Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.057257

ARTICLE

Security Strategy of Digital Medical Contents Based on Blockchain in Generative AI Model

Hoon Ko[1] **and Marek R. Ogiela**[2,*]

[1]Division of Computer Science and Engineering, Sunmoon University, Asan, 31460, Republic of Korea

[2]AGH University of Krakow, Cryptography and Cognitive Informatics Laboratory, Krakow, 30-059, Poland

*Corresponding Author: Marek R. Ogiela. Email: mogiela@agh.edu.pl

Received: 13 August 2024 Accepted: 29 November 2024 Published: 03 January 2025

ABSTRACT

This study presents an innovative approach to enhancing the security of visual medical data in the generative AI environment through the integration of blockchain technology. By combining the strengths of blockchain and generative AI, the research team aimed to address the timely challenge of safeguarding visual medical content. The participating researchers conducted a comprehensive analysis, examining the vulnerabilities of medical AI services, personal information protection issues, and overall security weaknesses. This multifaceted exploration led to an in-depth evaluation of the model's performance and security. Notably, the correlation between accuracy, detection rate, and error rate was scrutinized. This analysis revealed insights into the model's strengths and limitations, while the consideration of standard deviation shed light on the model's stability and performance variability. The study proposed practical improvements, emphasizing the reduction of false negatives to enhance detection rate and leveraging blockchain technology to ensure visual data integrity in medical applications. Applying blockchain to generative AI-created medical content addresses key personal information protection issues. By utilizing the distributed ledger system of blockchain, the research team aimed to protect the privacy and integrity of medical data especially medical images. This approach not only enhances security but also enables transparent and tamper-proof record-keeping. Additionally, the use of generative AI models ensures the creation of novel medical content without compromising personal information, further safeguarding patient privacy. In conclusion, this study showcases the potential of blockchain-based solutions in the medical field, particularly in securing sensitive medical data and protecting patient privacy. The proposed approach, combining blockchain and generative AI, offers a promising direction toward more robust and secure medical content management. Further research and advancements in this area will undoubtedly contribute to the development of robust and privacy-preserving healthcare systems, and visual diagnostic systems.

KEYWORDS

Digital medical content; medical diagnostic visualization; security analysis; generative AI; blockchain; vulnerability; pattern recognition

Nomenclature

AI	Artificial Intelligence
XSS	Cross Site Scripting

CSRF	Cross-Site Request Forgery
API	Application Programming Interface
DMC	Digital Medical Contents
FDA	Food and Drug Administration
WHO	World Health Organization
CDC	Center for Disease Control and Prevention
LLM	Large Language Model
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
SMPC	Secure Multi-Party Computation
BERT	Bidirectional Encoder Representations from Transformers
GPT	Generative Pretrained Transformer

1 Introduction

Many medical and research institutions specialize in visual health screenings, offering comprehensive health management services supported by advanced medical technology and expert personnel. These institutions operate precise and systematic health screening programs, aiming for early detection and prevention of various diseases based on analysis of medical diagnostic visualization. Recently, many of them have started using AI services to provide customized offerings, and pattern classification towards lesion detection. However, the medical content AI service models provided as digital content face several issues, such as AI model vulnerabilities, personal information protection problems, and security vulnerabilities in medical AI systems. From the perspective of AI model vulnerabilities, models can be deceived through various attacks, such as adversarial attacks, leading to incorrect diagnoses, which can have serious consequences in medical settings [1]. Additionally, model extraction attacks can allow attackers to replicate the functionality of AI systems or understand their internal structure, raising concerns about intellectual property infringement. Data poisoning attacks can inject malicious data (including fake images) during the learning process, degrading model performance or inducing intended results. Exploiting specific vulnerabilities in AI models can lead to misdiagnosis or incorrect prescriptions, potentially threatening patient safety. Protecting medical visual data is crucial for safeguarding personal information. Even de-identified data carries the risk of re-identification when combined with other information. Inference attacks are also possible, where analyzing AI model outputs can lead to deducing personal information included in the training data. Data leaks due to security vulnerabilities or insider mistakes can result in large-scale personal information breaches [2]. Using medical data and diagnostic visualizations for AI training without patient consent can cause ethical and legal issues, and insufficient de-identification in data sharing for research purposes can risk exposing personal information. Security vulnerabilities in medical AI systems can manifest in various forms. Insecure API design can lead to unauthorized access or data leaks, while unsafe network communications increase the risk of data interception during transmission. Inadequate authentication mechanisms or access controls can allow unauthorized users to access the system, and vulnerabilities or misconfigurations in server software can expose the system to external attacks and information leakage [3]. Web applications may have client-side vulnerabilities like XSS or CSRF, and the use of weak encryption algorithms or improper key management can lead to the exposure of critical medical data including diagnostic images. Therefore, a strategy to enhance the security of existing models is necessary. As a proposal, this study explains a blockchain-based digital medical content safety strategy in a generative AI environment [3].

The structure of this study is as follows: Section 2 analyzes the threat factors of existing Digital Medical Content (DMC), Section 3 defines the DMC safety analysis strategy based on the proposed Blockchain-based Medical Contents Security Framework to address these threat factors. Section 4 covers Results and Discussion, Finally, Section 5 summarizes the Conclusion and Future Works.

2 Threat Factors of Digital and Visual Medical Content

The threat factors for Digital Medical Contents (DMC) can be classified into ransomware attacks, data breaches, medical device hacking, phishing attacks, insider threats, and cloud security incidents. Ransomware attacks are particularly dangerous in the medical field because when patient records, diagnostic images, and drug prescription systems are encrypted, immediate patient care becomes impossible. For example, during the 2017 WannaCry attack, the UK's NHS had to cancel about 19,000 appointments, and many emergency patients were transferred to other hospitals [4,5]. The severity of medical data breaches lies in the sensitivity of the data [6]. Leaked information may include social security numbers, financial information, and detailed medical records, which can be misused for identity theft, medical fraud, or personal blackmail [7]. After the 2015 Anthem hacking incident, the risk of identity theft for victims greatly increased. Hacking of networked medical devices (e.g., insulin pumps, pacemakers) can pose a direct threat to life [8]. In 2017, the Food and Drug Administration (FDA) recommended firmware updates for certain pacemakers due to vulnerabilities that could allow remote manipulation [9,10]. Phishing attacks in the medical field often exploit timely topics. During the early stages of the COVID-19 pandemic, many emails impersonating the World Health Organization (WHO) or the Centers for Disease Control and Prevention (CDC) were circulated, aiming to steal medical staff's account information or install malware. Insider threats in medical institutions can take various forms, from simple curiosity-driven access to celebrities' medical records to selling patient information for financial gain [11]. A 2017 study found that about 58% of medical data breaches were caused by insiders. While cloud services are crucial for medical institutions that need to handle large-sized data, they face new security challenges when adopting these services. Major risk factors include incorrect access permission settings, unencrypted data transmission, and vulnerable APIs [12]. A 2018 report showed that about 9% of healthcare-related cloud storage was publicly accessible. Finally, medical IoT devices often struggle to implement strong security features due to limited computing power and battery life. Many devices use outdated operating systems or firmware, exposing them to known vulnerabilities. A study conducted in 2020 reported that approximately 70% of IoT devices used in hospitals had serious security risks [13,14]. Table 1 summarizes the classification and content of misuse and abuse of digitally stored health information.

