International journal of statistics **(2051-8285)**

A Layered Feature Resolution Model with Intelligent Hashing for Scalable Cloud-Based Video Redundancy Elimination

Dr. Chinedu Okafor

Department of Artificial Intelligence and Data Science

ABSTRACT

The exponential growth of digital video content has intensified storage requirements in cloud environments, creating significant challenges related to redundancy, scalability, and resource optimization. Traditional deduplication mechanisms primarily rely on exact matching techniques, which often fail to identify semantically similar video segments due to variations in resolution, compression, encoding formats, and content representation. This research proposes a Layered Feature Resolution Model with Intelligent Hashing (LFRM-IH) for scalable cloud-based video redundancy elimination. The proposed approach integrates hierarchical feature extraction, adaptive resolution analysis, clustering-assisted similarity identification, and intelligent hashing mechanisms to improve redundancy detection efficiency while maintaining data integrity and storage security. The theoretical foundation of the model is derived from existing cloud deduplication studies focusing on hashing, clustering, content-defined chunking, and secure distributed storage mechanisms. The framework introduces multiple feature-resolution layers where coarse-level visual characteristics are initially analyzed before applying fine-grained hashing comparisons, reducing computational overhead during large-scale video processing. Existing research demonstrates the effectiveness of hashing and clustering strategies in cloud deduplication environments, while video-specific approaches highlight the importance of structural segmentation and adaptive processing. The proposed model addresses limitations in existing systems by combining scalable feature representation with intelligent hash decision-making. Furthermore, integration considerations with cloud-edge environments and secure deduplication mechanisms are analyzed to improve practical applicability. The research contributes a conceptual architecture for efficient video redundancy elimination capable of supporting modern cloud storage ecosystems with improved scalability, reduced storage consumption, and optimized computational performance.

Keywords: Video Deduplication, Cloud Storage, Intelligent Hashing, Feature Resolution, Clustering Model, Data Redundancy Elimination, Storage Optimization, Cloud Computing.

INTRODUCTION

1.1 Background

The rapid expansion of video-based applications, including online learning platforms, surveillance systems, social media services, and multimedia repositories, has created unprecedented demand for cloud storage infrastructure. Unlike conventional data objects, video files contain large volumes of information and frequently exist in multiple modified forms due to compression variations, resolution changes, editing operations, and encoding differences. Consequently, cloud providers face increasing challenges in managing redundant video content while maintaining high availability, processing efficiency, and cost effectiveness.

Data deduplication has emerged as an essential storage optimization technique by eliminating duplicate data objects and maintaining only unique representations. Conventional deduplication systems generally rely on hash-based comparison mechanisms where generated fingerprints are used to identify identical data segments. However, video data introduces additional complexity because visually similar or partially modified videos may not produce identical hash values. Therefore, conventional exact matching approaches may overlook significant opportunities for redundancy elimination.

Hashing and clustering-based approaches have demonstrated potential for improving redundancy detection in cloud environments. Chen et al. (2021) investigated image deduplication using hashing and clustering methods and highlighted the effectiveness of combining similarity analysis with hash-based identification for reducing duplicate storage. Similarly, Periasamy and Latha (2021) proposed an efficient hash-function-based duplication detection mechanism emphasizing optimized fingerprint generation for improving deduplication efficiency.

Video-oriented deduplication requires more advanced processing because videos consist of sequential frames, temporal relationships, and multi-level visual characteristics. Sujatha et al. (2023) explored GOP-level video deduplication using adaptive GOP structures, demonstrating that structural video segmentation can improve redundancy identification by analyzing meaningful video components rather than treating entire files as independent objects.

The increasing adoption of cloud-edge computing further requires secure and scalable deduplication frameworks. Ke et al. (2025) introduced secure deduplication with partial migration between cloud and edge servers, showing the importance of balancing computational distribution and data security. Similarly, Wu and Fu (2024) examined cloud-edge collaborative secure deduplication using trusted execution approaches, emphasizing the need for secure processing mechanisms in distributed environments.

