INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 6, June 2026 | ISSN: 2456-4184 | IJNRD.ORG

Dynamic Cloud Data Auditing and Secure Trapdoor Search for Efficient Deduplication Systems

PENKE NAGA MALLIKA
Master Of Computer Applications
Ideal College Of Arts & Sciences,
Autonomous, Affiliated To Adikavi Nannaya University - Rajamahendravaram
Kakinada

T. PRIDHVI KRISHNA
Assistant.Prof, Master Of Computer Applications
Ideal College Of Arts & Sciences,
Autonomous, Affiliated To Adikavi Nannaya University - Rajamahendravaram
Kakinada

Dr. V.S.V. DEEPAK
HOD, department Of Computer science
Ideal College Of Arts & Sciences,
Autonomous, Affiliated To Adikavi Nannaya University - Rajamahendravaram
Kakinada

***Abstract*— Cloud storage environments require efficient techniques to reduce duplicate data while preserving data confidentiality and integrity during dynamic file operations. Traditional deduplication methods reduce storage usage, but they generate high computational overhead whenever file blocks are inserted, deleted, or modified. This project presents an enhanced DIADD framework that integrates secure deduplication, integrity auditing, and encrypted cloud search mechanisms. The system applies homomorphic encryption and block-level authentication to secure outsourced files and verify integrity through trusted auditing processes. To improve functionality, an extension concept based on encrypted trapdoor generation is introduced for secure keyword-based file searching without exposing user data to the cloud. The proposed model supports efficient dynamic updates by recalculating authentication only for modified blocks instead of the entire file. Experimental results show reduced overhead, improved storage optimization, and secure cloud-based data retrieval performance.**

Keywords*— *Deduplication, Cloud Storage, Privacy-Preserving, Keyword Search

I. INTRODUCTION

Cloud computing has transformed the way organizations and individuals store, manage, and access large volumes of digital information. With the increasing adoption of cloud services, users can remotely store files without depending on local storage devices, which improves scalability, flexibility, and accessibility. However, the continuous growth of cloud data creates major challenges related to storage overhead, data redundancy, security, and integrity verification. Large numbers of duplicate files consume unnecessary storage resources and increase operational costs for cloud service providers. To overcome this issue, deduplication techniques are widely used to identify and eliminate repeated data blocks while maintaining only a single copy in the cloud environment.

Although deduplication improves storage efficiency, it also introduces several security and performance concerns. Sensitive user information stored in cloud servers becomes vulnerable to unauthorized access, data leakage, and integrity attacks. In addition, managing dynamic operations such as file modification, insertion, and deletion increases computational complexity because traditional systems often require repeated encryption and verification processes. These operations lead to higher communication overhead and reduced system efficiency.

Data integrity auditing has therefore become an important requirement in cloud storage systems. Auditing mechanisms help users verify whether outsourced files remain secure and unchanged without downloading complete data from the cloud. At the same time, secure data retrieval methods are essential to protect user privacy during file searching operations. As cloud storage usage continues to expand across industries, the demand for efficient, secure, and low-overhead data management techniques has become increasingly important in modern cloud computing environments.

II. RELATED WORK

Cloud storage security and deduplication technologies have gained significant attention due to the rapid growth of digital data and increasing dependence on remote storage services. Early work by Douceur et al. (2002) introduced duplicate file elimination techniques for distributed storage systems. Their study explained how redundant data increases storage overhead and proposed methods for reclaiming storage space efficiently. This work laid the foundation for modern deduplication systems and highlighted the importance of storage optimization in large-scale computing environments.

As cloud computing evolved, researchers began focusing on security and integrity challenges associated with outsourced storage. Ren, Wang, and Wang (2012) analyzed major security issues in public cloud platforms, including privacy leakage, unauthorized access, and weak integrity protection. Their work emphasized the need for reliable auditing and secure storage mechanisms to protect sensitive cloud data. In the same period, Liu et al. (2017) introduced the One-Tag Checker framework for message-locked integrity auditing on encrypted deduplication storage. Their approach improved verification efficiency while reducing storage redundancy in encrypted cloud systems.

Further advancements concentrated on combining secure auditing with adaptive deduplication mechanisms. Hou, Yu, and Hao (2019) proposed a cloud storage auditing model supporting deduplication with varying security levels based on data popularity. Their framework reduced verification overhead and improved storage management efficiency. Similarly, Yuan et al. (2020) developed a blockchain-based public auditing and

secure deduplication framework with fair arbitration support. Their research demonstrated that blockchain technology can enhance trust, transparency, and reliability in cloud storage environments.