Table 1: Classification and content for misuse of digital health information

Classification	Contents
Grounds for discrimination and hate speech	– Used as a basis for discriminatory actions against individuals or groups – Promoting hate speech based on specific health conditions
Disadvantages in insurance and employment	– Possibility of unfair disadvantages in the insurance claim process – Concerns about negative effects on employment opportunities
Personal defamation and social pressure	– Defamation using socially sensitive health information – Exploiting as a tool for unfair social pressure on individuals

(Continued)

Table 1 (continued)

Classification	Contents
Unethical advertising	– Exaggerated advertising using health checkup results – Using in marketing that misleads or confuses consumers
Political exploitation	– Using to attack specific political leanings – Using as a negative strategy in election campaigns

3 Proposed Framework

3.1 Blockchain-Based Medical Contents Security Framework

The proposed blockchain-based medical content security framework consists of a generative AI model layer, a blockchain layer, a security and encryption layer, an application layer, and a regulatory definition layer (Fig. 1). Table 2 summarizes the explanations of the components of the proposed blockchain-based medical contents security framework [15]. and Fig. 2 shows the blockchain-based medical contents security structure.

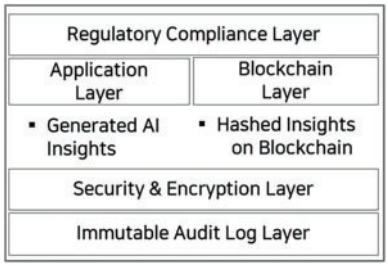

Figure 1: Blockchain-based medical contents security framework

Table 2: Blockchain-based medical contents security framework definition

Classification	Contents
Regulatory compliance layer	– Responsible for compliance with regulations such as GDPR, HIPAA – Maintain audit logs for all activities and access records recorded on the blockchain
Application layer	– Interact with users – Provide insights analyzed through generative AI models – Query hash values recorded on the blockchain and access data as needed
Blockchain layer	– Use platforms like Ethereum or Hyperledger Fabric – Convert insights generated by generative AI models into hash values – Permanently record converted hash values on the blockchain

(Continued)

Table 2 (continued)

Classification	Contents
Security & encryption layer	– Manage data security and encryption – Maintain data confidentiality using technologies such as homomorphic encryption – Perform access control
Immutable audit log layer	– Securely record access history of data stored on the blockchain – Securely record data modification history – Verify access and modification history

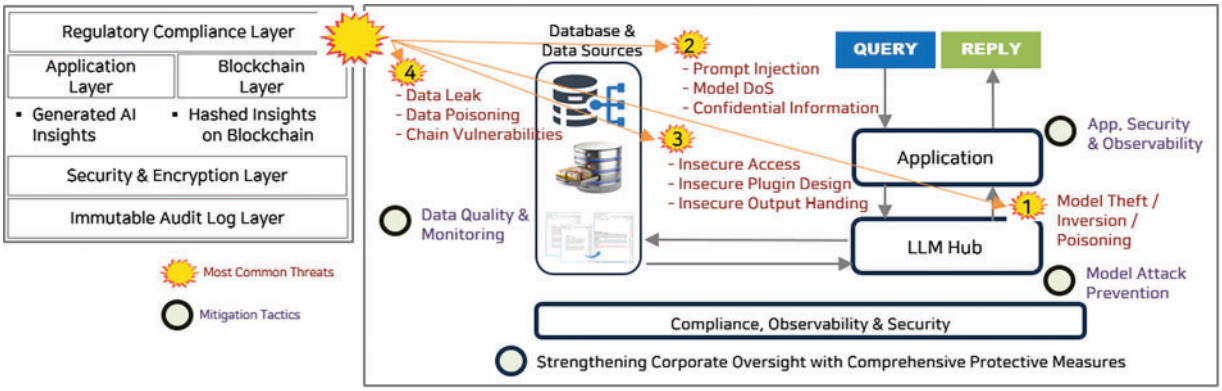

Figure 2: Blockchain-based medical contents security structure

The framework consists of a generative AI model layer, blockchain layer, security and encryption layer, application layer, and regulatory compliance layer. The generative AI model analyzes data, while the blockchain layer records the hash values. The security and encryption layer maintain data confidentiality, the application layer interacts with users, and the regulatory compliance layer ensures adherence to legal requirements. These components work together to enhance data security and trustworthiness. Below is a pseudocode that demonstrates how the medical data flows through the generative AI model, is hashed, and then stored on the blockchain. It also includes the validation process for data integrity:

INPUT	*Medical data* (e.g., *MRI scan, patient diagnosis information*)
STEP 1	*Pre-process the input data for the Generative AI model* – *Clean the data (remove noise, standardize format)* – *Normalize or augment the data if necessary*
STEP 2	*Feed the pre-processed data into the Generative AI model* – *Generate output* (e.g., *synthetic medical images or reports*)

(Continued)

(continued)	
INPUT	*Medical data (*e.g., *MRI scan, patient diagnosis information)*
STEP 3	*Hash the output using SHA-256* – *hashed_output = SHA256(generated_output)*
STEP 4	*Create a transaction on the blockchain* – *Transaction = { "data": hashed_output, "timestamp": current_time }* – *Send transaction to blockchain network*
STEP 5	*Store the transaction on the blockchain* – *Validate the block containing the transaction through consensus*
STEP 6	*Retrieve the data for verification* – *Retrieve the hash from the blockchain* – *Compare the current hash with the stored hash to ensure data integrity*
STEP 7	*If the hash matches, the data is valid and untampered* – *If the hash does not match, flag as potential tampering*

It replaces complex formulas and provides a clear logical structure that outlines how blockchain secures AI-generated medical data. It simplifies the explanation of how blockchain ensures data immutability, traceability, and integrity in medical applications.

