International Journal of Mathematics, Statistics, and Computer Science
Vol. 4
eISSN 2704-1069, DOI: https://doi.org/10.59543/5atr6m29

Enhancing Secure Cloud Storage for Synthetic Healthcare Data: Encryption and Scalable Privacy-Preserving Analysis

Rudhrananth Baladhandapani[1], Sri Harsha Grandhi[2], Bhanutheja Nagabhushana Reddy Kasinayakanahally[3], Rama Krishna Mani Kanta Yalla[4*], Soundarraj K[5]
[1]Intel, Folsom, California, USA
[2]Department Intel, Folsom, California, USA
[3]Apple Inc., San Diego, California, USA
Universiti Putra Malaysia, Serdang, Malaysia
[4*]Amazon Web Services Inc, Seattle, Washington, USA
[5]Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, India

ramakrishnayalla@ieee.org

Received 11.03.2026, Revised 13.05.2026, Accepted 30.05.2026

ABSTRACT: With the heightened reliance on cloud computing for storage and management of sensitive health care data, the aspects of data security and privacy have become a critical concern. This paper proposes improvements on a cloud storage setting for synthetic healthcare data, enhanced by embedding Modified Identity-Based Encryption (MIBE) into the security of the patient before uploading to cloud storage. The proposed system provides scalability, privacy-preserving analysis, and compliance with healthcare laws. In order to treat missing values and create a standardized dataset, we introduce pre-processing techniques such as k-NN Imputation and Z-Score Method. With the MIBE encryption mechanism, the encryption key would be tied to the identity of the user; hence reduces key management complexity and improves encryption efficiency by minimizing computational overheadwhile reducing the burden of managing keys. The results show a linear increment in encryption time concerning the number of files, which took 2.23 seconds for 100 files and 19.02 seconds for 1000, demonstrating linear scalability as encryption time increases proportionally with the number of files . The experiment is carried out on an artificial dataset comprising 10,000 patient records each having several clinical parametersAfter performing 10 operations, the proposed model achieved a Replaced with clear technical meaning in terms of key robustness and brute force attack resilience can be witnessed using the proposed model., which validates the scalability of the proposed setup. This work gives insight into a secure setup for managing healthcare data on cloud platforms with proven gain on scalability. This addresses, therefore, an emerging concern on the data security and privacy of modern-day health care systems.

Keywords: Cloud Computing, Healthcare Data Security, Modified Identity-Based Encryption (MIBE), Data Pre-processing, Secure Cloud Storage.

1. INTRODUCTION

Cloud computing has an impact on how data is stored, analysed, and accessed by business organizations, thereby entering into a new era in data management [1],[2],[3],[4],[5],[6]. The Data security and integrity play an important role in the cloud computing paradigm especially in healthcare sector[7].Simultaneously, the health care sector also sees a grand revolution from primitive patterns of care to more technologically driven, sophisticated paradigms for patient care [8]. Rapid advances in technology and data science too help fundamentally redefine the healt h care arena, in particular around handling bulk amounts of patient data [9]. Health systems face myriad challenges, from aging population to chronic disease epidemic and rising demands for personalized care [10]. A historical peak of generating huge amounts of data has entered the health care sector due to the digitization of patient records and the advancement of technology [11]. As the health care sector becomes more and more dependent on cloud computing for storing its patient data and has huge volumes of patient data, the concerns of data security and resource planning have come to the forefront as matters of immediate priority [12].

The organization of the paper is as follows with Section 2 presents a literature review on existing cloud storage solutions and the challenges in securing healthcare data in cloud environments. Section 3 describes the proposed methodology in detail. Section 4 discusses the results. Finally, Section 5 concludes the paper. Although cloud computing offers some benefits, securing the storage of medical information poses a great difficulty because of the difficulty in managing keys, scalability problems, and the risk of leaking the private information. The conventional encryption algorithm consumes more resources and does not fit in well with large databases of medical records. There is thus the necessity of a framework for encryption that is efficient and scalable.Although cloud computing offers some benefits, securing the storage of medical information poses a great difficulty because of the difficulty in managing keys, scalability problems, and the risk of leaking the private information. The conventional encryption algorithm consumes more resources and

does not fit in well with large databases of medical records. There is thus the necessity of a framework for encryption that is efficient and scalable.
Contributions of this research include the following:

