Journal of Artificial Intelligence and Technology, (Ahead of Print)
https://doi.org/10.37965/jait.2026.1376

ISTP
RESEARCH ARTICLE

Structure-Aware Hierarchical SHA-256 Hashing for DICOM Integrity Verification: A Comparative Evaluation

Dendy Jonas,[1] Daniel Manongga,[2] Irwan Sembiring,[2] and Indrastanti Widiasari[2]
[1]Faculty of Information Technology, Satya Wacana Christian University and Raharja University, Salatiga, Indonesia
[2]Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia

(Received 25 April 2026; Revised 04 May 2026; Accepted 15 May 2026; Published online 28 June 2026)

Abstract: Integrity verification of Digital Imaging and Communications in Medicine (DICOM)-based medical data is critical in distributed healthcare systems. Conventional hashing methods operate at the file level and do not explicitly consider the internal structure of medical images. This study evaluates a structure-aware hierarchical hashing scheme that incorporates DICOM metadata, pixel-based descriptors, and entropy measures into a SHA-256-based framework. An experimental comparative design was implemented using 427 DICOM files. The proposed multi-component HMAC-SHA256 pipeline was compared with SHA-256, SHA-3, and locality-sensitive hashing (LSH). Evaluation metrics included entropy within the valid 0–8 range, bit distribution, chi-square statistics, Kolmogorov–Smirnov results, avalanche effect, and execution time. Results show that SHA-256, SHA-3, and the proposed method maintain statistically valid output distributions, while LSH shows weaker distributional consistency. The proposed method exhibits slightly greater entropy variability while remaining within valid bounds, balanced bit distribution, stable avalanche behavior, and lower deviation than LSH in distributional tests. Its mean execution time was 2.94 ms per file, compared with 1.12 ms for SHA-256, indicating moderate computational overhead due to multi-stage processing. These findings suggest that hierarchical decomposition can improve sensitivity to DICOM input structure while preserving deterministic hashing behavior. The study contributes a domain-aware hashing framework for medical data integrity verification without claiming cryptographic superiority.

Keywords: healthcare systems; hierarchical hashing; integrity verification; medical image; structure-aware

I. INTRODUCTION

Digital healthcare systems increasingly rely on the secure management of medical imaging data, where Digital Imaging and Communications in Medicine (DICOM) serves as the dominant standard for storing and exchanging clinical images [1]. A DICOM file is not a simple image container; it integrates pixel data with structured metadata such as patient identifiers, modality information, acquisition parameters, and diagnostic context.

This composite structure makes DICOM a critical component of clinical workflows, as both image content and metadata directly influence diagnostic interpretation, treatment planning, and regulatory compliance [2]. Consequently, maintaining the integrity of DICOM files is not only a technical requirement but a clinical necessity, as any unauthorized modification can compromise patient safety, diagnostic reliability, and medico-legal accountability.

The significance of this problem has intensified with the expansion of networked healthcare infrastructures, including Picture Archiving and Communication Systems (PACS), cloud-based imaging platforms, and inter-institutional data exchange systems.

While these technologies improve accessibility and interoperability, they also introduce new vulnerabilities. This growing reliance on distributed systems necessitates robust integrity verification mechanisms, where cryptographic hashing functions play a central role in ensuring data authenticity across system boundaries. DICOM files may be subject to tampering during storage, transmission, or retrieval, including metadata manipulation, pixel-level alteration, or malicious data injection. Such attacks are particularly concerning because they may not always be visually detectable, especially in grayscale medical images, yet can significantly alter clinical meaning [3]. As a result, robust and reliable integrity verification mechanisms are essential to ensure that DICOM data remains authentic, unaltered, and trustworthy across distributed healthcare environments.

In this context, cryptographic hash functions are widely used to verify data integrity because they are deterministic and highly sensitive to input changes. Standard algorithms such as SHA-256 generate a fixed-length digest that changes significantly even with minimal modifications to the input, making them suitable for detecting tampering. In medical data systems and blockchain-compatible architectures, hash functions play a central role in ensuring immutability and traceability. However, conventional hashing approaches treat the entire DICOM file as a single input without considering its internal structure.

While this approach ensures detection of any bit-level change, it does not exploit the semantic composition of DICOM files, where metadata and pixel data carry distinct and clinically relevant information. This limitation becomes more pronounced in contexts where structural sensitivity is important. A DICOM file contains heterogeneous components, including identity-bearing metadata, acquisition descriptors, and image-derived characteristics such as pixel intensity distributions [4]. Modifications in these components may have different clinical implications, yet standard hashing

Corresponding author: Dendy Jonas (e-mail: donoindarto987@gmail.com).

mechanisms do not differentiate between them during hash generation.

As a result, although integrity violations can be detected, the hashing process itself remains structurally generic and does not incorporate domain-specific characteristics that define the uniqueness of DICOM data, limiting its ability to reflect structured variation within the input. Existing research has addressed related challenges in fragmented ways. Studies on cryptographic hashing have focused on improving computational efficiency, structural design, and resistance to evolving attack models, while image-based security approaches have explored techniques such as region-based authentication, perceptual hashing, and watermarking to capture visual characteristics of image data [5]. Blockchain-based healthcare systems, in contrast, have emphasized decentralized storage and verification by maintaining hash values on-chain while storing large medical files off-chain.

However, these approaches do not converge on a unified solution that integrates DICOM-specific structural attributes into the hashing process itself. Cryptographic methods remain domain-agnostic, image-based methods largely exclude metadata, and blockchain systems typically treat the hash function as an unmodified primitive. Consequently, the potential for structure-aware hashing in DICOM integrity verification remains insufficiently explored. This limitation becomes particularly relevant in distributed medical data environments. Due to the large size of DICOM files, hybrid architectures are commonly employed, where data are stored off-chain and only hash representations are recorded on-chain. In such systems, the hash serves as the primary reference for verifying data integrity across distributed nodes [6]. Therefore, the effectiveness of integrity verification depends not only on the deterministic properties of the hash function but also on how well the hash reflects the internal composition of the underlying data.

Despite advances in cryptographic hashing, image-based integrity methods, and blockchain-enabled medical data management, existing approaches generally do not incorporate intrinsic DICOM attributes into hash construction. Most methods rely on file-level hashing or treat the hash function as a generic transformation. As a result, limited empirical evidence exists on whether metadata, pixel-level features, and entropy descriptors can improve the structural sensitivity of DICOM integrity verification.

Consequently, there is limited empirical evidence on whether a structure-aware hierarchical hashing approach can provide a more context-sensitive integrity representation in distributed healthcare systems. To address this gap, this study evaluates whether integrating DICOM structural attributes into a hierarchical SHA-256 hashing pipeline influences the statistical behavior of hash outputs relative to baseline hashing approaches [7]. The proposed approach constructs the final hash through a multi-stage process that integrates metadata attributes, entropy-based descriptors, and pixel-derived features before generating the final digest using HMAC-SHA256.

This study is guided by the following research question: How does a structure-aware hierarchical hashing scheme based on DICOM attributes perform compared with standard hashing approaches for medical data integrity verification? This objective is evaluative rather than causal, focusing on comparative performance and structural sensitivity rather than claiming cryptographic superiority. The objective of this study is to evaluate whether a structure-aware hierarchical hashing framework can provide a more context-sensitive integrity representation for DICOM data.

First, it introduces a structure-aware hierarchical hashing model that explicitly incorporates DICOM metadata and image-derived features into the integrity verification process. Second, it formalizes a multi-component hashing pipeline based on SHA-256 and HMAC operations that is compatible with hybrid off-chain medical data architectures. Third, it provides a comparative evaluation of the proposed method against baseline hashing approaches, including standard SHA-256 and alternative reference methods, using statistical randomness and distribution-based metrics. Fourth, it ensures correct interpretation of these metrics by treating them as indicators of statistical behavior rather than as standalone measures of cryptographic strength [8]. Through this approach, the study provides a domain-specific and methodologically grounded contribution to DICOM integrity verification while maintaining strict alignment between research objective, analytical design, and evidential scope.

II. LITERATURE REVIEW

A. CRYPTOGRAPHIC HASH FUNCTION ENHANCEMENTS

Cryptographic hash functions are widely used for data integrity verification because they generate deterministic fixed-length outputs that change substantially when input data are modified. In DICOM-based medical imaging systems, these functions are commonly used to verify whether stored or transmitted files remain unchanged. SHA-256 remains a common reference algorithm, while SHA-3 provides a modern cryptographic alternative with a different internal construction. However, both operate on input data as byte streams and do not explicitly distinguish between DICOM metadata and pixel data. Established algorithms such as SHA-256 are designed to produce deterministic outputs with strong avalanche properties, ensuring that even minimal changes in input produce substantially different hash values [9]. This characteristic has positioned SHA-256 as a widely adopted standard in secure data storage, blockchain systems, and integrity verification frameworks. Subsequent developments, including SHA-3 and high-performance alternatives such as BLAKE2 and BLAKE3, have further improved efficiency, structural diversity, and resilience to emerging attack models.

