ISSN : 2455-9679
SJIF Impact Factor : 6.190

IJRTSM
INTERNATIONAL JOURNAL OF RECENT TECHNOLOGY SCIENCE & MANAGEMENT

"OPTIMAL CLASSIFICATION AND DEDUPLICATION OF ENCRYPTED BIG DATA FOR SECURITY ENHANCEMENT"

Anita Sahu [1], Pratima Gautam [2]

[1] *Research Scholar, Department of Computer Science and Application, Rabindranath Tagore University, Raisen, Madhya Pradesh, India*

[2] *Assistant Professor, Department of Computer Science and Application, Rabindranath Tagore University, Raisen, Madhya Pradesh, India*

ani.sahu2@gmail.com

pratima.gautam@aisectuniversity.ac.in

Corresponding Author: ani.sahu2@gmail.com

ABSTRACT

The quick growth of big data, attached with increasing concerns over information security, has necessitated the progress of robust methods for management and protecting sensitive information. This paper discovers the optimal organization and deduplication of encrypted big data to improve security procedures in data storage and programme. This paper proposes a comprehensive review that influences advanced machine learning algorithms for effective data organisation while preserving data confidentiality through encryption methods. Additionally, the deduplication development minimizes storage requirements, safeguarding efficient use of resources. This procedure is designed to balance presentation and security, providing a methodical approach to handling encoded data in various applications, as well as cloud computing and IoT atmospheres. Through experimental validation, this work determines the effectiveness of different frameworks in accomplishing high classification correctness, reducing redundancy, and maintaining information integrity, thereby contributing to enhanced sanctuary in the era of big data.

Key Words: *Machine Learning, Big Data, Information Security, Deduplication, Classification.*

I. INTRODUCTION

In an age where information is created at an unprecedented percentage, the challenge of managing big data successfully and securely has developed increasingly critical. Organizations across areas are faced with the twin responsibility of leveraging vast amounts of information for strategic visions while ensuring the confidentiality and honesty of sensitive information. With the growth of cybersecurity threats and stringent documents protection guidelines, the need for secure records management solutions is more pressing than forever [5, 8].

Single of the fundamental challenges in big data management is the arrangement and deduplication of information. Classification enables organizations to categorize information based on its consideration and relevance, facilitating effective access and compliance with controlling requirements. Simultaneously, deduplication eliminates terminated copies of data, optimizing storage prices and improving data recovery performance. However, these processes developed more complex when dealing with encrypted information, as traditional organization methods may not be straight applicable [3, 7].

ISSN : 2455-9679
SJIF Impact Factor : 6.190

This paper suggests several solutions aimed at the optimal classification and deduplication of scrambled big data, pointing to enhance security while preserving efficiency. By employing progressive machine learning techniques, the schemes goal to accurately classify data deprived of compromising its encrypted state, safeguarding that sensitive information remains protected throughout the data lifecycle. Furthermore, these methods to deduplication not only diminish storage usage then also aligns with best practices for records security [11].

In the subsequent sections, this work resolves review the current outstanding of research in big data management, encryption methods, and the tasks associated with classifying and deduplicating encrypted information. Ultimately, this research targets to provide a wide-ranging solution that addresses the increasing need for secure and resourceful big data management strategies.

II. Literature Review

Here's a comprehensive summary of several research papers are described in detail.

J. Thomason (2021) [1] discusses the connection of big tech, big data, and digital fitness, highlighting the transformative potential of large-scale information analytics in healthcare. The paper highpoints how major knowledge firms are leveraging vast amounts of health information to improve innovative solutions that improve persistent outcomes and streamline healthcare service area. Thomason also raises concerns nearby privacy, data ownership, and the ethical suggestions of using personal health records for profit. The author supporters for regulatory frameworks that safeguard data security and persistent privacy while promoting technical advancements in health.