Recent developments in digital ecosystems also indicate the importance of intelligent data management architectures. Hussain et al. (2026) highlighted the role of generative AI-driven sensor fusion frameworks for secure digital twin ecosystems, where intelligent data processing and standardized cyber-physical interactions become essential for managing complex information flows. Such intelligent processing principles are relevant for designing adaptive storage optimization models capable of dynamically analyzing large-scale multimedia data.

1.2 Problem Statement

Existing cloud-based video deduplication systems face several limitations. First, traditional hash-based methods are computationally expensive when applied directly to massive video collections because every object requires complete fingerprint generation and comparison. Second, exact hash matching cannot efficiently detect near-duplicate videos containing minor modifications such as resolution changes, frame alterations, or compression differences. Third, current approaches often lack adaptive mechanisms that determine the appropriate level of feature analysis according to video similarity characteristics.

Although clustering-based and content-aware methods improve redundancy detection, they still require optimization for large-scale cloud environments. Xia et al. (2020) demonstrated that efficient content-defined chunking improves deduplication-based storage systems by reducing unnecessary processing during data segmentation. However, applying similar principles to complex video structures requires additional feature-resolution strategies.

Furthermore, secure cloud storage environments require deduplication mechanisms that preserve privacy and prevent unauthorized data exposure. While secure deduplication approaches have improved cloud-edge collaboration, achieving a balance between security, computational efficiency, and redundancy elimination remains a significant research challenge.

1.3 Research Relevance

The development of an intelligent layered feature-resolution model is relevant because it addresses the limitations of existing single-stage deduplication approaches. Instead of performing expensive fine-grained comparisons on all video objects, a layered architecture enables progressive analysis. Initial coarse-level features can quickly identify potential similarity groups, while detailed hashing operations can be applied only to selected candidates.

This research direction aligns with previous findings regarding adaptive storage management. Tan et al. (2020) demonstrated that content-driven cache management improves the performance of deduplication-based storage systems by incorporating data characteristics into optimization decisions. Extending this principle toward video redundancy elimination enables more intelligent resource allocation.

Additionally, machine learning-assisted architectural optimization concepts can support adaptive system organization. Hebbar (2022) examined machine learning-assisted service boundary detection for modularizing legacy systems, demonstrating how intelligent analysis can improve complex computational structures. Similar principles can support modular feature-processing layers in cloud video deduplication frameworks.

1.4 Objectives

The primary objectives of this research are:

1. To design a layered feature-resolution architecture for scalable cloud-based video redundancy elimination.

2. To integrate intelligent hashing mechanisms with

clustering-based similarity analysis.

3. To reduce computational overhead through progressive feature comparison.

4. To improve storage efficiency by identifying exact and near-duplicate video content.

5. To analyze the applicability of secure and distributed deduplication strategies in cloud environments.

1.5 Scope and Significance

The proposed model focuses on cloud-based video storage systems where large-scale multimedia datasets require efficient redundancy management. The framework considers video segmentation, feature extraction, similarity clustering, intelligent hashing, and secure cloud integration.

The significance of this research lies in developing a scalable approach that moves beyond traditional duplicate detection. By combining layered analysis with intelligent decision mechanisms, the proposed model provides a foundation for next-generation storage optimization systems capable of handling increasing multimedia workloads.

LITERATURE REVIEW

2.1 Hash-Based Deduplication Approaches

Hash-based deduplication remains one of the most widely adopted techniques for identifying duplicate data. The fundamental principle involves generating unique fingerprints for data objects and comparing these fingerprints to determine redundancy. Chen et al. (2021) demonstrated that combining hashing with clustering improves duplicate identification by grouping similar data objects before performing detailed comparisons. Their findings indicate that similarity-aware hashing mechanisms can overcome limitations of direct hash comparison.

Periasamy and Latha (2021) further emphasized the importance of efficient hash function design for reducing computational requirements. Their duplication detection algorithm focused on improving fingerprint generation and comparison efficiency, providing evidence that optimized hashing strategies are essential for scalable storage environments.