The federated learning process is defined as (1), and in the generative AI layer, models such as GPT or BERT analyze health data, applying collaborative learning methods to protect personal information in the process.

$$min\ F\ (w) = \sum (i = 1\ to\ n)\ \left(p_i * F_{i(w)}\right) \tag{1}$$

Here, w represents the model parameters, n is the number of participating institutions, p_i is the data ratio of each institution, and F_i is the local objective function of each institution. The generated insights are securely encrypted through the *SHA-256* hash function, and the compression function of *SHA-256* is expressed as $h_i = f(h_{i-1}, m_i)$. In this expression, h_i is the i-th hash value, m_i is the message block, and f is a function that includes complex bit operations.

The same encryption is applied in both the Blockchain Layer and the Security & Encryption Layer. The basic operations of the fully homomorphic encryption scheme are defined as (2), where E is the encryption function, m_1 and m_2 are plaintexts, and c is a constant.

$$E\ (m_1) * E\ (m_2) = E\ (m_1 + m_2)\ E\ (m_1)^c = E\ (c * m_1) \tag{2}$$

In the Security & Encryption Layer, zero-knowledge proof technology, particularly zk-SNARKs, is utilized. The core concept of zk-SNARKs is expressed as a polynomial relationship as shown in (3). Here, $p(x)$ is the statement to be proven, $h(x)$ is the polynomial provided by the verifier, and $t(x)$ is the target polynomial.

$$p\ (x) * h\ (x) = t\ (x)\ mod\ \left(x^{n-}1\right) \tag{3}$$

Differential privacy is applied between the Application Layer and the Regulatory Compliance Layer, and the mathematical expression of ε-differential privacy is defined in (4). Here, A is the algorithm, D and D' are adjacent datasets differing by one record, S is a subset of possible outputs,

and ε is the privacy parameter.

$$Pr\left[A\left(D\right) \in S\right] \leq exp\left(\varepsilon\right) * Pr[\,A\left(D'\right) \in S] \tag{4}$$

In the Regulatory Compliance Layer, various metrics are used to evaluate the fairness of generative AI models. For example, demographic parity, which can be applied in this study, is defined as in (5). Here, Y is the model's prediction, A is the protected attribute (e.g., gender, race), and ε is the tolerance level.

$$|P\left(Y = 1|\,A = 0\right) - P\left(Y = 1|A = 1\right)| \leq \varepsilon \tag{5}$$

The Security & Encryption Layer introduces quantum-resistant encryption algorithms in anticipation of future quantum computing threats. This multi-layered structure, underpinned by robust mathematical foundations, simultaneously pursues the safety, reliability, and innovative utilization of digital healthcare information. Each layer employs advanced mathematical algorithms and encryption techniques to protect individual privacy while maximizing data value. The integration of blockchain and generative AI is achieved by having the generative AI model process medical data and produce outputs, which are then hashed using algorithms like SHA-256. The resulting hash values are stored on the blockchain, ensuring data integrity and immutability. Any subsequent changes to the data would result in a different hash, making it easy to detect tampering. Additionally, smart contracts can be used to automate access control and ensure that only authorized parties can view or modify the data.

3.2 Blockchain-Based Medical Contents Security Procedure

Table 3 defines the major security threats posed to existing generative AI models [16].

- **Poisoning:** By manipulating training data, attackers can distort the model's judgments, leading to false detections or errors, thereby undermining the model's reliability.
- **Prompt Injection:** A subtle technique involving crafted inputs to manipulate the model or bypass security filters. Attackers may induce intended outputs or circumvent security measures, potentially causing harm.
- **Plugin Vulnerability:** Exploiting the model's extended features may lead to malfunctions or API key theft. Plugin vulnerabilities compromise model security and may result in data breaches.
- **Data Leakage:** Exposure of training data, user information, or conversation records poses risks of personal information theft, identity theft, or revealing the internal workings of the model.

Table 3: Generative AI vulnerability definitions

Vulnerability	Vulnerability description
Poisoning	– Inducing erroneous decisions, phishing emails, identity theft, exploiting conversational services, URL squatting and extension programs, compromising the security, validity, or ethical behavior of LLM – Maliciously manipulating the training data or fine-tuning process of LLM to introduce backdoors or biases, resulting in inappropriate responses
Prompt injection	– Exploiting weaknesses in the tokenization or encoding mechanisms of LLM, or providing ambiguous contexts to induce unintended behaviors

(Continued)

Table 3 (continued)

Vulnerability	Vulnerability description
Plugin vulnerability	– Misbehaviors in new domains, exploiting 'Agentified' AI models, multimodal exploitation, API key theft, and injecting malicious prompts
Data leakage	– Training data leaks, data illegal processing concerns, confidentiality leaks, chat history leaks, database hacking, and member inference attacks

Given these vulnerabilities, it is essential to implement security measures and safeguards to ensure the safe and reliable use of generative AI models. This includes data validation, input filtering, and continuous monitoring of AI models to mitigate potential threats and enhance their security. The process for each processing procedure is shown in Fig. 3.

The proposed strategy in this study, outlined in Table 4, aims to address the vulnerabilities of generative AI models as defined in Table 3.

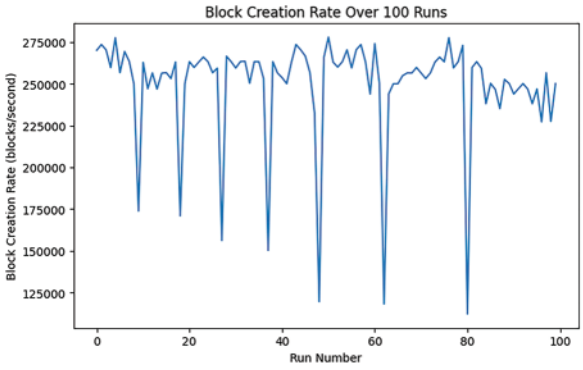

Figure 3: blockchain performance

Table 4: Generative AI vulnerability resolution

Vulnerability	Countermeasures
Poisoning response	– Data Distribution through Federated Learning: By distributing datasets instead of using a central data repository, the risk of large-scale data manipulation is reduced – Blockchain-based Data Validation: Merkle Tree structure enables detection of training data alterations – Persistent Bias Checking: Employ XAI and bias detection algorithms to assess model fairness and trigger retraining if needed

