ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 10, Issue 1, 2022

OPTIMIZING HEALTHCARE DATA COLLECTION AND SECURITY THROUGH CLOUD COMPUTING

Priyadarshini_Radhakrishnan,
Technical Lead, IBM,
Columbu, Ohio, United States
priyadarshinir990@gmail.com

Sathiyendran Ganesan,
Atos Syntel, California, USA
sathiyendranganesan87@gmail.com

Venkata Sivakumar Musam,
Astute Solutions LLC, California, USA
venkatasivakumarmusam@gmail.com

Nagendra Kumar Musham,
Celer Systems Inc, California, USA
nagendramusham9@gmail.com

Karthick M,
Associate Professor, Department of Information Technology,
Nandha college of Technology,
Erode, Tamilnadu-638052, India
magukarthik@gmail.com

ABSTRACT

The rapid growth of healthcare data and the increasing need for efficient management have led to challenges in cloud-based healthcare systems, including scalability, data security, and integration. Existing systems often struggle to manage large data volumes while ensuring secure transmission and storage. The aim of this work is to develop a secure and scalable cloud-based framework for efficient healthcare data collection and monitoring. The framework begins with collecting healthcare data from various sources, followed by preprocessing steps such as K-Nearest Neighbors (k-NN) imputation to handle missing values and Min-Max scaling for normalization. The data is then encrypted using Salsa 20 to ensure security, and Transport Layer Security is applied for secure data transmission to the cloud. The processed data is stored in cloud-based solutions for efficient management and real-time access. The results show that the latency of cloud systems increases with system load, from 1000 ms at lower loads to 4500 ms at higher loads, demonstrating the challenge of maintaining low latency as demand grows. Additionally, the Salsa 20 encryption achieves near 100% security strength with key sizes of 1024 bits. The contribution of this work lies in developing a robust, efficient, and secure framework that enhances healthcare data management while ensuring both performance and data security.

Keywords: Healthcare data, Cloud Storage, Encryption, Data transmission, Salsa20 and Transport Layer Security.

1 INTRODUCTION

The increasing volume and complexity of healthcare data have emphasized the need for optimized systems that can efficiently handle large datasets across various healthcare entities [1]. Traditional healthcare data management systems often face challenges in scalability, data integration, and accessibility, which can limit their effectiveness in improving patient care [2]. Cloud computing presents a powerful solution to these challenges by providing scalable infrastructure for data storage, processing, and sharing [3]. The proposed framework aims to optimize healthcare data collection and monitoring by leveraging cloud computing technologies to enhance the management, accessibility, and analysis of healthcare data, ultimately improving healthcare outcomes [4] [5].

Existing methods for healthcare data collection and monitoring primarily focus on centralized systems or on-premise servers, which struggle with scalability and efficient data integration [6]. Techniques like cloud-based Electronic Health Records (EHR) management, data warehousing, and cloud storage

solutions have been widely implemented [7]. However, these systems often encounter issues such as limited data accessibility across platforms, lack of seamless integration among different data types, and challenges with data security [8]. Despite the adoption of cloud solutions, managing heterogeneous healthcare data in a unified and secure manner remains a significant limitation in these existing systems [9].

The proposed framework addresses these drawbacks by utilizing advanced cloud computing architectures designed for enhanced scalability, data integration, and security [11]. By incorporating Transport Layer Security (TLS) for secure data transmission and leveraging flexible cloud storage solutions, the framework ensures data privacy and integrity [12]. Additionally, it overcomes the integration challenges faced by existing methods by providing a unified platform for efficient data storage and sharing across healthcare systems [13] [14]. The novelty of this approach lies in its ability to seamlessly manage healthcare data through cloud computing while ensuring both security and accessibility, optimizing data collection and monitoring for improved healthcare management.

The paper is structured as follows: Section 2 presents a literature survey, discussing existing works and their limitations. Section 3 outlines the methodology, followed by Section 4, which presents the results and Section 5 concludes the work.

