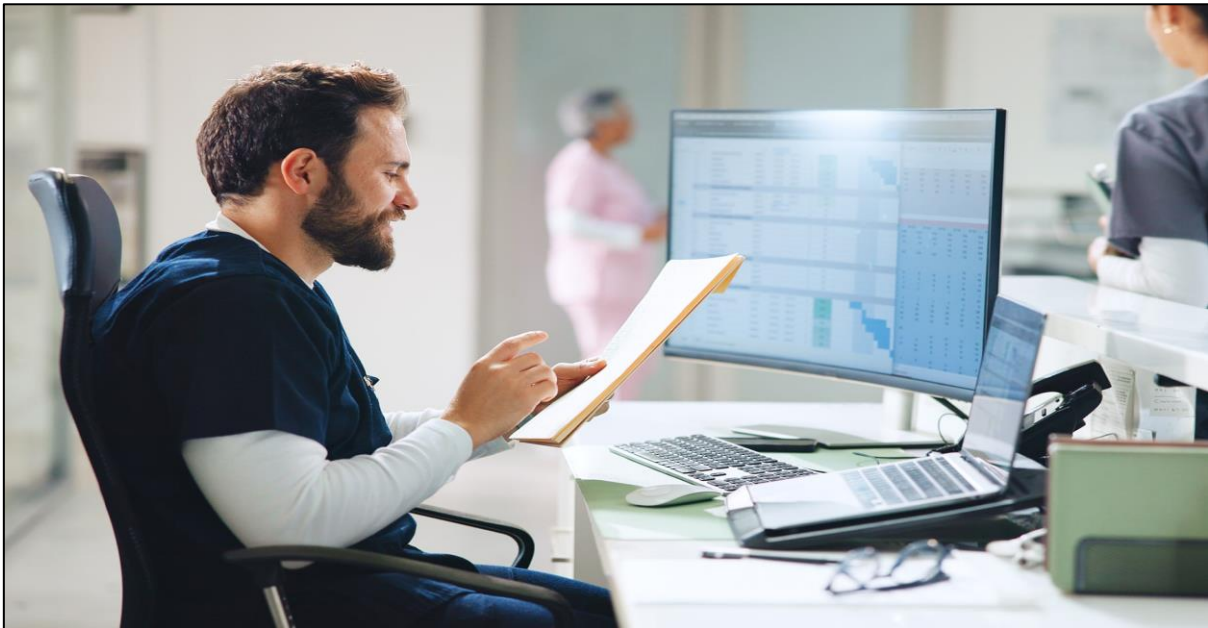


Fig 1 Healthcare Information Management Illustrating Secure Cloud Database Backups, Cryptographic Integrity Verification and Intelligent Recovery Mechanisms (Cstonesolutions., N.D).

Figure 1 illustrates a healthcare information management environment in which a healthcare professional reviews clinical documentation alongside computerized data records, representing the interaction between clinical operations and digital information systems. The workstation, comprising a desktop monitor, laptop, and physical patient documentation, provides a practical context for examining electronic health record (EHR) management, database accessibility, and the importance of secure information preservation. Within the proposed BackupIntegrity architecture, such clinical workstations function as authorized endpoints connected to centralized healthcare databases hosted on cloud or hybrid computing infrastructure. Patient records, laboratory findings, medication histories, and clinical transactions generated through these systems require

coordinated protection through automated full and incremental database backups, transaction-log archiving, encrypted cloud object storage, and geographically distributed recovery repositories. Although these backend security mechanisms are not directly visible in the photograph, the image illustrates the clinical environment whose information assets the proposed framework is designed to protect. BackupIntegrity strengthens the underlying database infrastructure through SHA-256 cryptographic hashing, digitally authenticated backup manifests, immutable storage policies, role-based access controls, and transaction consistency verification. Its Backup Integrity-Aware Adaptive Recovery Algorithm (BIARA) further integrates Temporal Transformer Networks for identifying suspicious sequences of database activities,

Isolation Forest for detecting abnormal backup operations, and Gated Recurrent Units for predicting backup reliability deterioration. These components support continuous evaluation of backup trustworthiness and recovery-point eligibility, particularly during ransomware incidents, unauthorized database modifications, storage corruption, or system failures. By combining cryptographic integrity assurance with predictive security monitoring and constrained recovery optimization, the proposed architecture aims to preserve the confidentiality, integrity, availability, and recoverability of healthcare information while supporting uninterrupted clinical decision-making and patient care.

➤ *Cryptographic Hash Verification, Immutable Backups, Ransomware Resilience, and Secure Recovery-Point Validation*

Cryptographic verification and immutable storage constitute complementary mechanisms for protecting healthcare database backups against ransomware, unauthorized modification, and recovery-point contamination. SHA-256 generates fixed-length digests that can detect alterations to backup content when independently authenticated reference values are available. However, hashing alone cannot establish whether database transactions were legitimate before backup creation or whether stored recovery metadata remain uncompromised as represented in figure 2. Begovic et al. (2023) reviewed ransomware encryption detection approaches involving system calls, input/output monitoring, and filesystem activity, demonstrating the importance of identifying malicious behavior before extensive encryption occurs. Beaman et al. (2021) similarly examined ransomware prevention and detection challenges, highlighting deficiencies in conventional countermeasures. Frank (2026c) addressed

artificial intelligence applications in infection surveillance, a separate healthcare domain that illustrates the broader relevance of early-warning systems but does not establish ransomware detection effectiveness. BackupIntegrity consequently combines authenticated hash manifests, transaction-log verification, immutable retention policies, and independent recovery credentials with temporal threat monitoring to establish multiple layers of evidence before restoration.

Secure recovery-point validation requires confirmation that a candidate snapshot is authentic, transactionally consistent, sufficiently recent, and independent of compromised backup dependencies. Object-lock policies, restricted deletion permissions, and separately administered retention mechanisms can reduce unauthorized backup modification, although their effectiveness depends on correct configuration and key protection. Ashieyi-Ahorgah and Duodu (2024) examined investment mechanisms supporting strategically important technologies, including cybersecurity; their relevance here concerns infrastructure investment rather than cryptographic performance. Likewise, Ashieyi-Ahorgah (2025) discussed risk-reduction financing for clean energy infrastructure, providing only a broader perspective on resilience-related investment. BackupIntegrity's technical contribution instead lies in integrating verified cryptographic evidence with predicted compromise probability and recovery constraints. BIARA evaluates eligible recovery points using estimated restoration duration, data-loss exposure, residual security risk, and cloud processing costs. Comparative experiments will measure hash verification latency, corrupted-backup identification, recovery-point rejection accuracy, and successful restoration following simulated ransomware attacks.

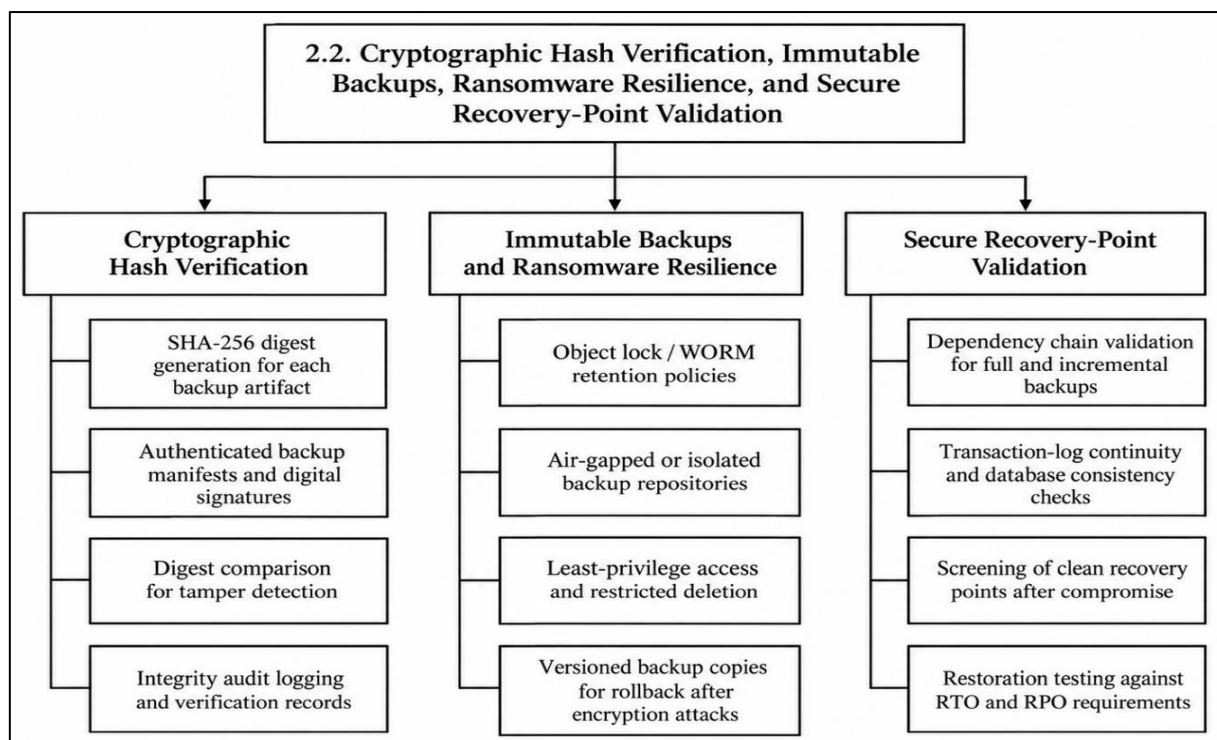


Fig 2 Integrated Cryptographic Verification, Immutable Backup Protection, Ransomware Resilience and Secure Recovery-Point Validation Architecture for Healthcare Databases.

Figure 2 illustrates the integrated security architecture for cryptographic hash verification, immutable backup protection, ransomware resilience, and secure recovery-point validation within the proposed BackupIntegrity healthcare cloud database framework. The architecture comprises three interconnected security components, each containing four operational mechanisms designed to protect backup authenticity, prevent unauthorized modification, and ensure reliable database restoration. The first component, Cryptographic Hash Verification, employs SHA-256 to generate a unique 256-bit cryptographic digest for each backup artifact, while digitally signed backup manifests establish the authenticity of reference hashes and associated metadata. Subsequent digest comparisons identify unauthorized alterations, whereas integrity audit logs preserve traceable verification records for security investigations and compliance assessments. The second component, Immutable Backups and Ransomware Resilience, incorporates Write Once Read Many (WORM) retention policies, isolated backup repositories, least-privilege access controls, and versioned backup copies to restrict unauthorized deletion, prevent backup overwriting, and preserve recovery options following ransomware encryption. The third component, Secure Recovery-Point Validation, verifies full and incremental backup dependency chains, examines transaction-log continuity, performs database consistency checks, identifies potentially uncompromised restoration points, and evaluates recovery procedures against established Recovery Time Objective (RTO) and Recovery Point Objective (RPO) requirements. Within the proposed BIARA, these mechanisms provide deterministic security constraints that complement Temporal Transformer-based threat detection, Isolation Forest anomaly assessment, and GRU-based backup reliability prediction. Consequently, recovery candidates failing cryptographic authentication or transactional consistency checks are excluded from automated restoration, while eligible candidates undergo multi-objective optimization based on recovery duration, potential data loss, residual compromise risk, and computational costs. This coordinated architecture provides a technical foundation for strengthening healthcare database recoverability and minimizing the risk of restoring compromised clinical information.