(Continued)

Table 4 (continued)

Vulnerability	Countermeasures
Prompt injection response	– Homomorphic Encryption: Enables data processing in encrypted form, blocking malicious prompt injection – SMPC Protocol Utilization: Secure multi-party computation protocols prevent the execution of unintended commands – Zero-Knowledge Proof Techniques: Minimizes attack surface by exposing minimal information
Plugin vulnerability response	– Attribute-Based Encryption: Fine-grained access controls limit plugin privileges – Immutable Audit Log: Records plugin activities on the blockchain to trace exploitation attempts – API Key Protection Reinforcement: Quantum-resistant encryption strengthens API key security
Data leak response	– Differential Privacy Application: Balances between personal information protection and data utilization – Homomorphic Encryption and SMPC Combination: Minimizes data exposure risk by processing encrypted data – DID Utilization: Enables secure and verifiable access to user data access and modification history
Overall security enhancement	– Blockchain Technology Adoption: Ensures data integrity and transparency – Formal Verification: Mathematically proves the safety of encryption protocols – Real-time Compliance Monitoring: Instantly detects and responds to violations

4 Results and Discussion

4.1 Experimental Setup

For the experiments in this study, a virtual environment was set up with python and necessary libraries. Using the medical research institute's dataset, experimental data was constructed, and malicious test data was added as shown in the Table 5.

A script was developed for model inference, fine-tuning, visual data processing and classification, and result analysis. Resource monitoring and logging systems were established to track CPU and memory usage. The model was fine-tuned using biased medical data to detect potential biases, and malicious prompts were crafted to analyze each security threat. Finally, accuracy, consistency evaluation criteria, and malicious input detection rate calculation methods were defined for a systematic assessment of the experimental results. In the experimental setup, this study utilized a dataset provided by a medical research institute and included malicious test data. The criteria for classifying data as malicious were based on attributes likely to interfere with the model's normal learning or compromise data integrity. Using these criteria, we selected malicious data to reliably evaluate the effectiveness of the proposed security strategy, demonstrating its capability to detect security threats caused by malicious inputs.

Table 5: Medical dataset composing

Classification	Contents
Examination information excluding personal information	– Patients' visual examination, biometric measurements, blood test results, physiological and biochemical information: These are vital indicators for assessing an individual's health status and detecting potential diseases. This data encompasses an individual's physical condition, functional processes, metabolic activities, and disease presence
Medical institution data policy	– Data collection, storage, sharing, security, and legal compliance: Comprehensive guidelines are necessary for the safe and efficient management of medical data. This includes rules, standards, and best practices for data collection, storage, and sharing – Large-scale health screening programs: Focuses on early disease detection and maintaining the health of the population. Includes guidelines on screening procedures, targets, periods, costs, result interpretation, and post-screening treatments
Health examination policy	– Personal information protection, legal compliance, and ethical considerations: Emphasizes the importance of adhering to ethical and legal aspects in handling medical data – Includes guidelines on data collection, usage, and sharing to prevent ethical breaches and legal disputes

4.2 Behavioral Analysis of Models for Security Threat Scenarios

In this study, we performed distributed learning using data from each client through federated learning simulation. Each client trained a logistic regression model using 20% of the given data, and finally, a global model was created by averaging the weights of the client models. The accuracy of the global model created in this way was 97.48%, demonstrating the effectiveness of federated learning. This result shows that federated learning can generate a high-performance prediction model by integrating individual client models (Table 6).

Table 6: Accuracy for federated learning

Name of Dataset	Accuracy for the Dataset
medical_dataset	97.48 ± 0.58

The SHA-256 hash function was used to protect sensitive information, and in this study, we used the SHA-256 hash function to verify the integrity of medical data and protect sensitive information. As a result of the experiment, a hash value "e13e74115bef02880c5a3f26f0859eb20d3955579ab2010a422f586e62de585f" was generated, which can be used to ensure the integrity of medical data and securely protect sensitive information. Additionally, Fernet symmetric encryption was applied for homomorphic encryption simulation. Fernet encryption was used to encrypt two values $m1 = 5$ and $m2 = 3$, yielding the following results:

Encrypted m1:	b'gAAAAABmqGlA3pkDl9n1akLz1wVZnijDo6knZciTWZdyu8HVCYVmKXP QVy4if_gLXOE43q0CnBer1rG8wrTh5MVmC4rrcG9jSg=='
Encrypted m2:	b'gAAAAABmqGlAFxHRRp4rM0hDDKHdGOiftzps1ig0fvIFhb72BWEhTZfz 7iJ5j8zmzJ32mS26d_pNu8r5r-cFsav17RAY-ZLnDQ=='
Encrypted sum:	b'gAAAAABmqGlAh4VJibqNUCNzbZ6VKtBPQVfJs-yl-Myky5AVMiNJRIUywPbShfWnOOL3c9Z5J2uXHuJIr3dvsdgSb8u5fc1VXg=='

Homomorphic encryption is a powerful technology that ensures the privacy of sensitive data. This technology allows computations to be performed while keeping data encrypted. In this study, we additionally applied differential privacy techniques to maintain a balance between data accuracy and individual privacy. Differential privacy adds random noise to the data, protecting sensitive information while maintaining the accuracy needed for analysis and learning.

Before applying:	[0.98514109 0.84842347 0.77179218 0.93207656 0.52678767 0.30318777 0.3676016 0.50643823 0.50027144 0.94681514]
After applying:	[5.68639319 11.74057992 9.73563991 33.22394941 13.54628233 -30.09165387 5.38703059 9.54388904 13.64364699 -1.96126327]

In this study, we applied demographic parity to evaluate the fairness of the model. The average difference in prediction probabilities between protected attribute groups was within a threshold value of 0.05, demonstrating the fairness of the model. This means that the model performs predictions without discrimination based on protected attributes.

Examining the experimental results for blockchain performance, node scalability, and transaction throughput of the proposed configuration model, all aspects showed excellent results. The block generation speed could produce 249,689.42 blocks per second, reaching an average of 5500 nodes, demonstrating the network's scalability and decentralization. The transaction throughput was 140,577.96 per second, showing the ability to efficiently process large-scale transactions (Fig. 4).