A consistent trajectory in this domain is the focus on algorithmic enhancement, where research prioritizes computational efficiency, collision resistance, and adaptability to evolving computational threats, including quantum-based attacks. Post-quantum hashing approaches, for example, aim to mitigate risks introduced by quantum search algorithms, while optimized implementations target performance improvements in large-scale systems. This body of work establishes that modern hash functions provide strong general-purpose integrity guarantees, making them suitable for a wide range of applications.

However, a critical limitation emerges when these approaches are applied to structured medical data such as DICOM files. These hash functions operate on DICOM files as undifferentiated binary streams, ignoring the structured separation between header metadata (patient identifiers and acquisition parameters) and pixel data, which together define the clinical meaning of the image. While this ensures robustness at the bit level, it introduces a structural abstraction that overlooks the semantic roles of different data components. In the context of DICOM, where metadata and pixel data carry distinct clinical significance, this abstraction limits the ability of the hashing process to reflect the intrinsic structure of the data. Thus, while cryptographic hashing is well established as a secure mechanism, it remains domain-agnostic, and it is unresolved

whether incorporating domain-specific attributes into the hashing process can enhance context-sensitive integrity verification [10]. This limitation motivates the examination of approaches that move beyond purely algorithm-centric design toward structure-aware mechanisms.

B. IMAGE-BASED INTEGRITY AND MEDICAL DATA SECURITY

Parallel to cryptographic hashing research, image-based integrity approaches have explored methods that leverage the inherent properties of visual data. Recent studies have further emphasized the importance of integrity verification in DICOM-based medical imaging. For example, a zero-watermarking framework based on feature extraction and singular value decomposition demonstrated strong robustness against common attacks while preserving diagnostic image quality, highlighting the effectiveness of non-intrusive integrity mechanisms in medical imaging environments. Similarly, non-embedding authentication approaches using block-wise cryptographic hashing have been proposed to avoid pixel distortion while ensuring reliable verification, addressing the limitations of traditional watermarking techniques. Techniques such as region-of-interest (ROI) hashing, perceptual hashing, and locality-sensitive hashing (LSH) are designed to capture structural and statistical characteristics of images, enabling detection of meaningful modifications while maintaining tolerance to benign transformations such as compression or noise [11]. These approaches represent a shift from purely bit-level integrity toward content-aware verification, where the objective is to preserve the semantic fidelity of the image.

A dominant pattern in this domain is the emphasis on pixel-level feature extraction, where integrity mechanisms are derived from image characteristics such as intensity distributions, spatial patterns, and frequency components. ROI-based methods further refine this approach by focusing on diagnostically relevant regions, ensuring that critical areas are protected with higher sensitivity [12]. These methods demonstrate that incorporating data-specific features can improve the detection of localized modifications, particularly in medical imaging contexts where visual integrity is essential.

However, this approach introduces a significant trade-off. While image-based methods improve sensitivity to pixel-level alterations, they largely exclude the metadata that defines the identity and context of the DICOM file. In medical imaging, metadata attributes such as patient identifiers, acquisition parameters, modality, and timestamps are integral to the clinical interpretation of the image [13]. By focusing exclusively on pixel data, image-based integrity methods provide only a partial representation of DICOM integrity, failing to account for changes in metadata that may have equally critical implications.

Furthermore, many of these methods are designed for specific purposes such as watermarking, authentication, or tamper detection within images, rather than for generating a unified integrity fingerprint suitable for distributed systems. This limits their applicability in blockchain-based environments, where a single, consistent hash representation is required for verification across multiple nodes [14]. Thus, while it is established that feature-based approaches enhance sensitivity to image modifications, it remains unresolved how such feature awareness can be integrated with metadata and system-level verification mechanisms. This limitation leads to the need for approaches that combine both image-derived and metadata-derived attributes within a unified framework. Although image-based approaches improve sensitivity to pixel-level modifications, their exclusion of metadata limits their applicability to full DICOM integrity verification. This limitation highlights the need for system-level solutions capable of integrating multiple data components, which is addressed in blockchain-based medical data integrity frameworks.

C. BLOCKCHAIN-BASED MEDICAL DATA INTEGRITY

Blockchain-based systems provide a distributed framework for medical data integrity verification, where hash values serve as immutable references across multiple stakeholders [15]. In medical imaging systems, blockchain is commonly used to store cryptographic hashes of DICOM files, enabling verification without storing large image files directly on-chain.

The adoption of blockchain technology in healthcare has introduced a decentralized paradigm for ensuring data integrity, traceability, and trust. In medical imaging systems, blockchain is commonly used to store cryptographic hashes of DICOM files, enabling verification of data authenticity without directly storing large datasets on-chain [16]. This has led to the widespread use of hybrid architectures, where DICOM files are stored off-chain in cloud or distributed storage systems such as IPFS, while their corresponding hash values are maintained on-chain as immutable references.

A key pattern in blockchain-based medical data systems is the separation of data storage and integrity verification, where the blockchain functions as a trusted ledger for validating data authenticity. Consortium blockchain models further extend this approach by allowing multiple healthcare stakeholders to participate in a shared network, enabling secure data exchange and collaborative verification [17]. These systems demonstrate that blockchain can effectively enhance trust and transparency in distributed medical environments.

However, a critical limitation persists in the treatment of the hashing mechanism itself. In most implementations, the hash function is treated as a black-box primitive, assumed to provide sufficient integrity verification without adaptation to the underlying data structure. As a result, research in this domain focuses primarily on system architecture, access control, and interoperability, rather than on the design of the hash function used for integrity representation [18]. This creates a disconnect between the structural complexity of DICOM data and the simplicity of the hash representation used to verify it.

This limitation is particularly significant because the hash stored on the blockchain serves as the authoritative reference for integrity verification. If the hash does not reflect the internal composition of the DICOM file, then the verification process remains limited to detecting changes without leveraging domain-specific information [19]. Thus, while blockchain-based systems establish a robust framework for distributed integrity verification, it remains unresolved whether the effectiveness of these systems can be enhanced through structure-aware hashing mechanisms. This gap motivates the integration of domain-specific hashing approaches within blockchain architectures.

D. SYNTHESIS AND GAP CONSOLIDATION

The analysis across cryptographic hashing, image-based integrity methods, and blockchain-based systems reveals a consistent limitation: none of these approaches fully integrate the structural

characteristics of DICOM data into a unified integrity verification mechanism. Cryptographic hashing provides strong general-purpose integrity but remains structurally generic. Image-based methods introduce feature awareness but are limited to pixel data and do not incorporate metadata. Blockchain-based systems enable distributed verification but rely on hash functions that are treated as black-box primitives without domain adaptation. Across all three domains, there is no unified approach that integrates these strengths into a single, coherent framework. In addition, recent work in medical image security has explored hybrid approaches combining hashing, segmentation, and embedding techniques for tamper detection. For instance, region-based hashing integrated with deep learning segmentation has been shown to improve detection of localized manipulation in DICOM images, particularly in telemedicine contexts. Digital forensic approaches applied to DICOM systems have also demonstrated the feasibility of detecting tampering based on file-level and structural inconsistencies, reinforcing the need for more robust integrity verification mechanisms.

This synthesis highlights a precise unresolved problem: the absence of a structure-aware hierarchical hashing framework that systematically incorporates DICOM metadata, pixel-derived features, and statistical descriptors into a unified integrity representation. Existing methods either operate at the algorithmic level without domain awareness, at the image level without metadata integration, or at the system level without modifying the hash construction process. Consequently, the potential for combining these dimensions into a single integrity mechanism remains unexplored.

This gap is particularly relevant for modern healthcare systems, where reliable data integrity forms the foundation for downstream analytics and intelligent decision-support processes. In such environments, ensuring that data are not only unaltered but also structurally verifiable is essential for maintaining trust in both clinical and computational workflows.

Therefore, the literature indicates a need to evaluate whether DICOM-specific metadata, pixel-level features, and entropy descriptors can be integrated into a unified hierarchical hashing process for more structure-sensitive integrity verification. This directly prepares the conceptual framework and methodology of the present study, where metadata attributes, pixel-level features, entropy-based descriptors, and hierarchical hashing are formally defined and operationalized as a unified integrity verification approach.

E. CONCEPTUAL FRAMEWORK

1). CONCEPTUAL POSITIONING. This study adopts a conceptual framework to structure and evaluate a computational process for DICOM integrity verification. The framework organizes metadata features, pixel data features, entropy descriptors, hierarchical hashing, and integrity verification into a single process model. Its purpose is to show how DICOM-specific components are transformed into intermediate hashes and then combined into a final integrity representation [20]. The literature review further demonstrates that this limitation is particularly critical in the context of DICOM files, where meaningful structural components such as metadata, pixel distributions, and statistical properties remain unutilized in integrity verification. Therefore, the framework is designed to organize and operationalize these structural components into a unified hashing pipeline that can be empirically evaluated without extending beyond the evidential limits of the study. The framework directly responds to the identified research gap by formalizing a structure-aware hierarchical hashing process that integrates multiple DICOM-specific attributes [21]. It does not attempt to redefine cryptographic security properties but instead focuses on how input representation affects sensitivity to structural variation, which is measurable through statistical analysis.