K. Ahmad et al. (2022) [2] judgmentally analyses the security, information management, and ethical tasks associated with developing human-cantered smart cities. The authors discover the character of big data in town planning and service delivery, highlighting the need for robust records governance frameworks to guard citizen privacy. They identify sanctuary vulnerabilities in smart city organization and highlight the ethical insinuations of data collection and custom. The paper calls for a cooperative approach involving stakeholders, officials, and technologists to address these tasks and create sustainable, protected smart cities.

W. Xing and Y. Bei (2019) [3] current a learning on classifying medical health big data by means of the K-nearest neighbours (KNN) procedure. The authors detail their organization, including data preprocessing and feature collection, to enhance classification correctness. The results demonstrate that KNN effectually categorizes health data into pertinent classes, providing insights for healthcare specialists to make knowledgeable decisions. The paper concludes by means of discussing the algorithm's potential solicitations in healthcare analytics, emphasizing the standing of accurate data classification in enlightening patient care.

Y. Luo et al. (2023) [4] presents a robust key update machinery utilizing Support Vector Machine (SVM) for secure Orthogonal Frequency Division Multiple Access Passive Optical Networks (OFDMA-PON). The authors suggest a method that balances safety and performance by permitting controlled accuracy in key updates. They appraise their approach complete simulations, demonstrating its effectiveness in cultivating security without compromising information transmission efficiency. The paper contributes near the ongoing improvement of secure communication protocols in visual networks.

I. V. Pustokhina et al. (2023) [5] offerings a novel method for intrusion detection in programme big data using a hybrid model merging Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM). The author's attention on optimizing hyperparameters to improve model performance. Investigates show that the future model effectively identifies intrusions in multimedia information, demonstrating high correctness and robustness. The study highpoints the importance of deep learning procedures in developing effective cybersecurity measures for multimedia atmospheres.

S. Srivastava et al. (2022) [6] employ an ensemble learning method to classify chronic kidney disease (CKD) from medicinal datasets. They compare numerous ensemble techniques, such as trapping and boosting, to determine their

ISSN : 2455-9679
SJIF Impact Factor : 6.190

efficiency in improving classification correctness. The results indicate that the collective methods significantly outperform single classifiers, as long as a reliable tool aimed at early diagnosis of CKD. The authors highlight the clinical significance of their findings in cultivating patient outcomes through timely intervention.

O. J. Ayodele et al. (2020) [7] contributions an ensemble-based procedure for diagnosing chronic kidney infection. Ayodele and colleagues integrate several classifiers to enhance diagnostic correctness. The authors perform a proportional analysis of different models and appraise their performance using several metrics. The findings reveal that the collective approach significantly progresses the accuracy of CKD analysis, making it a valuable tool for healthcare specialists. The learning underscores the potential of ensemble methods in medical diagnostics.

A. Ahmad et al. (2020) [8] discover the presentation of classification and association rule removal techniques to predict chronic kidney infection. They analyse patient information to identify patterns and connections associated with CKD. The learning demonstrates that combining organisation with association rule mining can increase predictive capabilities, providing visions that can inform preventive approaches. The authors discuss the suggestions of their findings for scientific practice, emphasizing the importance of data-driven methods in healthcare.

S. Bhattacharyya et al. (2023) [9] propositions an intelligent and protected manufacturing model that integrates Internet of Things (IoT) knowledge with blockchain in a hybrid cloud atmosphere. Bhattacharyya and generations discuss how the model enhances information security, traceability, and operative efficiency in manufacturing processes. They highpoint the potential of blockchain to safe IoT data and simplify seamless communication across the developed supply chain. The author's campaigner for the adoption of such models to recover general manufacturing security and performance.

T. M. Selvi and V. Kavitha, (2022) [10] current a privacy-aware deep learning framework for emerging health recommendation systems created on big data analysis. The author's attention on ensuring persevering privacy while providing personalized health commendations. Their framework employs forward-thinking deep learning techniques to analyse health information while incorporating privacy-preserving mechanisms. The outcomes indicate that their method effectively balances privacy and recommendation correctness, highlighting the potential for protected health informatics applications.