However, conventional hashing approaches remain limited when dealing with multimedia content. Small changes in video encoding or visual structure may produce completely different hash values. Therefore, intelligent hashing mechanisms capable of incorporating feature-level similarity are required for advanced video deduplication.

2.2 Clustering and Feature-Based Redundancy Detection

Clustering provides an effective mechanism for reducing unnecessary comparisons by organizing similar objects into groups. Instead of comparing every video against all stored content, clustering enables selective analysis among potentially related objects.

The combination of clustering and hashing provides a theoretical foundation for the proposed layered model. Chen et al. (2021) demonstrated that clustering improves the efficiency of hashing-based duplicate detection, while Sujatha et al. (2023) showed that video structural segmentation improves redundancy identification.

The GOP-level approach proposed by Sujatha et al. (2023) represents an important advancement because videos contain internal structures that can be independently analyzed. However, GOP-based methods may require further optimization when processing heterogeneous cloud-scale video repositories containing diverse resolutions and formats.

2.3 Content-Aware Storage Optimization

Content-aware processing has become increasingly important in modern storage systems. Xia et al. (2020) introduced fast content-defined chunking techniques for deduplication-based storage systems, demonstrating that adaptive segmentation improves processing efficiency.

Tan et al. (2020) extended this concept by proposing content-driven cache management methods that improve deduplication-based storage performance. Their findings indicate that storage optimization should consider data characteristics rather than relying solely on static processing strategies.

The proposed layered feature-resolution model builds upon these concepts by introducing adaptive feature analysis levels. Instead of applying identical processing operations to every video, the framework dynamically determines whether coarse or detailed analysis is required.

METHODOLOGY

3.1 Proposed Layered Feature Resolution Model with Intelligent Hashing (LFRM-IH)

This research proposes a Layered Feature Resolution

Model with Intelligent Hashing (LFRM-IH) designed for scalable cloud-based video redundancy elimination. The proposed framework is developed based on the observation that existing deduplication systems frequently perform either complete file-level comparison or fixed-resolution fingerprint analysis, resulting in unnecessary computational consumption. The proposed model introduces a hierarchical processing architecture where video data is progressively analyzed through multiple feature-resolution layers.

The theoretical foundation of the model is derived from three major concepts: similarity-aware clustering, intelligent hashing, and adaptive content analysis. Chen et al. (2021) established that combining clustering and hashing improves duplicate identification efficiency, while Sujatha et al. (2023) demonstrated the effectiveness of structural video segmentation for redundancy elimination. The proposed framework integrates these concepts into a unified architecture suitable for large-scale cloud video repositories.

The LFRM-IH framework consists of five primary layers:

1. Video preprocessing and structural decomposition layer

2. Coarse feature resolution analysis layer

3. Similarity clustering and candidate selection layer

4. Intelligent hashing and fine-resolution verification layer

5. Cloud storage optimization and secure management layer

Each layer performs a specific function while reducing unnecessary processing before reaching computationally expensive operations.

3.2 Video Preprocessing and Structural Decomposition Layer

The first stage focuses on transforming raw video data into analyzable components. Unlike conventional file-level deduplication, which treats videos as complete objects, the proposed framework considers the internal structure of video content.

Video preprocessing includes:

- Metadata extraction
- Frame sampling
- Resolution normalization
- Temporal segmentation
- Structural feature generation

The objective of this layer is to create representative video segments while maintaining essential visual information. The approach follows the concept of adaptive video segmentation demonstrated by Sujatha et al. (2023), where GOP-level structures were utilized to improve video redundancy detection.

For example, two surveillance videos recorded from the same camera may contain identical background scenes but differ in duration or encoding format. File-level hashing would identify them as different objects, whereas structural decomposition allows the system to analyze common segments and identify redundancy.

This layer reduces the complexity of later processing stages by creating meaningful video representations rather than processing entire video files repeatedly.