2 LITERATURE SURVEY

The integration of cloud computing in healthcare has revolutionized the way patient data is managed, stored, and processed. However, the security and privacy of healthcare data have always been major concerns. A secure cloud-based system for safeguarding sensitive medical data by using extended privacy homomorphism and modified RSA encryption [18]. Their approach focused on ensuring that even when the data is stored in the cloud, it remains protected from unauthorized access. Despite its promising results, the authors noted challenges in integrating Internet of Medical Things (IoMT) technologies, suggesting that future systems could benefit from further research into implementing SMPC techniques to safeguard the confidentiality of healthcare data during the computation phase.

In a similar vein, Altowaijri (2020) proposed a multi-layered security architecture for cloud-based healthcare applications, which emphasized the importance of data encryption, access control, and secure communication protocols [19]. The system provided a robust framework for managing healthcare data privacy, leveraging both cloud storage solutions and encryption techniques [20]. However, the study highlighted issues related to data sharing across multiple healthcare entities, where the encryption methods alone were insufficient to prevent unauthorized access during real-time data processing. Altowaijri suggested that integrating SMPC could offer a stronger privacy-preserving layer to mitigate these issues by ensuring that data remains encrypted even during collaborative computations [21].

Another study by Giannopoulos and Mouris (2018) explored the application of SMPC in medical data analytics, focusing on privacy-preserving computations that allow healthcare professionals to collaborate without exposing sensitive patient data [22]. Their research highlighted the potential of SMPC to enable secure computations on encrypted data, which aligns with the growing need for collaborative analysis among healthcare providers without compromising privacy [23]. The study demonstrated that SMPC could facilitate a secure exchange of healthcare data across institutions, offering a promising direction for future cloud-based healthcare solutions. However, the authors also pointed out the computational challenges involved in implementing SMPC at scale, particularly in large healthcare systems that generate vast amounts of data [24].

Panga (2021) focused on the detection of financial fraud using hybrid machine learning models, a methodology that could be adapted to healthcare data security [25]. The framework combined multiple machine learning algorithms such as neural networks, decision trees, and support vector machines to identify fraudulent behavior in e-commerce transactions [26]. By applying similar hybrid models to healthcare data, it is possible to detect anomalies in patient records and clinical data. While the proposed

ISSN 2321-2152
www.ijmece.com
Vol 10, Issue 1, 2022

approach offered improved detection accuracy, it also raised concerns about the quality of data used in the training process [27]. The challenge of ensuring clean, high-quality data for accurate fraud detection in healthcare systems is paramount, and SMPC could be an essential tool in protecting the data while performing such analytics.

Finally, Ayyadurai (2021) proposed a hybrid recommendation system for e-commerce product recommendations, which integrated clustering techniques with evolutionary algorithms to provide personalized suggestions [28]. In the healthcare domain, a similar hybrid framework could be used to recommend personalized treatment plans or healthcare services by analyzing vast amounts of patient data. The study emphasized the role of advanced clustering and optimization methods in improving recommendation accuracy, which can be beneficial when applied to healthcare data systems . However, just as with e-commerce systems, the challenge remains to ensure the security of patient data during the recommendation process. Incorporating SMPC into such systems would ensure that patient data remains private while still benefiting from personalized healthcare recommendations.

2.1 Problem Statement

While significant progress has been made in cloud computing and data security for healthcare, several critical challenges remain unresolved. These challenges include scalability issues, inadequate protection of sensitive data and lack of seamless integration [31]. Current systems often face difficulties in handling large-scale data as healthcare demands increase, leading to performance degradation [33]. Furthermore, the protection of sensitive patient information is still not foolproof, with gaps in security measures during data transmission and storage [34]. Additionally, the integration of heterogeneous data from various healthcare sources remains complex and inefficient [35]. The work is proposed to overcome these challenges by providing an optimized, secure, and integrated cloud-based framework for effective healthcare data management and processing.

