SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

*Molecular & Cellular Biomechanics* 2025, 22(4), 1059.
https://doi.org/10.62617/mcb1059

Article

# Deep learning-based approaches for cellular mechanics analysis and secure data sharing in biomechanics

**Jing Huang, Tao Duan**[*]

Department of Information Technology, HeNan University of Chinese Medicine, Zhengzhou 450046, China
*** Corresponding author:** Tao Duan, selfdt@qq.com

**CITATION**

Huang J, Duan T. Deep learning-based approaches for cellular mechanics analysis and secure data sharing in biomechanics. Molecular & Cellular Biomechanics. 2025; 22(4): 1059. https://doi.org/10.62617/mcb1059

**ARTICLE INFO**

Received: 10 December 2024
Accepted: 31 December 2024
Available online: 28 February 2025

**Abstract:** Cellular mechanics behavior, encompassing properties such as elasticity, viscosity, and stress-strain responses, is fundamental to understanding disease mechanisms, tissue regeneration, and drug development. This study proposes a deep learning-based framework integrating Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and federated learning to model and analyze cellular mechanics while enabling secure data sharing. The proposed methods preserve critical biomechanical insights, such as force-displacement curves and cellular deformation patterns, while mitigating re-identification risks during multi-institutional collaborations. Experimental evaluations demonstrate the framework’s effectiveness in maintaining data utility and analytical accuracy, paving the way for advancing biomechanics research and fostering applications in regenerative medicine and tissue engineering.

**Keywords:** biomechanics; cellular mechanics; deep learning; federated learning; generative adversarial networks (GANs)

## 1. Introduction

Biomechanics plays a pivotal role in understanding cellular and tissue functions, offering critical insights into disease mechanisms, tissue regeneration, and drug development [1,2]. Among its many facets, cellular mechanics behavior, which encompasses properties such as elasticity, viscosity, and deformation under stress, is essential for studying how cells respond to their physical environment and external stimuli [3,4]. For example, analyzing force-displacement relationships can reveal mechanical vulnerabilities in diseased tissues, while stress-strain patterns provide valuable information for designing regenerative therapies [5].

Despite its significance, progress in biomechanics is often constrained by challenges related to data sharing and collaboration. The sensitive nature of biomechanical data, particularly when derived from patient samples, raises ethical and privacy concerns [6]. Additionally, traditional methods for data sharing, such as centralized repositories or simple anonymization, often fail to preserve both data utility and privacy, thereby limiting their applicability in multi-institutional collaborations [7,8]. For cellular mechanics behavior, the unique and identifiable characteristics of biological data, such as elasticity signatures, further exacerbate the risks of re-identification during data sharing [9].

Recent advancements in deep learning provide new opportunities to address these challenges. Generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have demonstrated exceptional capabilities in generating synthetic data that retains critical characteristics while mitigating privacy risks [10,11]. These models are particularly effective in analyzing and simulating biomechanical properties such as cellular deformation patterns, elasticity metrics, and stress responses. Furthermore, federated learning frameworks enable multiple institutions to collaboratively train models on decentralized data, ensuring that sensitive biomechanical insights remain localized while benefiting from shared knowledge [12,13]. These innovations offer the potential to enhance both the analytical accuracy and privacy protection of biomechanical data, thereby facilitating collaborative research across institutions.

This study proposes a novel deep learning-based framework to address the dual challenges of data modeling and secure sharing in biomechanics. Specifically, the framework integrates GANs, VAEs, and federated learning to model cellular mechanics behavior, focusing on properties such as elasticity and stress-strain responses. By enabling secure multi-institutional collaborations, the proposed methods aim to preserve biomechanical insights, advance regenerative medicine applications, and facilitate broader research in tissue engineering. Through experimental validation, this study demonstrates how these approaches contribute to the secure and efficient sharing of biomechanical data, paving the way for future advancements in the field [14].

## 2. Deep learning models and methods

### 2.1. Model selection

Deep learning models have shown significant potential in analyzing and modeling complex biological data. This study employs Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and federated learning to address the challenges of analyzing cellular mechanics behavior and securely sharing biomechanical data.