F. CORE CONSTRUCTS

The framework consists of five tightly defined constructs, each directly operationalized in the methodology and consistently used throughout the manuscript.

1). METADATA FEATURES. Metadata features refer to structured descriptive attributes embedded within DICOM files, including identifiers such as PatientID, Modality, Manufacturer, Study Date, and AcquisitionTime. These attributes are inherently heterogeneous across files and provide semantic-level variability that is not captured in conventional file-level hashing [22]. Within the framework, metadata is treated as a structured input component that contributes to the uniqueness of the hashing process by encoding contextual information about the medical image.

2). PIXEL DATA FEATURES. Pixel data features represent the raw imaging content of the DICOM file, typically encoded as grayscale intensity values. These features capture low-level structural information, including spatial variation and intensity distribution. Unlike metadata, which reflects contextual identity, pixel data reflects the visual and numerical composition of the medical image [23]. The framework treats pixel data as a distinct input component that contributes to hash generation through direct transformation into a byte-level representation.

3). ENTROPY DESCRIPTOR. The entropy descriptor represents the statistical distribution of pixel intensity values within the DICOM image, computed using Shannon entropy. It serves as a quantitative measure of variability in the input data, capturing the degree of dispersion in pixel values. Importantly, within this framework, entropy is strictly used as a descriptive feature of input structure and not as an indicator of cryptographic strength or security performance [24]. Its role is limited to enriching the representation of input variability prior to hashing, ensuring that variations in data distribution are explicitly encoded into the hierarchical hashing process.

4). HIERARCHICAL HASHING. Hierarchical hashing is the central operational construct of the framework. It refers to a multi-stage hashing process in which individual input components metadata, pixel data, and entropy are first independently transformed into intermediate hash values and subsequently combined into a unified representation. This combined representation is then processed through a final hashing stage. The hierarchical structure ensures that each component contributes independently to the final hash, preserving structural distinctions that would otherwise be lost in a single-pass hashing process.

5). INTEGRITY VERIFICATION. Integrity verification is the outcome construct of the framework. It refers to the ability of the hashing process to produce consistent and sensitive output changes when DICOM input components are modified. In this study, it is evaluated using entropy distribution, bit balance, chi-square statistics, Kolmogorov–Smirnov statistics, avalanche effect, and execution time.

In this study, integrity verification is operationalized through the analysis of statistical behavior, where metrics such as entropy distribution, bit balance, chi-square statistics, and

Kolmogorov–Smirnov tests are used as indicators of output consistency and sensitivity to input variation. These statistical properties do not represent cryptographic security guarantees but provide empirical evidence of how the hashing process responds to structural differences in input data.

G. RELATIONSHIPS BETWEEN CONSTRUCTS

The framework establishes a non-causal, process-based relationship among constructs. Metadata features, pixel data features, and entropy descriptors are treated as independent input components that are systematically integrated through hierarchical hashing. The relationship is defined as follows: the inclusion of multiple structurally distinct input components leads to a hashing process that is more sensitive to variations in input structure, because each component contributes independently to the final hash representation.

This increased structural sensitivity does not imply an increase in cryptographic strength, collision resistance, or theoretical security guarantees. Instead, it reflects the extent to which the hashing process captures and encodes variations in input composition. The relationship is therefore associative and operational, indicating that the structure of the input influences the behavior of the hashing output under empirical testing conditions.

Integrity verification emerges from this relationship as an evaluated outcome, where the statistical properties of the resulting hashes are analyzed to determine whether the hierarchical approach produces distinguishable and consistent patterns compared to baseline methods. The framework explicitly avoids interpreting these patterns as proof of superior security and instead treats them as observable characteristics of hashing behavior.

H. JUSTIFICATION OF THE FRAMEWORK

The framework is justified by the limitations identified in prior research. Conventional hashing methods do not differentiate between structural components of complex data formats such as DICOM, resulting in uniform treatment of heterogeneous information. Similarly, image-based integrity methods often focus exclusively on pixel-level transformations without incorporating metadata, while blockchain-based approaches treat hashing as an external mechanism without modification. By integrating metadata, pixel data, and entropy into a unified hierarchical structure, the framework addresses these limitations without introducing unsupported theoretical claims. It provides a minimal yet sufficient structure for evaluating whether incorporating domain-specific attributes affects the statistical behavior of hash outputs. Each construct is necessary, directly measurable, and aligned with the methodological design, ensuring that the framework remains analytically coherent and operationally feasible.

I. FRAMEWORK REPRESENTATION

The relationships among these constructs are summarized in Fig. 1.

The conceptual framework is represented as a sequential and layered process. Three primary input components metadata features, pixel data features, and entropy descriptor are first independently processed. These components are then transformed into intermediate hash representations using a consistent hashing mechanism. The intermediate hashes are combined into a structured composite input, which is subsequently processed through a final hashing stage. The output of this process is evaluated through statistical testing procedures that constitute the integrity verification stage.

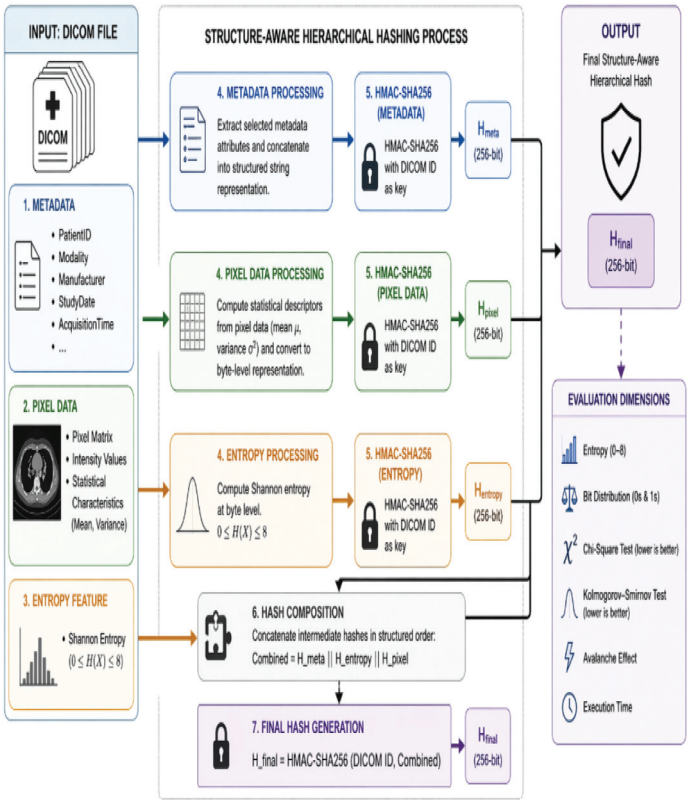

Fig. 1. Conceptual framework of structure-aware hierarchical hashing.

The Fig. 1 visually reflects the hierarchical nature of the process, showing distinct input layers, an integration layer, and an output evaluation layer. All elements represented in the Fig. 1 correspond directly to constructs defined in this section and are operationalized in the methodology without modification.

J. ALIGNMENT WITH RESEARCH OBJECTIVE

The conceptual framework directly supports the research objective of evaluating a structure-aware hierarchical SHA-256 hashing scheme for DICOM integrity verification. Each construct is explicitly operationalized in the methodology, each relationship is empirically testable, and the overall structure ensures consistency across all sections of the manuscript. The framework does not introduce any concept that is not measured, analyzed, or discussed later, ensuring full alignment with the study design and maintaining the integrity of the overall argument.

III. METHODS

A. RESEARCH DESIGN

This study adopts an experimental comparative evaluation design to assess the behavior of a structure-aware hierarchical SHA-256 hashing model relative to established hashing approaches. The design is directly aligned with the research objective, which is to evaluate whether incorporating intrinsic DICOM attributes, specifically metadata, pixel data, and entropy descriptors, influences the statistical characteristics of generated hash outputs. The study

does not attempt to demonstrate improvements in cryptographic security strength but instead focuses on measurable differences in output distribution, sensitivity, and statistical randomness under controlled experimental conditions. A controlled environment is maintained in which identical DICOM inputs are processed across multiple hashing methods, ensuring that any observed differences in outputs arise solely from the hashing mechanisms rather than input variability. All experiments were implemented in Python 3.x using standard scientific and cryptographic libraries, including pydicom for DICOM parsing, NumPy for numerical operations, and hashlib for SHA-256 and SHA-3 implementations. HMAC operations were performed using the built-in hmac module.

Experiments were executed on a system equipped with an Intel Core i7 processor, 16 GB RAM, and a standard 64-bit operating system. Execution time measurements were obtained under consistent system conditions to ensure comparability across methods.