3.3 Coarse Feature Resolution Analysis Layer

The second layer introduces coarse-level feature analysis to rapidly identify potentially similar video objects. The fundamental principle is that not all videos require detailed comparison. Many duplicate candidates can be eliminated through lightweight feature analysis.

The coarse feature layer extracts:

- Basic visual descriptors
- Frame-level statistical characteristics
- Temporal patterns
- Resolution-related information
- Compression characteristics

These features create an initial similarity representation. Videos with significant differences are discarded at this stage, while potentially similar videos are forwarded for advanced analysis.

This mechanism reduces computational overhead because expensive hashing operations are performed only on selected candidates. The approach is consistent with content-aware optimization strategies discussed by Xia et al. (2020), where efficient content-defined processing reduced unnecessary storage operations.

The major advantage of coarse resolution analysis is scalability. In cloud environments containing millions of videos, performing complete comparisons across all files is impractical. Layered filtering enables distributed systems to focus resources on high-probability duplicate groups.

3.4 Similarity Clustering and Candidate Selection Layer

After coarse feature extraction, the proposed framework applies clustering-based grouping to organize videos according to similarity characteristics.

The clustering process performs:

1. Feature vector generation
2. Similarity measurement
3. Cluster formation
4. Duplicate candidate identification

The theoretical basis of this layer originates from clustering-assisted deduplication approaches. Chen et al. (2021) showed that clustering improves hashing efficiency by reducing unnecessary comparisons among unrelated data objects.

For video storage environments, clustering provides additional benefits because videos frequently contain semantic relationships. For instance, multiple versions of educational videos may differ only in language tracks, resolution, or minor editing modifications. A clustering layer can group these related objects before detailed verification.

However, clustering introduces challenges related to computational complexity and dynamic cloud workloads. Large-scale video repositories continuously receive new content, requiring adaptive clustering strategies. Therefore, the proposed framework considers incremental cluster updating rather than rebuilding clusters from the beginning.

3.5 Intelligent Hashing and Fine-Resolution Verification Layer

The fourth layer performs detailed redundancy verification using intelligent hashing mechanisms. Unlike traditional cryptographic hashing, where minor modifications create entirely different fingerprints, intelligent hashing considers feature-level similarity.

The hashing process includes:

- Feature-aware fingerprint generation
- Multi-resolution hash creation
- Candidate fingerprint comparison
- Similarity confidence evaluation

The model combines multiple hash representations generated from different feature resolutions. A low-resolution hash provides fast similarity estimation, while higher-resolution hashes provide accurate verification.

Periasamy and Latha (2021) emphasized that efficient hash function design directly influences deduplication performance. Extending this principle, the proposed model introduces adaptive hash selection, where the required hashing depth depends on similarity confidence.

For example, two videos with identical structural features may require only limited verification, whereas videos with partial similarity undergo deeper feature comparison.

This adaptive mechanism improves efficiency by preventing unnecessary computational operations.

3.6 Cloud Storage Optimization and Secure Management Layer

The final layer integrates the deduplication framework with cloud storage management mechanisms. Once duplicate videos are identified, the system maintains a single optimized representation while creating reference mappings for duplicate requests.

The storage optimization process includes:

- Unique video object retention
- Duplicate reference management
- Storage space reduction
- Access consistency maintenance

Cloud security is an essential consideration because deduplication involves identifying relationships between stored objects. Secure processing mechanisms are therefore required to prevent unauthorized information disclosure.

Ke et al. (2025) investigated secure deduplication through partial migration between cloud and edge servers, demonstrating the importance of balancing security and computational efficiency. Similarly, Wu and Fu (2024) explored cloud-edge collaborative secure deduplication using trusted execution environments.

The proposed framework incorporates these principles by separating feature analysis and secure storage operations. Sensitive content processing can be distributed between cloud and edge environments while maintaining controlled access.

Additionally, Hussain et al. (2026) emphasized the importance of intelligent and secure data processing frameworks for complex cyber-physical ecosystems. This concept supports the requirement for adaptive and secure decision-making mechanisms in future cloud storage architectures.

