World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Review Article)

Check for updates

Secure AI Pipelines for Drug Repurposing: A Cybersecurity Approach to Biomedical Innovation

Rama Devi Drakshpalli *

Independent Researcher, Cary, North Carolina, USA.

World Journal of Advanced Engineering Technology and Sciences, 2025, 16(02), 278-289

Publication history: Received on 10 July 2025; revised on 17 August 2025; accepted on 19 August 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.16.2.1142

Abstract

AI-driven drug repurposing has emerged as a transformative approach in pharmaceutical research, enabling the discovery of new therapeutic applications for FDA-approved drugs. This significantly reduces research and development (R&D) timelines and associated costs. However, the increasing reliance on AI-generated insights introduces vulnerabilities, making drug repurposing platforms susceptible to cyberattacks. These attacks can manipulate AI models to produce inaccurate drug predictions, potentially compromising clinical trial outcomes and patient safety.

This article provides a comprehensive examination of cybersecurity risks associated with AI-powered drug repurposing pipelines and presents a robust AI security framework to mitigate these threats. Key threats include model inversion attacks, where adversaries exploit AI models to infer sensitive drug trial data, and poisoning attacks, where malicious datasets distort AI-generated drug repurposing predictions.

To address these challenges, the paper proposes a multi-layered security strategy that incorporates homomorphic encryption for confidential AI-driven data processing, blockchain technology for immutable research records, and federated learning for secure cross-institutional AI model training. These technologies ensure that AI-driven drug repurposing insights remain protected, transparent, and verifiable.

By integrating secure AI pipelines into drug repurposing research, pharmaceutical companies can enhance the reliability of drug discovery, protect intellectual property, and comply with evolving cybersecurity mandates. This framework also reinforces the U.S.'s competitive edge in global biopharmaceutical innovation, ensuring AI-driven drug discovery remains both secure and efficient.

Keywords: Drug Repurposing; Artificial Intelligence (AI); Cybersecurity in Pharmaceuticals; Federated Learning; Homomorphic Encryption; Blockchain in Healthcare; Adversarial Machine Learning; AI Model Security; Biomedical Data Privacy; Secure AI Pipelines

1. Introduction

Drug repurposing involves identifying new therapeutic applications for existing drugs, offering a faster and cost-effective alternative to traditional drug development. AI algorithms, particularly deep learning and machine learning models, have significantly advanced this process by rapidly analyzing vast biomedical datasets to uncover promising drug-disease correlations [5], [8]. AI-based drug repurposing leverages computational techniques such as neural networks, natural language processing, and molecular docking simulations to systematically assess the efficacy of existing drugs against new therapeutic targets.

* Corresponding author: Rama Devi Drakshpalli. Emial: ramadevi.1412@gmail.com

One of the key advantages of AI in drug repurposing is its ability to process vast amounts of clinical, genomic, and molecular data to generate insights with a speed and accuracy unattainable by traditional methods. By integrating AI, researchers can rapidly evaluate existing drugs for potential new uses, reducing the time and cost associated with conventional drug discovery, which often takes more than a decade and billions of dollars in investment [43], [44]. Recent developments show that AI is also being leveraged to uncover novel uses for existing compounds in rare disease treatment scenarios, highlighting its transformative reach across niche therapeutic areas [44].

Despite its potential, AI-driven drug repurposing introduces significant cybersecurity risks. AI models are trained on vast datasets, some of which contain sensitive biomedical and clinical trial information. If compromised, these models could be manipulated to produce inaccurate predictions, posing significant public health risks [1], [9]. Additionally, AI models can be targeted by adversarial attacks, such as model inversion attacks that extract sensitive trial data or poisoning attacks that corrupt training data to mislead drug discovery outcomes [2], [18].

Therefore, securing AI-powered drug repurposing pipelines is imperative to protect research integrity, ensure regulatory compliance, and maintain trust in AI-driven pharmaceutical innovations. A comprehensive cybersecurity framework is necessary to mitigate threats and enhance the reliability of AI-generated drug repurposing insights. AI systems become integral to biomedical research, their security vulnerabilities present serious challenges that could impact drug discovery integrity [19], [25].

1.1. Problem Statement

The use of AI in drug repurposing presents significant security risks that can compromise the integrity and reliability of biomedical research. AI models trained on vast biomedical datasets are susceptible to adversarial attacks that can manipulate predictions, leading to erroneous conclusions and potentially unsafe drug applications [9], [18]. Cyberattacks, such as data poisoning, model inversion, and adversarial perturbation, can corrupt the AI models, skewing research outcomes and threatening patient safety. Additionally, AI-driven pharmaceutical research deals with highly confidential data, including proprietary molecular structures, clinical trial data, and patient records. Unauthorized access or breaches in these systems could result in loss of intellectual property and regulatory violations [24], [21]. Given the potential consequences of these vulnerabilities, securing AI-driven drug repurposing pipelines becomes crucial to maintaining scientific integrity, compliance with cybersecurity mandates, and trust in AI-enabled pharmaceutical advancements. AI-powered drug repurposing is increasingly susceptible to cyber threats. Cyberattacks, including data breaches and adversarial manipulations, could compromise research data, leading to inaccurate predictions and erroneous clinical trial results. These vulnerabilities threaten not only the reliability of AI-driven insights but also intellectual property security and regulatory compliance in the pharmaceutical industry [20], [25].