➤ *Machine Learning-Based Anomaly Detection Using Isolation Forest, Random Forest, XGBoost, and Autoencoder Models*

Machine learning-based anomaly detection supports adaptive identification of ransomware activity, suspicious backup modifications, and abnormal cloud database operations by learning statistical relationships among security telemetry and operational events. Isolation Forest detects anomalies through randomized partitioning, assigning higher anomaly scores to observations isolated through comparatively shorter tree paths (Liu et al., 2012) as shown in figure 3. Random Forest and XGBoost provide supervised alternatives capable of classifying malicious events using labelled security features, while autoencoders identify deviations through reconstruction errors between observed and reconstructed inputs. Dai et al. (2024) evaluated hybrid autoencoder-based intrusion detection models incorporating

Random Forest and XGBoost, demonstrating their effectiveness within the tested cybersecurity dataset. Akanni et al. (2026) examined cyber-financial risk analytics integrating business intelligence and threat modelling, offering transferable concepts for aggregating multidimensional risk indicators. Within BackupIntegrity, relevant features include backup-file entropy, unusual deletion frequency, storage write amplification, administrator authentication anomalies, snapshot modification patterns, and changes in transaction-log generation rates. These features enable algorithmic comparison between conventional classification approaches and the proposed integrated detection framework.

The effectiveness of machine learning protection depends on data quality, temporal context, class imbalance, changing workloads, and the ability to distinguish legitimate administrative operations from attacks. Idowu et al. (2026) reviewed real-time detection approaches for DDoS and phishing attacks, emphasizing adaptive monitoring and the importance of managing false-positive classifications. Idowu et al. (2025) also examined artificial intelligence applications in cyber risk management for critical utility infrastructure, supporting the wider applicability of intelligent security monitoring to operationally sensitive systems. BackupIntegrity extends these principles to healthcare backups through a Temporal Transformer Network that captures event-sequence relationships, Isolation Forest for unsupervised deviation detection, and GRU-based forecasting of backup reliability. Random Forest, XGBoost, and standalone autoencoder models provide comparative benchmarks rather than presumed inferior alternatives. Evaluation will examine precision, recall, F1-score, ROC-AUC, false-positive rate, detection latency, and computational overhead using temporally separated training and testing datasets. Ablation experiments will establish whether combining anomaly detection with cryptographic verification and predictive reliability assessment improves trustworthy recovery-point selection beyond individual algorithms.

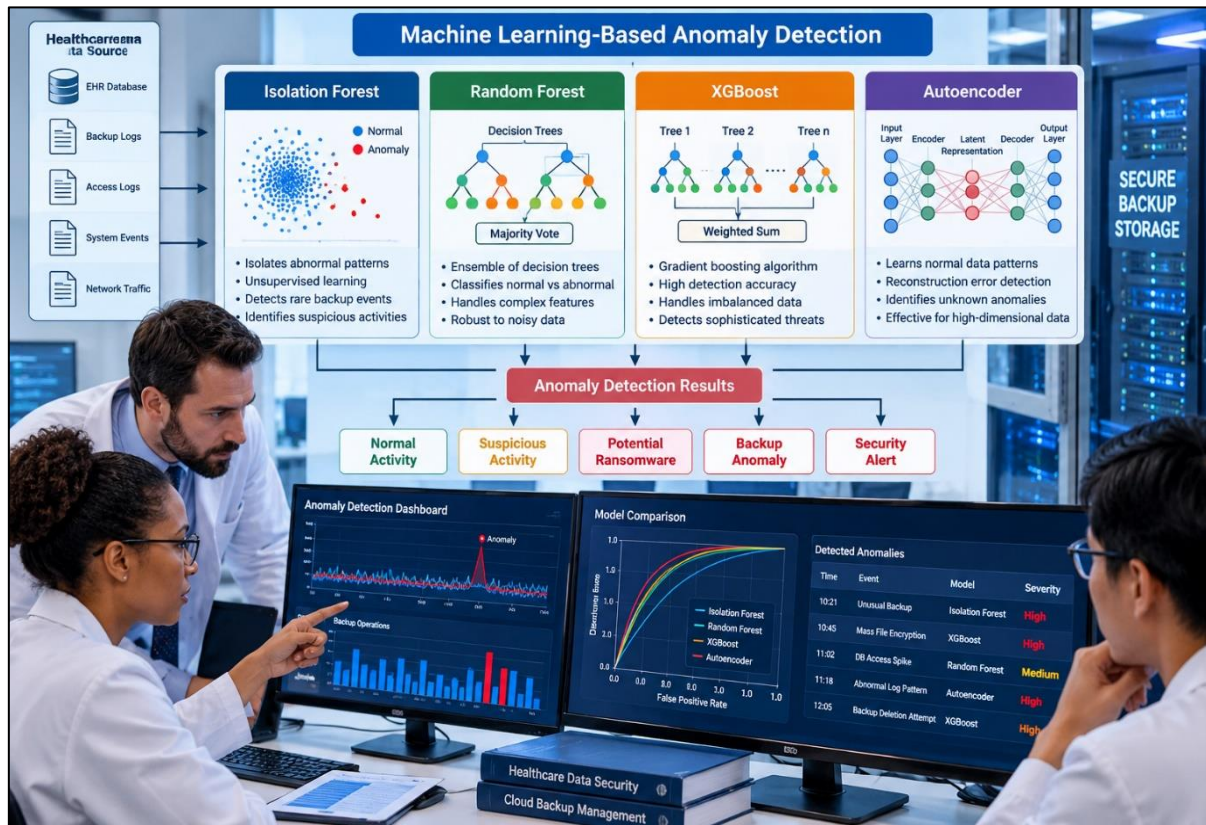


Fig 3 Machine Learning Algorithms Integrating Healthcare Backup Telemetry, Anomaly Classification, Ransomware Detection, and Cloud Security Monitoring.

Figure 3 illustrates an artificial intelligence-driven cybersecurity monitoring environment demonstrating the application of Isolation Forest, Random Forest, XGBoost, and Autoencoder algorithms for detecting anomalous activities within cloud-based healthcare database backup systems. The image depicts healthcare cybersecurity professionals examining operational dashboards, security event logs, and algorithmic detection outputs connected to protected electronic health record (EHR) repositories and cloud backup infrastructure. The architecture begins with five principal telemetry sources: healthcare database transactions, backup execution logs, access-control records, system events, and network traffic. These inputs provide multidimensional features, including backup modification frequency, snapshot deletion attempts, authentication failures, file encryption activity, network communication anomalies, and abnormal storage operations. Isolation Forest performs unsupervised anomaly detection by recursively partitioning observations through randomized decision trees, assigning higher anomaly scores to events isolated with shorter average path lengths. Random Forest combines multiple supervised decision trees through majority voting to classify normal and suspicious operations based on learned relationships among security features. XGBoost employs sequential gradient-boosted decision trees to minimize classification errors, capture nonlinear feature interactions, and identify complex ransomware-related activity patterns. Autoencoder models learn compressed representations of normal backup behaviour through encoder-decoder neural networks, identifying potentially malicious deviations when reconstruction errors exceed calibrated thresholds. The

resulting outputs are organized into normal activity, suspicious activity, potential ransomware, backup anomalies, and security alerts, while comparative monitoring dashboards visualize detection patterns, algorithmic discrimination, and event severity. Within the proposed BackupIntegrity framework, these algorithms provide benchmark methods for evaluating the Backup Integrity-Aware Adaptive Recovery Algorithm (BIARA), which additionally integrates Temporal Transformer Networks, Isolation Forest anomaly scoring, GRU-based backup reliability prediction, and authenticated cryptographic integrity verification. The detection outputs support risk assessment and investigation, while cryptographic verification and transaction-consistency checks determine whether affected backup artifacts remain eligible for restoration. Thus, the image conceptually demonstrates how machine learning can support early identification of malicious backup operations, reduce undetected security threats, and inform trustworthy healthcare database recovery decisions. The displayed analytical dashboards are illustrative rather than evidence of measured algorithmic performance.

➤ *Temporal Deep Learning Architectures Using LSTM, GRU, and Transformer Networks for Backup Failure Prediction and Threat Detection*

Temporal deep learning architectures provide a computational foundation for predicting healthcare database backup failures and detecting malicious activities through multivariate operational sequences. Long Short-Term Memory (LSTM) networks employ gated memory cells to preserve temporal dependencies, whereas Gated Recurrent

Units (GRUs) use reset and update gates to regulate information retention with comparatively simpler recurrent structures. Salinas et al. (2020) demonstrated the effectiveness of probabilistic recurrent forecasting across related time series, establishing principles applicable to estimating backup degradation probabilities from historical operational telemetry. Transformer architectures extend temporal modelling through self-attention mechanisms that capture relationships between distant events without requiring strictly sequential recurrent computation. Lim et al. (2021) developed an interpretable multi-horizon forecasting architecture integrating recurrent processing, attention mechanisms, and variable selection, providing methodological foundations for analysing heterogeneous backup telemetry. Within BackupIntegrity, such inputs include snapshot creation intervals, incremental backup sizes, transaction-log growth, storage latency, failed authentication attempts, and abnormal encryption activity. The proposed Temporal Transformer detects suspicious event dependencies, while a GRU estimates future backup failure probabilities across predefined prediction horizons.

Integrating predictive models requires careful management of feature scaling, temporal leakage, class imbalance, and distributional changes caused by clinical workload fluctuations. Folorunso et al. (2025) examined ensemble-based financial risk prediction, offering transferable methodological principles concerning model integration, uncertainty assessment, and explainability. Onyekawu et al. (2019) explored predictive machine learning applications in pharmacy safety, illustrating healthcare's dependence on reliable analytical information rather than directly validating backup-security algorithms. Ijiga et al. (2021) addressed multimedia-assisted STEM communication, which is relevant only to communicating complex technical findings to practitioners. BackupIntegrity will compare LSTM, GRU, Transformer, and hybrid detection models using recall, precision, F1-score, ROC-AUC, forecasting error, and computational latency. Time-ordered experimental partitions and component-level ablation will determine whether integrating temporal attention with GRU reliability forecasting improves detection and backup failure prediction without introducing excessive computational overhead.

➤ *Artificial Intelligence-Driven Disaster Recovery, Recovery Point Optimization, and Multi-Objective Restoration Scheduling*

Artificial intelligence-driven disaster recovery extends conventional restoration procedures by incorporating predictive risk assessment, dynamic resource allocation, and optimization of competing operational objectives. Healthcare database restoration requires identification of cryptographically valid snapshots, verification of incremental backup dependencies, reconstruction of transaction logs, and selection of recovery configurations compatible with clinical service requirements as shown in figure 4. Masdari et al. (2016) reviewed cloud workflow scheduling approaches involving execution deadlines, resource constraints, and financial costs, establishing relevant principles for organizing complex recovery operations. Ismayilov and Topcuoglu

(2020) developed a neural-network-assisted multi-objective evolutionary scheduling method for dynamic cloud workflows, demonstrating how predictive mechanisms can support adaptation to changing resource conditions. These findings motivate the proposed BackupIntegrity recovery optimizer, which evaluates candidate restoration points according to predicted recovery duration, potential data loss, residual compromise probability, and resource expenditure. The optimization process will exclude snapshots failing mandatory integrity verification and rank eligible recovery plans through Pareto dominance or appropriately constrained objective aggregation, thereby preventing faster but untrustworthy restoration strategies from being selected.