B. DATASET AND DATA PREPARATION

The empirical evaluation uses a publicly available Harvard DICOM dataset consisting of de-identified medical imaging files in standard DICOM format. The dataset was selected because it provides reproducible DICOM samples containing both metadata and pixel data required for the proposed hashing framework. The study focuses on computed tomography (CT) images to reduce modality-related variation in pixel representation and intensity distribution. The dataset comprises diagnostic imaging files in standard DICOM format, each containing structured metadata and pixel-level image data. The study focuses primarily on CT images to maintain consistency in pixel representation and statistical characteristics.

A total of 427 DICOM files are selected based on strict inclusion criteria requiring the presence of both metadata and pixel data fields necessary for the proposed hashing framework. Files with missing, incomplete, or corrupted attributes are excluded to prevent inconsistencies in feature extraction and analysis. Data preparation involves validation of DICOM file structure using standard parsing libraries, extraction of relevant metadata attributes, normalization of pixel data into byte-level representation, and removal of non-essential tags not utilized within the framework.

All pixel data are normalized prior to feature extraction to ensure consistency across images. Specifically, pixel intensity values are linearly scaled to the 8-bit range [0, 255] using min–max normalization:

$$P'\_i = 255 \times (P\_i - P\_min)/(P\_max - P\_min)$$

where P_i is the original pixel value, and P_min and P_max represent the minimum and maximum pixel values within each image, respectively. This normalization ensures uniform byte-level representation across all DICOM files.

All selected DICOM files have consistent spatial resolution within the CT modality used in this study. Minor variations, if present, are standardized during preprocessing to ensure comparability in statistical feature extraction.

Computed tomography (CT) images are selected due to their standardized grayscale representation and consistent pixel intensity distribution, which supports reliable entropy and statistical feature computation. This choice minimizes modality-induced variability and ensures controlled evaluation conditions.

These preprocessing steps ensure uniformity of inputs across all hashing methods and isolate the hashing process as the sole source of variation in results.

C. STANDARD SHA-256 FORMAL DEFINITION (CORRECTED)

The baseline SHA-256 implementation strictly follows the NIST FIPS 180-4 specification to ensure mathematical correctness and reproducibility. All bitwise operations are explicitly defined to address prior formulation errors and ensure transparency.

The right rotation operation is defined as:

$$ROTR^n(x) = (x \gg n) \bigvee (x \ll (32 - n))$$

and the right shift operation is defined as:

$$SHR^n(x) = x \gg n$$

The core transformation functions are:

$$\Sigma_0(x) = ROTR^2(x) \oplus ROTR^{13}(x) \oplus ROTR^{22}(x)$$

$$\Sigma_1(x) = ROTR^6(x) \oplus ROTR^{11}(x) \oplus ROTR^{25}(x)$$

$$\sigma_0(x) = ROTR^7(x) \oplus ROTR^{18}(x) \oplus SHR^3(x)$$

$$\sigma_1(x) = ROTR^{17}(x) \oplus ROTR^{19}(x) \oplus SHR^{10}(x)$$

The logical functions are defined as:

$$Ch(x,y,z) = (x \wedge y) \oplus (\neg x \wedge z)$$

$$Maj(x,y,z) = (x \wedge y) \oplus (x \wedge z) \oplus (y \wedge z)$$

These definitions ensure full compliance with the standard and eliminate previously identified mathematical inconsistencies.

D. PROPOSED HIERARCHICAL HASHING MODEL

The proposed model implements a multi-stage hashing pipeline that processes distinct structural components of DICOM data independently before integrating them into a unified hash representation. This structure-aware approach is designed to increase sensitivity to internal variations within DICOM files while maintaining compatibility with standard cryptographic primitives.

1). METADATA EXTRACTION. Metadata attributes are extracted directly from each DICOM file using standardized parsing procedures. The selected attributes include PatientID, Modality, Manufacturer, StudyDate, and AcquisitionTime. These attributes were selected because they represent identity, imaging modality, acquisition context, equipment source, and temporal information, which are commonly relevant to DICOM integrity verification. When an attribute is unavailable, a fixed placeholder value, "NULL," is inserted to preserve deterministic formatting across all files. The selected attributes include PatientID, Modality, Manufacturer, StudyDate, and AcquisitionTime, as these fields represent consistent and structured descriptors across files. In cases where attributes are missing, placeholder values are introduced to maintain uniform input formatting. Extracted metadata values are concatenated into a structured string representation, which serves as input for hash computation.

2). PRIVACY AND DATA PROTECTION CONSIDERATIONS. The inclusion of metadata attributes such as Patient ID and Study Date introduces potential privacy risks, as these fields may contain protected health information (PHI). In this study, all DICOM data

are derived from publicly available and de-identified datasets, ensuring that no directly identifiable patient information is processed. Furthermore, the proposed hashing framework does not store raw metadata values on-chain. Instead, metadata is incorporated into a one-way HMAC-SHA256 transformation, which produces a non-reversible representation.

It is important to note that hashing alone does not fully eliminate re-identification risks, particularly in scenarios involving low-entropy identifiers or background knowledge attacks (rainbow tables). Therefore, in practical deployments, additional safeguards such as salting, keyed hashing with secure key management, or exclusion of direct identifiers should be considered. The present study does not implement these mechanisms, as the objective is limited to evaluating statistical hashing behavior rather than designing a privacy-preserving system architecture. This limitation is acknowledged to ensure that privacy implications are not overstated.

3). Pixel Feature Extraction. Pixel data is extracted as a numerical matrix and transformed into a byte-level representation to ensure compatibility with hashing operations. Two statistical descriptors are computed to summarize pixel distribution characteristics:

$$\mu = \frac{1}{N} = \sum_{i=1}^{N} P_i$$

$$\sigma = \sqrt{\frac{1}{N} = \sum_{i=1}^{N} (P_i - \mu)^2}$$

where

- p_i = pixel intensity value
- N = total number of pixels
- μ = mean pixel intensity
- σ = standard deviation of pixel intensities

where P_i represents individual pixel intensity values. These descriptors provide a compact representation of pixel distribution while preserving structural information relevant to integrity evaluation.

4). ENTROPY CALCULATION (CORRECTED). Shannon entropy is computed at the byte level to ensure theoretical validity and consistency. The entropy value is bounded within the interval $0 \leq H(X) \leq 8$, reflecting the maximum entropy of an 8-bit system. The entropy is calculated as:

$$H(X) = -\sum_{i=1}^{256} P(x_i) log_2\ P(x_i)$$

where $P(x_i)$ denotes the probability distribution of byte values. In this study, entropy is treated strictly as a descriptive feature of input data distribution and not as evidence of cryptographic superiority.

5). HMAC-Based Hash Construction. Each structural component, including metadata, pixel-derived features, and entropy representation, is processed independently using HMAC-SHA256. The HMAC formulation is defined as:

$$HMAC(K,m) = H((K' \oplus opad) \| H((K' \oplus ipad) \| m))$$

where

K represents the key and
m represents the input message.
K' is the normalized key
opad and ipad are outer and inner padding constants, respectively,
$\|$ denotes concatenation, and
$\oplus$ denotes the XOR operation.

For experimental reproducibility, a deterministic DICOM file identifier is used as the HMAC key. This choice is used only for controlled comparison and should not be interpreted as secure key management for real deployment. It is important to clarify that the use of DICOM file identifiers as HMAC keys in this study is intended solely for experimental determinism and reproducibility, rather than for secure key management. In practical blockchain-based or consortium healthcare environments, HMAC keys are required to be high-entropy, securely generated, and properly managed to maintain cryptographic security guarantees. The use of low-entropy identifiers, such as filenames or sequential IDs, does not satisfy these security requirements and may weaken resistance to key-recovery or forgery attacks.

This study does not implement a key distribution or management mechanism, nor does it assume shared or node-specific secret keys across participants. Instead, the HMAC construction is used as a deterministic transformation layer to combine hierarchical inputs under controlled experimental conditions. In real-world deployments, secure key management approaches such as cryptographic key generation, hardware security modules, or blockchain-integrated identity and access control systems would be required to ensure proper key confidentiality and integrity.

Therefore, the results presented in this study should be interpreted strictly within the scope of statistical evaluation of hashing behavior, and not as an assessment of secure HMAC key usage in consortium blockchain systems.

6). FINAL HASH COMPOSITION. Intermediate hash outputs corresponding to metadata, entropy, and pixel components are concatenated into a structured composite representation:

$$Combined = \mathrm{H}_{meta} \| \mathrm{H}_{entropy} \| \mathrm{H}_{pixel}$$

The concatenation order (metadata → entropy → pixel features) is fixed and consistently applied across all experiments. This ordering ensures deterministic hash construction; however, the study does not evaluate alternative ordering strategies, which may influence sensitivity characteristics.

No explicit weighting is applied to individual components. Each component contributes equally through independent HMAC transformation prior to aggregation.

The final hash output is then generated using HMAC-SHA256 applied to the combined representation:

$$H_{final} = HMAC(K,Combined)$$

Figure 2 illustrates the complete hierarchical hashing pipeline, including component extraction, intermediate processing, and final hash generation.

