Journal of Artificial Intelligence and Cyber Security (JAICS)
An International Open Access, Peer-Reviewed, Refereed Journal

Enhancing Cloud Storage Performance Through Advanced Deduplication Techniques

[1] Dr.P.Meenalochini

[1]*Associate Professor, Department of Electrical and Electronics Engineering, Sethu Institute of Technology, Virudhunagar.*
Corresponding Author e-mail: *meenalochinip@gmail.com*

Abstract: As the volume of data stored in cloud environments continues to grow exponentially, efficient data management becomes critical to ensuring scalability, cost-effectiveness, and performance. One of the most effective strategies to optimize cloud storage is data deduplication—a technique that eliminates redundant copies of data to reduce storage space and improve input/output (I/O) performance. This paper explores advanced deduplication techniques and their role in enhancing cloud storage performance. Traditional deduplication methods, such as fixed-size and variable-size chunking, have provided substantial space savings but face limitations in handling complex, large-scale cloud workloads. These methods often result in high computational overhead, latency, and limited adaptability to evolving data patterns. To address these challenges, we investigate next-generation deduplication techniques that leverage machine learning, content-defined chunking (CDC), and hybrid approaches combining inline and post-process deduplication. Our research emphasizes the importance of intelligent data chunking and indexing mechanisms that adapt to varying data types and usage patterns. By integrating similarity detection algorithms and predictive models, advanced deduplication can preemptively identify redundant data with greater accuracy, thereby reducing processing time and resource consumption. Furthermore, distributed deduplication frameworks are examined, highlighting their potential to scale across cloud infrastructures while maintaining deduplication efficiency and fault tolerance. We also analyze the performance implications of deduplication on cloud systems, including its impact on storage latency, throughput, and network bandwidth utilization. Experimental results from simulations and real-world deployments demonstrate that advanced deduplication can achieve up to 70–90% storage savings while improving read/write performance by minimizing data transfer and disk I/O. Additionally, we discuss the trade-offs between storage optimization and computational cost, as well as strategies to mitigate deduplication-induced latency. Security and privacy concerns are addressed by evaluating the risks of data leakage through fingerprinting and proposing encryption-compatible deduplication schemes. The paper concludes with recommendations for implementing scalable, secure, and efficient deduplication systems in cloud environments, alongside future research directions in deduplication-aware storage architectures and edge-cloud collaboration. In summary, advanced deduplication techniques offer a promising path toward more efficient and performant cloud storage solutions. By intelligently eliminating redundancy while preserving system responsiveness and data integrity, these methods support the growing demand for reliable, high-performance cloud infrastructure in diverse application domains.

Keywords: Cloud storage, data deduplication, storage optimization, performance enhancement, data redundancy elimination, cloud computing, efficient storage management, backup optimization, storage cost reduction, duplicate data detection, scalable storage solutions, data compression, storage efficiency, cloud data management, advanced deduplication algorithms.

1. INTRODUCTION

The rapid expansion of digital data, driven by the proliferation of cloud services, Internet of Things (IoT) devices, multimedia content, and enterprise applications, has placed unprecedented demands on cloud storage infrastructures. As organizations and individuals increasingly rely on cloud platforms to store and access large volumes of data, service providers face the challenge of maintaining scalability, performance, and cost-efficiency. One of the most pressing concerns in this context is the significant amount of redundant data that accumulates across storage systems. Redundant files, repeated backups, and duplicate content contribute to increased storage requirements, bandwidth consumption, and system latency. To combat these inefficiencies, data deduplication has emerged as a key optimization technique in cloud storage management.

Data deduplication is the process of identifying and eliminating duplicate data blocks or files, ensuring that only unique instances are stored. This technique not only reduces storage capacity requirements but also enhances performance by minimizing data transfer and disk I/O operations. Traditional deduplication methods—such as fixed-size chunking and post-process deduplication—have achieved moderate success in reducing storage footprints. However, with the growing complexity and scale of modern cloud workloads, these methods face limitations in terms of speed, scalability, and adaptability.

Advanced deduplication techniques are designed to overcome these limitations by incorporating intelligent algorithms, real-time data analysis, and distributed architectures. For example, content-defined chunking (CDC) offers improved granularity in data segmentation, making it more effective at identifying redundancy even in dynamic or modified datasets. Inline deduplication processes data before it is written to storage, reducing latency and improving efficiency. Additionally, the integration of machine learning and predictive analytics into deduplication workflows enables more accurate redundancy detection and adaptive optimization based on usage patterns.

The effectiveness of advanced deduplication not only translates to reduced storage costs but also directly impacts cloud performance metrics such as data access speed, bandwidth utilization, and system responsiveness. These improvements are especially critical in multi-tenant cloud environments where performance consistency and resource sharing are paramount. Moreover, the rise of hybrid and edge-cloud models has introduced new challenges and opportunities for deduplication, demanding innovative solutions that can operate across distributed nodes while preserving data integrity and security.

This paper explores the landscape of advanced deduplication techniques, analyzing their architecture, performance implications, and practical applications in modern cloud environments. Through detailed evaluation and case studies, we aim to highlight how intelligent deduplication can significantly enhance the efficiency and performance of cloud storage systems.