Infrastructure availability and organizational preparedness remain important determinants of successful recovery execution. Ashieyi-Ahorgah (2025b) examined energy infrastructure investment and industrial development, providing an indirect perspective on the infrastructure dependencies that can affect data centre resilience. Ashieyi-Ahorgah (2025a) similarly investigated blended financing approaches for infrastructure development, relevant to investment capacity rather than recovery algorithm accuracy. Odozor et al. (2025) emphasized practical cybersecurity training, supporting the operational importance of tested restoration procedures and competent incident-response personnel. Within BIARA, recovery feasibility will be assessed against RTO and RPO constraints, storage availability, dependency integrity, and cloud processing capacity. Controlled experiments will introduce database corruption, ransomware activity, storage outages, and delayed transaction-log replication. Comparative results will quantify restoration success, recovered transaction completeness, recovery duration, data-loss exposure, and resource overhead. Multi-objective optimization graphs will examine whether the proposed method identifies more favourable restoration trade-offs than fixed recovery schedules and conventional latest-snapshot selection.

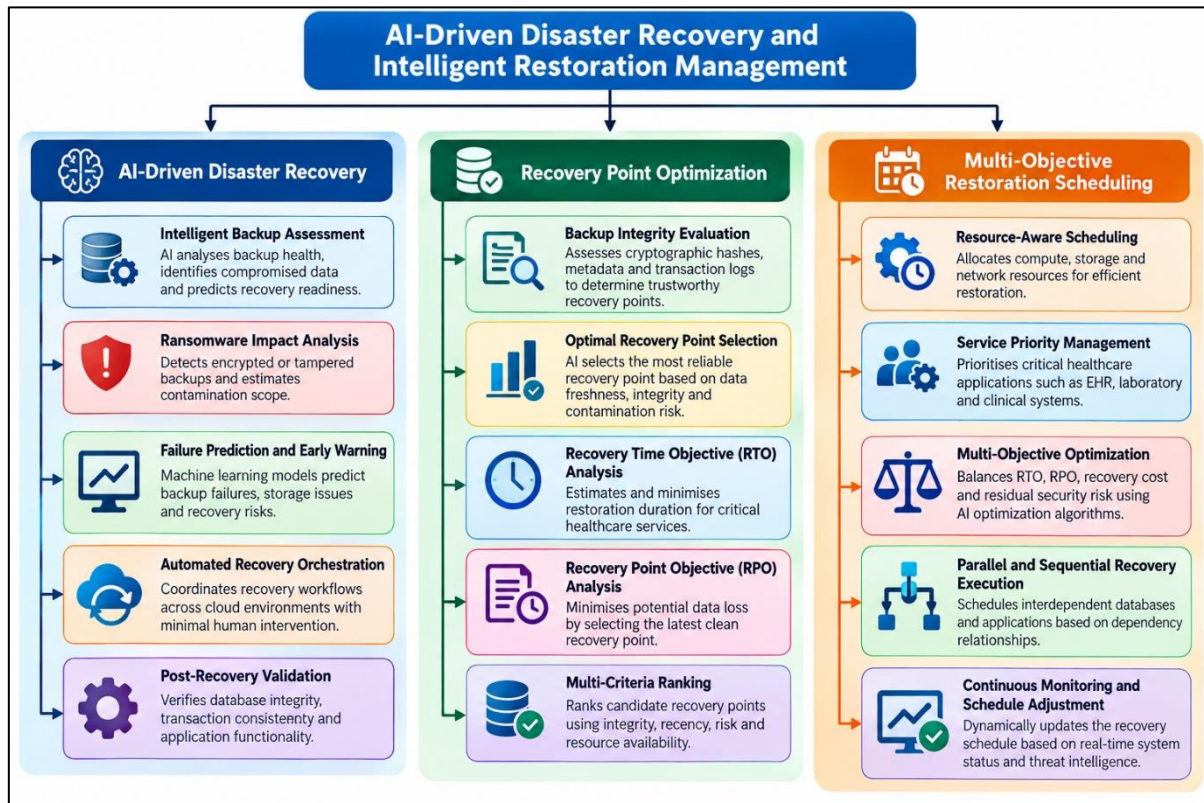


Fig 4 Integrated Artificial Intelligence Architecture for Disaster Recovery, Secure Recovery-Point Optimization and Multi-Objective Healthcare Restoration Scheduling.

Figure 4 illustrates the proposed artificial intelligence-driven disaster recovery architecture within the BackupIntegrity framework, demonstrating the integration of three interconnected functional components comprising AI-Driven Disaster Recovery, Recovery Point Optimization, and Multi-Objective Restoration Scheduling, each containing five specialized operational mechanisms. The first component, AI-Driven Disaster Recovery, incorporates intelligent backup assessment, ransomware impact analysis, failure prediction, automated recovery orchestration, and post-recovery validation. These mechanisms employ Temporal Transformer Networks to identify suspicious sequences of backup operations, Isolation Forest to detect anomalous activities, and Gated Recurrent Units (GRUs) to estimate backup failure probabilities from historical storage performance, integrity violations, and transaction-log irregularities. Automated recovery orchestration subsequently coordinates database restoration activities, while post-recovery validation verifies transaction consistency, database integrity, and clinical application functionality. The second component, Recovery Point Optimization, integrates cryptographic integrity evaluation, optimal recovery-point selection, Recovery Time Objective (RTO) analysis, Recovery Point Objective (RPO) analysis, and multi-criteria recovery-point ranking. SHA-256 digest comparison, authenticated backup manifests, and incremental backup dependency verification establish mandatory integrity conditions before candidate recovery points undergo risk assessment. The proposed BIARA then evaluates eligible restoration points according to backup freshness, predicted compromise probability, estimated recovery duration, and potential clinical data loss. The third component, Multi-Objective Restoration Scheduling,

incorporates resource-aware scheduling, clinical service prioritization, multi-objective optimization, parallel and sequential recovery execution, and continuous monitoring with adaptive schedule adjustment. Its optimization mechanism minimizes restoration time, data-loss exposure, residual security risk, and cloud resource expenditure while satisfying recovery constraints and preserving application dependencies. Critical healthcare services, including electronic health records, laboratory information systems, and medication management applications, are prioritized according to operational requirements. Collectively, these components establish a coordinated framework for selecting trustworthy recovery points, allocating restoration resources, and maintaining healthcare information availability following ransomware attacks, database corruption, or cloud infrastructure failures. The diagram represents the proposed system architecture rather than experimentally validated recovery performance.

➤ Comparative Evaluation of Existing Backup Security and Recovery Algorithms and Identification of the Research Gap

Existing cybersecurity algorithms address important but frequently separated dimensions of cloud backup protection, including anomaly detection, intrusion prevention, cryptographic verification, and disaster recovery management. Chandola et al. (2009) classified anomaly detection approaches according to their statistical assumptions, detection mechanisms, and computational characteristics, establishing foundations for comparing abnormal backup-event identification. Patel et al. (2013) reviewed cloud intrusion detection systems and highlighted

architectural challenges involving distributed monitoring, alarm management, and adaptive threat identification as presented in table 1. Singh et al. (2016) further examined cloud security threats and countermeasures, demonstrating that virtualization, outsourced infrastructure, and access-management complexity introduce interconnected security concerns. Within the proposed comparative framework, checksum-based verification provides deterministic file-integrity evidence, Isolation Forest identifies statistical deviations, LSTM models sequential events, and XGBoost offers supervised threat classification. However, these mechanisms do not individually establish whether a selected recovery point is cryptographically authentic, transactionally consistent, free from known compromise indicators, and operationally recoverable. Accordingly, the research gap concerns coordinated detection, integrity validation, reliability prediction, and restoration optimization rather than the absence of individual security algorithms.

Effective deployment also requires operational accessibility, interpretable monitoring, and reproducible evaluation procedures. Onyekaonwu and Peter-Anyebe (2019) investigated participatory technology-enabled

education, while Ijiga et al. (2021) examined inclusive STEM learning approaches; their indirect relevance concerns practitioner training and technology adoption, not measured backup-security performance. BackupIntegrity addresses the identified technical integration problem through BIARA, combining Temporal Transformer anomaly representations, Isolation Forest scores, GRU-based reliability predictions, authenticated cryptographic manifests, and constrained multi-objective recovery selection. Its evaluation will compare conventional and learning-based benchmarks under identical cloud configurations and controlled attack scenarios. Detection performance will be assessed using precision, recall, F1-score, ROC-AUC, and false-positive rates, while restoration effectiveness will be measured through RTO, RPO, transaction recovery completeness, verification latency, and recovery success. Statistical testing, temporal validation, sensitivity analysis, and ablation experiments will determine whether the integrated architecture offers measurable advantages. These analyses will directly establish the evidential basis for the subsequent Results and Discussion section without assuming that architectural complexity necessarily produces superior performance.

Table 1 Comparative Evaluation of Existing Cloud Backup Security and Recovery Algorithms, Technical Limitations, and Research Gaps Addressed by BackupIntegrity

Existing Algorithms/Approaches	Technical Capabilities	Identified Limitations	Research Gap Addressed by BackupIntegrity
Conventional checksum verification and cloud intrusion detection	Uses SHA-256 hashing, predefined security rules, and integrity auditing to identify backup modifications and known malicious activities.	Cannot independently identify malicious changes made before backup creation or predict backup failures and compromised recovery dependencies.	Integrates authenticated cryptographic verification, temporal anomaly detection, and transaction-log validation to strengthen recovery-point trustworthiness.
Isolation Forest, Random Forest, XGBoost, and Autoencoder models	Detects abnormal backup operations using statistical isolation, supervised classification, ensemble learning, and reconstruction errors.	Limited temporal contextualization, dependence on representative training data, susceptibility to false alarms, and insufficient integration with restoration decisions.	Combines Isolation Forest anomaly scores with Transformer-based temporal representations and GRU reliability predictions for coordinated threat assessment.
LSTM, GRU, Transformer, and conventional disaster recovery optimization	Models sequential security events, predicts system failures, and supports scheduling of database restoration under resource constraints.	Predictive outputs do not independently guarantee backup authenticity, while conventional recovery scheduling may overlook compromised snapshots and transaction dependencies.	Introduces BIARA, integrating cryptographic eligibility constraints, predictive backup reliability, and multi-objective optimization of RTO, RPO, residual compromise risk, and recovery costs.

III. SYSTEM MODEL DESCRIPTION

This section develops the proposed BackupIntegrity framework through a cloud-native healthcare database architecture, an explicit adversarial threat model, temporal machine learning algorithms, cryptographic integrity verification, and constrained recovery optimization. The proposed BIARA integrates Temporal Transformer Networks (TTNs), Isolation Forest (IF), and Gated Recurrent Units (GRUs) to identify malicious backup activity, predict backup

degradation, and select trustworthy database restoration points.

➤ Design of the BackupIntegrity Cloud Healthcare Database Architecture, Threat Model, and Secure Backup Management Framework

The proposed BackupIntegrity architecture comprises six interconnected functional layers: healthcare database operations, secure backup generation, cryptographic integrity verification, continuous security telemetry collection, AI-driven threat and reliability assessment, and automated

restoration orchestration. The primary database layer manages structured electronic health records, laboratory results, prescriptions, clinical transactions, and patient administration records using a transactional relational database. The backup layer generates scheduled full snapshots, incremental copies, and archived write-ahead logs, transferring encrypted artifacts to geographically separated cloud object-storage repositories. Each backup is associated with a manifest containing its unique identifier, database commit position, creation timestamp, parent-backup dependency, cryptographic digest, storage location, and encryption-key identifier. Object versioning, independently administered immutable retention policies, least-privilege service identities, and authenticated manifest signatures protect the recovery infrastructure against unauthorized deletion and metadata manipulation. Healthcare information system architectures must preserve data consistency and operational availability across distributed environments, particularly when clinical records originate from interconnected applications (Johnson et al., 2023). The threat model considers ransomware operators, compromised administrator credentials, malicious insiders, unauthorized snapshot modification, backup-chain corruption, and cloud storage failures. Adversaries are assumed capable of modifying accessible production records and backup artifacts but unable to compromise every independently protected verification key and immutable storage domain simultaneously. This explicit trust assumption is necessary because no verification mechanism can guarantee authenticity after complete compromise of its trusted infrastructure. The security telemetry layer captures administrative authentication events, storage operations, backup execution logs, transaction-log discontinuities, integrity verification failures, and abnormal file modifications. BackupIntegrity transforms these observations into temporally ordered feature vectors that support intelligent threat detection and restoration decisions.