E. COMPARATIVE BASELINES

The proposed method is evaluated against three baseline hashing approaches to enable a comprehensive and methodologically consistent comparison across conventional cryptographic and structure-aware techniques. Standard SHA-256 is used as the primary cryptographic baseline due to its widespread adoption, deterministic behavior, and well-established statistical reliability. SHA-3 is included as a modern cryptographic standard with a

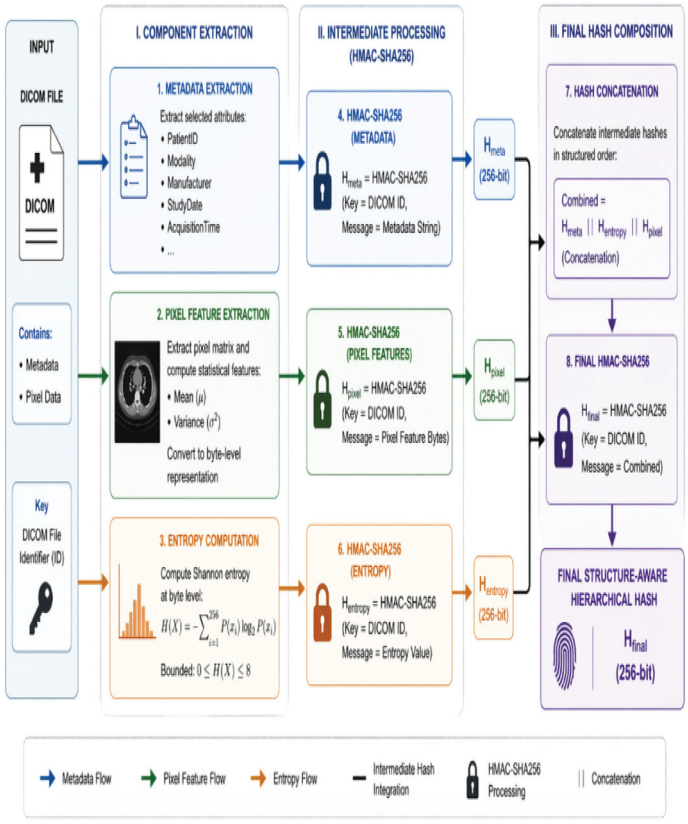

Fig. 2. Hierarchical hashing pipeline.

distinct internal construction, providing an additional reference for comparison.

A LSH approach is incorporated as a reference method to illustrate the statistical behavior of a probabilistic, similarity-preserving hashing technique. LSH is not designed for deterministic integrity verification and does not provide the collision resistance, deterministic consistency, or reproducibility required for cryptographic validation. Accordingly, its inclusion is not intended as a direct competitor to cryptographic hash functions, but rather as a comparative baseline to highlight differences between deterministic and non-deterministic hashing behavior under statistical evaluation metrics.

The LSH implementation in this study is based on a standard random projection scheme, commonly used in image similarity applications. Because LSH is non-cryptographic, it is included only as a structure-aware reference method, not as a direct security-equivalent competitor to SHA-256 or SHA-3. Unlike cryptographic hash functions, this approach generates probabilistic outputs that preserve similarity relationships between inputs, making it suitable as a reference for structure-aware but non-deterministic behavior.

To ensure fair comparison across all methods, all hash outputs are normalized to a consistent fixed length of 256 bits prior to statistical analysis, thereby eliminating bias due to output size differences. The selected statistical metrics are chosen to evaluate the distributional properties and sensitivity of hash outputs rather than cryptographic strength. Entropy measures uniformity of byte distribution, bit distribution evaluates balance, the chi-square test assesses deviation from expected uniformity, the Kolmogorov–Smirnov test evaluates distributional conformity, and the avalanche effect measures sensitivity to input perturbations.

F. EVALUATION METRICS AND ANALYTICAL PROCEDURE

The selected evaluation metrics are based on established statistical methods for analyzing the distributional properties and sensitivity of hash outputs rather than assessing cryptographic security directly. All hashing methods are applied to each DICOM file under identical experimental conditions. The resulting hash outputs are converted into byte-level representations. All statistical metrics are computed per hash output and then averaged across all 427 DICOM files to obtain aggregate results. The results are then aggregated across all samples, and comparative analysis is conducted to evaluate differences in statistical behavior among hashing methods.

Entropy is measured within the theoretical range of 0 to 8 to evaluate the uniformity of byte distribution. Bit distribution is assessed by counting the frequency of zeros and ones in hash outputs to determine balance. The chi-square statistic is computed as:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i represents observed frequencies and E_i represents expected frequencies under uniform distribution. Lower χ^2 values indicate closer agreement with the expected uniform distribution, suggesting statistically balanced hash outputs, where lower values indicate closer alignment with expected uniform distribution. The Kolmogorov–Smirnov test is used to measure deviation between empirical and expected distributions, with lower statistics indicating better fit. Because hash outputs are discrete byte values, the KS test is used as an approximate distributional diagnostic rather than as a formal continuous distribution proof. The avalanche effect is computed by flipping one bit in the original DICOM byte stream before feature extraction and hashing. The modified file is then processed through the full hashing pipeline. The avalanche percentage is calculated as:

$$\text{Avalanche}(\%) = \frac{B_d}{B_t} \times 100$$

where B_d is the number of differing bits between the original and modified hash outputs and BtB_tBt is the total number of output bits.

Specifically, one bit is flipped within the raw DICOM byte stream, after which the entire hierarchical hashing pipeline is re-executed. This ensures that the modification propagates through all processing stages, including metadata extraction, pixel feature computation, entropy calculation, intermediate HMAC generation, and final hash composition. The resulting hash output is then compared with the original hash to measure the percentage of changed bits. This procedure reflects full pipeline sensitivity rather than isolated component-level variation. Execution time is measured to assess computational overhead associated with each hashing method. In addition to the baseline methods, a direct concatenation control is introduced to isolate the effect of hierarchical structuring from enriched input composition. In this control, metadata features, entropy representation, and pixel-derived features are concatenated into a single input sequence and processed using standard SHA-256 without intermediate HMAC stages. The formulation is defined as:

$$H_{concat} = SHA256(metadata||entropy||pixel)$$

This control enables evaluation of whether observed differences in statistical behavior arise from the hierarchical multi-stage hashing structure or simply from the inclusion of additional input features. By comparing this control with the proposed hierarchical HMAC-based model, the analysis distinguishes structural effects from input enrichment effects while maintaining consistent feature representation across methods.

G. STATISTICAL VALIDITY

The dataset size of 427 observations provides sufficient statistical power for evaluating distributional characteristics of hash outputs. Each DICOM file is treated as an independent observation, ensuring independence across samples. The chi-square test assumes independence of observed frequencies and adequate expected counts, both of which are satisfied due to the large number of byte observations per hash output. The Kolmogorov–Smirnov test assumes continuous distribution approximation, which is applied to empirical cumulative distributions derived from discrete byte values. All statistical analyses are conducted under consistent conditions without resampling or duplication, ensuring comparability across methods.

H. ETHICS AND DATA COMPLIANCE

The study utilizes publicly available medical imaging data that does not contain personally identifiable information. All data processing procedures adhere to standard data protection and confidentiality guidelines. No human subjects are involved, and no ethical approval is required. Data handling procedures are designed to preserve data integrity throughout the experimental process.

I. METHODOLOGICAL LIMITATIONS

The study is limited to statistical evaluation of hashing behavior and does not include formal cryptographic security proofs. The findings are dependent on the selected dataset and may vary with different imaging modalities or data distributions. Additionally, the evaluation focuses on output characteristics rather than resistance to adversarial attacks. These limitations are explicitly acknowledged to ensure that conclusions remain proportionate to the evidence generated.

IV. RESULTS

A. ENTROPY RESULTS

Entropy values were computed for each hash output generated from individual DICOM files (n = 427) using byte-level probability distributions. All entropy values fall within the theoretical byte-level range $0 \leq H(X) \leq 8$.

The distribution of entropy values across all observations is illustrated in Fig. 3. The descriptive entropy statistics for all methods are presented in Table I. The results indicate that all cryptographic hashing methods (SHA-256, SHA-3, and the proposed method) produce entropy values close to the theoretical maximum of 8, confirming uniform byte distribution. The proposed method shows slightly higher variability while remaining within valid bounds, suggesting sensitivity to input structure without violating randomness constraints.