3.7 Integration of Intelligent Decision Mechanism

The proposed model incorporates an intelligent decision mechanism responsible for selecting appropriate processing levels. Instead of applying identical operations to every video object, the system evaluates:

- Content similarity probability
- Feature complexity
- Storage priority
- Processing cost
- Security requirements

Based on these factors, the framework dynamically determines whether a video requires coarse analysis, clustering verification, or complete fine-resolution hashing.

This adaptive decision capability improves resource utilization and aligns with machine-learning-assisted modular processing concepts. Hebbar (2022) demonstrated that intelligent analysis can support efficient system organization by identifying functional boundaries within complex environments. Similarly, the proposed framework uses intelligent classification to determine optimal processing pathways.

3.8 Algorithmic Workflow of LFRM-IH

The operational workflow of the proposed framework can be summarized as follows:

Step 1: Upload incoming video data into the cloud storage environment.

Step 2: Extract metadata and divide videos into meaningful structural segments.

Step 3: Generate coarse-level feature representations.

Step 4: Identify similarity groups through clustering analysis.

Step 5: Apply intelligent hashing only to selected candidate groups.

Step 6: Perform fine-resolution verification for duplicate confirmation.

Step 7: Store unique content and create reference links for redundant objects.

Step 8: Continuously update similarity clusters for future optimization.

This workflow minimizes unnecessary computation while maintaining accurate redundancy elimination.

RESULTS

The proposed Layered Feature Resolution Model with Intelligent Hashing demonstrates several conceptual improvements over conventional cloud video deduplication approaches. The primary finding is that hierarchical analysis significantly reduces the computational burden associated with large-scale redundancy detection. By introducing multiple feature-resolution levels, the framework avoids performing expensive hash comparisons across the entire video repository.

The first major outcome is improved scalability. Traditional deduplication approaches often experience performance degradation as storage volume increases because comparison operations grow proportionally with the number of stored objects. The proposed model addresses this limitation through coarse filtering and clustering-based candidate reduction. Similar findings were observed by Chen et al. (2021), where clustering improved hashing-based duplicate identification efficiency.

The second finding relates to storage optimization. Video redundancy frequently occurs due to multiple versions of similar content. The proposed framework identifies both exact and near-duplicate video segments by combining structural analysis with intelligent hashing. This extends the capability of conventional deduplication mechanisms that primarily focus on identical objects.

The third finding concerns processing efficiency. The adaptive hashing mechanism reduces unnecessary fingerprint generation by applying detailed verification only when similarity probability is high. This principle aligns with the optimization strategies discussed by Xia et al. (2020), where efficient content-based processing improved deduplication performance.

The fourth finding highlights the importance of secure distributed operation. Integration with cloud-edge security mechanisms provides opportunities for implementing deduplication without compromising data confidentiality. The approaches discussed by Ke et al. (2025) and Wu and Fu (2024) demonstrate that secure deduplication architectures are increasingly necessary for modern distributed storage environments.

The conceptual evaluation also indicates that the proposed framework can support diverse applications, including multimedia archives, surveillance storage, educational repositories, and enterprise video management platforms. However, the model requires further experimental validation using large-scale real-world video datasets to quantify improvements in storage reduction rate, processing time, and computational efficiency.

DISCUSSION

The proposed Layered Feature Resolution Model with Intelligent Hashing (LFRM-IH) provides a conceptual advancement in cloud-based video redundancy elimination by addressing major limitations associated with conventional deduplication approaches. The discussion highlights how layered processing, intelligent hashing, and adaptive similarity evaluation contribute to more scalable and efficient multimedia storage management.

A key theoretical implication of the proposed framework is the transition from traditional binary duplicate detection toward progressive similarity assessment. Conventional deduplication systems generally determine whether two objects are identical based on generated fingerprints. Although effective for exact duplicates, this approach becomes inefficient for videos because multimedia content frequently undergoes transformations such as compression changes, resolution adaptation, frame modification, and format conversion. The proposed layered architecture addresses this limitation by introducing multiple levels of analysis, where computational complexity increases only when necessary.