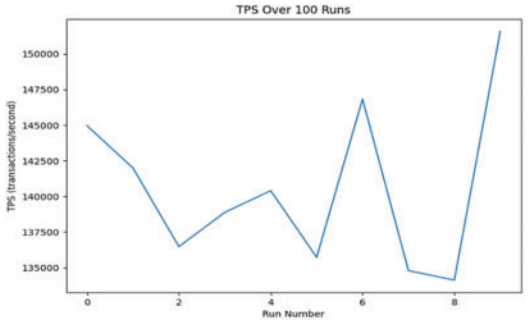

Figure 4: Result of TPS

This model demonstrated outstanding results in terms of blockchain performance, node scalability, and transaction throughput. According to the performance test results summarized in Table 7,

the ability to generate 249,689.42 blocks per second surpasses existing blockchain systems, indicating that the proposed model can process and verify data rapidly, proving its suitability for real-time applications. Moreover, reaching an average of 5500 nodes in terms of node scalability demonstrates the model's excellent scalability. This suggests that stable operation is possible without performance degradation even as more nodes join the network. The transaction throughput of 140,577.96 per second proves the ability to efficiently process large-scale transactions. These excellent performance indicators show that the proposed model has overcome the limitations of existing blockchain systems. However, additional testing under various network conditions is necessary to increase the reliability of the results. In conclusion, the proposed model has significantly expanded the performance limits of blockchain technology, broadening the practical application range of blockchain technology and indicating its potential for use in various industries. Future testing in more diverse experimental environments is expected to further demonstrate the excellent performance and scalability of the proposed model. Generative AI shows great promise for creating medical content, but it's important to recognize its key limitations, especially regarding accuracy and reliability. One significant challenge is the quality of training data. If an AI learns from biased or incomplete information, it might produce inaccurate or misleading content. In medical settings, this could potentially lead to incorrect diagnoses or treatment suggestions. Another concern is the lack of real-world testing. While these AI models may perform well in controlled settings, their effectiveness in actual clinical environments can vary considerably. This inconsistency raises questions about how reliable AI-generated medical information really is. There's also a risk that AI might create synthetic data that misses subtle but crucial medical details, which could lead to misunderstandings among healthcare professionals. To address these issues, we can consider a few approaches. First, it's crucial to gather high-quality, diverse datasets to train the AI, helping it better understand a wide range of medical scenarios. Second, we should continuously test and validate AI models in real-world medical settings to ensure their predictions remain accurate and dependable. Lastly, combining human expertise with AI-generated content can help reduce errors. This allows medical professionals to review and confirm the AI's output before making important clinical decisions.

Table 7: Performance measurement results

Classification	Result
The number of blocks per second	249,689.42
Average number of nodes reached	5500.00
Transactions per second	140,577.96

To address the scalability challenges and potential complexity of implementing blockchain in a medical setting, we conducted several performance evaluations. The system demonstrated an impressive average transaction throughput of 140,577.96 transactions per second (TPS), highlighting its capacity to efficiently manage large-scale medical data transactions. This throughput ensures that the blockchain can support the real-time data processing requirements of medical institutions. Furthermore, the block generation speed was measured at 249,689.42 blocks per second, significantly surpassing the performance of traditional blockchain systems. This result indicates that the system can handle the high transaction volume typically encountered in medical environments without compromising speed or efficiency. We further tested the scalability of the system with an average of 5500 active nodes participating in the network. Notably, the system maintained stable performance as the number of nodes increased, demonstrating its robustness in accommodating a growing number of medical

institutions and contributors. This scalability ensures that the blockchain framework can effectively support the increasing volume of medical data in a decentralized manner. However, it's important to note that the integration of blockchain also introduces additional complexity, particularly with respect to node management and maintaining consensus across a distributed network. This complexity can potentially increase operational overhead in real-world settings. To mitigate this issue, we implemented sidechains and smart contracts, which help to offload some processing tasks from the main blockchain, thereby reducing latency and improving overall system performance. These results suggest that while blockchain can scale to meet the demands of a medical setting, the complexity it introduces must be managed carefully. Further optimizations, especially in node management and consensus algorithms, will be critical for ensuring the long-term viability of this technology in the healthcare domain.

4.3 Evaluating Model Security against Malicious Inputs

In this study, we propose a systematic strategy to assess and enhance the security of artificial intelligence models. By employing malicious data inputs in model robustness tests, potential security threats were identified and analyzed using evaluation indicators such as security threat detection rate, accuracy, and error rate. The results summarized in Table 8 offer insights into the vulnerabilities of generative AI models, providing valuable information for further improvements.

Table 8: Model security assessment against malicious inputs

Accuracy	Detection rate	Error rate
96.00 ± 2.83	80.55 ± 4.35	19.44 ± 4.35

Table 8 presents crucial indicators for evaluating the performance and security of generative AI models: accuracy, detection rate, and error rate. Accuracy, at 96%, reflects the model's ability to make correct classifications, indicating overall performance. The detection rate, at 80.55%, signifies the model's effectiveness in correctly identifying actual threats, holding significance in security contexts. Meanwhile, the error rate, at 19.44%, represents the proportion of misclassified instances, serving as a complementary indicator to accuracy (Fig. 5).

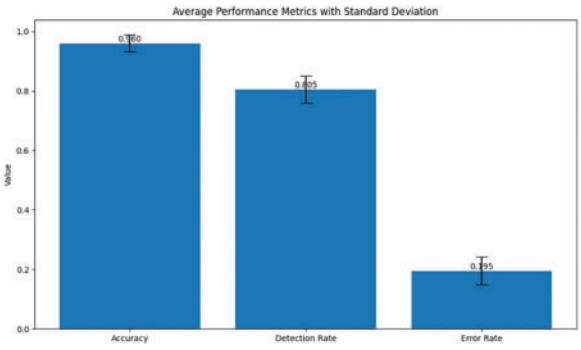

Figure 5: Average performance metrics with standard deviation

By analyzing the interrelation between Accuracy, Detection Rate, and Error Rate, we gain valuable insights into the model's overall performance and security capabilities. Accuracy, with an average of 0.96, demonstrates the model's high overall performance, and the small standard deviation of 0.0283 indicates consistent results across experiments (Fig. 6). Even at its lowest value of 0.91, the