2. LITERATURE SURVEY

Cloud storage systems have become an essential component of modern IT infrastructure, supporting a wide range of applications from personal file storage to enterprise-level data warehousing. With this widespread adoption comes an explosion of data, much of which is redundant. Deduplication has been extensively studied and implemented as a solution to improve cloud storage efficiency by eliminating such redundancy. This literature survey reviews significant contributions to deduplication research, focusing on methods, optimizations, and their impact on storage performance.

Traditional Deduplication Techniques

Early research in data deduplication focused on two main approaches: file-level and block-level deduplication. File-level deduplication eliminates entire duplicate files, which is straightforward but less effective when small differences exist between files. Block-level deduplication, on the other hand, breaks files into chunks and compares them using hash values. Fixed-size chunking, one of the earliest methods, segments data into blocks of a predefined size and compares their fingerprints (Policroniades & Pratt, 2004). While simple, it struggles with insertions or deletions in files, which cause misalignment and reduce deduplication efficiency.

To address these issues, **Content-Defined Chunking (CDC)** was introduced (Muthitacharoen et al., 2001). In CDC, boundaries are determined by the data content rather than fixed sizes. Rabin fingerprinting is commonly used in this technique, making it more resilient to small changes in data. CDC significantly increases deduplication ratios compared to fixed-size chunking, especially in environments with versioned or slightly modified files, such as backup systems.

Inline vs. Post-Process Deduplication

Inline deduplication processes data before it is written to disk, thereby saving I/O operations and reducing bandwidth usage. However, it can introduce latency due to the time needed to compute hashes and lookups in real time (Meyer & Bolosky, 2001). In contrast, post-process deduplication scans stored data at scheduled intervals. While less time-sensitive, it consumes additional storage temporarily and introduces delay before space savings are realized. Hybrid approaches are now being explored to combine the responsiveness of inline deduplication with the thoroughness of post-process techniques.

Distributed Deduplication in Cloud Environments

As cloud systems scale across multiple nodes and data centers, centralized deduplication techniques face limitations in scalability and metadata management. **Distributed deduplication** addresses these concerns by spreading the deduplication workload across multiple storage nodes. Xia et al. (2011) proposed a scalable deduplication framework for cloud backup systems that employs a two-level deduplication mechanism: local deduplication at client side and global deduplication at the cloud side. However, distributed schemes introduce complexity in maintaining deduplication indices and ensuring consistency.

Extreme Binning (Lillibridge et al., 2009) was another notable effort to improve scalability. This technique reduces indexing overhead by classifying data chunks into bins, each associated with a representative fingerprint. It enables approximate deduplication with significantly reduced memory and computation requirements, which is critical for cloud-scale storage.

Security and Privacy Concerns

While deduplication is beneficial, it also introduces privacy and security concerns. Deduplication can be exploited through **side-channel attacks**, such as deduplication-aware file scanning (Harnik et al., 2010), where attackers infer the existence of specific files in shared storage systems. To counter this, researchers have proposed **convergent encryption**, where identical data blocks produce the same encrypted output, enabling deduplication while maintaining confidentiality. However, this technique is vulnerable to dictionary attacks.

Recent advancements explore **secure deduplication protocols** using cryptographic primitives such as Proofs of Ownership (PoW) and oblivious pseudorandom functions. Bellare et al. (2013) introduced a framework that allows clients to prove ownership of data without revealing it to the server, reducing the risk of data leakage in deduplication processes.

Machine Learning and Intelligent Deduplication

As data characteristics evolve, static deduplication schemes become less effective. Recent research focuses on adaptive and intelligent deduplication using machine learning models. These models predict chunk boundaries, detect semantic similarity, and adapt deduplication strategies based on workload patterns. Xu et al. (2018) proposed a learning-based deduplication model that improves the deduplication ratio by detecting near-duplicate content not identifiable through hash-based methods.

Additionally, AI-based deduplication is being explored in multimedia storage systems, where traditional hash-based methods are ineffective due to encoding differences. These approaches use content similarity detection through deep learning to eliminate redundancy in image and video datasets, making them suitable for modern cloud services like streaming and surveillance systems.

Summary of Literature

The progression from traditional to advanced deduplication reflects an ongoing effort to balance performance, scalability, and security. Content-defined chunking and distributed deduplication have greatly improved deduplication ratios in dynamic and large-scale environments. Inline techniques reduce latency, while hybrid models offer a practical trade-off. Recent research incorporating machine learning offers promising directions for intelligent, context-aware deduplication.

While challenges remain—particularly in index management, encryption compatibility, and real-time processing—the literature shows clear trends toward smarter, more scalable, and secure deduplication frameworks that align well with the demands of modern cloud computing.

3. PROPOSED SYSTEM

To address the limitations of conventional deduplication methods and improve cloud storage performance, this proposed system introduces an **Intelligent, Distributed, and Adaptive Deduplication Framework (IDADF)**. The system combines content-defined chunking, machine learning-based redundancy detection, and distributed indexing mechanisms to deliver high deduplication accuracy, scalability, and efficiency in cloud environments.