3 METHODOLOGIES

The proposed framework begins by gathering healthcare data from various sources, including patient records, medical devices, and health monitoring systems. Data preprocessing follows, with K-Nearest Neighbors (k-NN) imputation to handle missing values and Min-Max scaling to normalize numerical data. After preprocessing, the data is encrypted using Salsa 20 to protect sensitive healthcare information during transmission and storage. The encrypted data is then sent to cloud integration, where Transport Layer Security (TLS) is used to ensure secure communication between healthcare systems and cloud server. Finally, processed and encrypted data is stored in cloud-based storage solutions, providing scalable, efficient, and flexible data management for further analysis and collaboration. The proposed framework is illustrated in Figure1.

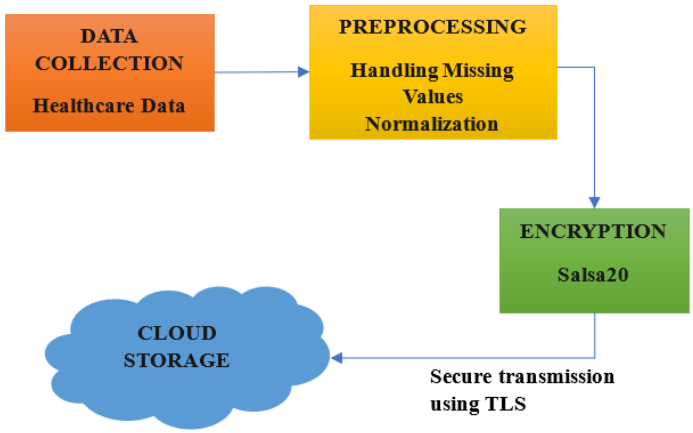

Figure 1: Workflow of Secure Healthcare Data Management

ISSN 2321-2152
www.ijmece.com
Vol 10, Issue 1, 2022

3.1 Data Collection

Data collection in the proposed framework involves gathering healthcare data from diverse sources, including patient records, medical devices, and health monitoring systems. This process ensures comprehensive data capture, including medical history, diagnostic results, and real-time health metrics. The collected data provides a complete view of patient health, enabling informed decision-making. Additionally, the framework ensures that the data is gathered securely and in compliance with privacy regulations. By integrating multiple data sources, the system supports robust healthcare analytics.

3.2 Preprocessing

The gathered healthcare data undergoes preprocessing to ensure it is ready for analysis.

3.2.1 Handling Missins values

The first step in preprocessing is missing value imputation, where K-Nearest Neighbors (k-NN) is used to fill in any missing data points by leveraging the similarities between neighboring instances. This process ensures that the dataset remains complete and that no valuable information is lost, which could negatively impact model accuracy.

3.2.2 Normalization

Following this, Min-Max scaling is applied to normalize the numerical data, ensuring all features are on the same scale. This step prevents features with larger ranges from dominating the analysis, allowing the model to treat each feature equally. These preprocessing steps ensure that the data is clean, consistent, and properly scaled, enhancing the effectiveness of subsequent analysis and model training.

3.3 Encryption

The pre-processed data is then encrypted using Salsa 20 to ensure the confidentiality and security of sensitive healthcare information. This encryption algorithm provides a lightweight yet robust method for protecting data during both storage and transmission. By encrypting the data before it is transferred to cloud storage, it prevents unauthorized access and ensures that even if intercepted, the data remains unreadable without the decryption key. Salsa 20 guarantees strong data protection, aligning with privacy regulations and maintaining the integrity of patient information. This step is essential for safeguarding healthcare data as it moves through the system.

The Salsa20 encryption algorithm is a stream cipher that encrypts data by combining a keystream with the plaintext through the bitwise XOR operation. The encryption process can be simplified as:

$$C_i = P_i \oplus K_i \tag{1}$$

Where, C_i is the ciphertext (encrypted data), P_i is the plaintext (original data), K_i is the keystream (generated by the Salsa20 algorithm) and $\oplus$ represents the XOR operation.

Key and Nonce Setup: The algorithm uses a 256 -bit key and a 64 -bit nonce (or counter) to initialize the Salsa20 state.

Keystream Generation: The Salsa20 algorithm generates a keystream by applying a series of mathematical transformations on the key and nonce.