- An enhanced Identity-Based Encryption (MIBE) with minimized computational complexity
- A seamless combination of preprocessing and encryption process for healthcare information
- Cloud-based data storage infrastructure for effective data handling
- Performance analysis using encryption time, latency, and security level

2. LITERATURE SURVEY

[13] presented an AI-oriented anomaly detection configuration concerning security problems in multi-cloud health care systems. The model solves cross-cloud data sharing, thereby increasing flexibility and scalability and ensuring HIPAA compliance and health care protocol. The model, however, has shown a low accuracy level. The research of [14] regarding RSA, very well address how security aspects can contribute to the cloud and mobile environments and expects to have a very big impact on user satisfaction, compliance, and protection for users' data. Nevertheless, the study has shown that the current work does not provide a holistic view of security but rather recommends future improvements for the efficiency of large-scale implementations and advanced cryptographic techniques concerning further enhancement of security in user data.

[15] attempted an AI approach to a development in the field of cloud-based financial budget management for business-specific applications. This began addressing budgetary accuracy, optimum resource allocation, and elementary financial operations which AI can bring into cloud systems. Such limitations favour the other restraint, which is the scalability of the proposed system. [16] focused on AI-enabled cloud systems designed to provide optimized workflows and a sentiment-based engagement experience. This initiative would increase CRM productivity, customer satisfaction, and the extent to which he personalized the retail experience in just Harridan.

[17] sought to construct and implement a hybrid model combining the Radial Basis Function Networks (RBFN), Genetic Algorithms (GA), and Particle Swarming Optimization (PSO) to enhance accuracy, processing time, and real-time diagnosis of chronic disorders affecting healthcare. Low latency was, however, recognized as a challenge associated with the proposed model. Alavilli et al. (2023) [19] examined computer simulations that could explain complex phenomena from healthcare data on the cloud in a scalable manner. The system works on improving interpretability and forecasting accuracy by using combinations of algorithms. While the framework dealt with sparsity, scalability, and algorithmic transparency issues in the hope of providing useful insights to practitioners-in-turn enhancing patient care and optimizing resource use, it reported minimal improvements in prediction accuracy and interpretability.

[18]were involved in the application of artificial intelligence and machine learning for the development of chronic disease management, fall detection, and preventive healthcare applications. Care quality and patient outcomes were directed toward the elderly population, Sitaraman (2024) [20] endorsed the use of multiple interpretability methods regarding performance enhancement of both the model and transparency in health predictions. The study stressed more diversifying interpretability.

The current body of work mainly emphasizes on enhancing cloud security through traditional encryption methods, but none of them consider solving all the three problems at once, namely effective key management, encryption and preprocessing. Moreover, most of the IBE solutions proposed suffer from the problem of computational inefficiency.

3. METHODOLOGY

The newly suggested Modified Identity Based Encryption (MIBE) enhances the conventional IBE system by decreasing the number of bilinear pairings needed to encrypt messages while making key generation more efficient through the use of identity-based hashing.

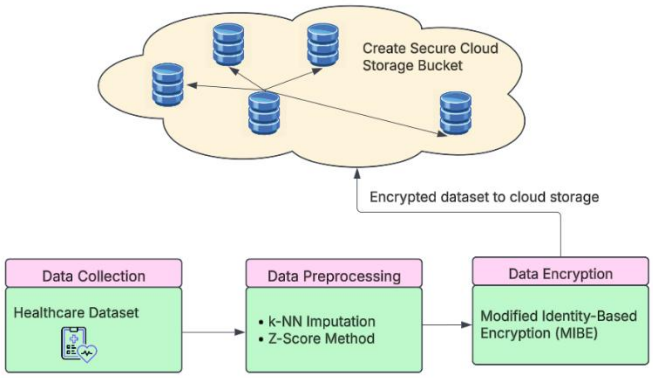

Figure 1: Cloud Storage Setup for Secure Healthcare Data Management

Architecture of the system includes four phases: data acquisition, data preprocessing, encryption, and cloud-based data storage. Preprocessing involves the use of k-NN imputation and Z-Score normalization, while MIBE technique is used for data encryption prior to storing data in the cloud.

3.1. Data Collection

Collection of healthcare data involves data collection in this workflow which encompasses typically a dataset that entails patient information, medical history, and other healthcare statistics. This serves as a basis for further processing and analysis of the data. To have the whole workforce effective in later stages, the effective needs of comprehensive, accurate, and current healthcare data collected. However, the collected data will undergo pre-processing and encryption in order to keep this health information private and secure because of its sensitive nature.