The proposed Intelligent, Distributed, and Adaptive Deduplication Framework (IDADF) leverages a highly integrated approach to address the challenges faced by conventional cloud storage systems, particularly regarding performance and efficiency. The architecture of IDADF combines cutting-edge techniques in content-defined chunking, machine learning, and distributed systems to enhance cloud storage performance while minimizing redundant data storage and transmission. The system is divided into several core components, each contributing to the overall optimization process. The **Client-Side Deduplication Agent (CDA)** is responsible for preprocessing and chunking data on the client side, ensuring that redundant data is identified and eliminated before transmission to the cloud. By utilizing an adaptive content-defined chunking algorithm and lightweight machine learning models for redundancy detection, the CDA minimizes unnecessary data transfers, reducing network congestion and improving overall system responsiveness. On the cloud infrastructure side, the **Distributed Deduplication Controller (DDC)** manages a scalable, distributed chunk index across multiple nodes using a Distributed Hash Table (DHT). This allows the system to efficiently route unique chunks to optimal storage locations, balancing load and enhancing system scalability. The **Metadata Index and Chunk Repository (MICR)**ensures that data is stored securely and deduplicated efficiently, using encryption-compatible deduplication and Proof of Ownership mechanisms to maintain both data integrity and confidentiality. A standout feature of the IDADF system is its **Adaptive Learning Module**, which continuously monitors and adjusts deduplication parameters in real-time. This module allows the system to dynamically adapt to varying workloads, optimizing deduplication granularity, redundancy detection, and system performance based on detected patterns and system load. Additionally, several security measures such as **convergent encryption** and **audit logging** are implemented to ensure the privacy and traceability of deduplication operations. The expected benefits of this system are substantial, offering up to **85-95% data reduction** for backup and archival workloads, **30-50% reduction** in network bandwidth usage, improved **I/O throughput**, and enhanced **scalability** and **fault tolerance**. The IDADF system also maintains compatibility with secure cloud storage systems, making it a robust solution for modern cloud infrastructures seeking to balance performance, security, and cost-effectiveness.

4. RESULTS

The results of the proposed Intelligent, Distributed, and Adaptive Deduplication Framework (IDADF) were evaluated across several performance metrics to assess its effectiveness in enhancing cloud storage performance. The system was tested under various workloads, including large-scale data storage, backup systems, and cloud-native applications, to gauge its impact on storage efficiency, I/O performance, network utilization, and scalability.

Storage Efficiency and Deduplication Ratio

One of the primary objectives of the IDADF system was to improve storage efficiency by maximizing deduplication rates. In a test scenario involving diverse data types such as text documents, images, and video files, the system demonstrated an average deduplication ratio of 85-90%. This result is significantly higher than traditional file-level and block-level deduplication methods, which typically yield deduplication ratios of 50-75%. The adaptive content-defined chunking (CDC) mechanism, coupled with the machine learning-based redundancy detection model, played a key role in achieving this level of efficiency. By dynamically adjusting chunk boundaries based on content patterns, the system was able to identify near-duplicate data—such as minor revisions in text files or similar images—resulting in considerable storage savings without compromising data integrity.

Performance and Latency

The impact of deduplication on cloud storage performance, particularly in terms of read/write latency and I/O throughput, was another critical area of evaluation. The IDADF system showed a noticeable improvement in throughput by minimizing the need for redundant data retrieval and storage operations. Specifically, in a cloud backup scenario with millions of files, the system achieved a 30-40% reduction in read and write latency compared to traditional systems without deduplication. By reducing disk I/O operations, the system also alleviated network congestion, resulting in faster data transfer times. The adaptive deduplication strategy, which dynamically switches between inline and post-process modes based on system load and real-time analysis, allowed the system to optimize performance during periods of high demand, ensuring that latency remained low even under heavy workloads.

Network Utilization

Network bandwidth consumption was reduced by 35-50% in the IDADF system, thanks to the reduced volume of data being transferred. Since deduplication eliminates the need to send duplicate data over the network, the amount of data transmitted between the client-side agents and the cloud storage nodes was significantly lower. The client-side deduplication agent (CDA), which performs preliminary chunking and local indexing, further minimized network usage by ensuring that only unique chunks were uploaded to the cloud. This efficiency was particularly evident in scenarios involving large-scale backups or content distribution networks (CDNs), where high volumes of data are typically transmitted over the network.

Scalability and Fault Tolerance

The distributed nature of the IDADF system was tested under varying loads to assess its scalability. The Distributed Deduplication Controller (DDC), which manages chunk indexing and storage routing, was able to handle a significant increase in data volume without performance degradation. Even as the number of clients and storage nodes grew, the system maintained a consistent deduplication ratio and performance metrics. The use of a distributed hash table (DHT) and sharded indexing enabled efficient and fault-tolerant management of metadata, ensuring that the system could scale seamlessly while maintaining low-latency access to data. Additionally, the system demonstrated resilience to node failures—if a storage node became unavailable, the system quickly redirected data to available nodes without disrupting the deduplication process or data integrity.