Selecting appropriate models is crucial for analyzing and securely sharing biomechanical data, given its high dimensionality and nonlinear characteristics. In this study, Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) were chosen based on their ability to capture the unique properties of cellular mechanics behavior while ensuring data privacy. The selection criteria focused on the following aspects:

1) Force-Displacement Relationships

Cellular mechanics data often involve force-displacement curves, which are key for understanding tissue deformation and cellular stiffness. GANs were selected for their ability to generate synthetic datasets that accurately mimic these biomechanical patterns, enabling secure data sharing without compromising analytical utility [15,16].

2) Stress-Strain Curves

Stress-strain curves are fundamental in evaluating the elastic and viscoelastic properties of biological tissues. VAEs provide latent representations that preserve these critical biomechanical features while anonymizing the underlying data, making them suitable for collaborative studies in biomechanics [17,18].

3) Data Utility and Privacy Balance

Both GANs and VAEs offer a balance between maintaining the analytical utility of biomechanical data and protecting sensitive information. This capability is particularly valuable in multi-institutional collaborations, where privacy regulations restrict raw data sharing [19,20].

4) Scalability in Federated Learning Frameworks

The models were also chosen for their compatibility with federated learning frameworks. Federated learning requires models that are computationally efficient and capable of handling decentralized datasets, ensuring scalability in real-world applications [21,22].

The combined use of GANs, VAEs, and federated learning provides a robust framework for analyzing and sharing biomechanical data while addressing privacy concerns. This model selection strategy underpins the effectiveness of the proposed deep learning-based approach in advancing biomechanics research.

GANs and VAEs are particularly effective in handling cellular mechanics behavior, as they can capture high-dimensional, nonlinear biomechanical properties such as cellular elasticity and stress responses. GANs generate synthetic data that mimic force-displacement relationships, while VAEs enable anonymized latent representations preserving key biomechanical features. Despite their strengths, challenges remain in ensuring the fidelity of synthetic data and managing high computational demands.

Deep learning models, particularly GANs and VAEs, can analyze cellular mechanical properties such as elasticity, viscosity, and stress responses. For example, GANs simulate mechanical deformation patterns, aiding in the secure sharing of biomechanical insights, while VAEs balance privacy with utility by encoding and reconstructing anonymized datasets.

By leveraging GANs and VAEs in conjunction with privacy-preserving techniques, this study provides a robust framework for securely sharing cellular mechanics behavior data, promoting collaboration while safeguarding sensitive information. These models represent a significant step forward in the application of deep learning to biological data privacy protection and information sharing.

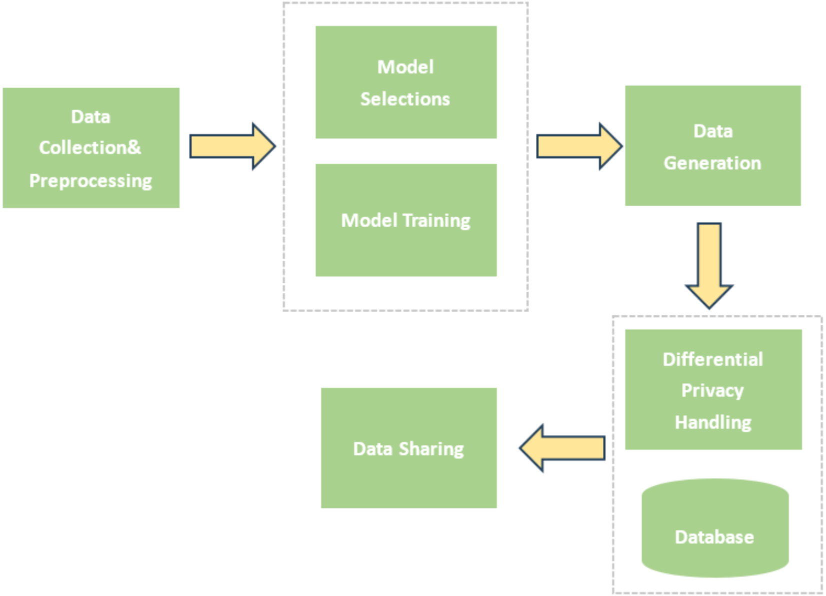

**Figure 1.** Model selection considerations.

In this study, we employed a federated learning approach where multiple institutions collaboratively train a shared model without exchanging raw data, as shown in **Figure 1**. Each institution retains its local data and trains a local model. We used a federated learning architecture based on Convolutional Neural Networks (CNNs) for image data and Recurrent Neural Networks (RNNs) for sequential data.