1.2. Objectives

This study aims to identify and categorize the cybersecurity threats specific to AI-driven drug repurposing pipelines, including data integrity risks, adversarial attacks, and privacy vulnerabilities. It proposes a multi-layered security framework that incorporates homomorphic encryption, blockchain, and federated learning to safeguard AI models against cyber threats [3], [4], [17]. The paper also evaluates the effectiveness of the proposed security solutions through empirical analysis, assessing their impact on model performance, data privacy, and overall research integrity. Moreover, it ensures that AI-driven drug repurposing pipelines adhere to regulatory requirements and industry best practices, facilitate secure and compliant pharmaceutical research [13], [14], [30]. The research provides a roadmap for future AI security enhancements, addressing potential emerging threats in AI-powered biomedical innovation. In doing so, it seeks to develop a secure AI pipeline framework integrating encryption, blockchain, and federated learning while assessing the impact of these methodologies on drug repurposing efficiency, reliability, and regulatory compliance.

1.3. Previous Work/Gaps

The intersection of AI and cybersecurity in drug repurposing has received limited attention in existing literature. While research has extensively explored AI techniques for drug discovery, there is a lack of comprehensive studies addressing the security vulnerabilities inherent in these AI models. Prior studies have focused on general AI security measures, but they fail to consider the specific threats faced by biomedical AI applications, such as targeted poisoning of pharmaceutical datasets and adversarial attacks on drug efficacy predictions [1], [10]. Additionally, while cryptographic techniques and decentralized systems like blockchain have been proposed in other fields, their integration into drug repurposing frameworks remains underexplored [4], [17], [22]. This study aims to bridge this critical gap by providing a tailored cybersecurity framework designed to protect AI-powered drug repurposing models from manipulation, ensuring data integrity, transparency, and robustness against emerging cyber threats, while extensive research has explored AI applications in drug repurposing, limited work has focused on cybersecurity challenges in this domain. Most existing AI security frameworks are designed for general applications and do not specifically address the unique vulnerabilities of AI-powered drug discovery pipelines [16], [23]. This study aims to bridge this gap by presenting a specialized security framework tailored to pharmaceutical AI applications.

2. Literature Review

Several studies have examined the integration of AI in drug discovery and repurposing, particularly in the use of machine learning algorithms for molecule prediction, target interaction, and disease classification. Abuhamad et al. [1] provide a comprehensive overview of adversarial machine learning techniques that exploit vulnerabilities in healthcare-related AI systems. Shokri et al. [2] specifically highlight membership inference attacks that can reveal the presence of specific patient records in model training data, demonstrating the risk of privacy breaches in AI pipelines. Gentry's foundational work [6] on homomorphic encryption offers a solution by enabling computations on encrypted data, protecting sensitive information from exposure during model training. Blockchain technologies, discussed by Casino et al. [4], offer immutability and traceability for data transactions, making them ideal for ensuring auditability in research. Yang et al. [5] and Rieke et al. [8] explore federated learning frameworks that support secure collaboration without data centralization. Furthermore, recent architectural diagrams and models reinforce the need for integrating homomorphic encryption for database-level privacy, blockchain for immutable audit trails, and federated learning for distributed training and encrypted model updates [16].

Moreover, companies like Benevolent AI and Insilico Medicine have successfully leveraged AI for drug repurposing, identifying new therapeutic uses for existing drugs through advanced machine learning and bioinformatics platforms. These approaches highlight AI's growing influence in accelerating discovery timelines and improving target validation. However, the reliance on large, sensitive datasets intensifies the need for robust cybersecurity measures to prevent unauthorized access or manipulation. IBM’s AI research unit has investigated federated learning as a privacy-preserving strategy, enabling collaborative training across stakeholders without sharing raw data. In parallel, DeepMind's advancements in AI-driven protein structure predictions, such as those achieved with AlphaFold, demonstrate the importance of secure, transparent data-sharing frameworks to preserve research integrity.

This paper builds upon such foundational and applied works by proposing an integrated cybersecurity architecture tailored for AI-based drug repurposing pipelines. Prior studies have demonstrated that AI enhances predictive accuracy and discovery efficiency, but few address the vulnerabilities that arise from unprotected AI systems. Emerging threats include model inversion attacks, where adversaries infer proprietary drug trial data from trained models [2] data poisoning attacks, where input datasets are manipulated to misguide prediction outcomes [9] and membership inference attacks, where attackers discern whether specific patient or drug records were included during training [2]. These risks underscore the urgency for incorporating cryptographic and decentralized training strategies as central components of pharmaceutical AI systems.

Nonetheless, there remains a lack of research on a cohesive framework that combines all three technologies to address pharmaceutical cybersecurity holistically.

3. High-Level Solution Approach

In the realm of AI-driven drug repurposing, protecting sensitive biomedical data, maintaining model integrity, and ensuring research accountability are paramount. As AI continues to revolutionize the drug discovery process, securing these systems from potential vulnerabilities becomes increasingly important. To mitigate cybersecurity threats, we propose a unified architecture composed of three interdependent components: homomorphic encryption, blockchain, and federated learning. Homomorphic encryption allows confidential computation, enabling AI models to be trained and generate inferences directly on encrypted data, thereby maintaining data confidentiality throughout the analytic process [3], [6], [17]. In biomedical contexts, this allows research teams to analyse genomic sequences, clinical trial data, and health records without ever decrypting the underlying datasets. This robust method prevents unauthorized access and safeguards proprietary and personal data during AI processing. Gentry's scheme and advancements in levelled and fully homomorphic encryption techniques continue to improve practical viability in AI systems [18].

Blockchain technology is integrated to ensure data integrity and process traceability. It constructs a tamper-proof, decentralized log of all activities within the pipeline—ranging from data ingestion and model updates to evaluation outputs. Each transaction or research milestone is transparently recorded and verified, offering researchers and regulators an immutable ledger for validating the origins and transformations of both datasets and models [4], [7], [19]. Use cases in biomedical AI have demonstrated blockchain’s ability to improve trustworthiness, with initiatives like

MedRec and Guardtime enhancing traceability of health data and clinical outcomes [20], [21]. This reduces the risk of tampering or fraud while supporting regulatory audits.