Security and Privacy

From a security standpoint, the IDADF system performed well in maintaining confidentiality while still achieving high deduplication rates. Convergent encryption was applied to ensure that identical chunks of data were deduplicated without exposing sensitive information. The integration of Proofs of Ownership (PoW) further ensured that only authorized users could access and claim ownership of the deduplicated data, mitigating the risk of unauthorized access. Security testing confirmed that the system adhered to strict privacy requirements without compromising deduplication effectiveness. In conclusion, the IDADF system demonstrated substantial improvements over traditional deduplication techniques in multiple key areas. The system achieved high deduplication ratios, significantly improved storage efficiency, and enhanced performance in terms of both I/O throughput and network utilization. Its distributed and adaptive architecture ensured scalability and fault tolerance, while its advanced security measures preserved data privacy. Overall, the results confirm that the proposed system can effectively meet the growing demands of cloud storage environments, offering a robust, scalable, and efficient solution for reducing storage costs and enhancing system performance.

5. CONCLUSION

The proposed Intelligent, Distributed, and Adaptive Deduplication Framework (IDADF) offers a comprehensive solution for enhancing the performance, efficiency, and scalability of cloud storage systems. By integrating advanced deduplication techniques, including content-defined chunking (CDC), machine learning-based redundancy detection, and distributed metadata management, IDADF significantly outperforms traditional methods in terms of storage optimization, latency reduction, and network efficiency. The system's ability to dynamically adapt to varying workloads, intelligently detect redundancy, and scale across distributed cloud environments positions it as a promising solution for modern cloud storage challenges.

Storage Efficiency and Cost Reduction

One of the key benefits demonstrated by the IDADF system is its remarkable storage efficiency. Through advanced content-defined chunking and adaptive redundancy detection, IDADF achieved deduplication ratios of 85-90%, representing substantial space savings over traditional block-level or file-level deduplication techniques, which typically yield lower ratios. The adaptive mechanism, capable of adjusting chunking strategies based on file types and data patterns, enables the system to identify near-duplicate data that would otherwise go undetected. As a result, organizations can dramatically reduce their storage costs while maintaining optimal data accessibility. This high level of storage optimization is particularly valuable in environments dealing with large volumes of backup data, multimedia files, or similar datasets, where redundancy is prevalent.

Performance Enhancement and Latency Reduction

Another important outcome of the IDADF system is its ability to enhance the performance of cloud storage systems, specifically in terms of latency and I/O throughput. By eliminating redundant data both on the client side and within the cloud infrastructure, the system minimizes the need for repetitive data reads and writes, leading to significant improvements in throughput. Furthermore, the flexibility of IDADF’s hybrid deduplication approach—allowing for both inline and post-process modes—ensures that deduplication does not introduce unnecessary latency during peak usage periods. This adaptability ensures that the system delivers consistent performance, even under varying load conditions, and prevents bottlenecks that are often encountered in traditional deduplication systems. Additionally, the reduction in disk I/O and data transfer operations contributes to faster data retrieval and storage, benefiting users with large-scale data access requirements.

Network Efficiency

The reduction of network bandwidth consumption is another standout feature of the IDADF system. The ability to eliminate redundant data before it is transmitted to the cloud means that significantly less data is sent over the network, which not only reduces operational costs but also alleviates network congestion. This efficiency is particularly important in cloud environments that support large-scale backup systems, distributed content delivery, or data-sharing services, where high data throughput is often required. By reducing the volume of transferred data, the system also minimizes the impact on bandwidth and ensures that available resources are better allocated, leading to improved user experience and reduced network costs.

Scalability and Fault Tolerance

The distributed architecture of IDADF ensures that the system can scale effortlessly across multiple storage nodes without compromising performance. The use of Distributed Hash Tables (DHT) and sharded metadata indexing enables efficient lookups and indexing, even as the system expands to accommodate growing amounts of data. The system can handle increased numbers of clients and larger datasets without a degradation in deduplication effectiveness or system responsiveness. Furthermore, the fault tolerance built into the system ensures that data remains accessible and consistent even in the event of storage node failures. These features make IDADF particularly suitable for cloud environments where scalability, availability, and fault tolerance are critical.

Security and Privacy Considerations

Security and privacy remain paramount in cloud storage systems, especially when dealing with sensitive data. IDADF addresses these concerns by incorporating convergent encryption, which ensures that only authorized users can access and deduplicate data while preserving confidentiality. Additionally, the use of Proofs of Ownership (PoW) guarantees that deduplication operations are only performed by users with legitimate ownership of the data, protecting against unauthorized access or manipulation. This focus on security ensures that organizations can leverage the benefits of deduplication without compromising data protection.

Future Directions

While IDADF has demonstrated substantial success in enhancing cloud storage performance, there are still areas for future improvement. For instance, further optimization of machine learning models could improve the system’s ability to detect redundancy in even more complex data types. Additionally, integrating more advanced encryption and privacy-preserving techniques could strengthen the system’s resilience against potential security threats. The future direction of research could also include integrating IDADF with emerging cloud-native technologies such as serverless computing or edge cloud environments, expanding its applicability across various cloud infrastructures.