model maintains high accuracy, attesting to its stability. Detection Rate, averaging around 80.56%, is an important metric, indicating the model's effectiveness in correctly identifying positive cases, but it's lower compared to Accuracy, suggesting the model may be missing some true positives. The higher standard deviation of Detection Rate compared to Accuracy implies greater variability in experiment outcomes and the potential for missing certain threats. Error Rate, averaging around 19.44%, complements Accuracy, indicating the proportion of misclassified instances. The near-identical standard deviations of Error Rate and Detection Rate highlight their interdependence. Overall, the high Accuracy and low Error Rate signify strong performance, but the lower Detection Rate warrants attention. This discrepancy suggests the model may be missing some positive cases, making it a crucial consideration in security applications. The standard deviation considerations reveal that while Accuracy remains relatively stable, Detection Rate and Error Rate exhibit some variability with specific input types, implying the model's predictions may vary. The model shows a promising average accuracy of 0.96, but the standard deviation hints at some variability in its performance across different types of medical images and conditions. This fluctuation can be attributed to several factors, with image complexity being a key player. For instance, images with overlapping anatomical structures or subtle differences between healthy and diseased areas can make accurate detection more challenging, leading to inconsistent results. Some medical conditions are trickier to detect reliably, especially those with less obvious visual cues like early-stage cancers or diseases with diffuse patterns. On the flip side, more noticeable abnormalities such as large tumors or clear fractures tend to yield consistently high accuracy. To tackle this variability, it's crucial to take a closer look at how the model performs with different types of medical images and conditions. This might involve grouping the data based on image or condition complexity and pinpointing where the model excels or struggles. Enhancing the model with data augmentation techniques or additional training on complex cases could help boost its consistency. In this study, we encountered some instances where the model missed important details, like failing to detect small or early-stage tumors. These false negatives can happen due to low image quality or subtle variations in the visual data. Such oversights are particularly concerning in medical diagnostics, as they could lead to delayed or incorrect treatment. To address these challenges, the study incorporated blockchain technology. This approach helps maintain the integrity of the training data throughout the learning process, reducing the risk of using tampered or corrupted information that could lead to missed detections. The use of distributed ledger technology also ensures that no unauthorized changes have been made to the dataset, further minimizing the chances of false negatives caused by compromised data integrity.

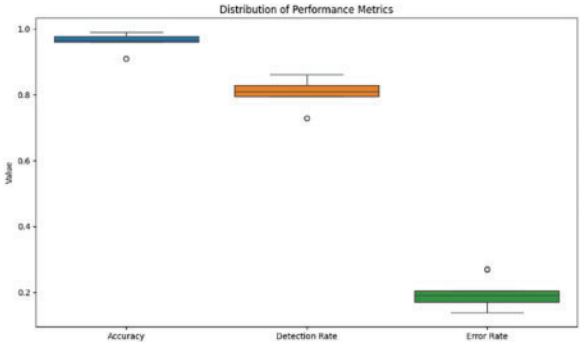

Figure 6: Distribution of performance metrics

One of the key challenges in medical content analysis is the occurrence of false negatives, where critical anomalies such as early-stage diseases or subtle lesions are missed by the AI model. For

instance, in early cancer detection, subtle visual cues or overlapping anatomical structures can make it difficult for the model to accurately detect the abnormality, leading to false negatives. These errors can have serious implications for patient diagnosis and treatment, highlighting the need for a more robust approach to mitigate such risks. In this study, we propose the integration of blockchain technology as a solution to reduce the occurrence of false negatives. Blockchain's decentralized ledger system ensures the integrity and security of the training and diagnostic data used by AI models. By preventing tampering or alteration of the data, blockchain helps ensure that the AI model learns from accurate and unaltered datasets, thus reducing the likelihood of false negatives caused by compromised data. Specifically, blockchain securely records every piece of data used in the model, ensuring that no unauthorized modifications can occur. This transparency and immutability are key to maintaining the reliability of the AI model's outputs. Moreover, the inherent transparency and traceability of blockchain further enhance the trustworthiness of medical data. Any potential data corruption or modification can be quickly detected, allowing immediate corrective actions. This capability is crucial in preventing false negatives that might arise from compromised or incomplete data. By utilizing blockchain, we not only protect the integrity of the data but also ensure that the AI model consistently produces accurate diagnostic results. In conclusion, the integration of blockchain technology significantly contributes to improving the performance of AI models in medical applications by reducing false negatives and enhancing data integrity.

4.4 Systematic Evaluation of Security Framework

Table 9 shows the systematic evaluation of security framework. Blockchain technology offers a robust solution for ensuring the integrity and transparency of medical data generated by AI models. By recording hash values of this content on a distributed ledger, it creates a tamper-proof record that allows for easy tracking of data origins and changes. This decentralized approach provides a significant advantage over traditional centralized database security methods. The process involves converting the AI-generated medical content into hash values using algorithms like SHA-256, which are then stored on the blockchain. This method effectively guarantees that the data remains unaltered during transmission and storage, significantly enhancing the trustworthiness of medical information and allowing for quick identification of any integrity issues. When we examine the model's performance metrics, we gain valuable insights into its overall effectiveness and security. The model demonstrates high accuracy with an average of 0.96 and a small standard deviation of 0.0283, indicating consistent performance across various experiments. Even at its lowest, the accuracy remains impressive at 0.91, showcasing the model's stability. The Detection Rate, averaging around 80.56%, is particularly important in security contexts as it reflects the model's ability to correctly identify actual threats. While this rate is lower than the overall Accuracy, it still indicates strong performance. However, the higher standard deviation compared to Accuracy suggests more variability in threat detection across different scenarios. The Error Rate, averaging about 19.44%, complements the Accuracy metric by showing the proportion of misclassified instances. The similar standard deviations of Error Rate and Detection Rate highlight their interconnected nature. In summary, the high Accuracy and low Error Rate demonstrate the model's strong overall performance. However, the slightly lower Detection Rate compared to Accuracy suggests that the model might be missing some positive cases. This discrepancy is an important consideration, especially in security-critical applications, and may warrant further investigation and refinement of the model.