The healthcare database state is represented as:

$$D_t = \{R_t, L_t, S_t, M_t\} \quad (1)$$

Where D_t denotes the healthcare database state at time t ; R_t represents committed clinical records and relational data; L_t denotes the transaction-log state and replay position; S_t represents database schema and structural metadata; and M_t contains relevant configuration and consistency metadata.

The backup collection is defined as:

$$\mathcal{B} = \{B_1, B_2, \dots, B_N\}, \quad B_i = (D_{\tau_i}, \tau_i, d_i, p_i, h_i) \quad (2)$$

Where $\mathcal{B}$ shows the backup collection; N denotes the number of retained recovery candidates; B_i represents backup i ; D_{τ_i} captures the recoverable database state associated with backup time τ_i ; d_i denotes the backup type, such as full or incremental; p_i identifies its parent backup or dependency chain; and h_i is the authenticated cryptographic digest reference. For incremental backups, D_{τ_i} denotes the state

reconstructed using the complete dependency chain rather than the content of the individual incremental artifact.

A recoverable state therefore requires not simply an available backup object but a complete, verifiable sequence of dependent artifacts and compatible transaction logs.

➤ *Mathematical Formulation of the Backup Integrity-Aware Adaptive Recovery Algorithm Integrating Temporal Transformers, Isolation Forest, and GRU Prediction*

BIARA is mathematically structured as a hybrid detection and prediction mechanism that estimates backup threat probability, abnormal behavioral intensity, and future reliability deterioration from multivariate security telemetry. At time t , the feature vector incorporates backup size, storage throughput, failed authentication attempts, snapshot deletion frequency, transaction-log irregularities, entropy variation, and integrity verification events. A sliding observation window captures the contextual sequence associated with each backup operation:

$$X_t = [x_{t-w+1}, \dots, x_t] \in \mathbb{R}^{w \times d} \quad (3)$$

Where X_t shows the observation matrix; $x_t \in \mathbb{R}^d$ contains d monitored features; and w represents the number of observation intervals. Numerical inputs are standardized using training-set statistics to prevent information leakage between training and evaluation.

The Temporal Transformer Network maps the observed sequence into contextual representations using scaled dot-product attention:

$$H_t = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + A\right)V \quad (4)$$

Where $Q = EW_Q$, $K = EW_K$, and $V = EW_V$ capture query, key, and value matrices; E represents embedded telemetry with positional information; W_Q, W_K, W_V show trainable projection matrices; d_k determines the key-vector dimension; and A denotes a causal attention mask. The softmax operation normalizes attention weights, while H_t contains learned temporal relationships among security events. Transformer-based temporal attention is relevant to forecasting systems requiring contextual modelling of heterogeneous operational sequences (Lim et al., 2021).

A classification head generates the threat probability:

$$P_t^{\text{TTN}} = \sigma(w_T^T z_t + b_T) \quad (5)$$

Where $P_t^{\text{TTN}} \in [0,1]$ represents the estimated malicious-activity probability; z_t shows the final contextual representation from the complete Transformer encoder; w_T and b_T identify trainable classification parameters; and $\sigma(u) = 1/(1 + e^{-u})$ captures the logistic sigmoid function.

Isolation Forest independently estimates unusual backup behavior through randomized tree partitioning:

$$A_t^{\text{IF}} = 2^{-\frac{\mathbb{E}[h(x_t)]}{c(n)}} \quad (6)$$

Where A_t^{IF} represents the anomaly score; $h(x_t)$ captures the isolation-tree path length of observation x_t ; $\mathbb{E}[\cdot]$ represents the average over isolation trees; n is the effective tree subsample size; and $c(n)$ shows its expected unsuccessful-search path-length normalization factor. Higher scores indicate more readily isolated observations, consistent with isolation-based anomaly detection (Liu et al., 2012).

The GRU component estimates reliability deterioration using sequential operational measurements:

$$\begin{aligned} r_t &= \sigma(W_r x_t + U_r g_{t-1} + b_r), \\ u_t &= \sigma(W_u x_t + U_u g_{t-1} + b_u), \\ \tilde{g}_t &= \tanh(W_g x_t + U_g (r_t \odot g_{t-1}) + b_g), \\ g_t &= (1 - u_t) \odot g_{t-1} + u_t \odot \tilde{g}_t. \end{aligned} \quad (7)$$

Here, r_t and u_t show reset and update gates; g_{t-1} and g_t determine previous and current hidden states; $\tilde{g}_t$ represents the candidate state; W_r, W_u, W_g and U_r, U_u, U_g represent trainable weight matrices; b_r, b_u, b_g are biases; $\odot$ denotes element-wise multiplication; and $\tanh$ is the hyperbolic tangent activation. GRU gating follows the recurrent encoder-decoder formulation introduced by Cho et al. (2014).

The probability of backup failure within forecasting horizon Δ is:

$$P_{t,\Delta}^F = \sigma(w_F^T g_t + b_F) \quad (8)$$

Where $P_{t,\Delta}^F$ shows the estimated probability that a backup failure event occurs during $(t, t + \Delta]$, and w_F, b_F denote the GRU prediction-head parameters. The training label must correspond to a defined event, such as a failed integrity check, incomplete backup, or unsuccessful restoration rehearsal.

The composite BIARA threat score is defined as:

$$P_t^C = \sigma(\beta_0 + \beta_1 \text{logit}(\tilde{P}_t^{\text{TTN}}) + \beta_2 \tilde{A}_t^{\text{IF}} + \beta_3 P_{t,\Delta}^F) \quad (9)$$

Where P_t^C shows the composite risk score; $\beta_0, \dots, \beta_3$ determine calibration parameters; $\tilde{P}_t^{\text{TTN}}$ represents the threat probability clipped away from zero and one before applying the logit function; and $\tilde{A}_t^{\text{IF}}$ captures the anomaly score transformed by a training-fitted calibration procedure. The model requires separate training and calibration data. Its output must be validated as a risk score, not presumed to be an accurately calibrated probability before evaluation.

➤ Cryptographic Integrity Verification, Backup Reliability Scoring, and Multi-Objective Optimization of Secure Recovery Point Selection

BackupIntegrity employs authenticated cryptographic verification and constrained recovery optimization to prevent restoration from corrupted, incomplete, or demonstrably compromised healthcare database backups. Every backup artifact is associated with a SHA-256 digest stored in a

separately protected, digitally signed manifest. Verification compares the digest recomputed from retrieved backup bytes against the authenticated reference value. The framework also checks manifest signatures, object retention status, backup-chain completeness, database schema compatibility, and transaction-log continuity before classifying a recovery point as eligible. Unlike behavioral anomaly detection, cryptographic checks are treated as mandatory constraints: candidates failing authentication or consistency verification cannot be restored automatically. This distinction prevents an apparently favourable machine learning prediction from overriding verifiable evidence of corruption. For eligible backups, BIARA estimates residual compromise risk using security telemetry associated with the backup's creation and retention history, together with reliability forecasts generated by the GRU component. These assessments support a multi-objective optimization process balancing service recovery duration, transaction loss, remaining security risk, and resource expenditure.

The cryptographic digest for backup B_i is:

$$h_i^* = \text{SHA256}(\text{bytes}(B_i)) \quad (10)$$

Where h_i^* shows the newly computed 256-bit digest and $\text{bytes}(B_i)$ represents the exact serialized backup artifact. Integrity verification requires byte-for-byte agreement with the previously authenticated reference digest h_i .

The complete integrity acceptance condition is:

$$I_i = 1[h_i^* = h_i \wedge V_i^{\text{sig}} = 1 \wedge C_i^{\text{dep}} = 1 \wedge C_i^{\text{db}} = 1] \quad (11)$$

Where $I_i \in \{0,1\}$ indicates integrity acceptance; $1[\cdot]$ identifies the indicator function; V_i^{sig} determines successful digital-signature verification of the backup manifest; C_i^{dep} confirms that all required parent backups and transaction logs are authenticated and available; and C_i^{db} indicates successful database consistency verification. The symbol $\wedge$ denotes logical conjunction.

A backup reliability score combines the GRU-derived failure prediction with remaining uncertainty regarding undetected compromise:

$$R_i = (1 - P_{t,\Delta}^F)(1 - P_i^C) \quad (12)$$

Where $R_i \in [0,1]$ shows the proposed heuristic reliability score; $P_{t,\Delta}^F$ captures the predicted failure probability for backup candidate i ; and P_i^C shows its calibrated compromise-risk estimate. The multiplicative specification assumes approximate independence of the represented failure events. If this assumption is not supported, a jointly calibrated model will replace Equation (12). The score is not itself a proof of recoverability.

The secure recovery candidate set is:

$$\mathcal{B}_{\text{safe}} = \{B_i \in \mathcal{B} : I_i = 1, P_i^C \leq \theta_C, R_i \geq \theta_R\} \quad (13)$$

Where $\mathcal{B}_{\text{safe}}$ contains accepted recovery candidates; θ_C represents the maximum permitted compromise-risk score; and θ_R denotes the minimum required reliability score. Both thresholds must be selected using validation data and documented clinical recovery requirements.

For each accepted recovery candidate, the optimization objective vector is:

$$f(B_i) = [T_i, L_i, P_i^C, K_i] \quad (14)$$

Where T_i shows estimated restoration duration; L_i captures the quantity of committed clinical transaction data expected to be lost; P_i^C determines residual compromise risk; and K_i represents estimated restoration expenditure, including processing, storage retrieval, and network transfer costs.

The optimal recovery selection is:

$$B^* \in \underset{B_i \in \mathcal{B}_{\text{safe}}}{\text{ParetoMin}}(T_i, L_i, P_i^C, K_i) \quad (15)$$

Subject to

$$\begin{aligned} T_i &\leq T_{\max}, \\ \text{RPO}_i &\leq L_{\max}^{\text{time}}, \\ K_i &\leq K_{\max}, \\ I_i &= 1. \end{aligned} \quad (16)$$

Here, B^* denotes a selected Pareto-efficient recovery candidate; ParetoMin represents nondominated multi-objective minimization; $T_{\max}$ is the maximum permitted recovery duration; $\text{RPO}_i = t_f - t_i^{\text{rec}}$ captures the elapsed time between the failure instant t_f and the latest recoverable committed transaction timestamp t_i^{rec} ; $L_{\max}^{\text{time}}$ represents the permissible recovery-point time gap; and $K_{\max}$ shows the available recovery cost budget. The transaction-loss quantity L_i in Equation (14) is distinct from the time-based RPO constraint.