The integration of clustering and hashing represents an important contribution of the proposed model. Chen et al. (2021) demonstrated that clustering-assisted hashing improves duplicate identification efficiency by reducing unnecessary comparisons. The proposed framework extends this concept to video environments by using clustering not only for grouping similar content but also for controlling the depth of feature analysis. This approach creates a more adaptive deduplication process where storage resources are allocated according to similarity probability and content characteristics.

The findings also demonstrate the importance of structural video analysis. Sujatha et al. (2023) highlighted that GOP-level processing improves video deduplication by considering internal video organization rather than treating a video as a single data object. The proposed model follows a similar principle but expands it through multi-resolution feature analysis. This enables the system to identify redundancy at different granularities, including complete videos, video segments, and visually similar portions.

From a practical perspective, the proposed framework can significantly benefit cloud service providers managing large-scale multimedia repositories. Video platforms, enterprise surveillance systems, and digital learning environments frequently store multiple copies of similar content. A scalable redundancy elimination mechanism can reduce storage requirements, improve data management efficiency, and optimize cloud resource utilization.

The intelligent hashing component introduces additional advantages by reducing unnecessary fingerprint computation. Periasamy and Latha (2021) emphasized the importance of efficient hashing mechanisms in improving deduplication performance. However, intelligent hashing also introduces challenges because similarity-based fingerprinting requires careful calibration between accuracy and computational cost. Excessive similarity tolerance may result in incorrect duplicate identification, while strict thresholds may fail to detect meaningful redundancy.

The framework also presents implications for secure cloud computing environments. Modern storage systems require deduplication mechanisms that maintain confidentiality while optimizing resources. Ke et al. (2025) and Wu and Fu (2024) demonstrated that secure cloud-edge deduplication approaches can improve distributed storage efficiency while addressing security concerns. The proposed model complements these approaches by separating similarity analysis from secure storage operations, enabling flexible deployment across cloud and edge infrastructures.

Furthermore, the proposed architecture aligns with intelligent system management principles. Hussain et al. (2026) emphasized the importance of intelligent data processing frameworks for secure digital ecosystems. Similar requirements exist in cloud storage environments where massive volumes of multimedia information require automated decision-making. The adaptive decision mechanism in LFRM-IH supports this requirement by dynamically selecting suitable processing strategies.

Despite these advantages, several limitations must be considered. First, the proposed model requires comprehensive experimental validation using diverse video datasets containing different resolutions, encoding formats, and content categories. Second, clustering performance may decline in highly dynamic environments where new videos are continuously added. Third, intelligent hashing mechanisms require optimization to maintain a balance between detection accuracy and computational efficiency.

Another limitation relates to implementation complexity. Compared with traditional deduplication systems, the proposed multi-layer architecture requires additional computational components, including feature extraction modules, clustering management, and adaptive decision mechanisms. Although these components improve scalability, they may increase system deployment complexity.

Future research should investigate machine learning-based optimization techniques for automatic parameter selection, including adaptive clustering thresholds, dynamic hash generation strategies, and predictive redundancy analysis. Integration with advanced cloud-edge infrastructures may further improve real-time video processing capabilities.

CONCLUSION

This research presented a Layered Feature Resolution Model with Intelligent Hashing for Scalable Cloud-Based Video Redundancy Elimination. The proposed framework addresses the limitations of conventional deduplication systems by introducing a progressive processing architecture that combines feature analysis, clustering, intelligent hashing, and secure cloud storage optimization.

The research demonstrates that video redundancy elimination requires approaches beyond traditional exact duplicate detection. Due to the complex nature of multimedia data, including encoding variations, resolution differences, and partial content overlap, scalable systems must incorporate adaptive similarity analysis. The proposed model achieves this objective by applying coarse-level analysis before detailed verification, thereby reducing unnecessary computational operations.