Encryption: The plaintext P_i is combined with the keystream K_i through XOR to produce the ciphertext C_i. In simple terms, Salsa20 encrypts data by generating a keystream from a key and nonce, then combining it with the data using the XOR operation to produce encrypted output.

3.4 Cloud Integration

ISSN 2321-2152
www.ijmece.com
Vol 10, Issue 1, 2022

After encryption, cloud integration ensures secure and efficient data storage and transmission. Transport Layer Security (TLS) is used to secure communication between healthcare systems and cloud servers, preventing unauthorized access during data exchange. The encrypted healthcare data is securely transmitted to the cloud, where it can be stored and accessed by authorized parties. Cloud-based storage solutions provide scalability, allowing healthcare organizations to manage large volumes of data efficiently. This integration facilitates seamless collaboration across different healthcare entities while maintaining data privacy and compliance with regulations.

4 RESULTS

The results focus on the impact of system load on cloud latency and the relationship between key size and security strength in the Salsa20 encryption algorithm. These findings emphasize the need for efficient performance and strong encryption in cloud-based systems to handle varying loads and ensure data security.

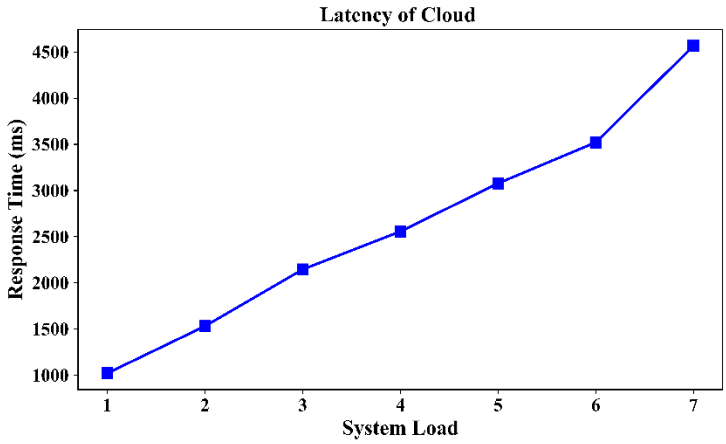

Figure 2: Latency of cloud

Figure 2 demonstrates the latency of cloud in response to varying system loads. As the system load increases from 1 to 7, the response time also rises, indicating a direct relationship between system load and latency. The graph shows a steady increase in latency, from 1000 ms at a low system load to 4500 ms at higher loads, highlighting how increased demand on cloud resources affects processing time. This trend reflects the challenge of maintaining low latency in cloud-based systems as they scale. The results emphasize the need for optimization strategies to mitigate latency under higher system loads.

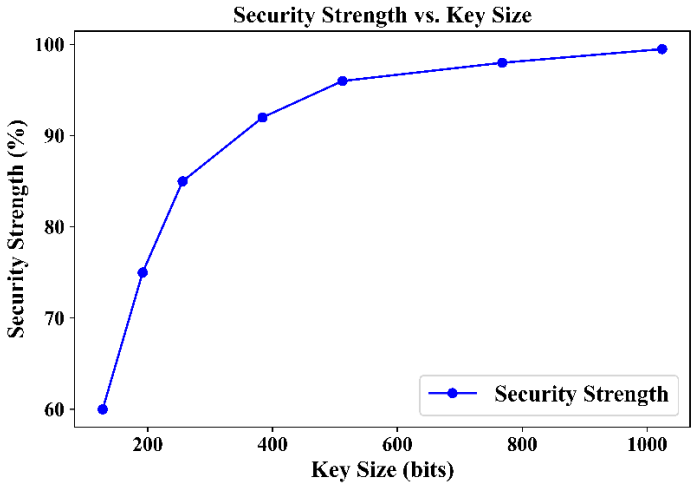

Figure 3: Security strength

Figure 3 illustrates the relationship between security strength and key size for the Salsa20 encryption algorithm. As the key size increases from 128 bits to 1024 bits, the security strength improves, reaching close to 100%. The graph demonstrates that larger key sizes provide exponentially stronger encryption, enhancing resistance to attacks. At lower key sizes, security strength is weaker, while higher key sizes offer significantly better protection. This shows the importance of choosing an appropriate key size in Salsa20 for achieving optimal encryption security.