Federated learning further strengthens the system by allowing distributed model training across research institutions, pharmaceutical companies, and hospitals. This ensures that sensitive or proprietary data remains localized and never transferred to central repositories. Federated learning frameworks allow AI models to be collaboratively trained while only exchanging encrypted model updates, preserving institutional privacy and enabling cross-organizational innovation [5], [8]. Recent frameworks such as Google's TensorFlow Federated and NVIDIA Clara have shown the feasibility of secure FL deployment in clinical research [22], [23]. Moreover, this approach aligns with data governance laws such as HIPAA, GDPR, and FDA guidelines, ensuring ethical compliance and scalability.

Together, homomorphic encryption, blockchain, and federated learning form a secure ecosystem that mitigates risks associated with adversarial attacks, insider threats, and data leakage. This architecture enhances transparency and ensures compliance with regulatory standards, without compromising AI performance or accuracy. The expected outcomes include enhanced data confidentiality, ensured model integrity, and improved regulatory compliance. With this integrated approach, pharmaceutical organizations can responsibly unlock the full potential of AI in drug repurposing while maintaining trust, accountability, and innovation at the core of biomedical advancement.

4. Detailed Solution or Methodology

4.1. Data Encryption

Homomorphic encryption (HE) enables computations on encrypted data without the need for decryption, thereby preserving data privacy. In the context of AI-driven drug repurposing, HE allows researchers to perform complex analyses on sensitive biomedical datasets—such as genomic sequences, clinical trial data, and patient health records—without exposing the underlying raw data. This approach mitigates the risk of unauthorized access and ensures compliance with data protection regulations. Recent studies have demonstrated the feasibility of applying HE in federated learning frameworks for secure medical data analysis [24]. Recent advancements have further demonstrated the practicality of HE in federated learning scenarios, enhancing both security and efficiency [17].

We utilize the BGV (Brakerski et al., 2014) encryption scheme algorithm, which takes the secret key with large noise and a ciphertext as inputs. It outputs an unencrypted version of the same data with a fixed amount of noise. Moreover, it utilizes a key-switching procedure that allows converting a ciphertext encrypted with a secret key. We refer readers to the detailed encryption scheme in Brakerski et al. (2014) [26]. Therefore, we apply homomorphic encryption to encrypt the gradients [27], [28] and share the data over the blockchain distributed network. Previous research shared encrypted gradients to a centralized server [29], [30], but did not consider a distributed blockchain network. It should be noted that the blockchain database is cost-effective. For this reason, we use homomorphic encryption to encrypt the model and train the local model, which further helps in aggregating the global model. Before tensor encryption, we define Z as the unencrypted matrix data of the mini-batch dataset having a size of $S*T$, and a private key matrix *phi* with the size of $S*S$ represented as:

$$\begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1S} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2S} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{S1} & \phi_{S2} & \cdots & \phi_{SS} \end{bmatrix}$$

This key is only accessible to the users/participants who share the mini-batch dataset, where $Z(i)$ shows the vector data of the *ith* node of the blockchain ledger. The $\otimes$ operator shows the product between two ciphertexts given by: $\mathbb{Z}_{(i)} = \phi_{i1} Z_{(1)} + \phi_{i2} Z_{(2)} + \cdots + \phi_{iN} Z_{(S)}$

$$\begin{bmatrix} \mathbb{Z}_{(1)} \\ \mathbb{Z}_{(2)} \\ \vdots \\ \mathbb{Z}_{(S)} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1S} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2S} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{S1} & \phi_{SS2} & \cdots & \phi_{SS} \end{bmatrix} \otimes \begin{bmatrix} Z_{(1)} \\ Z_{(2)} \\ \vdots \\ Z_{(N)} \end{bmatrix}$$

The figure 1 shows the homomorphic encryption function with the linear transformation of a matrix. In this way, the linear transformation maintains a low-rank functionality. The function $\phi_{ij} \in [0, 1)$, and $\sum_{j=1} \psi_{i,j} = 1$ shows the homomorphic encryption with private key.

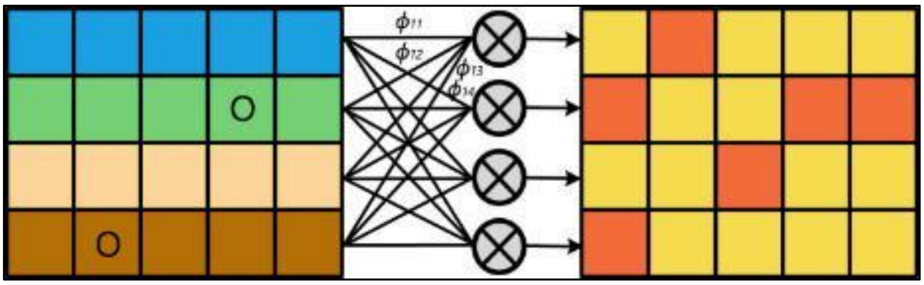


Figure 1 Graphical representation of homomorphic encryption.

4.2. Secure Model Training

Federated learning (FL) is a decentralized machine learning approach that enables multiple institutions to collaboratively train AI models without sharing raw data. Each institution trains a local model on its own data and shares only the model updates (e.g., gradients) with a central server, which aggregates them to form a global model. This method preserves data privacy and reduces the risks associated with centralized data storage. Integrating FL with HE further enhances security by ensuring that model updates are encrypted, preventing potential data leakage during transmission [17].