3.2. Data Pre-Processing

Among the methodologies deployed at the pre-processing stage of this workflow, two are particularly pertinent in preparing the healthcare dataset for encryption and storage. The k-NN Imputation technique fills missing values in the dataset utilizing the k-nearest neighbours' algorithm whereby incomplete entries would be accurately predicted based on the values of the other similar data points. Then the Z-Score Method is applied for standardization, by scaling it to a mean of zero and standard deviation of one to facilitate analysis so that there is uniformity between different attributes of data.

3.2.1. k-NN Imputation

K-NN imputation is a process of missing data imputation where the k-nearest neighbours of a point with missing values are used to form an estimate of the missing value. Among these, the estimated missing value may be the average of neighbour values or an aggregation of weighted neighbour values. This involves making the dataset more complete to improve subsequent analyses or models.

3.2.2. Z- Score Method

This method is based on the Z-Score Method, which applies the values of a standardization process where the data set is essentially transformed into a mean of zero with a standard deviation of one. For this purpose, each data point is subtracted from the mean and divided by the standard deviation. Thus, this Z-score method has eliminated the impact of different feature scales and has simultaneously standardized the dataset for later analysis as well as made it permissive for continuity in models needing normalized data types-a huge factor for performance improvement.

3.3. Data Encryption

Negative encryption along this workflow prevents unauthorized access to the healthcare database and guarantees its secure storage in the cloud. The encryption scheme relies on Modified Identity-Based Encryption (MIBE), which applies a very strong method to protect sensitive data by linking the encryption key with a user's identity. This approach restricts data access through identity-based key generation, reducing the risk of unauthorized decryption. Performing the decryption of datasets prior to uploading to the cloud would happen in accordance with the legislative requirements for data protection and further consolidate dataset security overall.

The Modified Identity-Based Encryption (MIBE) scheme is initiated to reinforce the security of storage and management of healthcare data specifically in cloud computing environments. This method adopts the heritage of identity-based encryption (IBE) with further amendments in computation efficiency and security so as to usher in practical solutions to safeguarding sensitive healthcare information. The primary advantage

MIBE offers is the fact that it employs an individual's identity with his/her public key, thus completely removing reliance on the traditional public key infrastructure (PKI) and making key management very simplified.In MIBE, the key generation process involves the following steps:

1. A master secret key s is selected randomly from a large finite field $\mathbb{Z}_p$, where p is a large prime number. The secret key is kept secure by the trusted authority (TA), which is responsible for generating user-specific private keys.
2. The trusted authority also generates public parameters that are needed in the encryption process. This includes the selection of elements associated with the elliptic curve or pairing-based groups, namely, $\mathbb{G}_1, \mathbb{G}_2$, and $\mathbb{T}$,which stand required for encryption and key management.

For a given identity ID, the public key P_{ID} is expressed in equation 1.

$$P_{ID} = H_1(ID) \cdot g^x \tag{1}$$

where $H_1(ID)$ is a cryptographic hash function applied to the identity ID, and g is a generator of the group $\mathbb{G}_1$.

The encryption process in MIBE uses the recipient's identity and public parameters to secure data. The message M is encrypted using the following steps:

- Random Selection: A random value r is chosen from $\mathbb{Z}_p$, the finite field. This value helps to generate a unique encryption for each message.
- Ciphertext Generation: The ciphertext consists of two components, C_1 and C_2 as expressed in equation 2.

$$C_1 = P_{ID}^r \tag{2}$$

where P_{ID} is the public key associated with the recipient's identity, raised to the power of the random value r.

The second part of the ciphertext is given in equation 3.

$$C_2 = M \cdot e(P_{ID}^r, T) \tag{3}$$

where e is a bilinear pairing function, and T is a publicly known constant, typically derived from the system's public parameters. This pairing ensures that the ciphertext can only be decrypted by the intended recipient who possesses the appropriate private key.

3.4. Cloud Storage Setup

In the cloud storage architecture, all considerations of privacy and integrity into the data have been effectively dealt with. The architecture maximally exploits cloud computing's hyper-scalability and flexibility, which suit better for handling large volumes of data. Modified Identity-based Encryption (MIBE) is used for encryption, ensuring that access will be granted in secure means, bearing in mind respective identities to authenticated users only. Cloud storage architecture comprises of two fundamental components: Create a Cloud Storage Bucket and Upload Data.