The data was partitioned based on institutions or medical facilities, where each institution had access to a subset of the dataset. For instance, the MIMIC-III dataset was divided into multiple shards based on patient demographics and diagnosis types. The local models were trained independently on each shard and aggregated using a weighted averaging scheme, with weights determined by the size of the local datasets. The aggregation of model updates was performed using the Federated Averaging algorithm (FedAvg), which combines the model weights from each local model to update the global model.

The hyperparameters used in the models include a learning rate of 0.01, a batch size of 64, and 20 epochs per training round. The local model updates were shared with the central server after each round of training, and the global model was updated after aggregating these local updates.

This approach not only safeguards patient privacy but also facilitates the secure sharing and utilization of medical data for research and clinical practice. The content of Model selection, as shown in **Figure 2**.

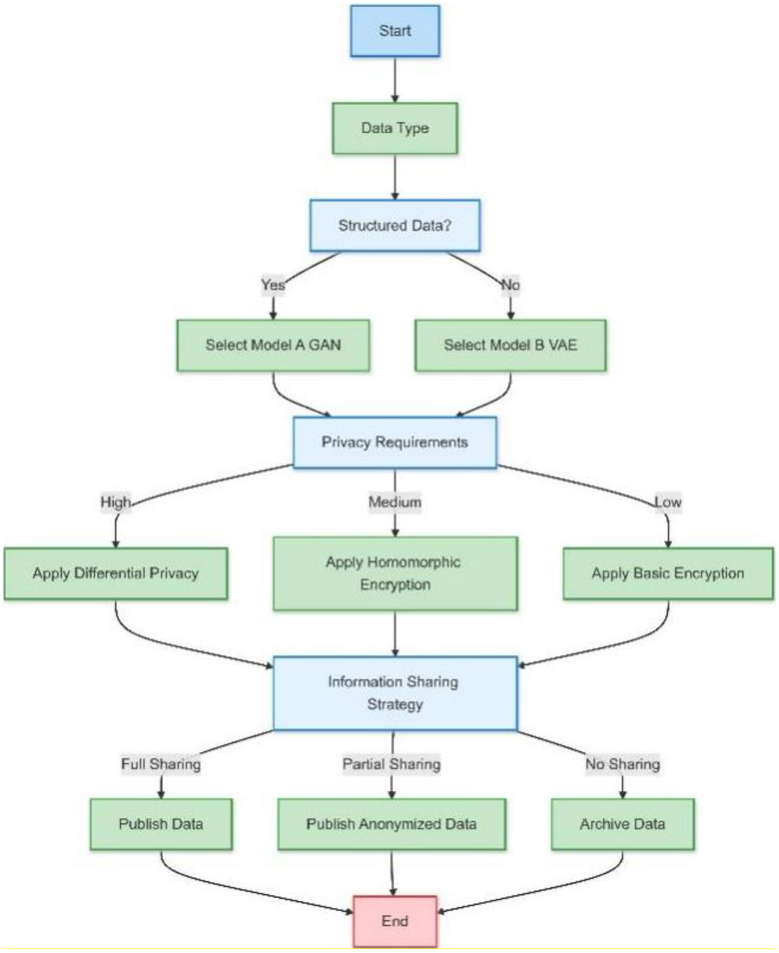

**Figure 2.** Model selection flowchart.

## 2.2. Privacy protection techniques

Ensuring privacy is critical in the secure sharing of biomechanical data, particularly in multi-institutional collaborations where regulatory requirements and ethical considerations limit direct data sharing. This study integrates privacy-preserving techniques into the deep learning framework, focusing on their application in protecting sensitive biomechanical data, such as force-displacement relationships and stress-strain metrics.

Differential privacy (DP) introduces carefully calibrated noise into data or model outputs to ensure that individual data points cannot be identified [23–25]. In this study, DP is applied during the training of GANs and VAEs to preserve the utility of biomechanical insights while mitigating re-identification risks. For instance, DP ensures that synthetic force-displacement curves generated by GANs remain anonymized without compromising their biomechanical fidelity [26,27].