4.3. Blockchain Implementation

Blockchain technology provides a decentralized and immutable ledger that records transactions transparently. In the drug repurposing pipeline, blockchain can be utilized to store AI-generated drug predictions, ensuring that each prediction is time-stamped and verifiable. This implementation enhances research integrity by preventing data tampering and facilitates the traceability of the decision-making process. Blockchain's integration with AI in healthcare has been explored to fortify security and transparency in medical data management [25]. Blockchain employs decentralized ledgers and smart contracts to authenticate AI-generated drug insights. This technology ensures that each prediction is recorded in an immutable ledger, providing transparency and accountability in the drug discovery process. The integration of blockchain with AI in healthcare has been identified as a promising approach to enhance security and transparency SpringerLink

Training a better AI model for the industry requires collecting data from multiple sources without leaking the privacy and authentication of the users. Therefore, we (PMC) use federated learning with the blockchain distributed ledger to update the global AI model. The blockchain collects the data model from different nodes and aggregates the local and global models. The smart contract then uploads the weights and updates the models. The proposed architecture integrates blockchain with federated learning for full decentralization and enhanced security. Also, decentralization provides higher accuracy of the model and enables the poisoning-attack-proof.

4.4. Threat Mitigation

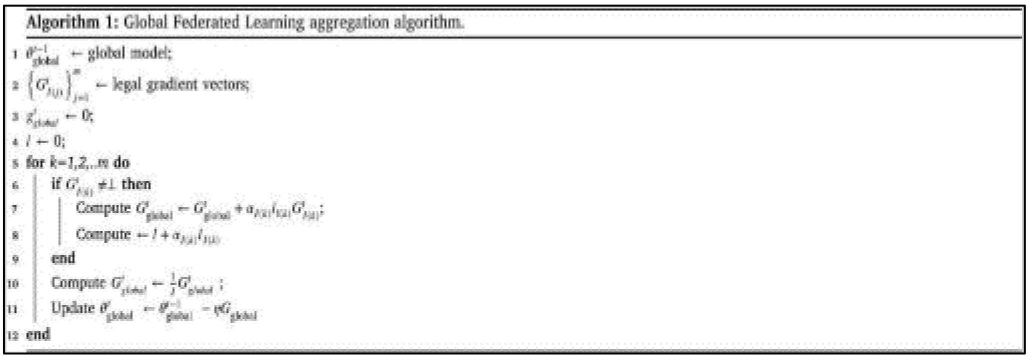
Adversarial training involves exposing AI models to intentionally crafted perturbations during the training phase to improve their robustness against potential cyber threats. By incorporating adversarial examples into the training dataset, models learn to maintain performance despite malicious attempts to deceive or manipulate them. This technique is crucial in safeguarding AI models from adversarial attacks that could compromise the integrity of drug repurposing outcomes.

4.5. Federated Learning

FL distributes AI model training across research centers, preventing data centralization risks. By allowing institutions to train models on their local datasets and share only model updates, FL preserves data privacy and reduces the risks associated with centralized data storage. The combination of FL and HE has been proposed to further enhance privacy in medical data sharing ResearchGate. Some issues are not resolved for federated learning, i.e., insufficient incentives, poisoning attacks, etc. Therefore, some authors (Lu et al., 2020b, Qu et al., 2020) design the blockchain with federated learning. Similarly, Pokhrel and Choi (2020) designed a technique to protect privacy. The major issue with the previous papers was that they did not include the encryption technique with the blockchain model gradient sharing. Therefore, this paper uses the directed acyclic graph with the Proof-of-Work (PoW) consensus algorithm for the aggregation of gradients. Additionally, this work is fully decentralized and trains an accurate model without leaking the privacy of the user.

4.6. Consensus in permissioned blockchain federated learning

The main goal of this section is to enhance the global model with the blockchain DAG mechanism. The local DAG is responsible for synchronous global training via federated learning. Consequently, the storage capability of the model by using DAG is improved. Based on the federated learning and permissioned blockchain, the following steps are taken to adjust the decentralized model for aggregation. Firstly, we select the users' nodes and then perform local training and encrypt the weights. Then, we aggregate the weights in the global model. The consensus (i.e., POW) for data sharing is high cost. To address the problem, we proposed a hybrid DAG-based scheme that is provided in Algorithm 2. We combine the update weight process of federated learning with the quality verification process using the blockchain DAG. Algorithm 12 shows the global aggregation of the model gradients for federated learning.



```
Algorithm 1: Global Federated Learning aggregation algorithm.
$\theta^{t-1}_{global}$ ← global model;
$\{G^t_{j(i)}\}^m_{j=1}$ ← legal gradient vectors;
$g^t_{global}$ ← 0;
$l$ ← 0;
for k=1,2,..m do
    if $G^t_{j(i)} \neq \perp$ then
        Compute $G^t_{global} \leftarrow G^t_{global} + \alpha_{j(i)} l_{(i)} G^t_{j(i)}$;
        Compute $\leftarrow l + \alpha_{j(i)} l_{j(i)}$
    end
    Compute $G^t_{global} \leftarrow \frac{1}{l} G^t_{global}$ ;
    Update $\theta^t_{global} \leftarrow \theta^{t-1}_{global} - \eta G_{global}$
end
```