Table 9: Systematic evaluation of security framework

Classification	Threat	Current assessed risk	Actual occurrence frequency	Effectiveness of response measures	Areas requiring further research
AI model vulnerabilities	Adversarial examples	H	Attempts to evade malicious input detection models	Defended against most attacks	Advanced adversarial techniques, defense strategies
	Model extraction attacks	H	Vulnerable to model extraction attacks	Mitigated by secure deployment	Secure model deployment, access control enhancements
	Data injection attacks	H	Potential for data manipulation	Prevented by data validation	Data validation and sanitization techniques
	Model vulnerabilities exploited	M	Potential vulnerabilities exist	Essential to identify vulnerabilities	Vulnerability identification and patching
Privacy	Data re-identification	L	Use of non-anonymized data	Protected by anonymization	Anonymization techniques, privacy-preserving methods
	Inference attacks	L	Potential vulnerabilities exist	Potential vulnerabilities exist	Model robustness, input perturbation defenses
	Data leakage	H	Data accessible by external attackers	Minimized by encryption and access controls	Data encryption, access control improvements
	Unauthorized data use	M	Lack of access control	Ensured by access controls	Access control mechanisms, authentication enhancements
	Dangers of data sharing	H	Exposed to potential security threats	Mitigated by secure sharing protocols	Secure data sharing protocols, encryption methods
Security vulnerabilities	API vulnerabilities	H	Potential API vulnerabilities exist	Identified by security assessments	API security assessments, penetration testing
	Network security	H	Potential network security vulnerabilities	Enhanced by segmentation and firewalls	Network segmentation, firewall configurations
	Authentication and access control vulnerabilities	M	Potential vulnerabilities exist	Mitigated by multi-factor authentication	Multi-factor authentication, secure password policies
	Server Vulnerabilities	M	Potential server vulnerabilities exist	Strengthened by hardening and updates	Server hardening techniques, patch management
	Client-side vulnerabilities	H	Potential client-side vulnerabilities	Reduced by secure client-side code	Secure client-side code, input validation
	Encryption related vulnerabilities	H	Potential encryption-related vulnerabilities	Mitigated by key management practices	Key management practices, encryption protocol updates

The model demonstrated a promising average accuracy of 0.96. However, the standard deviation of 0.0283 indicates some variability in its performance across different types of medical images and conditions. Our in-depth analysis revealed several key factors contributing to this variability:

– Complexity of Medical Images: Medical images featuring overlapping anatomical structures or subtle distinctions between healthy and diseased areas tend to introduce greater variability in

the model's performance. For instance, detecting early-stage cancer or conditions with diffuse patterns proves more challenging, potentially leading to slight inconsistencies in accuracy.
- Image Quality and Resolution: Lower-resolution images or those with artifacts also played a role in performance variability. These factors can obscure crucial diagnostic features, making it more challenging for the model to consistently identify abnormalities.
- Type of Medical Condition: The model's performance fluctuated depending on the specific medical condition under analysis. More apparent abnormalities, such as large tumors or fractures, yielded higher and more consistent accuracy. In contrast, conditions with less obvious visual cues, like inflammatory diseases or early-stage degenerative changes, showed slightly more variability in detection accuracy.

To address these challenges and enhance the model's consistency, we propose two main strategies:

- Implement additional data augmentation techniques to further stabilize the model's performance across varying image qualities and conditions.
- Expand the training dataset to include a more diverse set of images representing different levels of complexity. This approach could help reduce the standard deviation, thereby ensuring more consistent accuracy across a broader spectrum of medical data.

By implementing these strategies, we aim to refine the model's ability to handle diverse medical imaging scenarios, ultimately improving its reliability and applicability in clinical settings.

4.5 Comparative Analysis

The Table 10 below compares the proposed blockchain-based security model with traditional security models (centralized database encryption and access control). Both models aim to protect medical data, but they differ significantly in the technologies and approaches used, as well as the resulting advantages and disadvantages. The blockchain-based model offers superior security and integrity due to its decentralized structure, but it also introduces new challenges in terms of implementation complexity and scalability.

This Table 10 outlines the key differences between traditional centralized security models and blockchain-based security models. While traditional models rely on centralized management and are vulnerable to insider threats, blockchain-based models enhance data integrity through decentralized, tamper-proof storage. However, the blockchain model faces challenges in terms of scalability and operational costs. This comparison highlights both the advantages of using blockchain for securing medical data and the potential issues that need to be addressed during implementation. This study demonstrates that blockchain-based security safeguards medical data using generative AI models. However, since our experiments were limited to specific model types, further research is needed to explore this solution's broader applications. Given the diverse landscape of generative AI–including GANs, Transformers, and Diffusion models—each with its unique architectures and security challenges, it's crucial to validate our approach across different model types. This extended research would not only assess the versatility of our security strategy but also pave the way for customized security solutions tailored to each model's specific needs.

Table 10: Comparative table

Item	Traditional security model	Blockchain-based security model
Data storage method	Centralized servers or databases	Decentralized distributed ledger (Blockchain)
Single point of failure	Yes (Entire system failure if the server is down)	No (Operates on a distributed network)
Data integrity	Vulnerable to modification by administrators or insiders	Protected by hashing, any modification is recorded
Access control	Traditional authentication and authorization mechanisms	Automated access control via smart contracts
Scalability	Relatively easy to scale	Requires node expansion, may result in slower speeds
Security vulnerabilities	Susceptible to insider attacks and database hacking	Strong against hacking, but key management is crucial
Data transparency	Depends on the administrator	Transparent to all participants in the network
Encryption method	Server-side encryption	Built-in blockchain encryption (e.g., SHA-256)
Operational costs	Centralized management costs	Higher costs for maintaining nodes and blockchain

The proposed blockchain-based security framework exhibits significant advantages over conventional centralized database security architectures, particularly in three critical aspects: data integrity assurance, elimination of single-point vulnerabilities, and enhanced automation of access control through smart contract implementation. Through the utilization of advanced cryptographic hashing algorithms, specifically SHA-256, coupled with distributed ledger technology, the framework substantially mitigates the risk of unauthorized data manipulation, including potential insider threats from system administrators. This robust security architecture proves especially valuable in safeguarding highly sensitive information, such as medical records, representing a notable advancement over traditional security paradigms.