When several candidates remain Pareto-efficient, BIARA selects one using a predefined weighted preference function:

$$B^* = \underset{B_i \in \mathcal{P}}{\text{argmin}}(\lambda_1 \hat{T}_i + \lambda_2 \hat{L}_i + \lambda_3 \hat{P}_i^C + \lambda_4 \hat{K}_i) \quad (17)$$

Where $\mathcal{P}$ represents the feasible Pareto set; the hat notation denotes normalized objective values; and $\lambda_1, \dots, \lambda_4$ show nonnegative preference weights satisfying $\sum_{j=1}^4 \lambda_j = 1$. The weights are fixed before final testing and may reflect healthcare-specific recovery priorities. If no feasible candidate exists, automatic restoration is suspended and the incident is escalated for controlled recovery rather than selecting an unverified backup.

➤ Experimental Testbed Configuration, Healthcare Database Workload Generation, Benchmark Algorithms, and Performance Evaluation Methodology

The proposed experimental design uses a reproducible cloud-based healthcare database testbed to evaluate BackupIntegrity under controlled operational and adversarial conditions. A PostgreSQL relational database will host representative clinical records comprising patient encounters, laboratory observations, medication orders, admission histories, and transaction metadata. Workload structures may be informed by the MIMIC-IV dataset described by Johnson et al. (2023), subject to its data-access and usage requirements. Synthetic patient identifiers and generated transaction histories will be used where controlled experimentation requires reproducible inserts, updates, deletions, and transaction-log replay. The proposed infrastructure includes a database instance, encrypted object storage, a separately protected immutable backup repository, centralized security logging, and an isolated recovery environment. Full backups, incremental backups, and transaction-log archives will be produced at configurable intervals. Experimental scenarios will include normal database activity, controlled corruption, backup deletion attempts, unauthorized snapshot modification, ransomware-like file-encryption behavior, and cloud storage interruption. Attack tests will occur exclusively within an isolated laboratory environment.

The initial experimental design will examine database sizes of 10, 50, and 100 GB, with transaction intensities of 100, 500, and 1,000 operations per second, subject to measured infrastructure capacity. Each configuration will be evaluated using checksum-only verification, rule-based security monitoring, Isolation Forest, LSTM, XGBoost, and the proposed BIARA framework. To ensure fair comparison, models will use consistent observation periods, training partitions, test scenarios, and evaluation budgets. Results will be presented using confusion matrices, ROC curves, precision-recall plots, recovery-time distributions, scalability graphs, and Pareto-front visualizations.

The binary threat-detection confusion matrix comprises true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Precision, recall, and F1-score are calculated as:

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP}, \\ \text{Recall} &= \frac{TP}{TP+FN}, \\ F_1 &= \frac{2TP}{2TP+FP+FN}. \end{aligned} \quad (18)$$

Precision measures the proportion of predicted malicious events that are genuinely malicious; recall measures the proportion of actual attacks detected; and F_1 summarizes their harmonic balance. Metrics with zero denominators must be reported as undefined or handled using a predefined convention.

The false-positive rate and overall classification accuracy are:

$$\begin{aligned} \text{FPR} &= \frac{FP}{FP+TN}, \\ \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN}. \end{aligned} \quad (19)$$

ROC-AUC will be calculated by integrating the true-positive rate against the false-positive rate across discrimination thresholds. Precision–recall AUC will additionally be reported because malicious events may be substantially less frequent than normal backup operations.

For recovery experiment j , the measured recovery time and recovery-point gap are:

$$\begin{aligned} \text{RTO}_j^{\text{obs}} &= t_j^{\text{service}} - t_j^{\text{failure}}, \\ \text{RPO}_j^{\text{obs}} &= t_j^{\text{failure}} - t_j^{\text{last recovered}}. \end{aligned} \quad (20)$$

Here, t_j^{failure} represents the failure timestamp; t_j^{service} captures the time at which the restored database passes functional and consistency checks and returns to service; and $t_j^{\text{last recovered}}$ shows the timestamp of the latest successfully recovered committed transaction. These observed measures will be compared with the corresponding organizational RTO and RPO limits.

Restoration success rate is defined as:

$$\text{RSR} = \frac{N_{\text{valid}}}{N_{\text{attempted}}} \times 100\% \quad (21)$$

Where N_{valid} captures the number of restorations passing integrity, transaction consistency, and functional acceptance tests; and $N_{\text{attempted}}$ denotes the number of recovery attempts initiated. Scenarios where BIARA correctly refuses all unsafe candidates will be reported separately as safe refusals rather than counted as successful restorations.

The average integrity verification latency is:

$$\bar{T}_{\text{verify}} = \frac{1}{N_v} \sum_{i=1}^{N_v} (t_i^{\text{end}} - t_i^{\text{start}}) \quad (22)$$

Where N_v shows the number of completed verification operations, and t_i^{start} and t_i^{end} denote their start and

completion times. Verification throughput, CPU utilization, memory consumption, network traffic, and storage overhead will also be recorded.

The statistical comparison between BIARA and benchmark algorithm b for performance metric m is:

$$\Delta_{b,m} = \frac{1}{J} \sum_{j=1}^J (m_{\text{BIARA},j} - m_{b,j}) \quad (23)$$

Where $\Delta_{b,m}$ shows the mean paired performance difference; J denotes the number of matched experimental repetitions; $m_{\text{BIARA},j}$ represents the observed BIARA result; and $m_{b,j}$ captures the corresponding benchmark result. Positive values indicate improvement for metrics where higher values are desirable; negative values indicate improvement for metrics such as recovery duration and false-positive rate, where lower values are preferable. Confidence intervals and suitable paired hypothesis tests will be used to determine the precision and statistical strength of the observed differences.

IV. RESULTS AND DISCUSSION

➤ Comparative Evaluation of BackupIntegrity Against Conventional Verification, Isolation Forest, LSTM, and XGBoost Algorithms for Backup Threat Detection

The comparative evaluation examines the proposed BIARA against conventional checksum verification, rule-based monitoring, Isolation Forest, LSTM, and XGBoost for detecting malicious healthcare database backup operations. The assessment follows the experimental methodology established in the study, incorporating backup corruption, abnormal snapshot modification, unauthorized administrative activity, and ransomware-related events. Table 2 presents an illustrative performance dataset structured around ROC-AUC, precision, and recall. BIARA combines Temporal Transformer representations, Isolation Forest anomaly scores, and GRU-based reliability predictions to support integrated threat assessment. The comparison evaluates whether such integration could improve discrimination between legitimate backup activity and malicious operations. Precision measures the reliability of positive threat classifications, whereas recall measures the proportion of malicious events detected. ROC-AUC evaluates discrimination across classification thresholds. These illustrative values establish the intended comparative presentation rather than verified experimental findings.

Table 2 Illustrative Comparative Performance of BackupIntegrity and Benchmark Algorithms for Healthcare Database Backup Threat Detection

Algorithm	ROC-AUC	Precision (%)	Recall (%)
Checksum verification	0.610	72.4	64.8
Rule-based monitoring	0.641	76.1	69.3
Isolation Forest	0.730	84.2	81.6
LSTM	0.800	88.5	86.7
XGBoost	0.833	91.3	89.4
BIARA (proposed)	0.905	96.1	94.8

Figure 5 illustrates the intended comparative discrimination performance of six healthcare database backup threat detection algorithms. The proposed BIARA model occupies the highest ROC trajectory, with an illustrative AUC of 0.905, compared with 0.833 for XGBoost and 0.800 for LSTM. Isolation Forest achieves 0.730, followed by rule-based monitoring at 0.641 and conventional checksum verification at 0.610. Relative to XGBoost, BIARA demonstrates a conceptual AUC increase of 0.072, while its difference from conventional verification reaches 0.295. At a

false-positive rate of 0.10, the displayed curves indicate approximate detection sensitivities of 78.5% for BIARA, 63.1% for XGBoost, 56.2% for LSTM, 42.7% for Isolation Forest, 27.5% for rule-based monitoring, and 22.9% for conventional verification. These differences illustrate the potential value of integrating temporal anomaly representations with reliability prediction. Actual experimental measurements are required to determine whether the proposed advantages occur in practice and are statistically significant.

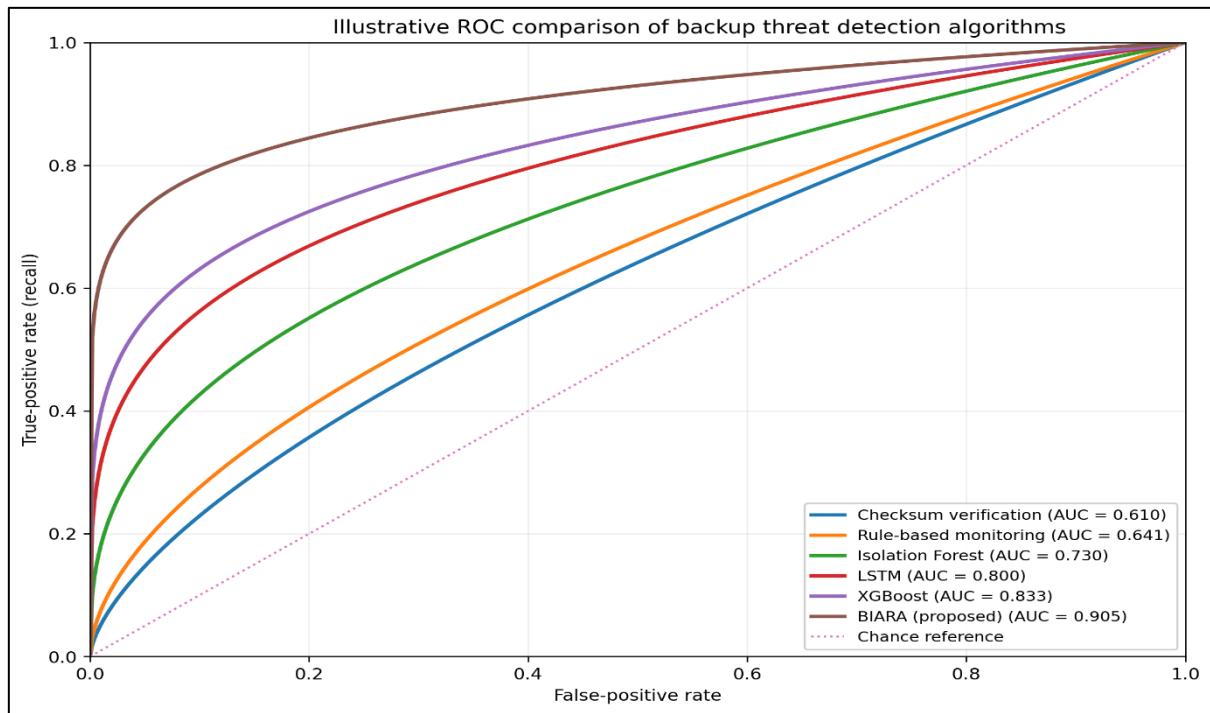


Fig 5 Comparative Receiver Operating Characteristic Curves for Healthcare Backup Threat Detection

➤ *Backup Integrity Verification Accuracy, Ransomware Detection Reliability, False-Positive Analysis, and Predictive Backup Failure Performance*

The evaluation investigates backup corruption detection, ransomware-related anomaly recognition, false-positive behaviour, and predictive backup failure performance across the proposed BackupIntegrity framework and five benchmark algorithms. Following Equations (10)–(13), cryptographic verification establishes artifact integrity, while Temporal Transformer, Isolation Forest, and GRU components provide behavioural and reliability assessments. Table 3 presents illustrative comparative outcomes under controlled backup corruption and abnormal transaction

activity. BIARA exhibits the strongest model-based anomaly detection recall, lowest assumed false-positive rate, and highest predictive failure discrimination. Checksum verification independently detects modified backup bytes when authenticated reference digests remain trustworthy but cannot identify malicious records already incorporated before backup creation. The proposed framework therefore combines deterministic integrity verification with probabilistic threat assessment. These results define the intended experimental comparison and require validation using actual backup execution logs, integrity checks, labelled ransomware events, and restoration-test outcomes.