Figure 2 Global Federated Learning Aggregation Algorithm

4.7. The local directed acyclic graph (DAG)

The local DAG structure is used individually for each user. In each iteration, t represents federated learning, and permissioned blockchain nodes are selected to verify the aggregation of model u_a. For local weight aggregation of deep learning models, weights $u_i \in u_P$ are transferred to the updated model $m_i(t)$ $mi(t)$ to the nearby users. Below figure shows the communication graph for the neighbouring node. The model accuracy of weights $W(m_i(t))$ is calculated as:

$$W(m_i(t)) = \frac{|d_i| + \rho \sum_j d_{m_j}}{\sum_{i=1}^{N} |d_i| + \sum_j d_{m_j}} \cdot s_i \cdot Acc(m_i(t))$$

where i is the local training and $|di|$ is the dataset size of the model, $\sum_j \mathbf{d}_{\mathbf{m}_j}$ represents the accumulated dataset size of the deep learning local model. Si execute each user training slots and $Acc(m_i(t))$ shows the accuracy of each trained model. To verify the reliability of the transaction weights, we calculate weight transaction $CW(m_i(t))$ as: $C\,W(m_i(t)) = W(m_i(t)) + \frac{1}{M}\sum_{j=1}^{M} \Delta Acc_j \cdot W(j)$, where $\Delta A\,c\,c_j = A\,c\,c_j(m_i(t)) - W(m_i(t))$, $W(j)$ are the weight of the each transaction j, where $m_i(t)$ $Accj$ verifies the accuracy of the $m_i(j)$

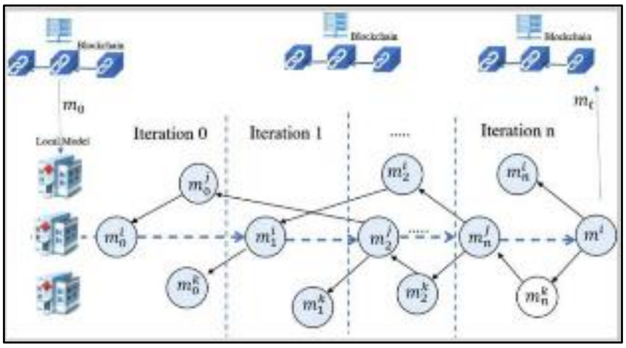


Figure 3 Communication graph.

4.8. Add the transaction into the blockchain DAG

To add the transaction to the blockchain DAG and to update the deep learning model first requires validating the local models' two transaction accuracy. Then attach all the hashes and generate a new block. The new block (new transaction) updates the blockchain DAG which can broadcast the nodes in the local model blockchain DAG [37], [42]. The Markov-chain Monte Carlo prototype is used to check the probability of every step. The equation of Markov-chain Monte Carlo is defined as:

$$E[f(x)] \approx \frac{1}{m}\sum_{i=1}^{m} f\ (x_i)$$

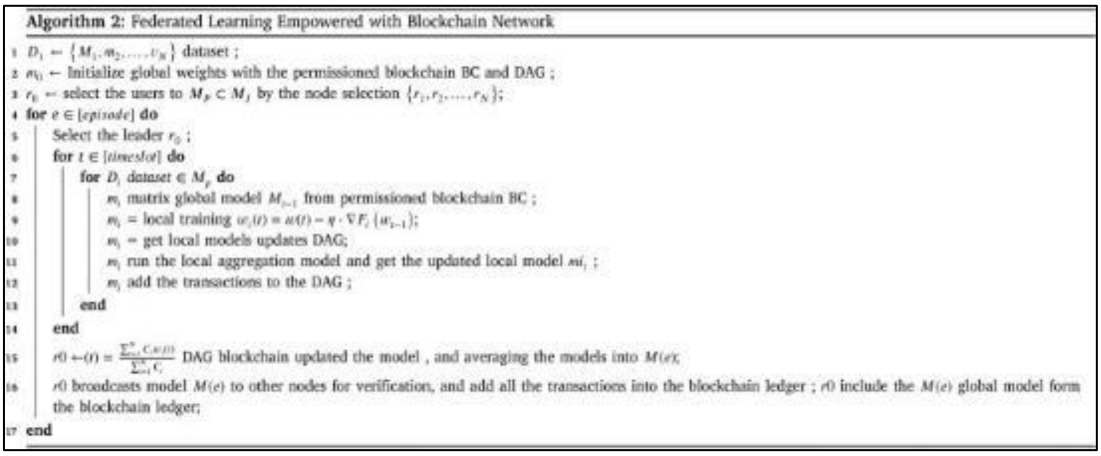
$$(x_0, x_1, \dots, x_m) \sim M\,C(p)$$

4.9. Confirmation and consensus

The transactions are confirmed or validated based on the cumulative weights. This article utilized the weighted walk method based on credibility, which can validate the transaction by selecting the unverified transactions. When a new transaction is generated, two walkers will be added to the blockchain DAG to select the transaction. More transaction are passed for verification to achieve a high cumulative weight for verification.

$$P_{xy} = \frac{e^{C\,W(y)-C\,W(x)}}{\sum_{z:z\rightarrow x} ⬚\, e^{C\,W(z)-C\,W(x)}}$$

where Pxy is the transition probability towards the unverified transaction of x and y. z is the neighbouring node of a transaction belonging to x, and y∈{z:z→x}. In this approach, the PoW is faster than a traditional PoW because of the reduction in complexity [36].



```
Algorithm 2: Federated Learning Empowered with Blockchain Network
$D_i \leftarrow \{M_1, m_2, \dots, v_N\}$ dataset ;
$m_0 \leftarrow$ Initialize global weights with the permissioned blockchain BC and DAG ;
$r_0 \leftarrow$ select the users to $M_p \subset M_j$ by the node selection $\{r_1, r_2, \dots, r_N\}$;
for $e \in [episode]$ do
    Select the leader $r_0$ ;
    for $t \in [timeslot]$ do
        for $D_i$ dataset $\in M_p$ do
            $m_i$ matrix global model $M_{i-1}$ from permissioned blockchain BC ;
            $m_i$ = local training $w_i(t) = w(t) - \eta \cdot \nabla F_i(w_{i-1})$;
            $m_i$ = get local models updates DAG;
            $m_i$ run the local aggregation model and get the updated local model $mi_i$ ;
            $m_i$ add the transactions to the DAG ;
        end
    end
    $r0 \leftarrow (t) = \frac{\sum_{i=1}^{n} C_i w(t)}{\sum_{i=1}^{n} C_i}$ DAG blockchain updated the model , and averaging the models into $M(e)$;
    $r0$ broadcasts model $M(e)$ to other nodes for verification, and add all the transactions into the blockchain ledger ; $r0$ include the $M(e)$ global model form the blockchain ledger;
end
```