T. Munirathinam et al. (2020) [11] suggests a privacy-preserved e-healthcare system that powers cloud computing and IoT machineries. Munirathinam and colleagues appliance a secured storage procedure to protect patient information while using deep learning aimed at health analytics. The system aims to deliver secure access to health data while maintaining information confidentiality. The authors validate their method through experimental results, demonstrating its efficiency in ensuring discretion and enhancing healthcare delivery.

A. Ponmalar and V. Dhanakoti, (2022) [12] present a hybrid architecture for imposition detection that combines Convolutional Neural Networks (CNN) with the Whale Optimization Algorithm and Tabu Search. The projected approach aims to improve the detection capabilities in big data atmospheres. The authors evaluate their model through wide-ranging experiments, showcasing this one superior performance in detecting several types of intrusions associated to traditional methods. This research contributes to educating cybersecurity events in big data applications.

K. Xiu et al. (2021) [13] suggest a novel imposition detection method for IoT networks that syndicates Particle Swarm Optimization (PSO) with Convolutional Neural Networks (CNN). The adaptive PSO is applied to improve the CNN parameters for improved detection accuracy. The authors determine through experimental results that their method effectively identifies impositions in IoT environments, addressing the unique tasks posed by IoT vulnerabilities. The learning emphasizes the consequence of integrating optimization methods in cybersecurity solutions.

A. Ponmalar and V. Dhanakoti (2022) [14] offerings an ensemble method for intrusion detection using Support Vector Machines (SVM) improved with a Chaos Game Optimization procedure. Ponmalar and Dhanakoti analyse several datasets to evaluate the presentation of their proposed method. The discoveries indicate that the collective SVM approach outperforms individual classifiers, attaining higher accuracy in noticing intrusions. The study underscores the efficiency of ensemble techniques in attractive cybersecurity in big data atmospheres.

ISSN : 2455-9679
SJIF Impact Factor : 6.190

D. Weiping et al. (2020) [15] discover the usage of fuzzy logic and chemical response optimization algorithms for imposition detection in industrial big data settings. They proposition a novel framework that participates these techniques to progress detection accuracy and decrease false positives. The results of their research validate the effectiveness of their method in recognising intrusions, contributing to the ongoing progress of secure industrial systems. The author's highpoint the potential for further exploration in applying progressive optimization methods to cybersecurity.

D. Viji and S. Revathy (2022) [16] presents a hash-indexing block-based deduplication procedure aimed at dropping storage requirements in cloud atmospheres. Viji and Revathy discuss the tasks of data redundancy and recommend a method that professionally identifies and eliminates duplicate information blocks. The authors evaluate the presentation of their algorithm concluded experiments, demonstrating its effectiveness in improving storage usage while maintaining information integrity. The study highpoints the importance of deduplication systems in enhancing cloud storage efficiency.

N. Mageshkumar et al. (2023) [17] suggest a hybrid cloud storage system that integrates a multilayer cryptosystem to enhance sanctuary during the deduplication development. The authors discuss the tasks of maintaining data discretion while optimizing storage efficiency. Their planned system is validated complete experiments, showing that it efficiently balances security and storage necessities. The study emphasizes the implication of secure deduplication methods in cloud computing atmospheres.

M. Akbar et al. (2023) [18] current a machine learning-based approach for attractive authentication during the deduplication of big data in cloud storage systems. They investigate the current challenges in safeguarding secure deduplication and suggest a framework that incorporates machine learning algorithms for enhanced authentication correctness. The authors validate their methodology through experiments, demonstrating its efficiency in securing cloud storage atmospheres. The study contributes to the ongoing improvement of secure information management practices in the cloud.