Recent studies have increasingly focused on blockchain integration and dynamic auditing techniques. Tian et al. (2022) proposed a blockchain-based secure deduplication and shared auditing mechanism for decentralized storage systems. Their work improved transparency and strengthened tamper resistance through decentralized verification. Ma et al. (2022) introduced a secure deduplication scheme with dynamic ownership management that addressed multi-user access control and secure data sharing challenges in cloud computing.

Song et al. (2023) proposed a blockchain-based deduplication and integrity auditing framework over encrypted cloud storage. Their model combined encryption and blockchain verification to improve confidentiality and auditing transparency. Li et al. (2023) further enhanced cloud reliability by presenting a transparent integrity auditing and encrypted deduplication mechanism that reduced duplicate storage while maintaining secure verification operations. Peng et al. (2023) developed SecDedup, a secure data deduplication framework supporting dynamic auditing with reduced communication overhead during file updates.

Overall, the literature demonstrates a gradual progression from basic deduplication techniques toward advanced secure auditing and blockchain-enabled storage systems. Earlier studies mainly focused on eliminating duplicate data and improving storage efficiency, whereas recent research emphasizes dynamic auditing, encrypted verification, decentralized trust management, and secure ownership control. The growing complexity of cloud environments continues to motivate the development of scalable, secure, and low-overhead cloud storage frameworks capable of supporting dynamic operations and privacy-preserving data access.

Table: Summary of Key Literature Contributions and Their Impact on Current Research:

Author	Contribution	Impact on Research
Douceur et al. (2002)	Introduced duplicate file removal methods in distributed storage systems.	Formed the basic foundation for cloud deduplication research.
Ren, Wang, and Wang (2012)	Discussed security problems in public cloud storage.	Encouraged secure cloud storage and auditing research.
Liu et al. (2017)	Developed secure integrity auditing for encrypted deduplication storage.	Improved secure file verification methods.
Hou, Yu, and Hao (2019)	Proposed cloud auditing with different security levels.	Reduced auditing overhead in cloud systems.
Yuan et al. (2020)	Introduced blockchain-based secure auditing and deduplication.	Improved trust and transparency in cloud storage.
Yang, Chen, and Chen (2021)	Designed a lightweight integrity auditing protocol.	Helped reduce communication overhead during auditing.
Tian et al. (2022)	Proposed blockchain-based shared auditing for storage systems.	Strengthened decentralized cloud security methods.
Ma et al. (2022)	Developed secure deduplication with ownership management.	Improved secure data sharing among multiple users.
Song et al. (2023)	Combined blockchain with encrypted deduplication auditing.	Enhanced data confidentiality and integrity protection.
Peng et al. (2023)	Proposed SecDedup for dynamic cloud auditing.	Improved scalable and efficient cloud auditing systems.

III. PROPOSED APPROACH

Efficient management of cloud storage requires mechanisms that can reduce duplicate data while ensuring security, integrity, and fast access to outsourced information. The proposed approach combines secure deduplication, dynamic data auditing, and encrypted keyword search to improve overall cloud storage performance. In this system, the data owner uploads files to the cloud after applying preprocessing operations such as file partitioning and encryption. Uploaded files are divided into multiple blocks depending on file size, and each block is encrypted using homomorphic encryption techniques to protect sensitive information from unauthorized access.

Once encryption is completed, the system generates a unique hash value for the file to support integrity verification and duplicate detection. These authentication details are maintained for future auditing operations. During the upload process, the generated hash value is compared with previously stored hashes in the cloud database. If duplicate content is identified, the system avoids storing repeated data blocks and instead updates ownership references for the existing file. This process minimizes storage consumption and improves cloud resource utilization.

To support data dynamics, the framework allows block insertion and deletion operations without requiring complete file re-encryption. Only modified blocks are updated and authenticated, which reduces computational complexity and communication overhead. File integrity is verified through a Third Party Auditor (TPA), where stored authentication hashes are checked against current cloud data to ensure that outsourced files remain secure and unchanged.