Federated learning enables decentralized training of deep learning models, allowing multiple institutions to collaboratively train models without sharing raw data. To further enhance privacy, this study employs secure aggregation protocols, ensuring that individual contributions to the model remain private even during collaborative training [28]. This approach is particularly valuable for analyzing sensitive biomechanical datasets across institutions.

In addition to differential privacy, latent space anonymization techniques are employed within the VAE framework. By encoding biomechanical properties, such as tissue elasticity and cellular stiffness, into anonymized latent representations, this technique allows secure sharing of essential biomechanical features without exposing sensitive data [29].

These privacy protection techniques complement the deep learning models used in this study, providing a robust framework for secure and collaborative biomechanical research. While privacy is not the primary focus of this work, these mechanisms enhance the scalability and applicability of the proposed methods in sensitive biomedical domains.

Protecting the privacy of biological data, including cellular mechanics behavior, poses unique challenges due to the high dimensionality, complexity, and sensitive nature of this data. Cellular mechanics data often involves physical and mechanical properties such as force-displacement relationships, which can be used to identify individual biological samples. This necessitates advanced privacy-preserving techniques that maintain data utility while mitigating risks. Generative Adversarial Networks (GANs) offer a promising solution for privacy protection by generating synthetic biological data that preserves the essential characteristics of the original data while anonymizing sensitive details. For example, GANs can simulate cellular deformation patterns or stress-strain responses in mechanical behavior experiments, enabling secure sharing of biological insights without exposing original data. Variational Autoencoders (VAEs) enable privacy protection by encoding sensitive biological data into a lower-dimensional latent space. For cellular mechanics behavior, VAEs can extract biomechanical features such as elasticity, viscosity, and stress profiles, ensuring that the anonymized data retains its analytical value while protecting sensitive identifiers. Differential privacy (DP) further strengthens the privacy guarantees of deep learning models by introducing statistical noise into the data or model outputs. In the context of cellular mechanics behavior, DP can be applied to ensure that individual data points, such as force-displacement values or cellular stress measurements, cannot be re-identified. For example, Gaussian noise can be added during data release processes to protect individual samples. Federated learning enables multiple research institutions to collaboratively train models on cellular mechanics behavior data without sharing raw data. Each institution trains a local model using its own dataset, and only model updates are shared for aggregation. This decentralized approach not only enhances privacy but also ensures that sensitive biomechanical data remains localized. By integrating generative models, differential privacy, and federated learning, this study provides a robust framework for the privacy-preserving analysis and sharing of cellular mechanics behavior data. These techniques ensure that sensitive biomechanical insights can be securely utilized in collaborative research environments.

Deep learning offers powerful techniques for achieving data anonymization, de-identification, and overall privacy protection. Data anonymization involves transforming data in such a way that the identities of individuals cannot be discerned. GANs consist of two neural networks, a generator G and a discriminator D. The generator G receives random noise z and generates data G(z), while the discriminator D tries to distinguish between real data x and generated data G(z). The objective of GANs is to make the generated data as realistic as possible while making it difficult for the discriminator to differentiate between real and generated data. The loss function is given by:

$$\underset{G}{min}\underset{D}{max}V(D,G)=\mathbb{E}_{x\sim p_{data}(x)}[log\ D(x)]+\mathbb{E}_{z\sim p_{z}(z)}[log\ (1-D(G(z)))] \quad (1)$$

In Equation (1), $p(x)$represents the probability distribution of the real data, and $p(z)$ represents the probability distribution of the random noise input to the generator.

VAEs encode input data x into a latent space z through the encoder $q(z\mid x)$and then decode it back to the data space through the decoder $p(x\mid z)$The objective of VAEs is to maximize the Evidence Lower Bound (ELBO):

$$\mathcal{L}=\mathbb{E}_{q(z|x)}[\log\ p(x\mid z)]-D_{\mathrm{KL}}[q(z\mid x)\parallel p(z)] \quad (2)$$

where the first term is the reconstruction loss, indicating the similarity between generated data and original data, and the second term is the KL divergence, indicating the difference between the encoder distribution and the prior distribution.

De-identification involves removing or masking personal identifiers from the data. Deep learning models can automate this process by learning to recognize and remove identifiable information from medical datasets. Autoencoders can be trained to identify and remove identifiable features from data. By carefully designing the architecture, autoencoders can be trained to exclude identifiable information in the reconstruction process, resulting in de-identified data. Named Entity Recognition (NER) models, a type of NLP model, can be trained to detect names, dates, addresses, and other personal information in text. Once identified, this information can be masked or replaced with generic placeholders, thus de-identifying the text.