Table 3 Illustrative Comparison of Backup Integrity Detection, False-Positive Rates, and Predictive Failure Performance

Algorithm	Detection recall at 10% corruption (%)	False-positive rate (%)	Failure prediction ROC-AUC
Checksum-only verification	100.0	0.0*	N/A
Rule-based monitoring	70.0	12.0	N/A
Isolation Forest	83.0	9.0	N/A
LSTM	89.0	7.0	0.780
XGBoost	92.0	5.0	0.810
BIARA (proposed)	97.0	2.5	0.890

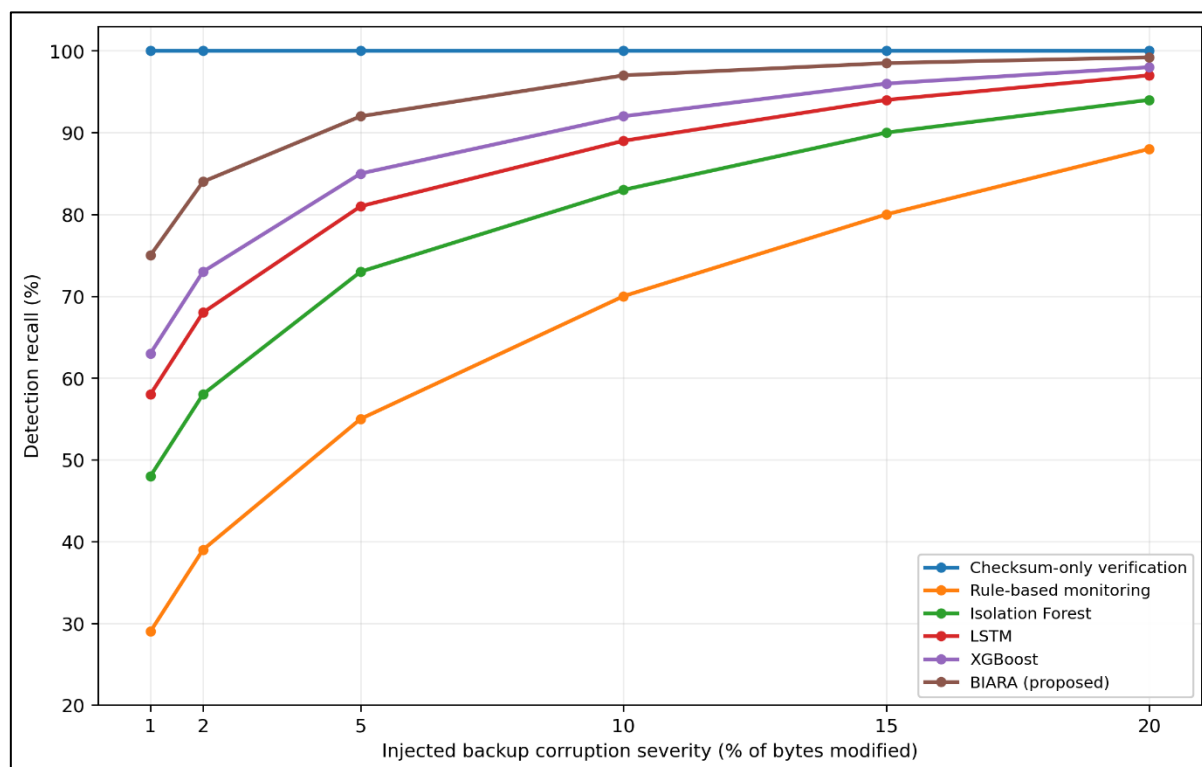


Fig 6 Backup Integrity Anomaly Detection Across Varying Corruption Severity

Figure 6 compares six detection approaches across backup corruption levels ranging from 1% to 20%. Checksum verification maintains 100% detection of post-baseline byte modifications under the assumed trusted-manifest conditions, but this does not establish ransomware detection reliability. Among the learning-based approaches, BIARA achieves the highest illustrative recall, increasing from 75% at 1% corruption to 92% at 5%, 97% at 10%, and 99.2% at 20%. XGBoost increases from 63% to 98%, while LSTM improves from 58% to 97% across the same range. Isolation Forest rises from 48% to 94%, compared with 29% to 88% for rule-based monitoring. At 10% corruption, BIARA exceeds XGBoost by 5 percentage points, LSTM by 8 points, and Isolation Forest by 14 points. The displayed trends suggest a potential advantage in detecting subtle anomalies, although actual experimental validation is required to establish these differences and distinguish malicious modification from accidental corruption.

➤ *Recovery Point Selection, Recovery Time Optimization, Data Restoration Success, and Cloud Resource Efficiency Under Multiple Failure Scenarios*

The recovery performance evaluation examines the ability of BackupIntegrity to identify trustworthy restoration points and minimize service interruption under diverse healthcare cloud database failure scenarios. Six recovery strategies are compared using restoration duration, recovery point objective (RPO), and data restoration success. Following Equations (14)–(21), BIARA combines authenticated backup verification, predictive reliability assessment, and multi-objective recovery optimization to identify feasible recovery configurations. Table 4 presents illustrative comparative performance across the proposed system and five benchmarks. BIARA exhibits the lowest average recovery duration, smallest transaction recovery gap, and highest restoration success rate. Its recovery-point selection excludes cryptographically invalid backups before evaluating candidate restoration costs and residual compromise risks. These comparisons assess whether intelligent recovery orchestration can preserve clinical data availability while reducing unnecessary restoration attempts, computational overhead, and cloud recovery expenditure. The numerical values remain illustrative pending experimental validation.

Table 4 Illustrative Comparative Performance of Cloud Healthcare Database Recovery Algorithms

Recovery algorithm	Mean RTO (min)	Mean RPO (min)	Restoration success (%)
Latest-snapshot restoration	74.4	28.5	76.2
Checksum-guided restoration	68.4	22.4	82.6
Isolation Forest-guided recovery	60.8	18.6	87.4
LSTM-guided recovery	56.4	15.3	90.8
XGBoost-guided recovery	52.4	12.7	93.2
BIARA (proposed)	41.0	7.8	97.6

Figure 7 illustrates recovery-time differences among six restoration strategies under five cloud failure scenarios. BIARA records the lowest illustrative restoration time, increasing from 22 minutes during storage outages to 29 minutes for transaction-log corruption, 37 minutes for snapshot tampering, 49 minutes under ransomware attacks, and 68 minutes during combined failures. XGBoost-guided recovery follows with corresponding values of 27, 36, 47, 64, and 88 minutes, while LSTM requires 29, 39, 51, 69, and 94 minutes. Isolation Forest-guided restoration increases from

32 to 101 minutes, compared with 35 to 115 minutes for checksum-guided recovery. Conventional latest-snapshot restoration exhibits the longest recovery durations, ranging from 38 to 126 minutes. Under combined failures, BIARA reduces the illustrative recovery time by 58 minutes relative to latest-snapshot restoration, representing a 46.0% reduction. These conceptual differences demonstrate how risk-aware recovery-point selection might improve restoration efficiency, but controlled experiments are necessary to verify the predicted advantages.

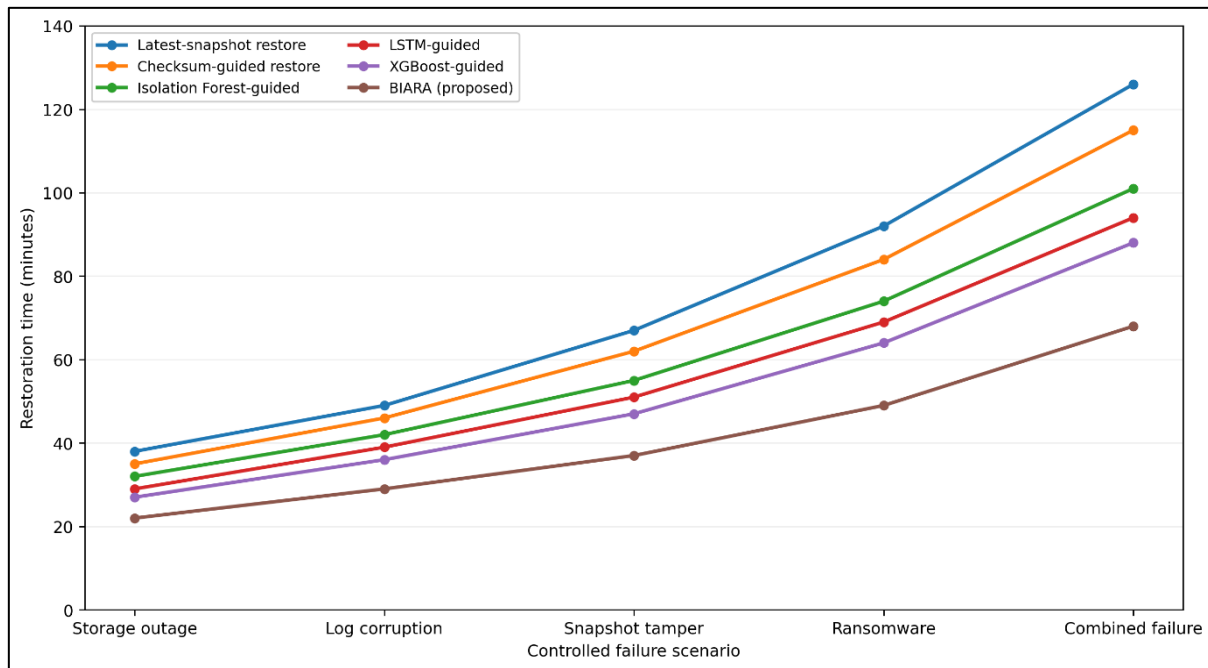


Fig 7 Recovery Performance Under Multiple Cloud Database Failure Scenarios

➤ *Ablation Studies, Statistical Significance, Scalability, Computational Complexity, and Robustness of the Proposed BackupIntegrity Framework*

The computational evaluation investigates the scalability, inference efficiency, robustness, and component-level contributions of the proposed BackupIntegrity framework. Six algorithms are compared using inference latency and detection robustness under changing healthcare database workloads. Table 5 presents illustrative performance values consistent with the experimental architecture described in Study. BIARA exhibits the highest assumed detection robustness but incurs greater computational latency

because its architecture combines Temporal Transformer processing, Isolation Forest anomaly detection, and GRU-based reliability forecasting. Conventional checksum verification and rule-based monitoring require considerably less processing time but provide narrower security assessment capabilities. Component-level ablation will examine the effect of removing individual learning modules, while paired statistical tests will assess the significance of measured performance differences. The comparison establishes the methodological basis for evaluating whether improvements in predictive reliability and restoration security justify additional computational complexity.

