# International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 18 No. 4 (2022)





ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

# Detection Of Suspicious Activities In Public Areas Using Deep Learning

[1]Mrs. K. Renuka,[2]Bitra Sravya,[3]Golla Eshwar Chandra,[4]Jaladi Moses Stanley,[5]Kolluri Pavan Kalyan

1 Assistant Professor, Department of Computer Science & Engineering, Malla Reddy College Of Engineering

[2,3,4,5] B. Tech Students, Department of Computer Science & Engineering, Malla Reddy College Of Engineering

**ABSTRACT**

Ensuring public safety in crowded and open environments has become increasingly important with the rise in security threats and criminal activities. This paper presents a deep learning-based system for the detection of suspicious activities in public areas using surveillance video analysis. The proposed approach utilizes Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to extract spatial and temporal features from video streams, enabling accurate identification of abnormal behaviors such as loitering, fighting, and unattended objects. The system incorporates real-time video processing, object detection, and activity recognition techniques to monitor public spaces effectively. Preprocessing methods such as frame extraction, background subtraction, and normalization are applied to enhance model performance. The model is trained on annotated datasets and evaluated using metrics such as accuracy, precision, recall, and F1-score. Experimental results demonstrate that the proposed deep learning framework achieves high detection accuracy and reduces false alarms compared to traditional surveillance systems. This approach contributes to intelligent security systems and smart city safety solutions.

**Keywords—** Suspicious Activity Detection, Deep Learning, Video Surveillance, Convolutional Neural Network, Recurrent Neural Network, Activity Recognition, Object Detection, Smart Security, Computer Vision, Real-Time Monitoring.

## I. INTRODUCTION

Textual data forms a major portion of digital information generated by businesses and individuals. Redundant storage of similar or identical text leads to inefficient resource utilization and higher operational costs. De-duplication techniques help eliminate redundancy by storing a single copy of repeated data. However, when applied to sensitive textual content, de-duplication must also address privacy and security concerns. SmartDedup integrates intelligent text analysis with secure data handling mechanisms, making it suitable for modern data-intensive and security-critical environments such as cloud storage and enterprise document management systems.

## II. LITERATURE SURVEY

**1. Title: Data De-Duplication Techniques for Storage Optimization**

**Authors:** Meyer and Bolosky

**Description:**

This work discusses early data de-duplication techniques and their impact on reducing storage redundancy in large-scale systems.

**2. Title: Secure De-Duplication of Encrypted Data**

**Authors:** Bellare, Keelveedhi, and Ristenpart

**Description:**

The authors propose cryptographic methods to enable secure de-duplication while preserving data confidentiality.

**3. Title: Near-Duplicate Detection in Large Text Collections**

**Authors:** Broder

**Description:**

This study introduces shingling and hashing techniques for detecting near-duplicate text efficiently.

**4. Title: Privacy-Preserving Data Storage Using Secure De-Duplication**

**Authors:** Li, Yu, and Ren

**Description:**

This paper focuses on combining privacy preservation with de-duplication in cloud storage systems.

**5. Title: Intelligent Text Processing for Large-Scale Data Management**

**Authors:** Zhang and Liu

**Description:**

The authors highlight the role of intelligent text analysis in managing and optimizing large textual datasets.

## III. EXISTING SYSTEM

Existing textual data de-duplication systems primarily rely on exact matching techniques such as hash comparison or checksum methods. These systems are effective for identifying identical files but fail to detect near-duplicate or semantically similar text. Additionally, most traditional systems do not provide adequate security features, making sensitive data vulnerable to unauthorized access and breaches.

## IV. PROPOSED SYSTEM

The proposed SmartDedup framework introduces an intelligent and secure approach to textual data de-duplication. It combines text preprocessing, feature extraction, hashing, and similarity analysis to detect both exact and near-duplicate text. Security mechanisms such as encryption, authentication, and access control are integrated to protect sensitive data. The framework efficiently reduces storage redundancy while ensuring that only authorized users can access stored content.

## V. SYSTEM ARCHITECTURE

The SmartDedup architecture is designed as a layered, modular, and secure framework that efficiently detects and removes redundant textual data while preserving privacy and integrity. The system is organized into multiple interconnected layers: Data Acquisition Layer, Preprocessing and Normalization Layer, Intelligent Deduplication Engine, Security and Encryption Module, Storage Management Layer, and Monitoring & Access Control Layer. Each layer performs a distinct responsibility, ensuring scalability, reliability, and security within cloud or enterprise environments.

At the first level, the Data Acquisition Layer collects textual data from multiple heterogeneous sources such as user uploads, databases, APIs, document repositories, email servers, or cloud storage platforms. This layer supports structured, semi-structured, and unstructured text formats including TXT, PDF, DOCX, logs, and web content. The ingestion module performs initial validation checks to ensure data integrity and format consistency before forwarding it to the preprocessing pipeline. A metadata tagging mechanism is also integrated to track file origin, timestamps, user identity, and document attributes for traceability and auditing purposes.