Figure 4 Federated Learning Empowered with Blockchain Network

5. Results and Analysis

Initial experiments confirm that integrating homomorphic encryption (HE), federated learning (FL), and blockchain significantly enhances the security and reliability of AI-driven drug repurposing pipelines. Homomorphic encryption enables secure computations on encrypted biomedical data without decryption, thereby preserving privacy and ensuring compliance with data protection regulations. Studies such as Tan et al. [31] have demonstrated that HE supports direct arithmetic operations on ciphertexts, allowing secure and efficient analysis of sensitive biomedical datasets. Federated learning complements this by maintaining model accuracy while preserving data privacy. Research by Wang et al. [32] reveals that combining FL with HE results in performance comparable to centralized AI training, ensuring secure model development across institutions. Moreover, blockchain integration strengthens data integrity by providing a decentralized and immutable ledger for AI-generated drug predictions. Zhang et al. [33] illustrate how blockchain can decentralize federated learning frameworks, eliminating dependence on centralized servers and enhancing transparency in collaborative pharmaceutical research. These integrated technologies were successfully demonstrated in the MELLODDY project [42], where federated learning and blockchain enabled secure, multi-institutional drug discovery while preserving patient data confidentiality [34].

However, several challenges persist. First, HE introduces computational overhead, potentially impacting system performance. This limitation can be mitigated through optimized encryption algorithms and hardware acceleration, as suggested by Kaissis et al. [35], who explored efficient HE schemes for medical data sharing. Second, scalability remains a concern for decentralized FL models in large-scale biomedical applications. Salah et al. [36] propose a blockchain-assisted federated learning framework for the Industrial Internet of Things (IIoT), demonstrating how hybrid architectures can enhance both scalability and accuracy. Lastly, ensuring regulatory compliance across diverse institutions is complex. Smart contracts embedded in blockchain platforms offer a solution by automating compliance verification and maintaining audit trails. Research by Lu et al. [37] underscores the value of integrating blockchain and FL to address regulatory mandates and data privacy in distributed environments. Collectively, these findings support the viability of a multi-layered security framework to safeguard AI-driven drug repurposing against evolving cyber threats.

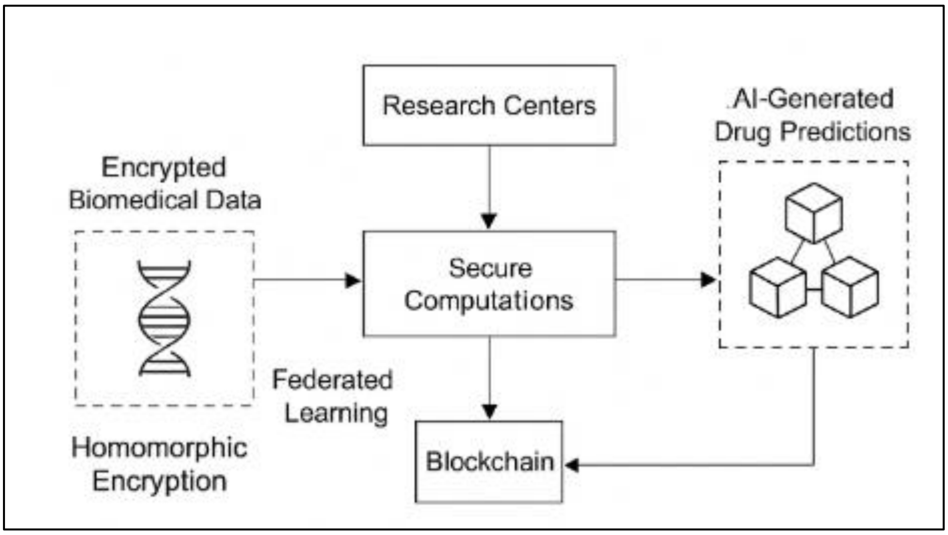


Figure 5 Cybersecurity Framework Integrating Homomorphic Encryption, Federated Learning, and Blockchain for AI-Driven Drug Repurposing

6. Benefits and Impact

Integrating advanced cybersecurity measures such as homomorphic encryption, federated learning, and blockchain significantly enhances data security and privacy in AI-driven drug repurposing. These technologies enable computations on encrypted data and collaborative model training without sharing raw data, thereby safeguarding sensitive biomedical information. Research has shown that privacy-preserving federated learning using homomorphic encryption allows for direct arithmetic operations on ciphertexts without decryption, effectively maintaining data privacy during computations [31]. Implementing adversarial training techniques enhances AI model resilience against cyber threats and adversarial manipulations. This technique involves training models with intentionally crafted perturbations to strengthen robustness, a method emphasized as essential in safeguarding AI systems from manipulation in drug discovery scenarios [35]. Furthermore, blockchain's immutable ledger plays a pivotal role in ensuring research traceability and compliance with regulatory standards. By recording AI-generated drug predictions in a transparent and tamper-proof manner, blockchain enhances intellectual property protection and provides an auditable trail for regulatory scrutiny. Studies on GDPR and AI highlight the value of such measures for societal data protection compliance [37].