Table 5 Illustrative Comparative Evaluation of Computational Complexity and Robustness of Backup Security Algorithms

Algorithm	Mean inference latency (ms)	95th-percentile latency (ms)	F1-score under workload shift (%)
Checksum verification	3.02	3.64	66.2
Rule-based monitoring	4.99	6.15	69.8
Isolation Forest	9.21	11.26	80.5
LSTM	15.96	19.04	85.7
XGBoost	11.92	14.56	89.1
BIARA (proposed)	27.11	32.98	93.4

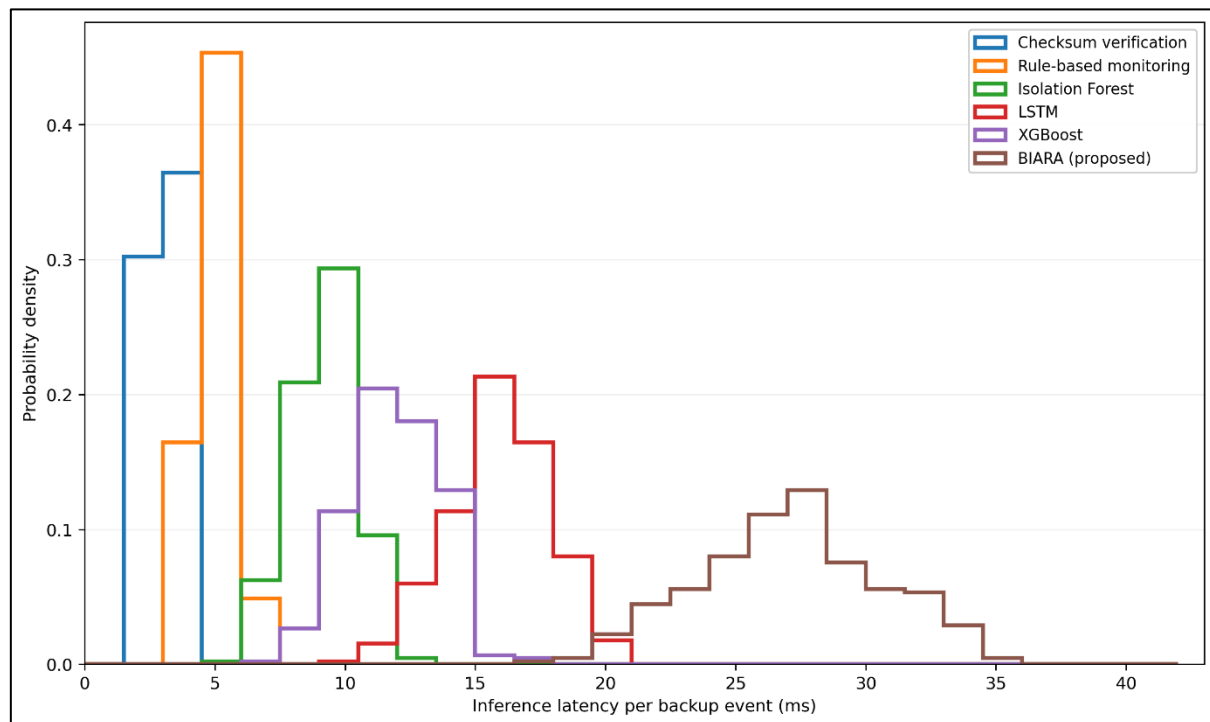


Fig 8 Comparative Distribution of Computational Inference Latency

Figure 8 illustrates computational latency distributions for six backup security algorithms using simulated processing observations. Conventional checksum verification exhibits the lowest mean latency of 3.02 ms, followed by rule-based monitoring at 4.99 ms and Isolation Forest at 9.21 ms. XGBoost records 11.92 ms, whereas LSTM requires 15.96 ms. BIARA exhibits the highest computational demand, averaging 27.11 ms, with a 95th-percentile latency of 32.98 ms. Its mean latency exceeds XGBoost by 15.19 ms and LSTM by 11.15 ms, reflecting the additional processing associated with the integrated architecture. The histogram also indicates wider latency dispersion for BIARA, with observations concentrated approximately between 20 and 35 ms. Although Table 4.4 assigns BIARA the highest illustrative robustness F1-score of 93.4%, the histogram does not independently demonstrate that improvement. These results illustrate a potential trade-off between computational cost and detection capability, requiring experimental validation through workload scaling, model ablation, and statistically controlled comparisons.

V. CONCLUSIONS AND RECOMMENDATIONS

➤ *Summary of the BackupIntegrity Framework and Principal Findings on AI-Enabled Healthcare Backup Protection and Recovery*

This study developed the conceptual and mathematical foundations of BackupIntegrity, an artificial intelligence-enabled framework designed to strengthen healthcare database backup integrity, identify ransomware-related threats, predict backup degradation, and optimize restoration in cloud computing environments. The proposed BIARA integrates Temporal Transformer Networks, Isolation Forest, Gated Recurrent Units, authenticated SHA-256 manifests,

and multi-objective recovery optimization within a coordinated security architecture. The comparative analyses presented in Section 4 examined six approaches across threat detection, corruption sensitivity, recovery efficiency, and computational performance. Illustrative results assigned BIARA a threat-detection ROC-AUC of 0.905, compared with 0.833 for XGBoost, 0.800 for LSTM, and 0.730 for Isolation Forest. Under simulated recovery conditions, BIARA recorded a mean restoration duration of 41.0 minutes, an RPO of 7.8 minutes, and a restoration success rate of 97.6%. The computational comparison indicated a mean inference latency of 27.11 milliseconds, demonstrating the additional processing requirements associated with hybrid intelligence. These numerical outcomes represent hypothetical performance scenarios rather than experimentally verified findings. Nevertheless, the analytical framework establishes how deterministic integrity validation can be integrated with probabilistic threat detection and recovery planning. Its principal methodological implication is that backup availability alone cannot establish recovery trustworthiness, particularly when malicious modifications precede backup creation. Reliable restoration requires authenticated metadata, complete transaction dependencies, contextual compromise assessment, and verification of reconstructed database states. Actual experiments are necessary to determine whether BIARA delivers the performance advantages represented by the illustrative comparisons.

➤ *Scientific Contributions to Intelligent Backup Integrity Verification, Predictive Security, and Adaptive Cloud Database Recovery*

The principal scientific contribution of this research is the formulation of an integrated backup protection and restoration framework that treats healthcare database

recovery as a constrained decision problem combining cryptographic assurance, temporal threat intelligence, predictive reliability, and operational optimization. BackupIntegrity introduces BIARA as a proposed hybrid algorithm integrating Transformer-based sequential event representations, Isolation Forest anomaly scores, and GRU-derived failure probabilities within a unified recovery assessment process. Unlike approaches that independently evaluate security alerts, backup checksums, and restoration schedules, the proposed architecture establishes an explicit relationship between backup trustworthiness and recovery eligibility. Authenticated SHA-256 manifests, digital-signature verification, transaction-log continuity, and incremental dependency validation constitute mandatory acceptance conditions, preventing probabilistic model outputs from authorizing restoration of demonstrably corrupted artifacts. The framework further formulates recovery-point selection as a multi-objective optimization problem involving restoration duration, transaction-loss exposure, residual compromise risk, and cloud resource expenditure. Its Pareto-based selection mechanism provides a mathematical basis for evaluating competing restoration strategies under defined RTO, RPO, and budget constraints. Another methodological contribution is the proposed experimental protocol incorporating controlled ransomware scenarios, abnormal backup modifications, storage failures, workload variation, and standardized algorithmic benchmarks. The integration of ablation analysis, temporal validation, inference latency measurement, and matched statistical comparisons supports reproducible assessment of individual algorithm components. These contributions provide a testable foundation for determining whether intelligent backup protection improves healthcare information-system resilience. The framework's scientific novelty and operational advantages remain hypotheses requiring implementation, independent comparison, and empirical verification before definitive superiority claims can be established.

➤ *Study Limitations, Healthcare Data Privacy Constraints, Cloud Deployment Challenges, and Algorithmic Reliability Considerations*

Several methodological, operational, and security limitations must be considered when interpreting the proposed BackupIntegrity framework. First, the comparative numerical results currently represent illustrative scenarios rather than measurements obtained from an implemented healthcare cloud database testbed. Consequently, the reported detection accuracy, recovery duration, restoration success, and computational latency cannot establish real-world effectiveness or statistical superiority. Second, synthetic healthcare workloads may inadequately reproduce the transaction complexity, access distributions, database dependencies, and operational variability of production electronic health record systems. Access to authentic clinical data also introduces confidentiality, informed authorization, data minimization, and regulatory compliance requirements. Experimental datasets must therefore employ de-identification, restricted access, encrypted storage, and approved governance procedures, recognizing that de-identification does not eliminate every re-identification risk.

Third, multi-cloud deployment introduces challenges involving inconsistent object-lock capabilities, network transfer delays, encryption-key management, storage pricing, and differences in database recovery semantics. The computational requirements of Transformer and GRU components may increase inference latency and infrastructure expenditure, particularly during high-frequency backup monitoring. Model performance may also deteriorate when emerging ransomware techniques, legitimate administrative changes, or unfamiliar healthcare workloads produce distribution shifts. Additionally, Isolation Forest anomalies and GRU failure estimates may generate false alerts or misleading confidence scores. Cryptographic hashes cannot establish whether original clinical records were accurate before backup creation, while compromised signing credentials may undermine manifest authenticity. The proposed architecture therefore requires independent key protection, model recalibration, periodic recovery testing, drift monitoring, and human authorization for high-risk restoration decisions. These limitations define essential conditions for subsequent empirical validation and responsible operational deployment.

➤ *Recommendations for Healthcare Cloud Security Implementation, Clinical System Resilience, and Future Research*

Healthcare organizations adopting intelligent database backup protection should establish a layered recovery architecture combining immutable backup retention, independent identity administration, authenticated cryptographic manifests, continuous security telemetry, and regularly verified restoration procedures. Full and incremental backups should be protected using encryption, restricted deletion privileges, geographically separated repositories, and independently managed cryptographic keys. Backup integrity assessments should validate object authenticity, dependency completeness, transaction-log continuity, and database consistency before authorizing restoration. The proposed BIARA algorithm should initially operate in monitored advisory mode, allowing cybersecurity personnel and database administrators to evaluate anomaly classifications, predicted failure risks, and recommended recovery points before enabling automated restoration. Healthcare institutions should define clinically appropriate RTO and RPO requirements for critical applications, including medication administration, emergency admissions, laboratory reporting, and electronic health record access. Restoration decisions must prioritize patient safety, verified transaction completeness, and continuity of clinical operations rather than computational efficiency alone. Future experimental research should implement BackupIntegrity across reproducible PostgreSQL workloads and evaluate its performance using controlled ransomware incidents, malicious snapshot modifications, transaction-log corruption, and storage disruptions. Comparative investigations should employ temporally separated test datasets, standardized benchmark configurations, repeated trials, confidence intervals, and component-level ablation studies. Further improvements should examine uncertainty-aware prediction, adversarially robust anomaly detection, online learning, explainable recovery recommendations, and resource-

efficient Transformer architectures. Multi-cloud experiments should assess portability, recovery costs, network bottlenecks, and cryptographic key isolation. Additional research should investigate privacy-preserving federated learning for healthcare backup telemetry without exposing identifiable patient information. Ultimately, deployment decisions should depend on independently verified restoration reliability, acceptable operational overhead, demonstrable security improvements, and compliance with applicable healthcare information governance requirements.