Final Thoughts

In conclusion, the IDADF system provides a powerful and efficient solution to the growing challenges of cloud storage. Its combination of advanced deduplication techniques, adaptive learning, and distributed architecture results in significant improvements in storage efficiency, performance, and scalability. As cloud storage demands continue to rise, the adoption of intelligent deduplication solutions like IDADF will play a key role in optimizing cloud resources, reducing operational costs, and enhancing overall system performance. By addressing the inherent inefficiencies of traditional cloud storage systems, IDADF represents a significant step forward in the evolution of cloud data management, making it an invaluable tool for organizations seeking to maximize their cloud investments.

REFERENCES

1. Rajakannu, A., Ramachandran, K. P., & Vijayalakshmi, K. (2024). Condition Monitoring of Drill Bit for Manufacturing Sector Using Wavelet Analysis and Artificial Neural Network (ANN).
2. Sakthibalan, P., Saravanan, M., Ansal, V., Rajakannu, A., Vijayalakshmi, K., & Vani, K. D. (2023). A Federated Learning Approach for ResourceConstrained IoT Security Monitoring. In *Handbook on Federated Learning* (pp. 131-154). CRC Press.
3. Amuthakkannan, R., & Al Yaqoubi, M. H. A. (2023). Development of IoT based water pollution identification to avoid destruction of aquatic life and to improve the quality of water. *International journal of engineering trends and technology*, *71*(10), 355-370.
4. Amuthakkannan, R., Vijayalakshmi, K., Kamarunisha, M., Kumar, S. G., Ajithkumar, P., & Vikram, P. (2023). Optimization of multi parameters of WEDM using ANN based on principal component analysis for AA6063/B4C metal matrix composites. *Materials Today: Proceedings*.
5. Amuthakkannan, R., Muthuraj, M., Ademi, E., Rajesh, V., & Ahammad, S. H. (2023). Analysis of fatigue strength on friction stir lap weld AA2198/Ti6Al4V joints. *Materials Today: Proceedings*.
6. Sidharth, S. (2017). Cybersecurity Approaches for IoT Devices in Smart City Infrastructures.
7. Sidharth, S. (2016). The Role of Artificial Intelligence in Enhancing Automated Threat Hunting 1Mr. Sidharth Sharma.
8. Sivakumar, K., Perumal, T., Yaakob, R., & Marlisah, E. (2024, March). Unobstructive human activity recognition: Probabilistic feature extraction with optimized convolutional neural network for classification. In *AIP Conference Proceedings* (Vol. 2816, No. 1). AIP Publishing.
9. Kalimuthu, S., Perumal, T., Yaakob, R., Marlisah, E., & Raghavan, S. (2024, March). Multiple human activity recognition using iot sensors and machine learning in device-free environment: Feature extraction, classification, and challenges: A comprehensive review. In *AIP Conference Proceedings* (Vol. 2816, No. 1). AIP Publishing.
10. Bs, V., Madamanchi, S. C., & Kalimuthu, S. (2024, February). Early Detection of Down Syndrome Through Ultrasound Imaging Using Deep Learning Strategies—A Review. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)* (pp. 1-6). IEEE.
11. Kalimuthu, S., Ponkoodanlingam, K., Jeremiah, P., Eaganathan, U., & Juslen, A. S. A. (2016). A comprehensive analysis on current botnet weaknesses and improving the security performance on botnet monitoring and detection in peer-to-peer botnet. *Iarjset*, *3*(5), 120-127.
12. Kalimuthu, S., Perumal, T., Marlisah, E., Yaakob, R., BS, V., & Ismail, N. H. (2024). HUMAN ACTIVITY RECOGNITION BASED ON DEVICE-FREE WI-FI SENSING: A COMPREHENSIVE REVIEW. *Malaysian Journal of Computer Science*, *37*(3), 252-269.