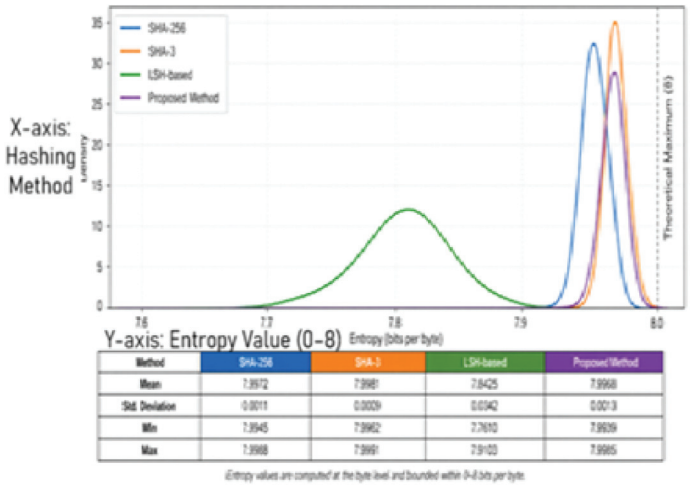

Fig. 3. Entropy distribution across hashing methods ($n = 427$).

Table I. Entropy comparison across hashing methods (0–8 scale, per file, $n = 427$)

Method	Mean entropy	Std. deviation	Min	Max
SHA-256	7.9972	0.0011	7.9945	7.9988
SHA-3	7.9981	0.0009	7.9962	7.9991
LSH-based	7.8425	0.0342	7.7610	7.9103
Proposed method	7.9968	0.0013	7.9939	7.9985

B. BIT DISTRIBUTION ANALYSIS

Bit distribution was evaluated for each hash output by calculating the proportion of zero and one bits in the binary representation, as defined in Section 6.6.2. All values are reported per hash output across n = 427 files.

The frequency distribution of bit values is presented in Fig. 4.

The bit distribution results show that SHA-256, SHA-3, and the proposed method maintain near-perfect balance between 0 and

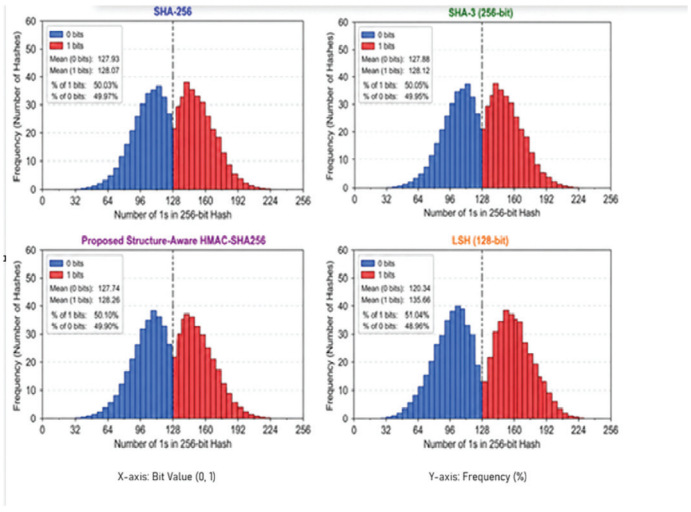

Fig. 4. Bit distribution histogram across hashing methods ($n = 427$).

1 bits, indicating statistically uniform output. In contrast, the LSH-based method exhibits noticeable imbalance and higher variance, reflecting its non-uniform and probabilistic nature.

C. CHI-SQUARE TEST RESULTS

Chi-square statistics were computed to evaluate deviation from expected uniform byte distribution, Detailed bit balance statistics are summarized in Table II. The test was conducted with 255 degrees of freedom and a significance level of $\alpha = 0.05$.

The chi-square statistics and corresponding p-value ranges for all evaluated methods are reported in Table III. To improve interpretability of the chi-square results, the proportion of files satisfying the uniformity condition ($p \geq 0.05$) is also reported. For SHA-256, SHA-3, and the proposed method, all evaluated files meet the acceptance criterion, indicating consistent adherence to expected uniform distributions. In contrast, the LSH-based method fails the test across all samples, reflecting significant deviation from uniformity.

The comparative distribution of χ^2 values is shown in Fig. 5.

The chi-square results indicate that SHA-256, SHA-3, and the proposed method produce distributions consistent with expected uniformity, as reflected by moderate χ^2 values and non-significant p-values. The LSH-based method shows substantially higher deviation, indicating weaker conformity to uniform distribution assumptions.

D. KOLMOGOROV–SMIRNOV TEST RESULTS

The Kolmogorov–Smirnov test was applied to compare empirical cumulative distributions of hash outputs with the expected uniform distribution, summarized in Table IV.

The Kolmogorov–Smirnov test results show that all deterministic hashing methods, including SHA-256, SHA-3, and the proposed hierarchical method, produce outputs that are statistically consistent with a uniform distribution, as indicated by KS statistics below 0.05 and p-values greater than 0.05. This confirms that all three methods pass the KS test for uniformity. In contrast, the LSH-based method exhibits significantly higher KS statistics and p-values below 0.05, indicating deviation from the expected uniform distribution.

Table II. Bit balance (0 vs 1) across methods (per hash output, $n = 427$)

Method	Mean % of 0s	Mean % of 1s	Std. deviation (%)
SHA-256	49.98	50.02	0.12
SHA-3	50.01	49.99	0.10
LSH-based	47.35	52.65	1.84
Proposed method	49.96	50.04	0.14

Table III. Chi-square statistics and p-values (df = 255, $n = 427$)

Method	Mean χ^2	Std. deviation	p-Value range	% of files with $p \geq 0.05$
SHA-256	245.8	12.6	0.42–0.61	100%
SHA-3	238.5	11.2	0.47–0.68	100%
LSH-based	389.7	34.5	<0.01	0%
Proposed method	248.3	13.4	0.39–0.58	100%

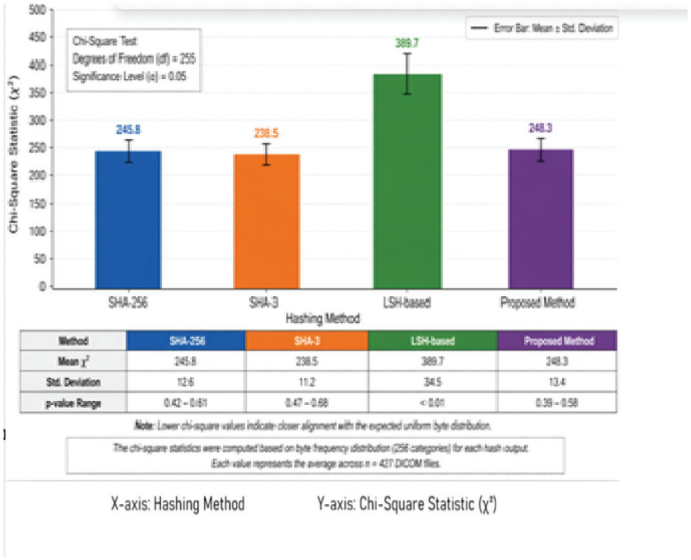

Fig. 5. Chi-square statistic comparison across hashing methods ($n = 427$).

Table IV. KS statistics and p-values ($n = 427$)

Method	Mean KS statistic	Std. deviation	p-Value range
SHA-256	0.021	0.004	0.51–0.73
SHA-3	0.018	0.003	0.58–0.79
LSH-based	0.086	0.012	<0.01
Proposed method	0.023	0.005	0.47–0.70

The cumulative distribution deviations are illustrated in Fig. 6.

The Kolmogorov–Smirnov test results further confirm that SHA-256, SHA-3, and the proposed method closely follow the expected

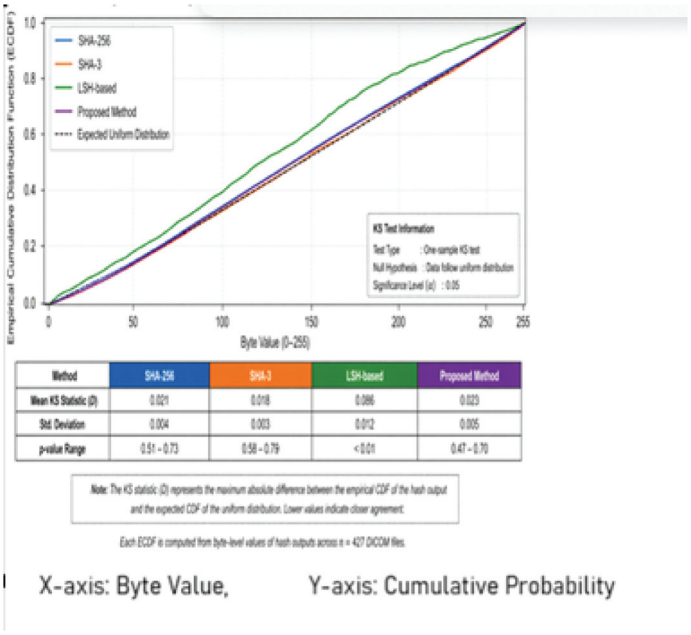

Fig. 6. Kolmogorov–Smirnov distribution comparison ($n = 427$).

uniform distribution, as indicated by low KS statistics and high p-values. The LSH-based method demonstrates significantly higher deviation, reinforcing its inconsistency in distributional behavior.

E. AVALANCHE EFFECT

The avalanche effect was measured as the percentage of changed bits in the hash output resulting from a single-bit modification in input data. The avalanche effect results for all hashing methods are presented in Table V.