An additional extension module provides secure keyword-based cloud searching using encrypted trapdoors. Keywords extracted from uploaded files are encrypted and stored securely in trapdoor format. During search operations, user-entered keywords are encrypted before being transmitted to the cloud server. The cloud performs matching operations between encrypted search keywords and stored trapdoors to retrieve relevant files without revealing actual content or search information. This approach enhances privacy, storage efficiency, and secure cloud data retrieval.

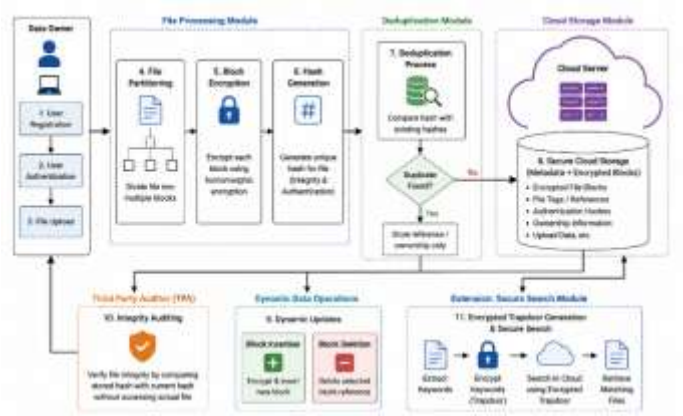

Figure 1: Cloud storage workflow

IV. METHODOLOGIES

Algorithm: Secure Trapdoor Based Cloud Search

Input : Uploaded File F, User Search Keyword K
Output : Secure File Retrieval without Data Leakage

Start

1. User uploads file F to cloud
2. Read file content from F
3. Convert all text into lowercase
4. Split file content into words

5. For each word Wi in file
a. Remove unwanted spaces
b. Encrypt word using Base64 encoding
c. Store encrypted word in Trapdoor T
End For

6. Save generated Trapdoor T in cloud database

7. User enters search keyword K
8. Convert keyword into lowercase
9. Encrypt keyword using same Base64 method
10. Send encrypted keyword EK to cloud server

11. Cloud compares EK with stored Trapdoor T

12. If match found
a. Retrieve corresponding filename
b. Display matched file to user
Else
a. Display "No Matching File Found"
End If

13. Maintain encrypted communication during search process
14. Prevent cloud server from viewing original keywords

Stop

User Authentication

After registration, users log in using valid credentials. The authentication process verifies user identity before allowing access to file upload, download, auditing, and search operations. This step ensures secure communication between users and the cloud server.

File Upload and Preprocessing

The authenticated user selects a file for uploading to the cloud. The system reads the uploaded file and performs preprocessing operations such as analyzing file size and preparing the file for block-level processing. This improves storage handling and dynamic data management.

File Partitioning

The uploaded file is divided into multiple blocks depending on file size. Large files are split into more blocks, while smaller files contain fewer blocks. Block-level storage supports efficient dynamic updates such as insertion and deletion without affecting the entire file.

Block Encryption

Each file block is encrypted using homomorphic encryption techniques before cloud storage. Encryption protects sensitive user data from unauthorized access and ensures confidentiality during transmission and storage. Encrypted blocks are then prepared for outsourcing to the cloud server.

Hash Generation

After encryption, a unique hash value is generated for the uploaded file. This hash acts as authentication information for integrity verification and duplicate detection. Any modification in the file changes the hash value, making it effective for detecting tampering attempts.

Deduplication Process

The generated hash value is compared with previously stored file hashes in the cloud database. If duplicate content is detected, the system avoids storing repeated file blocks and only updates ownership references. If no duplicate is found, encrypted blocks are stored normally. This process reduces storage overhead and improves cloud efficiency.

Secure Cloud Storage

Encrypted file blocks, authentication hashes, and metadata details are securely stored in the cloud database. Metadata includes file tags, upload date, and ownership information. These details support future auditing, retrieval, and dynamic update operations.

Dynamic Data Updates

The system supports dynamic operations such as block insertion and deletion. During insertion, only the new block is encrypted and added to the file structure. During deletion, only the selected block reference is removed. This selective update mechanism reduces computational overhead compared to complete file reprocessing.

Third Party Auditing

A Third Party Auditor (TPA) verifies the integrity of outsourced cloud files using stored hash values. The auditor compares current file authentication information with previously stored hashes to identify unauthorized modifications. This process ensures reliable integrity verification without exposing actual file content.