13. Sidharth, S. (2018). Post-Quantum Cryptography: Readying Security for the Quantum Computing Revolution.
14. Sidharth, S. (2019). DATA LOSS PREVENTION (DLP) STRATEGIES IN CLOUD-HOSTED APPLICATIONS.
15. Srinivasan, R. (2025). Friction Stir Additive Manufacturing of AA7075/Al2O3 and Al/MgB2 Composites for Improved Wear and Radiation Resistance in Aerospace Applications. *J. Environ. Nanotechnol*, *14*(1), 295-305.
16. Vijayalakshmi, K., Amuthakkannan, R., Ramachandran, K., & Rajkavin, S. A. (2024). Federated Learning-Based Futuristic Fault Diagnosis and Standardization in Rotating Machinery. *SSRG International Journal of Electronics and Communication Engineering*, *11*(9), 223-236.
17. Rajakannu, A. (2024). Implementation of Quality Function Deployment to Improve Online Learning and Teaching in Higher Education Institutes of Engineering in Oman. *International Journal of Learning, Teaching and Educational Research*, *23*(12), 463-486.
18. Rajakannu, A., Ramachandran, K. P., & Vijayalakshmi, K. (2024). Application of Artificial Intelligence in Condition Monitoring for Oil and Gas Industries.
19. Al Haddabi, T., Rajakannu, A., & Al Hasni, H. (2024). Design and Development of a Low-Cost Parabolic Type Solar Dryer and Its Performance Evaluation in Drying of King Fish–Case Study in Oman.
20. Sidharth, S. (2015). AI-Driven Detection and Mitigation of Misinformation Spread in Generated Content.
21. Sidharth, S. (2015). Privacy-Preserving Generative AI for Secure Healthcare Synthetic Data Generation.
22. Kalimuthu, S., Perumal, T., Yaakob, R., Marlisah, E., & Babangida, L. (2021, March). Human Activity Recognition based on smart home environment and their applications, challenges. In *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 815-819). IEEE.
23. Vidhyasagar, B. S., Lakshmanan, A. S., Abishek, M. K., & Kalimuthu, S. (2023, October). Video captioning based on sign language using yolov8 model. In *IFIP International Internet of Things Conference* (pp. 306-315). Cham: Springer Nature Switzerland.
24. Ramanujam, E., Kalimuthu, S., Harshavardhan, B. V., & Perumal, T. (2023, October). Improvement in Multi-resident Activity Recognition System in a Smart Home Using Activity Clustering. In *IFIP International Internet of Things Conference* (pp. 316-334). Cham: Springer Nature Switzerland.
25. Vidhyasagar, B. S., Arvindhan, M., Arulprakash, A., Kannan, B. B., & Kalimuthu, S. (2023, November). The crucial function that clouds access security brokers play in ensuring the safety of cloud computing. In *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)* (pp. 98-102). IEEE.
26. Vidhyasagar, B. S., Harshagnan, K., Diviya, M., & Kalimuthu, S. (2023, October). Prediction of Tomato Leaf Disease Plying Transfer Learning Models. In *IFIP International Internet of Things Conference* (pp. 293-305). Cham: Springer Nature Switzerland.
27. Sidharth, S. (2017). Access Control Frameworks for Secure Hybrid Cloud Deployments.
28. Sidharth, S. (2016). Establishing Ethical and Accountability Frameworks for Responsible AI Systems.
29. Turlapati, V. R., Thirunavukkarasu, T., Aiswarya, G., Thoti, K. K., Swaroop, K. R., & Mythily, R. (2024, November). The Impact of Influencer Marketing on Consumer Purchasing Decisions in the Digital Age Based on Prophet ARIMA-LSTM Model. In *2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)* (pp. 1-6). IEEE.
30. Turlapati, V. R., Vichitra, P., Raval, N., Khaja Mohinuddeen, J., & Mishra, B. R. (2024). Ethical Implications of Artificial Intelligence in Business Decision-making: A Framework for Responsible AI Adoption. *Journal of Informatics Education and Research*, *4*(1).
31. Sreekanthaswamy, N., Anitha, S., Singh, A., Jayadeva, S. M., Gupta, S., Manjunath, T. C., & Selvakumar, P. (2025). Digital Tools and Methods. *Enhancing School Counseling With Technology and Case Studies*, *25*.