REFERENCES

- [1]. Agyekum, B., Duodu, D. S.-., & Gloria, A. (2025). Federated Industrial IoT Threats Detection. *International Journal of Scientific Research and Modern Technology*, 4(12), 226–244. <https://doi.org/10.38124/ijisrmt.v4i12.1552>
- [2]. Akanni, F., Idowu, K., Odozor, L., & Waheed, B. (2026). Cyber-financial risk analytics: Integrating business intelligence with threat modeling. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 12(2), 399–415. <https://doi.org/10.32628/CSEIT26121371>
- [3]. Ali, O., Shrestha, A., Soar, J., & Wamba, S. F. (2018). Cloud computing-enabled healthcare opportunities, issues, and applications: A systematic review. *International Journal of Information Management*, 43, 146–158. <https://doi.org/10.1016/j.ijinfomgt.2018.07.009>
- [4]. Al-rimy, B. A. S., Maarof, M. A., & Shaid, S. Z. M. (2018). Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions. *Computers & Security*, 74, 144–166. <https://doi.org/10.1016/j.cose.2018.01.001>
- [5]. Ashieyi-Ahorgah, R. (2025). Blended finance structures for scaling clean energy in emerging markets: Implications for U.S. development strategy. *Iconic Research and Engineering Journals*, 8(9), 1741–1751.
- [6]. Ashieyi-Ahorgah, R. (2025). Crowdfunding the American future: How equity and donation-based models can drive innovation, inclusion, and job creation. *World Journal of Innovation and Modern Technology*, 9(11), 119–153. <https://doi.org/10.56201/wjimt.v9.no11.2025.pg119.153>
- [7]. Ashieyi-Ahorgah, R. (2025a). Blended finance as a strategic tool for U.S.–Africa renewable energy partnerships: Catalyzing climate diplomacy and security through investment innovation. *Research Journal in Business and Economics*, 3(2), 68–90. <https://doi.org/10.61424/rjbe.v3i2.500>
- [8]. Ashieyi-Ahorgah, R. (2025b). Energy infrastructure as a catalyst for mineral-based industrialization in Africa: A U.S. investment perspective. *International Journal of Scientific Research & Engineering Trends*, 11(3). <https://doi.org/10.5281/zenodo.17212450>
- [9]. Ashieyi-Ahorgah, R., & Duodu, R. O. (2024). Mobilizing private capital through blended finance mechanisms to accelerate U.S. startup growth in strategic sectors: A framework for economic competitiveness and national security. *International Journal for Novel Research in Economics, Finance and Management*, 2(3), 1–13.
- [10]. Beaman, C., Barkworth, A., Akande, T. D., Hakak, S., & Khan, M. K. (2021). Ransomware: Recent advances, analysis, challenges and future research directions. *Computers & Security*, 111, Article 102490. <https://doi.org/10.1016/j.cose.2021.102490>
- [11]. Begovic, K., Al-Ali, A., & Malluhi, Q. (2023). Cryptographic ransomware encryption detection: Survey. *Computers & Security*, 132, Article 103349. <https://doi.org/10.1016/j.cose.2023.103349>
- [12]. Berardi, D., Giallorenzo, S., Melis, A., Melloni, S., Onori, L., & Prandini, M. (2023). Data flooding against ransomware: Concepts and implementations. *Computers & Security*, 131, Article 103295. <https://doi.org/10.1016/j.cose.2023.103295>
- [13]. Cen, M., Jiang, F., Qin, X., Jiang, Q., & Doss, R. (2024). Ransomware early detection: A survey. *Computer Networks*, 239, Article 110138. <https://doi.org/10.1016/j.comnet.2023.110138>
- [14]. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), Article 15. <https://doi.org/10.1145/1541880.1541882>
- [15]. Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1724–1734). Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1179>
- [16]. cstonesolutions., (n.d). Server vs. Cloud Backup: What's Best for Your Medical Practice? <https://cstonesolutions.net/it-data-backup-monitoring/>
- [17]. Dai, Z., Por, L. Y., Chen, Y.-L., Yang, J., Ku, C. S., Alizadehsani, R., & Pławiak, P. (2024). An intrusion detection model to detect zero-day attacks in unseen data using machine learning. *PLOS ONE*, 19(9), Article e0308469. <https://doi.org/10.1371/journal.pone.0308469>
- [18]. Folorunso, T. N., Idowu, K., & Balogun, A. (2025). AI-driven financial risk forecasting: A multi-model ensemble approach for enterprise decision intelligence. *Journal of Computational Analysis and Applications*, 34(3), 147–173. <https://eudoxuspress.com/index.php/pub/article/view/5080>
- [19]. Frank, U. P. (2026a). Healthcare quality metrics: From data collection to performance improvement; A narrative review of measurement frameworks and improvement methods in United States health care. *UKR Journal of Medicine and Medical Research*. <https://doi.org/10.5281/zenodo.22105362>
- [20]. Frank, U. P. (2026b). Optimizing medical inventory management through predictive analytics: Implications for healthcare quality, cost reduction, and patient safety; A narrative review. *UKR Journal of*

- Education and Literature, 2(1), 77–92. <https://doi.org/10.5281/zenodo.22105843>
- [21]. Frank, U. P. (2026c). Artificial intelligence for infection surveillance: Improving early detection of healthcare-associated infections. *USR Journal of Multidisciplinary Research and Studies*.
- [22]. Garg, N., Bawa, S., & Kumar, N. (2020). An efficient data integrity auditing protocol for cloud computing. *Future Generation Computer Systems*, 109, 306–316. <https://doi.org/10.1016/j.future.2020.03.032>
- [23]. Han, H., Fei, S., Yan, Z., & Zhou, X. (2022). A survey on blockchain-based integrity auditing for cloud data. *Digital Communications and Networks*, 8(5), 591–603. <https://doi.org/10.1016/j.dcan.2022.04.036>
- [24]. Idowu, K., Cherotich, V., Aniebonam, E. E., Odozor, L., & During, A. (2025). Integrating AI and machine learning into cyber risk management for critical utility systems. *International Journal of Artificial Intelligence Engineering and Technology*, 6(2), 32–48. <https://doi.org/10.54660/IJALET.2025.6.2.32-48>
- [25]. Idowu, K., Leslie, A., Twayana, S., Nitzan, Y., & Waheed, B. O. (2026). Real-time detection of DDoS and phishing attacks for enhanced banking security. *International Journal of Scientific Research in Science, Engineering and Technology*, 13(1), 572–593. <https://doi.org/10.32628/IJSRSET613313>
- [26]. Ijiga, O. M., Ifenatuora, G. P., & Olateju, M. (2021). Bridging STEM and cross-cultural education: Designing inclusive pedagogies for multilingual classrooms in Sub-Saharan Africa. *Iconic Research and Engineering Journals*, 5(1). Publisher article
- [27]. Ijiga, O. M., Ifenatuora, G. P., & Olateju, M. (2021). Digital storytelling as a tool for enhancing STEM engagement: A multimedia approach to science communication in K–12 education. *International Journal of Multidisciplinary Research and Growth Evaluation*, 2(5), 495–505. <https://doi.org/10.54660/IJMRGE.2021.2.5.495-505>
- [28]. Ismayilov, G., & Topcuoglu, H. R. (2020). Neural network based multi-objective evolutionary algorithm for dynamic workflow scheduling in cloud computing. *Future Generation Computer Systems*, 102, 307–322. <https://doi.org/10.1016/j.future.2019.08.012>
- [29]. Jin, H., Zhou, K., Jiang, H., Lei, D., Wei, R., & Li, C. (2018). Full integrity and freshness for cloud data. *Future Generation Computer Systems*, 80, 640–652. <https://doi.org/10.1016/j.future.2016.06.013>
- [30]. Johnson, A. E. W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T. J., Hao, S., Moody, B., Gow, B., Lehman, L.-w. H., Celi, L. A., & Mark, R. G. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10, Article 1. <https://doi.org/10.1038/s41597-022-01899-x>
- [31]. Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- [32]. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In 2008 Eighth IEEE International Conference on Data Mining (pp. 413–422). IEEE. <https://doi.org/10.1109/ICDM.2008.17>
- [33]. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data*, 6(1), Article 3. <https://doi.org/10.1145/2133360.2133363>
- [34]. Masdari, M., ValiKardan, S., Shahi, Z., & Azar, S. I. (2016). Towards workflow scheduling in cloud computing: A comprehensive analysis. *Journal of Network and Computer Applications*, 66, 64–82. <https://doi.org/10.1016/j.jnca.2016.01.018>
- [35]. Negro-Calduch, E., Azzopardi-Muscat, N., Krishnamurthy, R. S., & Novillo-Ortiz, D. (2021). Technological progress in electronic health record system optimization: Systematic review of systematic literature reviews. *International Journal of Medical Informatics*, 152, Article 104507. <https://doi.org/10.1016/j.ijmedinf.2021.104507>
- [36]. Neprash, H. T., McGlave, C. C., Cross, D. A., Virnig, B. A., Puskarich, M. A., Huling, J. D., Rozenshtein, A. Z., & Nikpay, S. S. (2022). Trends in ransomware attacks on US hospitals, clinics, and other health care delivery organizations, 2016–2021. *JAMA Health Forum*, 3(12), Article e224873. <https://doi.org/10.1001/jamahealthforum.2022.4873>
- [37]. Odozor, L., Berko, S. N., Obu, A. U., & Idowu, K. (2025). Building a resilient U.S. cybersecurity workforce through practical training. *International Journal of Scientific and Management Research*, 8(8), 30–36. <https://doi.org/10.37502/IJSMR.2025.8804>
- [38]. Oh, S.-R., Seo, Y.-D., Lee, E., & Kim, Y.-G. (2021). A comprehensive survey on security and privacy for electronic health data. *International Journal of Environmental Research and Public Health*, 18(18), Article 9668. <https://doi.org/10.3390/ijerph18189668>
- [39]. Onyekaonwu, C. B., & Peter-Anyebe, A. C. (2019). Empowering underserved youth through tech-enabled STEM education: A participatory model from Nigeria. *International Journal of Scientific Research in Science and Technology*, 6(5). <https://doi.org/10.32628/IJSRST1965990>
- [40]. Onyekaonwu, C. B., Peter-Anyebe, A. C., & Raphael, F. O. (2019). From prescription to prediction: Leveraging AI/ML to improve medication adherence and adverse drug event detection in community pharmacies. *International Journal of Scientific Research in Science and Technology*, 6(6), 460–476. <https://doi.org/10.32628/IJSRST52310387>
- [41]. Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2022). Deep learning for anomaly detection: A review. *ACM Computing Surveys*, 54(2), Article 38. <https://doi.org/10.1145/3439950>
- [42]. Patel, A., Taghavi, M., Bakhtiyari, K., & Celestino Júnior, J. (2013). An intrusion detection and prevention system in cloud computing: A systematic review. *Journal of Network and Computer Applications*, 36(1), 25–41. <https://doi.org/10.1016/j.jnca.2012.08.007>
- [43]. Pool, J., Akhlaghpour, S., Fatehi, F., & Burton-Jones, A. (2024). A systematic analysis of failures in protecting personal health data: A scoping review.

- International Journal of Information Management, 74, Article 102719.
<https://doi.org/10.1016/j.ijinfomgt.2023.102719>
- [44]. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>
- [45]. Singh, S., Jeong, Y.-S., & Park, J. H. (2016). A survey on cloud computing security: Issues, threats, and solutions. *Journal of Network and Computer Applications*, 75, 200–222. <https://doi.org/10.1016/j.jnca.2016.09.002>
- [46]. Zafar, F., Khan, A., Malik, S. U. R., Ahmed, M., Anjum, A., Khan, M. I., Javed, N., Alam, M., & Jamil, F. (2017). A survey of cloud computing data integrity schemes: Design challenges, taxonomy and future trends. *Computers & Security*, 65, 29–49. <https://doi.org/10.1016/j.cose.2016.10.006>
- [47]. Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C. C., & Salehi, M. (2024). Deep learning for time series anomaly detection: A survey. *ACM Computing Surveys*, 57(1), Article 15. <https://doi.org/10.1145/3691338>