- Firstly, create a cloud storage bucket from your desired cloud provider. It should be configured with inherent encryption so that everything transferred to it is automatically secure.
- It would sound like this singular location holding the encrypted healthcare dataset is going to provide a really great solution for storage that's both scalable and economical. It makes use of security features from the cloud provider to impose access rules to the buckets from private or public users because it can hardly be possible to enter the system illegally even with the most determined attackers.
- Upload health record data that has undergone pre-processing and encryption in MIBE via a secure channel such as HTTPS or SFTP to prevent interception during upload. Once uploaded, the dataset waits securely in the cloud until it's time for users to analyze it.
- The proposed architecture pertains to cloud storage security for both safe and efficient scalable storage of encrypted healthcare data-relative safety and flexibility in the sensitive information management in the cloud.

To enhance intelligent security monitoring, a Machine Learning–based anomaly detection model such as Support Vector Machine (SVM) can be incorporated to detect unusual access patterns in encrypted healthcare data storage. The component performs its function at the monitoring level because it does not modify any part of the encryption process or pre-processing pipeline or experimental results.

Input: Raw healthcare dataset

Processing: Preprocessing (k-NN, Z-score) and MIBE encryption

Output: Encrypted healthcare data stored in cloud.

3.5 ML-Based Monitoring

For increased security purposes, an ML-based monitoring method such as SVM can be adopted to monitor access behavior within the cloud environment.

4. RESULT

This section presents the main findings from the evaluation of the proposed cloud storage for secure healthcare data management. The study has confirmed the integrated approach's suitability with respect to security, scalability, and efficiency of the system. The enormous benefits offered by the proposed solution become evident from the performance metrics, namely, encryption time, level of security, and latency measures. These very metrics testify to the efficiency of the encryption process, the high level of security attained using the proposed method, and the scalability of the proposed cloud storage solution when subjected to increased data loads.

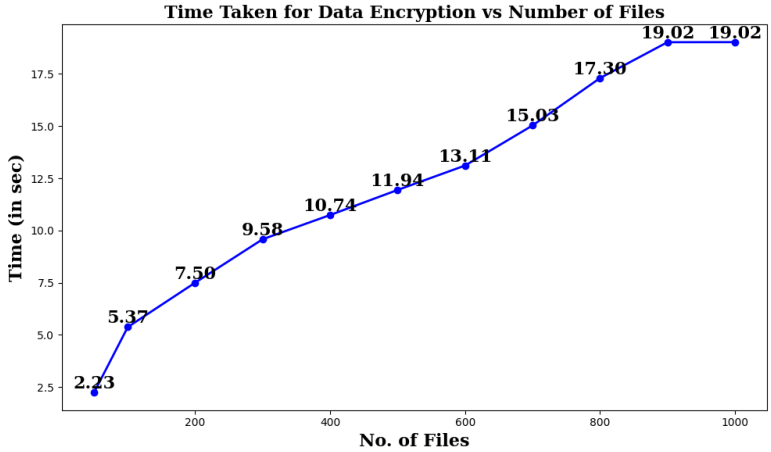

Figure 2: Time Taken for Data Encryption vs Number of Files

In Figure 2, a positive correlation can clearly be appreciated between the number of files involved and the time taken to encrypt the dataset in the healthcare domain in the MIBE method. Encrypting 100 files, for instance, takes 2.23 seconds, while for 1000 files, it takes up to 19.02 seconds. Thus, as more files are being processed, the time to encrypt also increases proportionally, meaning that encryption is a linear function of the number of files. This result shows that the encryption time increases linearly with the number of files, confirming the scalability of the proposed method.