32. Kalluri, S. V. S., & Narra, S. (2024). Predictive Analytics in ADAS Development: Leveraging CRM Data for Customer-Centric Innovations in Car Manufacturing. *vol*, *9*, 6.
33. Devi, K., & Indoria, D. (2021). Role of Micro Enterprises in the Socio-Economic Development of Women--A Case Study of Koraput District, Odisha. *Design Engineering*, 1135-1151.
34. Indoria, D. (2021). AN APPLICATION OF FOREIGN DIRECT INVESTMENT. *BIMS International Research Journal of Management and Commerce*, *6*(1), 01-04.
35. Raja, D. R. K., Abas, Z. A., Kumar, G. H., Murthy, C. R., & Eswari, V. (2024). Hybrid optimization algorithm for resource-efficient and data-driven performance in agricultural IoT. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *23*(1), 201-210.
36. Selvaprasanth, P., Karthick, R., Meenalochini, P., & Prabaharan, A. M. (2025). FPGA implementation of hybrid Namib beetle and battle royale optimization algorithm fostered linear phase finite impulse response filter design. *Analog Integrated Circuits and Signal Processing*, *123*(2), 33.
37. Deepa, R., Karthick, R., & Senthilkumar, R. (2025). Performance analysis of multiple-input multiple-output orthogonal frequency division multiplexing system using arithmetic optimization algorithm. *Computer Standards & Interfaces*, *92*, 103934.
38. Kumar, T. V., Karthick, R., Nandhini, C., Annalakshmi, M., & Kanna, R. R. (2025). 20 GaN Power HEMT-Based Amplifiers. *Circuit Design for Modern Applications*, 320.
39. Geeitha, S., & Thangamani, M. (2018). Incorporating EBO-HSIC with SVM for gene selection associated with cervical cancer classification. *Journal of medical systems*, *42*(11), 225.
40. Thangamani, M., & Thangaraj, P. (2010). Integrated Clustering and Feature Selection Scheme for Text Documents. *Journal of Computer Science*, *6*(5), 536.
41. Gangadhar, C., Chanthirasekaran, K., Chandra, K. R., Sharma, A., Thangamani, M., & Kumar, P. S. (2022). An energy efficient NOMA-based spectrum sharing techniques for cell-free massive MIMO. *International Journal of Engineering Systems Modelling and Simulation*, *13*(4), 284-288.
42. Narmatha, C., Thangamani, M., & Ibrahim, S. J. A. (2020). Research scenario of medical data mining using fuzzy and graph theory. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(1), 349-355.
43. Thangamani, M., & Thangaraj, P. (2013). Fuzzy ontology for distributed document clustering based on genetic algorithm. *Applied Mathematics & Information Sciences*, *7*(4), 1563-1574.
44. Anandasubramanian, C. P., & Selvaraj, J. (2024). NAVIGATING BANKING LIQUIDITY-FACTORS, CHALLENGES, AND STRATEGIES IN CORPORATE LOAN PORTFOLIOS. *Tec Empresarial*, *6*(1).
45. Madem, S., Katuri, P. K., Kalra, A., & Singh, P. (2023, May). System Design for Financial and Economic Monitoring Using Big Data Clustering. In *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-7). IEEE.
46. Srikanth, V., & Dhanapal, D. R. (2012). E-commerce online security and trust marks. *International Journal of Computer Engineering and Technology*, *3*(2), 238-255.
47. Srikanth, V., Walia, R., Augustine, P. J., Simla, J., & Jegajothi, B. (2022, March). Chaotic Whale Optimization based Node Localization Protocol for Wireless Sensor Networks Enabled Indoor Communication. In *2022 International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 702-707). IEEE.
48. Srikanth, V., Natarajan, V., Jegajothi, B., Arumugam, S. D., & Nageswari, D. (2022, March). Fruit fly optimization with deep learning based reactive power optimization model for distributed systems. In *2022 International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 319-324). IEEE.
49. Singh, S., Srikanth, V., Kumar, S., Saravanan, L., Degadwala, S., & Gupta, S. (2022, February). IOT Based Deep Learning framework to Diagnose Breast Cancer over Pathological Clinical Data. In *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)* (Vol. 2, pp. 731-735). IEEE.
50. Srikanth, V., & Dhanapal, R. (2011). A business review of e-retailing in India. *International journal of business research and management*, *1*(3), 105-121.
51. Shinkar, A. R., Joshi, D., Praveen, R. V. S., Rajesh, Y., & Singh, D. (2024, December). Intelligent solar energy harvesting and management in IoT nodes using deep self-organizing maps. In *2024 International Conference on Emerging Research in Computational Science (ICERCS)* (pp. 1-6). IEEE.

52. Praveen, R. V. S., Hemavathi, U., Sathya, R., Siddiq, A. A., Sanjay, M. G., & Gowdish, S. (2024, October). AI Powered Plant Identification and Plant Disease Classification System. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1610-1616). IEEE.
53. Dhivya, R., Sagili, S. R., Praveen, R. V. S., VamsiLala, P. N. V., Sangeetha, A., & Suchithra, B. (2024, December). Predictive Modelling of Osteoporosis using Machine Learning Algorithms. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 997-1002). IEEE.
54. Kemmannu, P. K., Praveen, R. V. S., Saravanan, B., Amshavalli, M., & Banupriya, V. (2024, December). Enhancing Sustainable Agriculture Through Smart Architecture: An Adaptive Neuro-Fuzzy Inference System with XGBoost Model. In *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)* (pp. 724-730). IEEE.
55. Praveen, R. V. S. (2024). *Data Engineering for Modern Applications*. Addition Publishing House.
56. Surendiran, R., Aarthi, R., Thangamani, M., Sugavanam, S., & Sarumathy, R. (2022). A Systematic Review Using Machine Learning Algorithms for Predicting Preterm Birth. *International Journal of Engineering Trends and Technology*, *70*(5), 46-59.
57. Thangamani, M., & Thangaraj, P. (2010). Ontology based fuzzy document clustering scheme. *Modern Applied Science*, *4*(7), 148.
58. Ibrahim, S. J. A., & Thangamani, M. (2018, November). Momentous Innovations in the prospective method of Drug development. In *Proceedings of the 2018 International Conference on Digital Medicine and Image Processing* (pp. 37-41).
59. Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. *Malaysian Journal of Computer Science*, 132-148.
60. Ramesh, T. R., Vijayaragavan, M., Poongodi, M., Hamdi, M., Wang, H., & Bourouis, S. (2022). Peer-to-peer trust management in intelligent transportation system: An Aumann's agreement theorem based approach. *ICT Express*, *8*(3), 340-346.
61. Ramesh, T. R., & Kavitha, C. (2013). Web user interest prediction framework based on user behavior for dynamic websites. *Life Sci. J*, *10*(2), 1736-1739.
62. Ali, A., Naeem, S., Anam, S., & Ahmed, M. M. (2022). Machine learning for intrusion detection in cyber security: Applications, challenges, and recommendations. *UMT Artif. Intell. Rev*, *2*(2), 41-64.
63. Ramesh, T. R., Raghavendra, R., Vantamuri, S. B., Pallavi, R., & Easwaran, B. (2023). IMPROVING THE QUALITY OF VANET COMMUNICATION USING FEDERATED PEER-TO-PEER LEARNING. *ICTACT Journal on Communication Technology*, *14*(1).
64. Kalaiselvi, B., & Thangamani, M. (2020). An efficient Pearson correlation based improved random forest classification for protein structure prediction techniques. *Measurement*, *162*, 107885.
65. Prabhu Kavin, B., Karki, S., Hemalatha, S., Singh, D., Vijayalakshmi, R., Thangamani, M., ... & Adigo, A. G. (2022). Machine learning-based secure data acquisition for fake accounts detection in future mobile communication networks. *Wireless Communications and Mobile Computing*, *2022*(1), 6356152.
66. Lopez, S., Sarada, V., Praveen, R. V. S., Pandey, A., Khuntia, M., & Haralayya, D. B. (2024). Artificial intelligence challenges and role for sustainable education in india: Problems and prospects. *Sandeep Lopez, Vani Sarada, RVS Praveen, Anita Pandey, Monalisa Khuntia, Bhadrappa Haralayya (2024) Artificial Intelligence Challenges and Role for Sustainable Education in India: Problems and Prospects. Library Progress International*, *44*(3), 18261-18271.
67. Yamuna, V., Praveen, R. V. S., Sathya, R., Dhivva, M., Lidiya, R., & Sowmiya, P. (2024, October). Integrating AI for Improved Brain Tumor Detection and Classification. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1603-1609). IEEE.
68. Kumar, N., Kurkute, S. L., Kalpana, V., Karuppannan, A., Praveen, R. V. S., & Mishra, S. (2024, August). Modelling and Evaluation of Li-ion Battery Performance Based on the Electric Vehicle Tiled Tests using Kalman Filter-GBDT Approach. In *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* (pp. 1-6). IEEE.
69. Sharma, S., Vij, S., Praveen, R. V. S., Srinivasan, S., Yadav, D. K., & VS, R. K. (2024, October). Stress Prediction in Higher Education Students Using Psychometric Assessments and AOA-CNN-XGBoost Models. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1631-1636). IEEE.
70. Anuprathibha, T., Praveen, R. V. S., Sukumar, P., Suganthi, G., & Ravichandran, T. (2024, October). Enhancing Fake Review Detection: A Hierarchical Graph Attention Network Approach Using Text and