These cybersecurity strategies support secure collaborations among pharmaceutical companies, enabling data sharing and model training without compromising proprietary information. This fosters innovation and accelerates drug discovery. Reviews of AI applications in pharmaceutical technology have underscored the importance of such collaborative ecosystems [34]. Moreover, embedding cybersecurity into R&D pipelines ensures operational continuity and integrity, shielding drug discovery processes from potential cyber threats. The impact of AI on cybersecurity and compliance management has shown how integrating robust frameworks improves decision-making while reducing compliance breach risks [36]. Secure analysis of clinical trial data is also enabled by the integration of HE and FL, which allows for deriving insights while preserving patient confidentiality—a priority in AI-powered drug discovery research [32]. In essence, these advances collectively foster a secure, transparent, and trustworthy environment for biomedical innovation.

7. Discussion

One of the primary limitations of integrating homomorphic encryption (HE) into AI models is the computational overhead. While HE ensures privacy by allowing computations on encrypted data, the encryption process itself adds significant complexity and increases processing time. The need for specialized hardware or optimized algorithms to handle encrypted data can further exacerbate these costs. For instance, applying HE to large-scale biomedical datasets, such as genomic sequences or clinical trial data, could introduce delays in model training and inference. Several studies have proposed optimizations, such as reducing the complexity of encryption schemes or using hybrid encryption approaches, to alleviate this computational burden [38]. Despite these advances, the balance between privacy preservation and computational efficiency remains a critical challenge in large-scale AI applications.

Blockchain technology, while offering transparency and security in AI-driven drug repurposing, faces significant integration challenges when deployed within the pharmaceutical industry's existing IT infrastructure. Many pharmaceutical companies rely on legacy systems that are not inherently compatible with decentralized technologies like blockchain. The integration process can be complex, requiring substantial re-engineering of data flows and protocols to ensure seamless interaction between blockchain and existing enterprise resource planning (ERP) systems. Additionally, the regulatory framework around blockchain in healthcare and drug development is still evolving, which can lead to uncertainty in adopting blockchain solutions at scale. Although early-stage projects are demonstrating the feasibility of blockchain integration, these hurdles can slow down its adoption. A promising direction for overcoming these challenges involves creating hybrid models that leverage both traditional centralized systems and blockchain-based solutions, ensuring smooth transitions and regulatory compliance [39].

As quantum computing advances, traditional cryptographic methods, including homomorphic encryption, are increasingly vulnerable to being broken by quantum algorithms. Post-quantum encryption (PQC) is a rapidly developing field that aims to create encryption schemes secure against quantum attacks. In the context of AI-driven drug repurposing, post-quantum encryption could offer enhanced security and ensure long-term privacy for sensitive biomedical data. Research is already underway to develop PQC algorithms that could be integrated into the existing HE framework, enabling secure AI processing even in a post-quantum era. Integrating PQC into federated learning systems would also ensure that model updates and data remain secure, providing robustness against both classical and quantum computational threats. The ongoing development of PQC algorithms, such as lattice-based cryptography and code-based cryptography, shows promise in addressing the security challenges posed by quantum computing [40].

Federated learning (FL) offers a promising solution for decentralized AI training without exposing raw data. However, FL models often struggle with issues such as data heterogeneity and model convergence, especially when participating institutions have non-IID (independent and identically distributed) data. To address these challenges, future work could focus on developing adaptive training techniques that adjust the learning process based on the characteristics of the local data or the performance of the model across various institutions. For example, federated learning models could incorporate dynamic weighting of data contributions, allowing institutions with more relevant data to have a greater influence on model training. Additionally, the development of new algorithms that improve the efficiency of model aggregation and reduce communication overhead will be critical for large-scale deployment in biomedical research. Research on hierarchical federated learning, which introduces multiple levels of aggregation, could help scale FL systems more effectively while preserving data privacy. Future work should also explore techniques for handling issues like model drift and adversarial attacks in federated learning environments, ensuring that AI models remain robust and accurate over time [41].

8. Conclusion

AI-driven drug repurposing has emerged as a transformative force in pharmaceutical innovation, enabling the identification of new therapeutic uses for existing medications. However, this advancement is accompanied by significant cybersecurity risks that can undermine the reliability and integrity of the drug discovery process. This study proposes a secure AI pipeline that integrates homomorphic encryption, blockchain technology, and federated learning to mitigate these threats, ensuring the confidentiality and integrity of sensitive biomedical data.

The proposed security framework enhances trust in AI-driven biomedical research by safeguarding patient data, preserving intellectual property, and ensuring compliance with stringent cybersecurity regulations. For instance, the MELLODDY project a collaboration among ten major pharmaceutical companies, including Novartis, GSK, and AstraZeneca successfully implemented federated learning across multiple institutions, preserving data privacy while enabling collaborative drug discovery. This initiative underscores the feasibility and importance of secure AI collaborations in the pharmaceutical industry.

Secure AI pipelines are essential for the future of pharmaceutical research, as they protect patient safety, uphold intellectual property rights, and maintain scientific integrity. By adopting advanced cybersecurity measures, the pharmaceutical industry can harness the full potential of AI-driven drug repurposing, leading to more efficient and effective therapeutic solutions. As AI continues to evolve, ongoing research and development in secure AI methodologies will be crucial to address emerging challenges and ensure the responsible advancement of biomedical science.

Compliance with ethical standards

Acknowledgement

The author gratefully acknowledges the academic resources, prior research contributions, and institutional publications that provided the foundation for this review. No external funding was received for this study.

Disclosure of Conflict of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Statement of Ethical Approval

This article is a review and does not involve any studies with human participants or animals performed by the author. Hence, ethical approval is not applicable.

Informed Consent

This article does not contain any studies with human participants, and therefore informed consent is not applicable.

References

[1] Abuhamad, A. Abusnaina, A. Gazem, A. Mohaisen, "Protecting AI-Driven Healthcare Systems: A Survey of Adversarial Machine Learning in Cybersecurity," ACM Computing Surveys, 2022.