Here's a table 1 summarizing the proposed methodology, performance parameters, advantages, and limitations for the latest papers:

Table 1: Comprehensive Review

Paper	Proposed Methodology	Performance Parameters	Advantages	Limitations
J. Thomason (2021) [1]	Logical review of big tech's character in digital health	N/A	Highlights probable for improved patient products; emphasizes ethical concerns	Limited empirical information; more theoretical discussion
K. Ahmad et al. (2022) [2]	Critical examination of smart city frameworks	N/A	Provides a all-inclusive view of sanctuary, data management, and ethical challenges	Lack of precise case studies; overall recommendations
W. Xing et al. (2019) [3]	KNN organization algorithm for health data	Classification correctness, precision, recall	Simple and operative for classification assignments; well-suited for healthcare applications	Searching to feature scaling; may struggle with huge datasets

ISSN : 2455-9679
SJIF Impact Factor : 6.190

Paper	Proposed Methodology	Performance Parameters	Advantages	Limitations
Y. Luo et al. (2023) [4]	SVM for important updates in OFDMA-PON	Key update accurateness, security level	Balances sanctuary and performance; adaptable to network situations	May necessitate fine-tuning of parameters; performance in need of on data quality
I. V. Pustokhina et al. (2023) [5]	Hybrid CNN and Bi-LSTM for intrusion discovery	Correctness, false positive rate	High detection correctness; suitable for multimedia big data atmospheres	Difficulty of model; may need extensive computational possessions
S. Srivastava et al. (2022) [6]	Ensemble knowledge methods for CKD classification	Classification correctness, F1-score	Progresses diagnostic accuracy compared to particular classifiers	Entails diverse training data; model complexity
O. J. Ayodele et al. (2020) [7]	Ensemble procedures for CKD diagnosis	Correctness, sensitivity, specificity	Enhanced accurateness and reliability for CKD diagnosis	Difficulty of model ensemble; computationally concentrated
A. Ahmad et al. (2020) [8]	Combination of organisation and association law mining	Predictive correctness, rule relevance	Recognises patterns and correlations; supports preventive approaches in healthcare	Requires quality information; may produce irrelevant rules
S. Bhattacharyya et al. (2023) [9]	Blockchain incorporation with IoT for manufacturing	Sanctuary level, operational efficiency	Enhances information security and traceability; progresses operational efficiency	Implementation difficulty; requires cooperation across stakeholders
T. M. Selvi et al. (2022) [10]	Deep learning agenda for health recommendations	Recommendation correctness, privacy metrics	Balances personalization and secrecy; applicable to fitness informatics	May not oversimplify well across changed datasets
T. Munirathinam et al. (2020) [11]	Cloud and IoT mixing with secured storage systems	Data access speed, sanctuary level	Safeguards data confidentiality; enhances healthcare distribution	Probable latency in access; dependency scheduled cloud infrastructure
A. Ponmalar et al. (2022) [12]	Hybrid methodology combining CNN with Whale Optimization and Tabu Examination	Detection correctness, processing time	Superior discovery capabilities in big information; adaptable to various types of occurrences	Difficulty in implementation; requires parameter tuning
K. Xiu et al. (2021) [13]	Adaptive Element Swarm Optimization for CNN	Discovery rate, false positive rate	Improves correctness in identifying intrusions in	Complexity of the procedure; sensitivity to

ISSN : 2455-9679
SJIF Impact Factor : 6.190

Paper	Proposed Methodology	Performance Parameters	Advantages	Limitations
	factor optimization		IoT; dynamic version	initial situations
A. Ponmalar et al. (2022) [14]	Collective SVM optimized with chaos game system	Detection correctness, response time	Improves intrusion detection; reduced false positives	Potential great computational cost; requires robust information
D. Weiping et al. (2020) [15]	Fuzzy logic and chemical feedback optimization for intrusion discovery	Detection correctness, false positive rate	Improves accurateness; suitable for industrial big data atmospheres	Complication of model design; may necessitate extensive tuning
D. Viji et al. (2022) [16]	Hash-indexing for deduplication in cloud loading	Storage productivity, deduplication ratio	Reduces storage necessities; maintains data integrity	Incomplete to specific types of information; performance may vary with workload
N. Mageshkumar et al. (2023) [17]	Improved multilayer cryptosystem for safe deduplication	Safety level, deduplication efficiency	Balances safety and storage efficiency; protects sensitive information	Potential performance above; complexity in application
M. Akbar et al. (2023) [18]	Machine learning-based verification framework for deduplication	Authentication accurateness, deduplication ratio	Improves sanctuary of cloud storage; enhances deduplication effectiveness	Requires extensive training information; potential model bias

This table 1 provides a designed overview of the procedures, performance metrics, recompenses, and limitations associated with individually paper.