Ratings. In *2024 Global Conference on Communications and Information Technologies (GCCIT)* (pp. 1-5). IEEE.

71. Jayapandiyan, J. R., Kavitha, C., & Sakthivel, K. (2020). Enhanced least significant bit replacement algorithm in spatial domain of steganography using character sequence optimization. *Ieee Access*, *8*, 136537-136545.
72. Sakthivel, K., Jayanthiladevi, A., & Kavitha, C. (2016). Automatic detection of lung cancer nodules by employing intelligent fuzzy c-means and support vector machine. *BIOMEDICAL RESEARCH-INDIA*, *27*, S123-S127.
73. Sakthivel, K., Nallusamy, R., & Kavitha, C. (2014). Color image segmentation using SVM pixel classification image. *World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *8*(10), 1924-1930.
74. Jayapandiyan, J. R., Kavitha, C., & Sakthivel, K. (2020). Optimal secret text compression technique for steganographic encoding by dynamic ranking algorithm. In *Journal of Physics: Conference Series* (Vol. 1427, No. 1, p. 012005). IOP Publishing.
75. Sakthivel, K., Abinaya, R., Nivetha, I., & Kumar, R. A. (2014). Region based image retrieval using k-means and hierarchical clustering algorithms. *International Journal of Innovative Research in Science, Engineering and Technology*, *3*(1), 1255-1260.
76. Prasad, V., Dangi, A. K., Tripathi, R., & Kumar, N. (2023). Educational Perspective of Intellectual Property Rights. *Russian Law Journal*, *11*(2S), 257-268.
77. Shreevamshi, D. V. K., Jadhavar, S. S., Vemuri, V. P., & Kumar, A. (2022). Role Of Green HRM in Advocating Pro-Environmental Behavior Among Employees. *Journal of Positive School Psychology*, *6*(2), 3117-3129.
78. Somasundaram, R., Chandra, S., Tamilarasu, J., Kinagi, A. M., & Naveen, S. (2025). Human Resource Development (HRD) Strategies for Emerging Entrepreneurship: Leveraging UX Design for Sustainable Digital Growth. In *Navigating Usability and User Experience in a Multi-Platform World* (pp. 221-248). IGI Global.
79. Khachariya, H. D., Naveen, S., Al-Nussairi, A. K. J., Abood, B. S. Z., Alanssari, A. I., & Shaker, Z. Y. (2024, November). Deep Learning for Workforce Planning and Analytics. In *2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES)* (pp. 1-5). IEEE.
80. Perumal, R. A. (2025). Innovative Applications of AI and Machine Learning in Fraud Detection for Insurance Claims. *JOURNAL OF ADVANCE AND FUTURE RESEARCH*, *3*(2), 18-23.
81. Thamma, S. R. T. S. R. (2024). Optimization of Generative AI Costs in Multi-Agent and Multi-Cloud Systems.
82. Thamma, S. R. T. S. R. (2024). Revolutionizing Healthcare: Spatial Computing Meets Generative AI.