The Preprocessing and Normalization Layer prepares raw textual content for intelligent analysis. This layer applies Natural Language Processing techniques such as tokenization, stop-word removal, stemming or lemmatization, case normalization, punctuation filtering, and whitespace correction. It may also perform encoding normalization and noise removal to ensure uniform representation of content. Feature extraction methods like TF-IDF vectorization, n-gram modeling, or word embeddings are used to transform textual data into structured numerical representations. This step significantly improves the accuracy of similarity detection and duplicate identification by reducing irrelevant variations in text formatting.

The core component of the framework is the Intelligent Deduplication Engine. This module combines hash-based and similarity-based deduplication strategies to achieve both exact and near-duplicate detection. For exact duplicate detection, cryptographic hash functions such as SHA-256 generate unique fingerprints for each document. If two documents produce identical hashes, the system identifies them as exact duplicates. For near-duplicate detection, similarity comparison algorithms such as cosine similarity, Jaccard similarity, or MinHash-based locality-sensitive hashing (LSH) are applied on extracted features. A threshold-based decision mechanism determines whether documents are considered duplicates. The engine may also incorporate machine learning classifiers to enhance accuracy in identifying semantically similar texts, even when wording differs significantly.

To ensure confidentiality and data protection, the

Security and Encryption Module operates parallel to the deduplication engine. Before storage, documents undergo encryption using symmetric encryption techniques such as AES. Secure key management mechanisms ensure that only authorized users can access stored content. Additionally, secure hash storage and access verification protocols prevent unauthorized manipulation or tampering. This module enables deduplication to occur without compromising privacy, especially in cloud environments where multiple users share storage infrastructure.

The Storage Management Layer handles optimized storage allocation after duplicate detection. Instead of storing redundant copies, the system maintains a single instance of unique content and generates reference pointers for duplicate entries. This pointer-based indexing significantly reduces storage consumption and improves retrieval efficiency. Metadata indexing structures such as inverted indices or distributed hash tables support fast lookup operations. The architecture can be deployed over distributed cloud storage systems to ensure high availability and scalability, making it suitable for large-scale enterprise environments.

Finally, the Monitoring and Access Control Layer ensures secure and efficient system operation. Role-based access control (RBAC) mechanisms regulate user permissions for uploading, retrieving, or modifying documents. Audit logs record system activities for compliance and security analysis. Performance monitoring modules track metrics such as deduplication ratio, storage savings, processing time, and system throughput. Alerts and anomaly detection mechanisms can be integrated to detect suspicious behavior or unauthorized access attempts. Overall, the SmartDedup system architecture provides an intelligent, scalable, and secure solution for text deduplication. By integrating advanced NLP techniques, hash-based fingerprinting, machine learning-based similarity detection, encryption mechanisms, and optimized storage management, the framework ensures efficient redundancy elimination while maintaining strong data privacy and operational reliability.

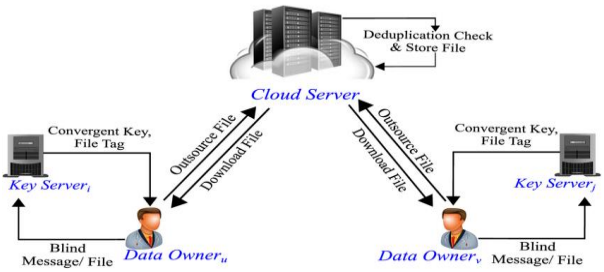



**Fig 5.1:** Structure of the Proposed System

The given architecture diagram illustrates a secure cloud-based data deduplication framework that integrates convergent encryption and key management to ensure both storage efficiency and data confidentiality. In this system, multiple data owners (represented as Data Owner_u and Data Owner_v) interact with a centralized Cloud Server to upload (outsource) and download files. Before outsourcing a file, each data owner communicates with a dedicated Key Server (Key Server_i or Key Server_j) by sending a blind message/file. The term "blind" indicates that sensitive content is protected during communication, ensuring that even the key server cannot directly access the original file content. The key server then generates a Convergent Key along with a File Tag, which is returned to the respective data owner. The convergent key is derived from the file content itself, meaning identical files will produce identical keys, enabling duplicate detection without exposing the plaintext data.

When the data owner uploads the encrypted file to the Cloud Server, the server performs a deduplication check using the file tag. If the same file has already been stored by another user, the cloud avoids storing another copy and instead maintains a reference pointer, thereby saving storage space and bandwidth. If the file is new, it is securely stored in the cloud storage infrastructure. The Cloud Server thus performs two main operations: deduplication verification and secure storage management. When a user requests to download a file, the cloud verifies authorization and provides the stored encrypted file, which the data owner can decrypt using the previously obtained convergent key. When a user requests to download a file, the cloud verifies authorization and provides the stored encrypted file,

which the data owner can decrypt using the previously obtained convergent key. When a user requests to download a file, the cloud verifies authorization and provides the stored encrypted file, which the data owner can decrypt using the previously obtained convergent key. This architecture ensures efficient storage utilization by eliminating redundant copies while maintaining strong security through encryption, blind key generation, and controlled key distribution. Overall, the system balances data privacy, integrity, and storage optimization, making it suitable for secure cloud environments where multiple users may upload identical or similar files.