5 CONCLUSIONS

The aim of this work was to develop a secure and scalable cloud-based framework for efficient healthcare data collection and monitoring. The framework enhances data management, privacy, and accessibility across healthcare systems. The results indicate that as system load increases from 1 to 7, the latency of cloud systems rises from 1000 ms to 4500 ms, reflecting the challenge of maintaining low latency under higher system loads. Additionally, the security strength of Salsa 20 encryption improves significantly with key size, reaching near 100% security at 1024 bits. This demonstrates the importance of choosing an appropriate key size for robust encryption in cloud-based healthcare systems. The proposed framework provides an efficient, secure, and scalable solution for healthcare data management, ensuring data privacy and optimizing performance. Future work will focus on implementing advanced load balancing and resource allocation techniques to reduce latency and improve the responsiveness of cloud-based healthcare systems under varying system loads.

REFERENCES

1. S. R. Sitaraman, “Optimizing Healthcare Data Streams Using Real-Time Big Data Analytics and AI Techniques,” *Int. J. Eng. Res. Sci. Technol.*, vol. 16, no. 3, Art. no. 3, Aug. 2020.
2. B. R. Gudivaka, “BIG DATA-DRIVEN SILICON CONTENT PREDICTION IN HOT METAL USING HADOOP IN BLAST FURNACE SMELTING,” *Int. J. Inf. Technol. Comput. Eng.*, vol. 7, no. 2, pp. 32–49, Apr. 2019.
3. Alagarsundaram, P. (2019). Implementing AES encryption algorithm to enhance data security in cloud computing. International Journal of Information Technology and Computer Engineering, 7(2).
4. N. S. Allur, “Genetic Algorithms for Superior Program Path Coverage in software testing related to Big Data,” *Int. J. Inf. Technol. Comput. Eng.*, vol. 7, no. 4, pp. 99–112, Dec. 2019.
5. S. S. Kethu, “AI-Enabled Customer Relationship Management: Developing Intelligence Frameworks, AI-FCS Integration, and Empirical Testing for Service Quality Improvement,” *Int. J. HRM Organ. Behav.*, vol. 7, no. 2, pp. 1–16, Apr. 2019.
6. R. K. Gudivaka, “ROBOTIC PROCESS AUTOMATION OPTIMIZATION IN CLOUD COMPUTING VIA TWO-TIER MAC AND LYAPUNOV TECHNIQUES”.
7. D. P. Deevi, “ARTIFICIAL NEURAL NETWORK ENHANCED REAL-TIME SIMULATION OF ELECTRIC TRACTION SYSTEMS INCORPORATING ELECTRO-THERMAL INVERTER MODELS AND FEA,” *Int. J. Eng.*, vol. 10, no. 3.
8. H. Chetlapalli, “ENHANCING TEST GENERATION THROUGH PRE-TRAINED LANGUAGE MODELS AND EVOLUTIONARY ALGORITHMS: AN EMPIRICAL STUDY”.
9. N. S. Allur, “Enhanced Performance Management in Mobile Networks: A Big Data Framework Incorporating DBSCAN Speed Anomaly Detection and CCR Efficiency Assessment,” vol. 8, no. 9726, 2020.
10. D. P. Deevi, “REAL-TIME MALWARE DETECTION VIA ADAPTIVE GRADIENT SUPPORT VECTOR REGRESSION COMBINED WITH LSTM AND HIDDEN MARKOV MODELS,” *J. Sci. Technol. JST*, vol. 5, no. 4, Art. no. 4, Aug. 2020.
11. S. Kodadi, “ADVANCED DATA ANALYTICS IN CLOUD COMPUTING: INTEGRATING IMMUNE CLONING ALGORITHM WITH D-TM FOR THREAT MITIGATION,” *Int. J. Eng. Res. Sci. Technol.*, vol. 16, no. 2, pp. 30–42, Jun. 2020.
12. K. Dondapati, “INTEGRATING NEURAL NETWORKS AND HEURISTIC METHODS IN TEST CASE PRIORITIZATION: A MACHINE LEARNING PERSPECTIVE,” *Int. J. Eng.*, vol. 10, no. 3.