Now are important research gaps acknowledged from the above papers, converging on various aspects of big data management, safety, and healthcare applications:

III. Research Gap

1. **Integration of Health Data Standards**:
   Several studies, including Thomason (2021), discourse big data in digital health but deficiency a comprehensive integration of health information standards that ensure interoperability between different systems, which is crucial for current health analytics.
2. **Human-Centered Security Protocols**:
   Ahmad et al. (2022) highpoint ethical tasks in smart city development then do not propose specific human-centered safety protocols that account for user secrecy and consent in records management practices.
3. **Scalability of Classification Algorithms**:
   Although Xing and Bei (2019) utilize KNN for medical information classification, they prepare not address the scalability of such procedures when functional to larger datasets, which is important for real-world applications.
4. **Real-Time Key Management**:
   Luo et al. (2023) attention on protected key updates but do not discover the challenges of real-time key

ISSN : 2455-9679
SJIF Impact Factor : 6.190

organization in dynamic atmospheres, which is critical for maintaining sanctuary in high-speed networks.

5. **Generalization of Intrusion Detection Models**:
   Pustokhina et al. (2023) existing a model for imposition detection in program big data but do not examine its generalization competences across different types of information and environments.
6. **Cross-Dataset Validation**:
   The trainings on chronic kidney disease arrangement (Srivastava et al. 2022; Ayodele et al. 2020; Ahmad et al. 2020) regularly rely on specific datasets, deficient cross-dataset validation that would increase the robustness and applicability of their discoveries.
7. **Ethical Implications of AI in Manufacturing**:
   Bhattacharyya et al. (2023) proposition an IoT-enabled manufacturing model but then do not address the principled implications of AI and information use, particularly regarding operative surveillance and data ownership.
8. **Privacy Preservation in Cloud Environments**:
   Selvi and Kavitha (2022) discourse a privacy-aware framework then again do not examine the trade-offs amongst privacy preservation and system concert in cloud environments, which can distress user adoption.
9. **Evolving Threat Models in Intrusion Detection**:
   The imposition detection methods proposed by Ponmalar and Dhanakoti (2022) and Weiping et al. (2020) fix not consider the evolving environment of cyber threats, which demands adaptive and resilient detection techniques.
10. **Impact of Deduplication Techniques on Data Integrity**:
    While trainings on deduplication (Viji & Revathy, 2023; Mageshkumar et al., 2023; Akbar et al., 2023) application on storage efficiency, they frequently overlook the potential impact of these methods on data honesty and recovery processes in cloud storage structures.

These gaps designate areas where additional research is needed to increase the effectiveness and security of big data systems crossways several domains, particularly in healthcare, smart cities, and cybersecurity.

Here are some proposed solutions for the identified research gaps related to the papers:

Proposed Solutions for Research Gaps

1. **Integration of Health Data Standards**:
   Solution: Progress a unified framework that combines existing health information standards (e.g., HL7, FHIR) into big data analytics stands. This framework should encompass guidelines for data interoperability and conversation, ensuring that health information from different sources can be effortlessly integrated and analysed.
2. **Human-Centred Security Protocols**:
   Solution: Enterprise security protocols that highlight user consent and privacy, including user feedback mechanisms. Conduct factories with stakeholders to understand user worries and integrate these visions into the design of safety measures, ensuring they are see-through and user-friendly.
3. **Scalability of Classification Algorithms**:
   Solution: Discover hybrid classification models that association KNN with other scalable procedures, such as decision trees or collective methods, to improve performance with larger datasets. Appliance techniques such as dimensionality discount and distributed computing to simplify scalability.
4. **Real-Time Key Management**:
   Solution: Progress dynamic key management resolutions that utilize blockchain technology for protected, real-time key updates. Instrument a protocol that automates key revolution based on predefined safety policies and network conditions to safeguard ongoing security in high-speed atmospheres.
5. **Generalization of Intrusion Detection Models**:
   Solution: Deportment extensive training of imposition detection models on diverse datasets to progress their generalization aptitudes. Utilize techniques like transfer knowledge to adapt models trained proceeding one dataset to perform successfully on others, ensuring robustness across several environments.
6. **Cross-Dataset Validation**:
   Solution: Generate a consortium of healthcare administrations to share datasets for concerted research.