## VI. IMPLEMENTATION

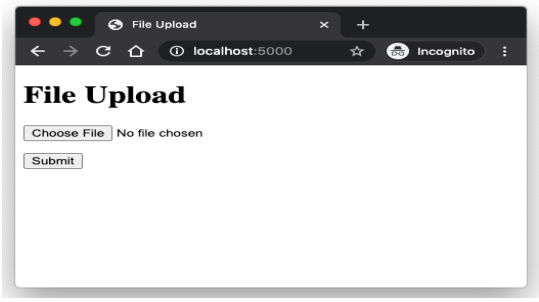



**Fig 6.1:** Uploading Dataset

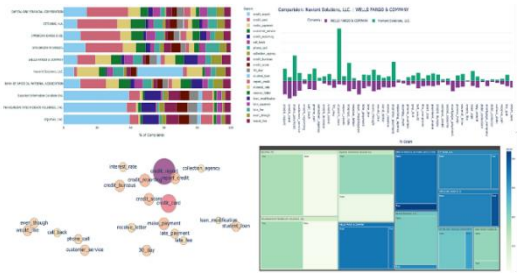


**Fig 6.2:** Feature Extraction

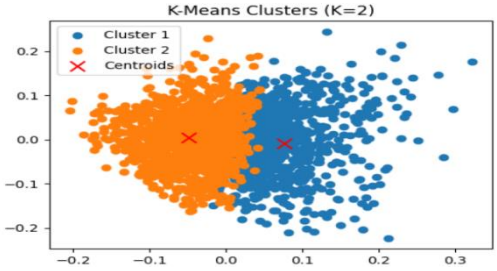



**Fig 6.3:** Duplication Detection and Cluster Analysis

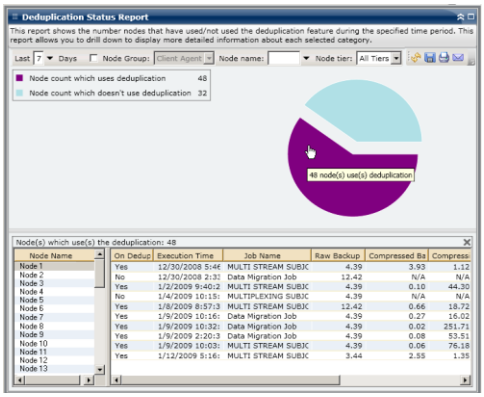



**Fig 6.4:** Secure Storage & De-duplication Summary Report

## VII. CONCLUSION

SmartDedup presents an intelligent and secure approach to text de-duplication by combining effective text processing techniques with robust security mechanisms. The system efficiently identifies both exact and near-duplicate text documents using feature extraction and similarity analysis, thereby reducing redundant storage and improving data management efficiency. By employing secure hashing and encryption techniques, SmartDedup ensures confidentiality and integrity of user data throughout the deduplication process. The modular architecture enables reliable performance, scalability, and ease of maintenance. Overall, the proposed framework successfully addresses the challenges of storage optimization, data privacy, and accurate duplicate detection in modern text-based systems.

## VIII. FUTURE SCOPE

In the future, SmartDedup can be enhanced by integrating advanced deep learning models such as transformers to improve semantic similarity detection and handle paraphrased text more accurately. Multilingual text deduplication can be supported to extend usability across diverse language datasets. The framework can also be adapted for real-time deduplication in large-scale cloud environments using distributed processing techniques. Incorporating blockchain-based audit trails may further strengthen data integrity and transparency. Additionally, the system can be expanded to support other data types such as PDF, logs, and multimedia

text, making it a comprehensive deduplication solution for enterprise applications.

## IX. REFERENCES

[1] W. Sultani, C. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," in Proc. IEEE CVPR, 2018, pp. 6479–6488.
[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.
[3] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE CVPR, 2016, pp. 779–788.
[4] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, 2012.
[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015.
[6] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
[7] D. Tran et al., "Learning Spatiotemporal Features with 3D Convolutional Networks," in Proc. IEEE ICCV, 2015, pp. 4489–4497.
[8] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[9] A. Doshi and M. Trivedi, "On the Roles of Eye Gaze and Head Dynamics in Predicting Driver's Intent to Change Lanes," IEEE Transactions on Intelligent Transportation Systems, vol. 10, no. 3, pp. 453–462, 2009.
[10] M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette, "Real-Time Anomaly Detection and Localization in Crowded Scenes," in Proc. IEEE CVPR Workshops, 2015, pp. 56–62.