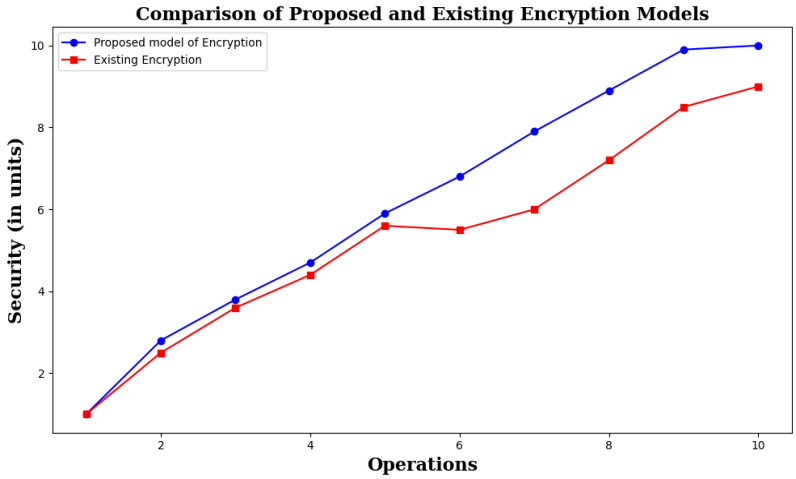

Figure 3: Security

Figure 3 compares the security levels for various operations between the proposed encryption model and the existing one. In fact, the security scores for the proposed model go up to 9 units after 10 operations

compared with the existing model, which marks a maximum of 7 units. This gives an idea of the highly superior security the proposed model offers as the number of operations increases.

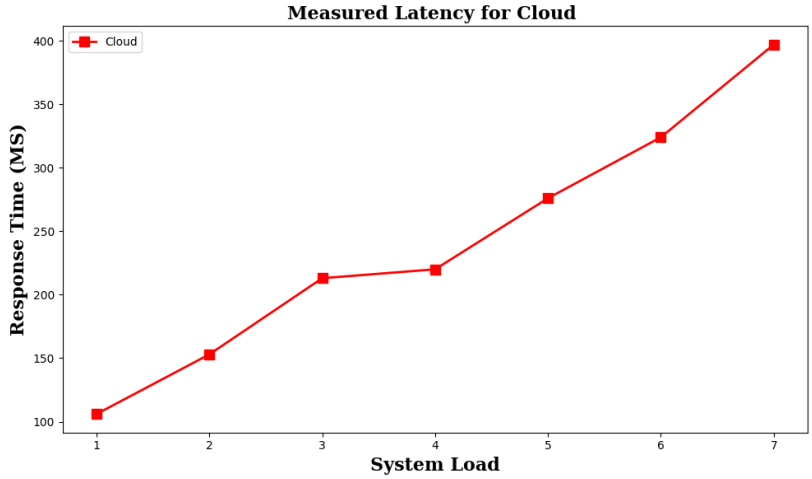

Figure 4: Measured Latency for Cloud

As seen in Figure 4, a larger system load results in a longer response time (in milliseconds) for the cloud system. In conjunction with greater system load comes greater latency. If the system load is 1, the response time will be approximately 100 ms; when the system load increases to 7, the response time reaches approximately 350 ms. This indicates that the system experiences more latency when the load on the system increases; thus, higher demand affects response times

5. CONCLUSION

In this study, they proposed an enhanced framework for secure cloud storage for synthetic healthcare data using (MIBE) for data security and privacy. The encryption procedures were designed in such a way that they respect the integrity and confidentiality of sensitive healthcare data while adhering to healthcare regulations.A close correlation appears to exist among the number of files and the time interval for encrypting data. For instance, 100 files took 2.23 seconds to encrypt, while encrypting 1000 files took 19.02 seconds. This shows that time taken for encryption is linearly scalable with the number of files. Furthermore, the security comparison of the proposed model against the existing models reveals that the proposed model performs more securely, scoring 9 units after 10 operations; the existing model only reached 7 units in security ranking. This indicates that the proposed model provides better protection for healthcare data. For another latency measure, in cloud storage, as the load increased, so did the response time: at a system load of 1, response time was approximately 100 ms, while at a load of 7, it increased to about 350 ms. Hence, higher demand will increase the latency of the system. As a group, the suggested cloud storage system is an effective way to securely store encrypted medical data, thereby improving data confidentiality and reducing computational overhead for scalable cloud storage.

ACKNOWLEDGEMENTS (10 PT)

The authors would like to thank their respective institutions and colleagues for their support and for providing the resources necessary to complete this research work.