ISSN : 2455-9679
SJIF Impact Factor : 6.190

Develop standardized procedures for cross-dataset validation to measure the performance of classification models, safeguarding they are robust and dependable across different populations and settings.

7. **Ethical Implications of AI in Manufacturing**:
   Solution: Create an ethical framework for AI placement in manufacturing that embraces guidelines for data possession, user consent, and transparency in AI decision-making. Participate stakeholders in negotiations about ethical practices to produce accountability and promote accountable AI use.
8. **Privacy Preservation in Cloud Environments**:
   Solution: Appliance advanced cryptographic techniques such as per homomorphic encryption and protected multi-party computation to enhance discretion while maintaining performance. Conduct presentation evaluations to find the optimum balance between privacy protection and system effectiveness.
9. **Evolving Threat Models in Intrusion Detection**:
   Solution: Advance adaptive intrusion detection structures that leverage machine learning procedures to continuously learn from innovative threat data. Implement mechanisms for actual updates to threat prototypes based on emerging attack patterns, improving resilience against evolving cyber threats.
10. **Impact of Deduplication Techniques on Data Integrity**:
    Solution: Proposal deduplication algorithms that include information integrity checks, such as cryptographic muddles, to ensure that the original information can be dependably restored. Conduct experiments to appraise the performance impact of these truth checks on deduplication effectiveness and develop best practices for information recovery in cloud environments.

These solutions goal to address the acknowledged research gaps by recommending innovative strategies and methodologies, eventually enhancing the effectiveness and sanctuary of big data systems in healthcare, smart cities, and cybersecurity domains.

IV. CONCLUSION

In conclusion, the comprehensive review for optimum classification and deduplication of scrambled big data addresses the critical necessity for enhanced security methods in the management of complex information. As the volume of big data remains to grow, the addition of advanced machine learning algorithms not only simplifies efficient data classification but moreover safeguards data confidentiality through robust encryption methods. The deduplication development further optimizes packing resources, reducing redundancy without compromising information integrity.

The investigational results validate the efficiency of machine learning approach, showcasing high classification correctness and significant enhancements in storage efficiency. By accomplishing this balance between performance and safety, this review work offers a methodical solution that is appropriate across various domains, including cloud computing and IoT atmospheres. As administrations increasingly rely on big data analytics, the consequence of secure information management becomes paramount. This study contributes to the ongoing dissertation on data security, providing a groundwork for future advancements in the arena and addressing the tasks posed by the evolving scenery of big data. Ultimately, the outcomes highlight the importance of integrating sanctuary measures within data administration practices, ensuring that sensitive data remains protected in a progressively interconnected world.

REFERENCES

[1] J. Thomason, “Big tech, big data and the new world of digital health”, Global Health Journal, **5**(4), pp. 165-168, 2021.
[2] K. Ahmad, M. Maabreh, M. Ghaly, K. Khan, J. Qadir, and A. Al-Fuqaha. "Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges", Computer Science Review, Vol. **43,** 100452, 2022.
[3] W. Xing, and Y. Bei. "Medical health big data classification based on KNN classification algorithm." IEEE Access, Vol. **8,** pp. 28808-28819, 2019.

[4] Y. Luo, C. Zhang, X. Wang, X. Liang, and K. Qiu, "Robust Key Update with Controllable Accuracy Using Support Vector Machine for Secure OFDMA- PON", Journal of Lightwave Technology, IEEE, pp. 1-18, 2023.