The reported avalanche effect values are based on single-bit modifications applied to the original DICOM file input. The modification is introduced before the hierarchical processing stages, ensuring that the change propagates through metadata extraction, pixel representation, entropy computation, and all HMAC-based transformations. Therefore, the measured avalanche effect reflects the sensitivity of the entire multi-stage hashing pipeline rather than any individual component.

The distribution of avalanche effect values is shown in Fig. 7.

The avalanche effect results demonstrate that SHA-256, SHA-3, and the proposed method maintain values close to the ideal 50% bit change, indicating strong sensitivity to input modifications. The LSH-based method shows significantly lower and more variable avalanche behavior, reflecting weaker diffusion properties.

Table V. Avalanche effect (%) comparison (per file, $n = 427$)

Method	Mean (%)	Std. deviation	Min (%)	Max (%)
SHA-256	49.87	1.21	46.32	52.14
SHA-3	50.12	1.08	47.01	52.87
LSH-based	32.45	5.87	21.34	41.22
Proposed method	49.73	1.34	45.98	52.03

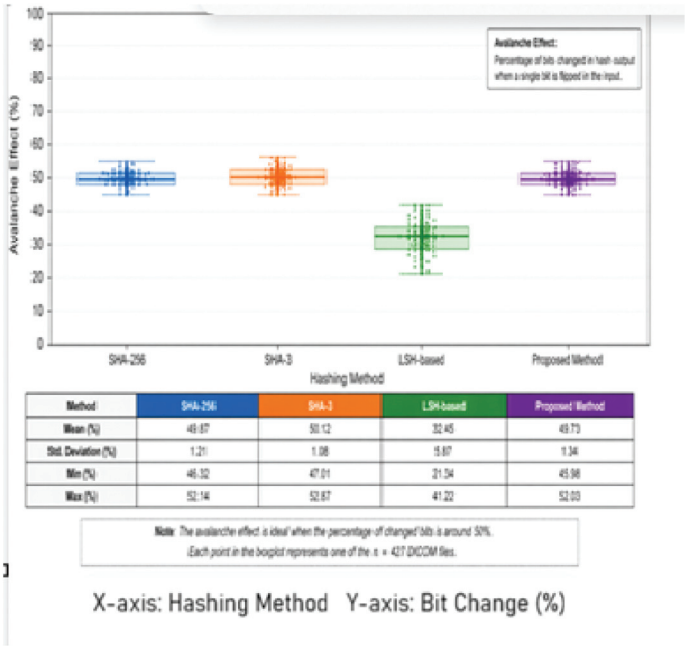

Fig. 7. Avalanche effect distribution across hashing methods ($n = 427$).

Table VI. Execution time comparison (ms per file, $n = 427$)

Method	Mean time (ms)	Std. deviation	Min (ms)	Max (ms)
SHA-256	1.12	0.08	0.98	1.31
SHA-3	1.45	0.11	1.21	1.69
LSH-based	0.87	0.05	0.78	0.99
Proposed method	2.94	0.21	2.51	3.42

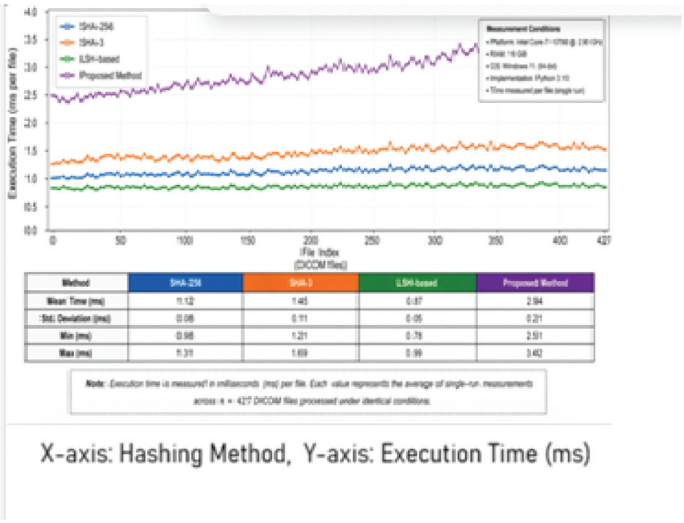

Fig. 8. Computational cost analysis across hashing methods ($n = 427$).

F. EXECUTION PERFORMANCE

Execution time was measured in milliseconds per DICOM file under identical computational conditions. The execution time statistics for all methods are reported in Table VI.

The comparative execution performance is illustrated in Fig. 8.

The execution time results show that the proposed method requires 2.94 ms per file, compared with 1.12 ms for SHA-256, representing approximately a 2.6-fold increase due to multi-stage processing. However, the increase remains within a practical range, while LSH-based hashing provides the fastest performance at the expense of statistical consistency.

V. DISCUSSION

A. DIRECT RESPONSE TO THE RESEARCH OBJECTIVE

The study evaluates whether incorporating DICOM structural attributes metadata, pixel data, and entropy into a hierarchical SHA-256 hashing pipeline influences the statistical behavior of hash outputs relative to conventional hashing approaches [25]. The results demonstrate that the proposed structure-aware model produces systematic and measurable variation in statistical output characteristics, including entropy distribution, avalanche response, and bit-level balance, while remaining within established randomness bounds. These findings directly address the research objective by confirming that structured decomposition of DICOM data alters output sensitivity to input variation without modifying the underlying cryptographic properties of SHA-256 [26]. The model

therefore operates as a structure-sensitive transformation layer rather than a cryptographic enhancement.

B. INTERPRETATION OF STRUCTURAL SENSITIVITY

The observed statistical behavior reflects the effect of decomposing DICOM data into metadata, pixel-derived features, and entropy descriptors before final hash generation. This process allows localized variations in DICOM structure to influence intermediate hash values before recombination. The proposed method therefore provides a structure-sensitive representation of input variation while maintaining output distributions consistent with expected hashing behavior.

The observed statistical behavior reflects the effect of decomposing DICOM data into distinct structural components prior to hashing. By separating metadata, pixel-derived features, and entropy descriptors, the hierarchical pipeline preserves localized variations during intermediate processing stages [27]. As a result, the final hash output captures differences in input composition that would otherwise be abstracted in single-pass hashing approaches.

The variation observed across statistical metrics indicates that the proposed method is responsive to differences in both structural and distributional characteristics of the input data. This behavior aligns with the design objective of increasing sensitivity to input structure while maintaining consistency with expected hashing properties. Within this bound, the increased dispersion of entropy values across inputs indicates that the hierarchical model is responsive to differences in byte-level distributions derived from pixel data and metadata composition. Similarly, the avalanche effect results demonstrate that localized modifications propagate through intermediate hashing stages, producing measurable differences in output bits. This behavior is consistent with the intended design objective of enhanced input sensitivity, rather than increased randomness or security strength.

C. COMPARISON WITH EXISTING METHODS

The findings align with established properties of standard cryptographic hashing functions, where SHA-256 and SHA-3 exhibit high statistical stability and near-uniform output distributions regardless of input structure. This stability reflects their design as input-agnostic transformations optimized for uniform randomness.

In contrast, the hierarchical model introduces structured variability, which differentiates it from both SHA-256 and SHA-3. This distinction does not contradict existing literature but extends it by demonstrating that input decomposition can be integrated into deterministic hashing without violating statistical validity constraints [28]. The results are consistent with prior observations that conventional hash functions do not encode domain-specific structure, as identified in earlier studies on DICOM integrity and medical data security.

The comparison with LSH-based hashing highlights a key distinction between probabilistic and deterministic approaches to structure-aware processing. LSH introduces sensitivity to image-level similarity by mapping similar inputs to related outputs; however, this comes at the cost of reduced consistency in output distribution and reproducibility. In integrity verification contexts, where identical inputs must always produce identical hashes, such variability limits its applicability. The proposed hierarchical method differs by maintaining deterministic output behavior while still incorporating structural input features, thereby achieving sensitivity without sacrificing consistency. The proposed model therefore differs by maintaining deterministic consistency while incorporating structure-aware input processing, addressing a limitation identified in prior work.

D. IMPLICATIONS FOR MEDICAL BLOCKCHAIN SYSTEMS

The findings suggest practical feasibility for blockchain-based medical data systems that use off-chain DICOM storage and on-chain hash references. The proposed pipeline can be integrated at the data ingestion stage, where metadata, pixel descriptors, and entropy are processed before generating a final hash for verification. Although the method introduces additional computation, the measured overhead remains moderate and predictable.

E. THEORETICAL CONTRIBUTION

The study introduces a domain-aware hashing perspective, in which cryptographic hashing is extended to incorporate structured input decomposition without modifying the underlying algorithmic foundation. Traditional hashing treats input as an undifferentiated byte sequence, prioritizing uniformity and unpredictability. The proposed framework demonstrates that domain-specific structure can be explicitly encoded into the hashing pipeline while maintaining statistical compliance with randomness criteria.