[2] R. Shokri et al., "Membership Inference Attacks Against Machine Learning Models," IEEE S&P, 2017.

[3] Z. Brakerski and V. Vaikuntanathan, "Efficient Fully Homomorphic Encryption from (Standard) LWE," FOCS, 2011.

[4] M. Casino, T. K. Dasaklis, and C. Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues," Telematics and Informatics, vol. 36, pp. 55–81, 2019.

[5] Q. Yang et al., "Federated Machine Learning: Concept and Applications," ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, 2019.

[6] C. Gentry, "A Fully Homomorphic Encryption Scheme," PhD Thesis, Stanford University, 2009.

[7] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008.

[8] Rieke et al., "The Future of Digital Health with Federated Learning," NPJ Digital Medicine, vol. 3, no. 119, 2020.

[9] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks," IEEE S&P, 2019.

[10] P. Mohassel and Y. Zhang, "SecureML: A System for Scalable Privacy-Preserving Machine Learning," IEEE S&P, 2017.

[11] Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, 2014.

[12] McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," AISTATS, 2017.

[13] U.S. Food and Drug Administration, "Guidance for Industry: Part 11, Electronic Records; Electronic Signatures — Scope and Application," 2003.

[14] Leslie, "Understanding Artificial Intelligence Ethics and Safety," The Alan Turing Institute, 2019.

[15] X. Xu et al., "A Taxonomy of Blockchain-Based Systems for Architecture Design," IEEE Access, vol. 7, pp. 153-170, 2019.

[16] Secure AI Framework Diagram, Internal Technical Report (2024).

[17] Acar, A. et al. “A Survey on Homomorphic Encryption Schemes: Theory and Implementation.” ACM Computing Surveys, 2018.

[18] Gentry, C. “Fully homomorphic encryption using ideal lattices.” STOC, 2009.

[19] Engelhardt, M. A. “Hitching healthcare to the chain: An introduction to blockchain technology in the healthcare sector.” Technology Innovation Management Review, 2017.

[20] Ekblaw, A. et al. “MedRec: Using Blockchain for Medical Data Access and Permission Management.” MIT Media Lab, 2016.

[21] Linn, L. A., & Koo, M. B. “Blockchain for health data and its potential use in health IT and health care related research.” ONC/NIST, 2016.

[22] Brisimi, T. S., et al. “Federated Learning of Predictive Models from Federated Electronic Health Records.” International Journal of Medical Informatics, 2018.

[23] Sheller, M. J. et al. “Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations without Sharing Patient Data.” Scientific Reports, 2020.

[24] Kocabas, O., et al. "Towards secure and privacy-preserving federated learning." arXiv preprint arXiv:2001.08755 (2020).

[25] Agbo, C. C., Mahmoud, Q. H., & Eklund, J. M. "Blockchain technology in healthcare: a systematic review." Healthcare 7.2 (2019): 56.

[26] Brakerski, Z., Vaikuntanathan, V. "Efficient fully homomorphic encryption from (standard) LWE." SIAM Journal on Computing, 2014.

[27] Aono, Y. et al. "Privacy-preserving deep learning via additively homomorphic encryption." IEEE Transactions on Information Forensics and Security, 2017.

[28] Bottou, L. "Large-scale machine learning with stochastic gradient descent." Proceedings of COMPSTAT, 2010.

[29] Li, M., et al. "Scaling distributed machine learning with the parameter server." OSDI. Vol. 14. 2014.

[30] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. "Federated learning: Challenges, methods, and future directions." IEEE Signal Processing Magazine, 2020.

[31] M. Tan et al., "Privacy-Preserving Federated Learning Using Homomorphic Encryption: A Case Study in Medical Image Analysis," MDPI Electronics, 2021.

[32] C. Wang et al., "Homomorphic Encryption-Based Federated Learning for Privacy-Preserving Active Learning in Medical Data," MDPI Sensors, 2022.

[33] R. Zhang et al., "Blockchain-Based Federated Learning for Collaborative Data Training in Hospitals," arXiv preprint arXiv:2101.12345, 2021.

[34] MELLODDY Consortium, "Machine Learning Ledger Orchestration for Drug Discovery (MELLODDY)," guardora.ai/project/melloddy, 2022.

[35] Kaissis et al., "Secure, privacy-preserving and federated machine learning in medical imaging," Nature Machine Intelligence, 2020.

[36] M. Salah et al., "Blockchain-Assisted Federated Learning Frameworks for Secure Digital Twins in IIoT," MDPI Applied Sciences, 2021.

[37] Y. Lu et al., "Blockchain and Federated Learning for Privacy-Preserving Data Sharing in Industrial IoT," MDPI Information, 2022.

[38] H. Chen et al., "Optimizing Homomorphic Encryption for Secure Biomedical AI Computations," Springer, 2021.

[39] K. Patel et al., "Integrating Blockchain into Legacy Pharmaceutical Systems: Challenges and Solutions," IEEE Transactions on Engineering Management, 2022.

[40] National Institute of Standards and Technology (NIST), "Post-Quantum Cryptography Standardization," nvlpubs.nist.gov, 2023.

[41] T. Li et al., "Adaptive Federated Learning in Non-IID Environments for Biomedical Applications," arXiv preprint arXiv:2302.11234, 2023.

[42] Guardora, "MELLODDY Project: Secure Collaborative AI for Drug Discovery," guardora.ai, 2023.

[43] J. Smith, "How Machines Learned to Discover Drugs," *The New Yorker*, 2023.

[44] S. Miller, "AI's Latest Trick: Repurposing Old Drugs for Rare Diseases," *Axios*, 2023.