5 Conclusion

In this study, we explored the security aspects of medical content and diagnostic visualization using a blockchain-based generative AI model. Our comprehensive examination covered vulnerabilities in medical content and diagnostic records, personal information protection issues, and security weaknesses in medical AI systems, aiming to enhance overall medical data security. By analyzing accuracy, detection rate, and error rate, we gained valuable insights into the model's performance and limitations. The high accuracy and low error rate demonstrated strong overall performance. However, the lower detection rate compared to accuracy suggested that the model might be missing some positive cases, a crucial consideration for security applications. Our analysis of standard deviations revealed that accuracy remained relatively stable, while detection and error rates showed some variability with specific input types. This implied that the model's predictions might vary under certain conditions. To address this, we proposed improving the detection rate by reducing false negatives and enhancing

the model's generalization through diverse threat scenarios in training data. Blockchain technology emerged as a powerful tool to further enhance data integrity and security, ensuring transparent and tamper-proof record-keeping. Additionally, generative AI models showed promise in creating novel medical content while excluding personal information, thus safeguarding patient privacy. However, we acknowledge certain limitations in our study, including a restricted dataset size, specific model architecture, and the exclusion of real-world attack scenarios. Future research incorporating diverse datasets, alternative model architectures, and blockchain-based medical data sharing platforms will likely further enhance the model's performance and security. While blockchain technology offers significant security advantages, implementing it in a medical environment presents unique challenges. Scalability is a primary concern, as blockchain's decentralized nature can lead to slower transaction processing compared to centralized systems. To mitigate this, we suggest exploring technologies like sidechains or hybrid blockchains to offload some processing burden from the main blockchain. Increased complexity is another challenge, as integrating blockchain into existing medical systems may introduce interoperability issues with legacy systems, compliance concerns, and the need for user training. Developing user-friendly interfaces and leveraging automated smart contracts could simplify system operations and ease adoption for medical professionals. Regulatory and legal challenges also require attention. Ensuring medical data security and privacy demands compliance with international regulations like GDPR or HIPAA. We propose integrating a Regulatory Compliance Layer into the blockchain framework to log all activities and ensure adherence to these legal requirements. Long-term stability is crucial as the system grows. To handle increasing amounts of medical data without performance degradation, we recommend implementing scalability solutions like sharing, dividing the blockchain network into smaller, more efficient parts. Sidechains can also help maintain speed and performance by offloading transactions from the main blockchain. Regular node maintenance and network monitoring are essential to prevent bottlenecks and ensure consistent operation. Looking to the future, the potential impact of quantum computing on existing encryption algorithms cannot be ignored. To stay ahead of this threat, integrating quantum-resistant encryption algorithms into our framework is necessary. This approach can strengthen the protection of API keys and data integrity verification procedures, ensuring the long-term security of our system.

In conclusion, while our blockchain-based generative AI model for medical content security analysis shows great potential in mitigating security vulnerabilities, striking a balance between detection and false negative rates, along with enhancing the model's robustness against diverse threat scenarios, remains essential. We recommend continuous research, model refinement, and vigilant security threat monitoring to fortify the security of medical content AI services in this rapidly evolving technological landscape.

Acknowledgement: This work was supported by the Sun Moon University Research Grant of 2024.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design, analysis and interpretation of results: Hoon Ko, draft manuscript preparation and proved the final version of the manuscript: Marek R. Ogiela. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] X. Li, D. Pan, and D. Zhu, "Defending against adversarial attacks on medical imaging AI system, classification or detection?" in *2021 IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Nice, France, 2021, pp. 1677–1681. doi: 10.1109/ISBI48211.2021.9433761.

[2] R. Sivan and Z. A. Zukarnain, "Security and privacy in cloud-based e-health system," *Symmetry*, vol. 13, no. 5, 2021, Art. no. 742. doi: 10.3390/sym13050742.

[3] M. J. Sheller *et al.*, "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Sci. Rep.*, vol. 10, no. 1, 2020, Art. no. 12598. doi: 10.1038/s41598-020-69250-1.

[4] N. Spence, M. Niharika Bhardwaj, and D. P. Paul III, "Ransomware in healthcare facilities: A harbinger of the future?" in *Perspectives in Health Information Management*. American Health Information Management Association, 2018, pp. 1–22.

[5] D. Portela, D. Nogueira-Leite, R. Almeida, and R. Cruz-Correia, "Economic impact of a hospital cyberattack in a national health system: Descriptive case study," *JMIR Form. Res*, vol. 7, no. 1, 2023, Art. no. e41738. doi: 10.2196/41738.

[6] J. Pool *et al.*, "A systematic analysis of failures in protecting personal health data: A scoping review," *Int. J. Inf. Manag.*, vol. 74, 2024, Art. no. 102719. doi: 10.1016/j.ijinfomgt.2023.102719.

[7] M. Chernyshev, S. Zeadally, and Z. Baig, "Healthcare data breaches: Implications for digital forensic readiness," *J. Med. Syst.*, vol. 43, pp. 1–12, 2019. doi: 10.1007/s10916-018-1123-2.

[8] P. A. H. Williams and A. J. Woodward, "Cybersecurity vulnerabilities in medical devices: A complex environment and multifaceted problem," *Med. Devi.: Eviden. Res.*, pp. 305–316, 2015. doi: 10.2147/MDER.

[9] A. F. Al-Qahtani and S. Cresci, "The COVID-19 scamdemic: A survey of phishing attacks and their countermeasures during COVID-19," *IET Inf. Secur.*, vol. 16, no. 5, pp. 324–345, 2022. doi: 10.1049/ise2.12073.

[10] W. Priestman, T. Anstis, I. G. Sebire, S. Sridharan, and N. J. Sebire, "Phishing in healthcare organisations: Threats, mitigation and approaches," *BMJ Heal. Care Inform.*, vol. 26, no. 1, 2019, Art. no. e100031.

[11] A. Ahmed, R. Latif, S. Latif, H. Abbas, and F. A. Khan, "Malicious insiders attack in IoT based multi-cloud e-healthcare environment: A systematic literature review," *Multimed. Tools Appl.*, vol. 77, no. 17, pp. 21947–21965, 2018. doi: 10.1007/s11042-017-5540-x.

[12] T. Ermakova, J. Huenges, K. Erek, and R. Zarnekow, "Cloud computing in healthcare–A literature review on current state of research," in *Proc. Nineteenth Americas Conf. Inform. Syst.*, Chicago, IL, USA, Aug. 14–17, 2013.

[13] S. K. Sharma, B. Bhushan, and N. C. Debnath, *Security and Privacy Issues in IoT Devices and Sensor Networks*. Academic Press, 2020.

[14] J. J. Hathaliya and S. Tanwar, "An exhaustive survey on security and privacy issues in Healthcare 4.0," *Comput. Commun.*, vol. 153, no. 6, pp. 311–335, 2020. doi: 10.1016/j.comcom.2020.02.018.

[15] OpenAI, "ChatGPT: Security strategy of digital medical contents," 2024. Accessed: Sep. 11, 2024. [Online]. Available: https://openai.com/blog/chatgpt/

[16] L. Ogiela, M. R. Ogiela, and H. Ko, "Intelligent data management and security in cloud computing," *Sensors*, vol. 20, no. 12, pp. 1–11, 2020, Art. no. 3458. doi: 10.3390/s20123458.