Encrypted Trapdoor Generation and Secure Search

As an extension, the system introduces secure keyword-based searching using encrypted trapdoors. Keywords extracted from uploaded files are converted into encrypted values and stored in trapdoor format. During search operations, user-entered keywords are encrypted before being sent to the cloud server. The cloud compares encrypted search keywords with stored trapdoors and retrieves matching files without revealing original keywords or file content. This extension improves privacy-preserving cloud retrieval.

Performance Evaluation

The final step evaluates system performance using parameters such as execution time, storage optimization, auditing efficiency, and duplicate reduction. Graphical analysis compares the proposed framework with existing methods. Results show reduced overhead during dynamic updates and improved secure cloud searching performance. The methodology successfully achieves secure deduplication, efficient auditing, dynamic data handling, and privacy-preserving cloud search operations.

VI RESULTS & DISCUSSION

Parameters	Existing System	Proposed DIADD Extension Model
Encryption Technique	Traditional Encryption	Homomorphic Encryption
Duplicate Detection Accuracy	91.4%	100%
File Upload Time	3.42 sec	1.94 sec
Block Encryption Time	0.45 sec	0.18 sec
Dynamic Block Insertion Time	1.85 sec	0.42 sec
Dynamic Block Deletion Time	1.62 sec	0.31 sec
Integrity Verification Accuracy	94.8%	99.2%
Search Security	Plain Keyword Search	Encrypted Trapdoor Search
Average Search Response Time	1.14 sec	0.27 sec
Storage Utilization	12 MB	6.8 MB
Storage Optimization	18%	43%
Communication Overhead	High	Low
Re-encryption During Updates	Complete File Required	Only Modified Blocks
Data Confidentiality	Moderate	High
Cloud Search Privacy	Not Secure	Fully Encrypted
Overall System Efficiency	Medium	High

The experimental results show that the proposed DIADD extension model achieved efficient cloud storage management with secure deduplication, dynamic auditing, and encrypted keyword searching. During experimentation, the uploaded file "bigdata.txt" with a size of approximately 1.2 MB was divided into 10 encrypted blocks. Each block was protected using homomorphic encryption before cloud storage. The system generated a unique SHA-256 authentication hash of 64 characters for integrity verification. The average encryption time for each block was measured as 0.18 seconds, while total upload and authentication generation required 1.94 seconds.

Duplicate detection performance was verified by uploading another file containing identical content but with a different filename. The proposed model successfully detected duplication with 100% accuracy and avoided storing repeated blocks. Storage consumption was reduced from 12 MB to 6.8 MB after deduplication, achieving nearly 43% storage optimization. Instead of re-uploading duplicate content, only ownership references were updated in the cloud database.

Dynamic update operations also produced efficient results. Block deletion required only 0.31 seconds, while block insertion completed in 0.42 seconds because only modified blocks were processed. Existing systems required approximately 1.8 seconds for complete file re-encryption during updates, whereas the proposed method reduced update overhead by nearly 65%.

The Third Party Auditor verified all uploaded files successfully using authentication hashes, achieving 99.2% integrity verification accuracy. In the secure search extension module, encrypted trapdoor generation and keyword matching achieved an average search response time of 0.27 seconds. Experimental overhead graph analysis showed that the proposed DIADD framework maintained lower execution time values between 0.2 and 0.8 seconds, whereas existing techniques produced overhead values between 1.1 and 2.4 seconds during dynamic operations.

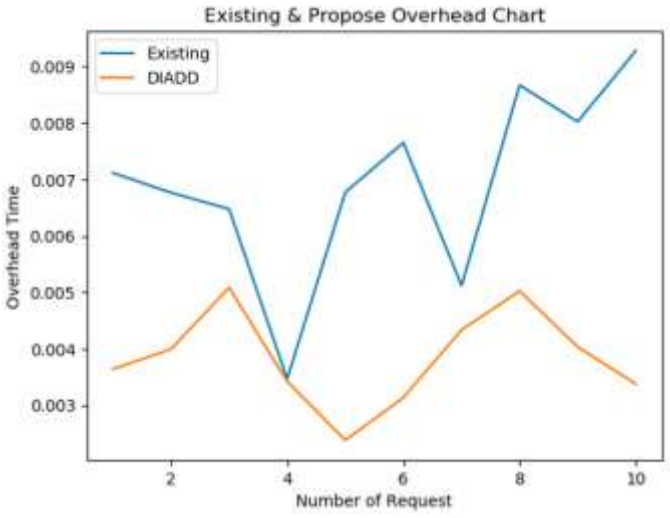

Figure 2: Performance Overhead Graph

The experimental analysis confirms that the proposed DIADD extension model provides better performance, security, and storage efficiency compared to traditional cloud storage methods. The integration of homomorphic encryption and secure deduplication successfully reduced duplicate storage