13. N. S. Allur and W. Victoria, “Big Data-Driven Agricultural Supply Chain Management: Trustworthy Scheduling Optimization with DSS and MILP Techniques,” *Curr. Sci.*, 2020.
14. S. S. Kethu, K. Corp, and S. Diego, “AI and IoT-Driven CRM with Cloud Computing: Intelligent Frameworks and Empirical Models for Banking Industry Applications,” vol. 8, no. 1, 2020.
15. N. S. Allur, “Phishing Website Detection Based on Multidimensional Features Driven by Deep Learning: Integrating Stacked Autoencoder and SVM,” *J. Sci. Technol. JST*, vol. 5, no. 6, Art. no. 6, Dec. 2020.
16. N. K. R. Panga, “Optimized Hybrid Machine Learning Framework for Enhanced Financial Fraud Detection Using E-Commerce Big Data,” vol. 11, no. 2.
17. R. Ayyadurai, “Big Data Analytics and Demand-Information Sharing in E- Commerce Supply Chains: Mitigating Manufacturer Encroachment and Channel Conflict,” vol. 15, no. 3, 2021.
18. S. R. Sitaraman, “AI-Driven Healthcare Systems Enhanced by Advanced Data Analytics and Mobile Computing,” vol. 12, no. 2, 2021.
19. A. R. G. Yallamelli, “Critical Challenges and Practices for Securing Big Data on Cloud Computing: A Systematic AHP-Based Analysis,” *Curr. Sci.*, 2021.
20. R. Ayyadurai, “ADVANCED RECOMMENDER SYSTEM USING HYBRID CLUSTERING AND EVOLUTIONARY ALGORITHMS FOR E-COMMERCE PRODUCT RECOMMENDATIONS,” *Int. J. Manag. Res. Bus. Strategy*, vol. 11, no. 1, pp. 17–27, Jun. 2021.
21. T. Ganesan, “INTEGRATING ARTIFICIAL INTELLIGENCE AND CLOUD COMPUTING FOR THE DEVELOPMENT OF A SMART EDUCATION MANAGEMENT PLATFORM: DESIGN, IMPLEMENTATION, AND PERFORMANCE ANALYSIS,” *Int. J. Eng.*, vol. 11, no. 2.
22. S. R. Sitaraman, “Crow Search Optimization in AI-Powered Smart Healthcare: A Novel Approach to Disease Diagnosis,” *Curr. Sci.*, 2021.
23. M. R. Sareddy, “THE FUTURE OF HRM: INTEGRATING MACHINE LEARNING ALGORITHMS FOR OPTIMAL WORKFORCE MANAGEMENT”.
24. A. R. G. Yallamelli, “CLOUD COMPUTING AND MANAGEMENT ACCOUNTING IN SMES: INSIGHTS FROM CONTENT ANALYSIS, PLS- SEM, AND CLASSIFICATION AND REGRESSION TREES,” *Int. J. Eng.*, vol. 11, no. 3.
25. K. Gattupalli and H. M. Khalid, “Revolutionizing Customer Relationship Management with Multi-Modal AI Interfaces and Predictive Analytics,” *J. Sci. Technol. JST*, vol. 6, no. 1, Art. no. 1, Jan. 2021.
26. D. K. R. Basani, “Leveraging Robotic Process Automation and Business Analytics in Digital Transformation: Insights from Machine Learning and AI,” *Int. J. Eng. Res. Sci. Technol.*, vol. 17, no. 3, pp. 115–133, Sep. 2021.
27. V. K. Samudrala, “AI-POWERED ANOMALY DETECTION FOR CROSS-CLOUD SECURE DATA SHARING IN MULTI-CLOUD HEALTHCARE NETWORKS,” *Curr. Sci.*, 2020.