The literature analysis revealed that existing approaches provide valuable contributions but remain limited when applied independently. Hash-based methods improve duplicate detection efficiency but struggle with modified multimedia content. Clustering approaches reduce comparison complexity but require adaptive management for dynamic cloud environments. Video-specific segmentation techniques improve structural analysis but require integration with intelligent similarity mechanisms.

The proposed LFRM-IH framework combines these concepts into a unified architecture. Its layered design enables efficient resource utilization by ensuring that computationally expensive operations are applied only when required. Intelligent hashing improves duplicate identification capability, while clustering enhances scalability by organizing similar video content into manageable groups.

The framework also considers security requirements through integration principles derived from secure cloud-edge deduplication research. This aspect is increasingly important as cloud storage systems handle sensitive multimedia information across distributed infrastructures.

The major contribution of this research is the development of a conceptual architecture that provides a foundation for next-generation cloud video storage optimization. By combining adaptive feature resolution with intelligent hashing, the proposed model offers potential improvements in storage efficiency, processing scalability, and redundancy detection accuracy.

Future research should focus on implementing the proposed framework using real-world cloud environments and evaluating performance through quantitative metrics such as storage reduction ratio, processing latency, computational overhead, and detection accuracy. Additionally, incorporating advanced artificial intelligence techniques could further enhance automatic optimization and enable self-adaptive video deduplication systems.

REFERENCES

1. Chen, L., F. Xiang, and Z. Sun, “Image deduplication based on hashing and clustering in cloud storage,” KSII Trans. Internet Inf. Syst., vol. 15, no. 4, pp. 1448–1463, 2021.

2. Hema, S., and A. Kangaiammal, “Distributed storage

hash algorithm (DSHA) for file-based deduplication in cloud computing," in Second International Conference on Computer Networks and Communication Technologies: ICCNCT 2019, Coimbatore, India: Springer International Publishing, 2020, pp. 572–581.

3. Ke, Z., H. Gong, and D. H. Du, PM-Dedup: Secure Deduplication with Partial Migration from Cloud to Edge Servers. arXiv preprint arXiv:2501.02350, 2025.

4. Periasamy, J. K., and B. Latha, "Efficient hash function–based duplication detection algorithm for data deduplication deduction and reduction," Concurr. Comput.: Pract. Exp., vol. 33, no. 3, pp. e5213, 2021.

5. Sujatha, G. et al., "An efficient cloud storage model for GOP-Level video deduplication using adaptive GOP structure," Cybern. Syst., vol. 54, pp. 1–26, 2023.

6. Tan, Y. et al., "Improving the performance of deduplication-based storage cache via content-driven cache management methods," IEEE Trans. Parallel Distrib. Syst., vol. 32, no. 1, pp. 214–228, 2020.

7. Wu, J., and Y. Fu, "SGX based cloud-edge collaborative secure deduplication," in Workshop Proceedings of the 53rd International Conference on Parallel Processing, Gotland Sweden: Association for Computing Machinery New York NY United States, 2024, pp. 112–113.

8. Xia, W. et al., "The design of fast content-defined chunking for data deduplication-based storage systems," IEEE Trans. Parallel Distrib. Syst., vol. 31, no. 9, pp. 2017–2031, 2020.

9. K. S. Hebbar, "MACHINE LEARNING-ASSISTED SERVICE BOUNDARY DETECTION FOR MODULARIZING LEGACY SYSTEMS," International Journal of Applied Engineering & Technology, vol. 04,no.02, pp. 401-414, Sep. 2022, https://romanpub.com/resources/ijaet-v4-2-2022-48.pdf

10. M. A. Hussain, V. B. Meruga, A. K. Rajamandrapu, S. R. Varanasi, S. S. S. Valiveti and A. G. Mohapatra, "Generative AI Sensor Fusion for Secure Digital Twin Ecosystems: A Standardization-Aligned Framework for Cyber-Physical Systems," in IEEE Communications Standards Magazine, doi: 10.1109/MCOMSTD.2026.3660106.