[5] I. V. Pustokhina, D. A. Pustokhin, E. L. Lydia, P. Garg, A. Kadian and K. Shankar, "Hyperparameter search based convolution neural network with Bi-LSTM model for intrusion detection system in multimedia big data environment", Multimedia Tools and Applications, Vol. **81**(24), pp. 34951-34968, 2023.

[6] S. Srivastava, R. K. Yadav, V. Narayan, and P. K. Mall, "An Ensemble Learning Approach for Chronic Kidney Disease Classification", Journal of Pharmaceutical Negative Results, pp. 2401-2409, 2022.

[7] O. J. Ayodele, A. O. Adetunmbi, R. B. Ogunrinde, and B. B. Ajisafe, "Development of an ensemble approach to chronic kidney disease diagnosis", Scientific African, Vol. **8**, pp. 1-10, 2020.

[8] A. Ahmad, H. Najadat, B. Mohsen, and K. Balhaf, "Classification and association rule mining technique for predicting chronic kidney disease". Journal of Information & Knowledge Management, Vol. **19**(01**),** pp. 1-15, 2020.

[9] S. Bhattacharyya, S. Athithan, S. Pal, B. Sarkar, D. Akila, S. Chowdhury, K. Chandran and S. Gurusamy, "An IoT-Enabled Intelligent and Secure Manufacturing Model Using Blockchain in Hybrid Cloud Communication System", Security and Communication Networks, pp. 1-1, 2023.

[10] T. M. Selvi, and V. Kavitha, "A privacy-aware deep learning framework for health recommendation system on analysis of big data", The Visual Computer, Vol. **38(**2), pp. 385-403, 2022.

[11] T. Munirathinam, S. Ganapathy, and A. Kannan, "Cloud and IoT based privacy preserved e- Healthcare system using secured storage algorithm and deep learning", Journal of Intelligent & Fuzzy Systems, Vol. **39**(3), pp. 3011-3023, 2020.

[12] A. Ponmalar, and V. Dhanakoti. "Hybrid Whale Tabu algorithm optimized convolutional neural network architecture for intrusion detection in big data", Concurrency and Computation: Practice and Experience, Vol. **34 (**19), pp. 1-12, 2022.

[13] K. Xiu, Y. Fan, Z. Fang, L. Cao, N. N. Xiong, D. Yang, and X. Li, "A novel IoT network intrusion detection approach based on adaptive particle swarm optimization convolutional neural network", Information Sciences, Vol. **568,** pp. 147-162, 2021.

[14] A. Ponmalar, and V. Dhanakoti, "An intrusion detection approach using ensemble support vector machine-based chaos game optimization algorithm in big data platform", Applied Soft Computing, Vol. **116**, pp. 1-10, 2022.

[15] D. Weiping, J. Nayak, B. Naik, D. Pelusi, and M. Mishra, "Fuzzy and real-coded chemical reaction optimization for intrusion detection in industrial big data environment", IEEE Transactions on Industrial Informatics, Vol. **17**(6), pp. 4298-4307, 2020.

[16] D. Viji and S. Revathy, "Hash-Indexing Block-Based Deduplication Algorithm for Reducing Storage in the Cloud", Computer Systems Sciences & Engineering, Tech Science Press, Vol. 46 (1), pp. 27-42, 2022.

[17] N. Mageshkumar, J. Swapna, A. Pandiaraj, R. Rajkumar, M. Krichen and V. Ravi, "Hybrid Cloud Storage System with Enhanced multilayer Cryptosystem for Secure Deduplication in Cloud", International Journal of Intelligent Networks, KeAi, Vol. 4, pp. 301-309, 2023.

[18] M. Akbar, I. Ahmad, M. Mirza, M. Ali and P. Barmavatu, "Enhanced Authentication for De-Duplication of Big Data on Cloud Storage System using Machine Learning Approach", Cluster Computing, pp. 1-21, 2023.