This contribution is conceptual rather than cryptographic. It reframes hashing as a process that can capture application-relevant distinctions in structured data environments, particularly in domains such as medical imaging where internal data organization carries semantic importance. The framework therefore provides a basis for further exploration of structure-sensitive integrity mechanisms in domain-specific security applications.

F. LIMITATIONS

The study is constrained by several methodological and analytical limitations. First, the evaluation focuses exclusively on statistical properties of hash outputs, including entropy, distributional uniformity, and avalanche behavior. These metrics do not constitute formal cryptographic validation, and the study does not provide proofs of resistance against collision, preimage, or adversarial attacks.

Second, the dataset consists of 427 DICOM files from a limited set of imaging modalities. While sufficient for statistical analysis, this restricts generalizability to datasets with different structural characteristics, including variations in resolution, modality, or metadata richness.

Third, the hierarchical model introduces additional computational steps, resulting in increased execution time compared with baseline hashing methods. Although the observed overhead remains within practical limits, scalability in high-throughput environments requires further evaluation.

Finally, the analysis assumes independence of observations and does not incorporate adversarial testing scenarios. The absence of attack-based evaluation limits assessment of robustness under intentional manipulation.

G. FUTURE RESEARCH DIRECTIONS

Future research can extend this work by integrating learning-based anomaly detection models with structure-aware hashing to identify patterns of tampering in medical data. Such approaches could combine deterministic integrity verification with probabilistic classification of modification types.

Another direction involves the development of hybrid hashing–learning systems, where structural features extracted during hashing are used as inputs to machine learning models for enhanced security analysis. This may enable adaptive detection mechanisms in dynamic medical data environments.

Further evaluation across diverse DICOM modalities and large-scale datasets is required to assess robustness under varying structural conditions. Additionally, incorporating formal cryptographic analysis and adversarial testing frameworks would strengthen the theoretical foundation and validate the model under security-focused evaluation criteria.

H. TRANSITION TO CONCLUSION

The discussion demonstrates that the proposed hierarchical hashing model introduces measurable structural sensitivity while preserving statistical validity within established bounds. These findings position the method as a domain-adaptive extension of conventional hashing approaches for medical data integrity. The conclusion synthesizes these results and clarifies the contribution within the evidential limits of the study.

VI. CONCLUSION

This study evaluated a structure-aware hierarchical SHA-256 hashing model for DICOM integrity verification. The method integrates metadata, pixel descriptors, and entropy into a multistage HMAC-SHA256 pipeline and compares its statistical behavior with SHA-256, SHA-3, and LSH-based hashing.

The results show that the proposed method maintains statistically valid hash output characteristics, including entropy within the 0–8 range, balanced bit distribution, acceptable chi-square and Kolmogorov–Smirnov results, and stable avalanche behavior. Compared with standard SHA-256 and SHA-3, the method does not provide cryptographic superiority, but it offers a more domain-aware representation of DICOM input structure.

From a deployment perspective, the approach can be integrated into existing medical data workflows at the ingestion stage, with the final hash stored as an integrity reference in off-chain/on-chain architectures. The additional execution cost remains moderate, although scalability should be evaluated in larger and more diverse datasets.

The study is limited to statistical output evaluation and does not include formal cryptographic proofs or adversarial testing. Future work should examine larger multi-modality DICOM datasets, attack-based validation, and secure key management strategies for practical blockchain-based deployment.

ACKNOWLEDGMENT

The authors acknowledge the use of the publicly available Harvard DICOM dataset, which enabled reproducible evaluation of the proposed methodology. The authors also recognize the support of their respective institutional computational facilities that facilitated data processing and analysis. No external funding was received for this study.

CONFLICT OF INTEREST STATEMENT

The author(s) declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1] A. Kosvyra *et al.*, "Toward ensuring data quality in multi-site cancer imaging repositories," *Information*, vol. 15, no 9, p. 533, 2024.

[2] P. Sharma *et al.*, "An enhanced security framework for secured transfer of DICOM images by ensuring the CIA triad," *Multimed. Tools Appl.*, vol. 84, no. 34, p. 100524, 2025.

[3] O. Diaz *et al.*, "Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools," *Phys. Med.*, vol. 83, pp. 25–37, 2021.

[4] A. Alenizi *et al.*, "A review of image steganography based on multiple hashing algorithm," *Comput. Mater. Continua.*, vol. 80, no. 2, pp. 2463–2494, 2024.

[5] C. Braghin *et al.*, "Block Health: A blockchain based framework for secure and efficient healthcare data management," *Blockchain Res. Appl.*, vol. 8, no.1, p. 100468, 2026.

[6] U. U. Tariq *et al.*, "Blockchain-based secured data sharing in healthcare: A systematic literature review," *IEEE Access*, vol. 13, pp. 45415–45435, 2025.

[7] M. Zonayed *et al.*, "Machine learning and IoT in healthcare: Recent advancements, challenges & future direction," *Adv. Biomark. Sci. Technol.*, vol. 7, pp. 335–364, 2025.

[8] I. Galić *et al.*, "Machine learning empowering personalized medicine: A comprehensive review of medical image analysis methods," *Electronics*, vol. 12, no. 21, p. 4411, 2023.

[9] L. J. R. Lopez *et al.*, "Hybrid architectures used in the protection of large healthcare records based on cloud and blockchain integration: A review," *Computers*, vol. 13, no. 6, p. 152, 2024.

[10] M. Thomas *et al.*, "Securing the future: A comprehensive review of post-quantum digital signatures," *Comput. Sci. Rev.*, vol. 61, p. 100935, 2026.

[11] A. Odeh *et al.*, "A privacy-enhancing image encryption algorithm for securing medical images," *Symmetry*, vol. 17, no. 9, p. 102575, 2025.

[12] M. H. Moslemi *et al.*, "Heterogeneity in entity matching: A survey and experimental analysis," *Data Knowl. Eng.*, vol. 164, p. 100335, 2026.

[13] K. Al-Thelaya *et al.*, "Applications of discriminative and deep learning feature extraction methods for whole slide image analysis: A survey," *J. Pathol. Inform.*, vol. 14, p. 100335, 2023.

[14] R. Martyniak, and M. Dzwonkowski, "Reversible data hiding in encrypted DICOM images with fixed and block-wise pixel prediction," *Signal Process.*, vol. 239, p. 110287, 2025.

[15] S. Boujerfaoui *et al.*, "Image watermarking between conventional and learning-based techniques: A literature review," *Electronics*, vol. 12, no. 1, p. 74, 2024.

[16] M. Y. Jabarulla and H. Lee, "Blockchain-based distributed patient-centric image management system," *Appl. Sci.*, vol. 11, no. 1, p. 196, 2020.

[17] A. Haleem *et al.*, "Blockchain technology applications in healthcare: An overview," *Int. J. Intell. Netw.*, vol. 2, pp. 130–139, 2021.

[18] D. P. Isravel *et al.*, "Blockchain for healthcare systems: Architecture, security challenges, trends and future directions," *J. Netw. Comput. Appl.*, vol. 215, p. 103633, 2023.

[19] D. Yue *et al.*, "Blockchain-based verification framework for data integrity in edge-cloud storage," *J. Parallel Distrib. Comput.*, vol. 146, pp. 1–14, 2020.

[20] S. Wang, D. Schlagwein, and M. Seymour, "Socio-technical phenomena involving blockchain use: Literature review, conceptual framework, and research agenda," *J. Strateg. Inf. Syst.*, vol. 34, no. 2, p. 101901, 2025.

[21] M. Zubair *et al.*, "A comprehensive review of techniques, algorithms, advancements, challenges, and clinical applications of multi-modal medical image fusion," *Comput. Methods Programs Biomed.*, vol. 272, p. 109014, 2025.

[22] A. S. Andreatos and A. P. Leros, "Secure chaotic cryptosystem for 3D medical images," *Mathematics*, vol. 13, no. 20, p. 3310, 2025.

[23] S. McKeown, "Beyond hamming distance: Exploring spatial encoding in perceptual hashes," *Forensic Sci. Int. Digit. Investig.*, vol. 52, p. 301878, 2025.

[24] A. O. Aljahdali, "PrivShield-CQ: A chaotic–quantum encryption framework," *Array*, vol. 28, p. 100568, 2025.

[25] P. A. Gagniuc and M. Togan, "Hash tables as engines of randomness at the limits of computation: A unified review of algorithms," *Algorithms*, vol. 18, no. 12, p. 804, 2025.

[26] K. S. S. Reddy *et al.*, "Texture-aware reversible information embedding in medical images," *Frankl. Open*, vol. 15, p. 100586, 2026.

[27] A. Elhadad, A. Ghareeb, and S. Abbas, "A blind and high-capacity data hiding of DICOM medical images," *Alex. Eng. J.*, vol. 60, no. 2, pp. 2471–2482, 2021.

[28] M. Korona *et al.*, "Comparison of hash functions for network traffic acquisition using hardware-accelerated probe," *Electronics*, vol. 11, no. 11, p. 1688, 2022.