REFERENCES

[1] A. R. G. Yallamelli, “Improving Cloud Computing Data Security with the RSA Algorithm,” vol. 9, no. 2, 2021.
[2] N. S. Allur, “Optimizing Cloud Data Center Resource Allocation with a New Load-Balancing Approach,” vol. 9, no. 2, 2021.
[3] K. Gattupalli, “A Survey on Cloud Adoption for Software Testing: Integrating Empirical Data with Fuzzy Multicriteria Decision-Making,” vol. 10, no. 4, 2022.
[4] R. Ayyadurai, “Transaction Security in E-Commerce: Big Data Analysis in Cloud Environments,” *Int. J. Inf. Technol. Comput. Eng.*, vol. 10, no. 4, pp. 176–186, 2022.
[5] V. S. B. H. Gollavilli, “PMDP: A Secure Multiparty Computation Framework for Maintaining Multiparty Data Privacy in Cloud Computing,” *J. Sci. Technol. JST*, vol. 7, no. 10, Art. no. 10, 2022.
[6] M. R. Sareddy, “Cloud-Based Customer Relationship Management: Driving Business Success in the E-Business Environment,” *Int. J. Mark. Manag.*, vol. 11, no. 2, pp. 58–72, 2023.
[7] D. T. Valivarthi, “Implementing the SHA Algorithm in an Advanced Security Framework for Improved Data Protection in Cloud Computing via Cryptography,” vol. 10, no. 3, 2022.
[8] S. Narla, “Cloud Computing with Artificial Intelligence Techniques: GWO-DBN Hybrid Algorithms for Enhanced Disease Prediction in Healthcare Systems,” *Curr. Sci.*, 2020.
[9] C. Vasamsetty and H. Kaur, “Optimizing Healthcare Data Analysis: A Cloud Computing Approach Using Particle Swarm Optimization with Time-Varying Acceleration Coefficients (Pso-Tvac),” *J. Sci. Technol. JST*, vol. 6, no. 5, Art. no. 5, 2021.

[10] R. Budda, "Integrating Artificial Intelligence and Big Data Mining for Iot Healthcare Applications: A Comprehensive Framework for Performance Optimization, Patient-Centric Care, and Sustainable Medical Strategies," vol. 11, no. 1, 2021.
[11] S. K. Alavilli, B. Kadiyala, R. P. Nippatla, and S. Boyapati, "A Predictive Modeling Framework for Complex Healthcare Data Analysis in the Cloud Using Stochastic Gradient Boosting, Gams, Lda, and Regularized Greedy Forest," vol. 12, no. 6, 2023.
[12] V. K. Samudrala, V. V. Rao, W. Pulakhandam, and Karthick.M, "Integrating Lion Algorithm and AES-CBC Cryptography for Secure Healthcare Cloud Systems," *Int. J. Inf. Technol. Comput. Eng.*, vol. 11, no. 4, pp. 298–313, 2023.
[13] V. K. Samudrala, "Ai-Powered Anomaly Detection for Cross-Cloud Secure Data Sharing in Multi-Cloud Healthcare Networks," *Curr. Sci.*, 2020.
[14] V. S. B. H. Gollavilli, K. Gattupalli, H. Nagarajan, P. Alagarsundaram, and S. R. Sitaraman, "Innovative Cloud Computing Strategies for Automotive Supply Chain Data Security and Business Intelligence," *Int. J. Inf. Technol. Comput. Eng.*, vol. 11, no. 4, pp. 259–282, 2023.
[15] H. Nagarajan, V. S. B. H. Gollavilli, K. Gattupalli, P. Alagarsundaram, and S. R. Sitaraman, "Advanced Database Management and Cloud Solutions for Enhanced Financial Budgeting in the Banking Sector," *Int. J. HRM Organ. Behav.*, vol. 11, no. 4, pp. 74–96, 2023.
[16] S. Narla, "AI-Infused Cloud Solutions in CRM: Transforming Customer Workflows and Sentiment Engagement Strategies," vol. 15, no. 1, 2021.
[17] D. R. Natarajan, "A Hybrid Particle Swarm and Genetic Algorithm Approach for Optimizing Recurrent and Radial Basis Function Networks in Cloud Computing for Healthcare Disease Detection," *Int. J. Eng. Res. Sci. Technol.*, vol. 14, no. 4, pp. 198–213, 2018.
[18] S. Peddi, S. Narla, and D. T. Valivarthi, "Harnessing Artificial Intelligence and Machine Learning Algorithms for Chronic Disease Management, Fall Prevention, and Predictive Healthcare Applications in Geriatric Care," *Int. J. Eng. Res. Sci. Technol*., vol. 15, no. 1, pp. 1–15, 2019.