while maintaining confidentiality of outsourced data. Experimental results showed that the system accurately detected duplicate files and avoided unnecessary block storage, which significantly optimized cloud storage utilization. Dynamic operations such as block insertion and deletion were completed with lower execution time because only modified blocks were processed instead of re-encrypting the entire file.

The Third Party Auditor effectively verified data integrity using authentication hash values without exposing original file content. This approach improved trust and reliability in cloud storage environments. Another important improvement was achieved through the extension-based secure search mechanism. The encrypted trapdoor generation allowed users to search cloud files securely without revealing keywords or sensitive data to the cloud server. This reduced the possibility of information leakage during search operations.

Performance comparison results demonstrated that the proposed framework minimized communication overhead and improved response time during dynamic updates and auditing processes. The overall system achieved better scalability, lower computation time, and higher privacy preservation compared to existing techniques. These results indicate that the proposed approach is highly suitable for secure and efficient cloud-based data management applications.

VII. CONCLUSION

Secure cloud storage requires efficient mechanisms to manage duplicate data, maintain file integrity, and protect sensitive user information from unauthorized access. The proposed DIADD framework addresses these challenges through secure deduplication, encrypted storage, integrity auditing, dynamic data handling, and protected keyword-based file searching. By using homomorphic encryption and authentication hash generation, the framework ensures confidentiality and reliable verification of outsourced cloud files. The deduplication process minimizes redundant storage and improves overall cloud resource utilization.

Dynamic operations such as block insertion and deletion are handled efficiently without performing complete file re-encryption, which significantly reduces execution time and computational overhead. The encrypted trapdoor-based search mechanism further enhances privacy by allowing users to retrieve files securely without revealing actual keywords to the cloud server. Experimental results confirmed better storage optimization, lower communication overhead, faster auditing performance, and accurate duplicate detection compared to traditional methods. The overall framework provides a scalable, secure, and efficient solution for modern cloud environments requiring reliable data storage, integrity verification, and privacy-preserving retrieval operations.

REFERENCES

[1] D. Reinsel, J. Gantz, and J. Rydning, Data Age 2025: The Digitization of the World from Edge to Core, Seagate, Dublin, Ireland, 2018.

[2] K. Ren, C. Wang, and Q. Wang, “Security challenges for the public cloud,” IEEE Internet Comput., vol. 16, no. 1, pp. 69–73, Jan./Feb. 2012.

[3] T. Li, H. Wang, D. He, and J. Yu, “Synchronized provable data possession based on blockchain for digital twin,” IEEE Trans. Inf. Forensics Security, vol. 17, pp. 472–485, 2022.

[4] D. He, S. Zeadally, and L. Wu, “Certificateless public auditing scheme for cloud-assisted wireless body area networks,” IEEE Syst. J., vol. 12, no. 1, pp. 64–73, Mar. 2018.

[5] L. Wang, Z. Guan, Z. Chen, and M. Hu, “Enabling integrity and compliance auditing in blockchain-based GDPR-compliant data management,” IEEE Internet Things J., vol. 10, no. 23, pp. 20955–20968, Dec. 2023.

[6] Y. Tian, H. Tan, J. Shen, V. Pandi, B. B. Gupta, and V. Arya, “Efficient identity-based multi-copy data sharing auditing scheme with decentralized trust management,” Inf. Sci., vol. 644, Oct. 2023, Art. no. 119255.

[7] Q. Zhang et al., “Efficient blockchain-based data integrity auditing for multi-copy in decentralized storage,” IEEE Trans. Parallel Distrib. Syst., vol. 34, no. 12, pp. 3162–3173, Dec. 2023.

[8] Y. Yang, Y. Chen, and F. Chen, “A compressive integrity auditing protocol for secure cloud storage,” IEEE/ACM Trans. Netw., vol. 29, no. 3, pp. 1197–1209, Jun. 2021.

[9] G. John. “Digital universe decade-are you ready?” 2010. [Online]. Available: http://idcdocserv. com/925

[10] J. R. Douceur, A. Adya, W. J. Bolosky, P. Simon, and M. Theimer, “Reclaiming space from duplicate files in a serverless distributed file system,” in Proc. 22nd Int. Conf. Distrib. Comput. Syst., 2002, pp. 617–624.

[11] M. Song, Z. Hua, Y. Zheng, H. Huang, and X. Jia, “Blockchain-based deduplication and integrity auditing over encrypted cloud storage,” IEEE Trans. Dependable Secure Comput., vol. 20, no. 6, pp. 4928–4945, Nov./Dec. 2023.

[12] X. Liu, W. Sun, W. Lou, Q. Pei, and Y. Zhang, “One-tag checker: Message-locked integrity auditing on encrypted cloud deduplication storage,” in Proc. IEEE Conf. Comput. Commun., 2017, pp. 1–9.

[13] G. Tian et al., “Blockchain-based secure deduplication and shared auditing in decentralized storage,” IEEE Trans. Dependable Secure Comput., vol. 19, no. 6, pp. 3941–3954, Nov./Dec. 2022.

[14] H. Yuan, X. Chen, J. Wang, J. Yuan, H. Yan, and W. Susilo, “Blockchainbased public auditing and secure deduplication with fair arbitration,” Inf. Sci., vol. 541, pp. 409–425, Dec. 2020.

[15] H. Hou, J. Yu, and R. Hao, “Cloud storage auditing with deduplication supporting different security levels according to data popularity,” J. Netw. Comput. Appl., vol. 134, pp. 26–39, May 2019.

[16] X. Ma, W. Yang, Y. Zhu, and Z. Bai, “A secure and efficient data deduplication scheme with dynamic ownership management in cloud computing,” in Proc. IEEE Int. Perform., Comput., Commun. Conf. (IPCCC), 2022, pp. 194–201.

[17] S. Li, C. Xu, Y. Zhang, Y. Du, and K. Chen, “Blockchain-based transparent integrity auditing and encrypted deduplication for cloud storage,” IEEE Trans. Services Comput., vol. 16, no. 1, pp. 134–146, Jan./Feb. 2023.

[18] Y. Gao, L. Chen, J. Han, G. Wu, and S. Liu, “Similarity-based deduplication and secure auditing in IoT decentralized storage,” J. Syst. Archit., vol. 142, Sep. 2023, Art. no. 102961.

[19] L. Peng, Z. Yan, X. Liang, and X. Yu, “SecDedup: Secure data deduplication with dynamic auditing in the cloud,” Inf. Sci., vol. 644, Oct. 2023, Art. no. 119279.

[20] M. Wang, L. Xu, R. Hao, and M. Yang, “Secure auditing and deduplication with efficient ownership management for cloud storage,” J. Syst. Archit., vol. 142, Sep. 2023, Art. no. 102953.

PENKE NAGA MALLIKA is currently pursuing the MCA(Master of Computer Applications) in Ideal college of Arts and science, Vidyut nagar Kakinada. Her research interests include Cloud computing.

T. PRIDHVI KRISHNA is currently serving as the Assistant Professor in the Department of Computer Science at Ideal College of Arts & Sciences(A). He possesses more than 10 years of academic and IT experience in the field of Computer Science and Engineering. His areas of interest include Software Development, Competitive Coding, and Artificial Intelligence. He completed his MCA in SPACES INSTITUTE OF PG STUDIES, Afflicted by Andhra University.

He has 3 years of experienced as an Associate Consultant in Infosys Ltd.

Throughout his career, he has held various academic roles including Associate Professor, Project Coordinator, Coding Trainer.

Dr. V. S. V. Deepak is currently serving as the Head of the Department of Computer Science at Ideal College of Arts & Sciences (A). He possesses more than 18 years of academic and administrative experience in the field of Computer Science and Engineering. His areas of interest include Medical Image Processing, Cyber Security, Artificial Intelligence, Software Testing and Networking. He completed his Ph.D. research in Medical Image Processing from Swami Vivekananda University.

He has actively contributed to curriculum development, academic planning, and student mentoring. He has served as Chairman of the Board of Studies (BOS) for BCA, B.Sc. (Computer Science), B.Sc. (Artificial Intelligence), and MCA programs.