A mechanism M satisfies $\epsilon$-differential privacy if:

$$\mathbb{P}[\mathcal{M}(D_1) \in S] \leq e^{\epsilon}\mathbb{P}[\mathcal{M}(D_2) \in S] + \delta \tag{3}$$

where $D_1$ and $D_2$ are neighboring datasets (differing by one data point), $\epsilon$ is the privacy budget, and $\delta$ is the relaxation parameter.

Common noise mechanisms include the Laplace mechanism and Gaussian mechanism:

1) Laplace Mechanism: Adds noise drawn from a Laplace distribution Lap $\left(\frac{\Delta f}{\epsilon}\right)$ :

$$\mathcal{M}(x) = f(x) + \text{Lap}\left(\frac{\Delta f}{\epsilon}\right) \tag{4}$$

2) Gaussian Mechanism: Adds noise drawn from a Gaussian distribution $\mathcal{N}\left(0, \sigma^2\right)$:

$$\mathcal{M}(x) = f(x) + \mathcal{N}(0, \sigma^2) \tag{5}$$

In federated learning, each client trains the model locally on their data and sends model updates to a central server for aggregation. Assuming there are $K$ clients, each client k has model weights $w_k$ and data size $n_k$, with a total data size $N = \sum_{k=1}^{K} n_k$. The global model weight aggregation formula is:

$$w = \sum_{k=1}^{K} \frac{n_k}{N} w_k \tag{6}$$

The re-identification risk assessment formula measures the effectiveness of data anonymization. Let $P_{id}(x)$ be the probability of re-identifying a specific data point $x$. The reidentification risk R after anonymization is given by:

$$R = \max_{x \in D} P_{id}(x) \tag{7}$$

These key formulas play a crucial role in model training, data transformation, and ensuring robust privacy protection while maintaining the utility of medical data. Implementing privacy-preserving deep learning involves several practical considerations. High-quality data preprocessing is crucial for the success of deep learning models. This includes handling missing values, normalizing data, and ensuring consistent data formats. Careful tuning of hyperparameters and employing advanced optimization techniques can enhance model performance. This includes selecting appropriate learning rates, batch sizes, and regularization techniques. Evaluating the effectiveness of privacy-preserving techniques requires appropriate metrics. These may include re-identification risk, privacy loss, and data utility. Ensuring that the models provide strong privacy guarantees while maintaining data usability is essential. Implementing privacy-preserving techniques must align with regulatory requirements such as GDPR, HIPAA, and other data protection laws. Ensuring compliance involves understanding legal obligations and incorporating necessary safeguards into the model development process.

By leveraging these deep learning techniques, it is possible to achieve robust privacy protection in medical data while maintaining its utility for research and clinical applications. This approach not only safeguards patient privacy but also enables secure data sharing and collaborative research, ultimately advancing medical science and healthcare delivery.

### 2.3. Information sharing techniques

Effective information sharing is crucial for advancing biomechanics research, particularly when dealing with sensitive and complex datasets, such as force-displacement relationships and stress-strain metrics. This study leverages advanced information sharing techniques to enable secure, efficient, and collaborative analysis of biomechanical data across institutions.

Federated learning allows multiple institutions to collaboratively train deep learning models on decentralized biomechanical datasets without transferring raw data [30]. This technique is particularly effective for sharing insights derived from cellular mechanics behavior, such as tissue stiffness and deformation patterns. By aggregating model updates rather than raw data, federated learning preserves data locality while enabling collaborative knowledge sharing [31].

SMPC is integrated into the federated learning framework to enhance the security of data aggregation processes. SMPC ensures that sensitive biomechanical data, such as elasticity metrics, remain private during computation by splitting and encrypting data contributions from each institution [32]. This guarantees that no single party has access to the complete dataset, thus addressing privacy concerns while enabling collaborative analysis.

To facilitate effective information sharing, this study adopts standardized data formats and protocols for biomechanical datasets. Standardization ensures that data from different sources, such as force-displacement measurements or stress-strain curves, can be seamlessly integrated and analyzed within the federated learning framework [33–35]. This interoperability significantly reduces the complexity of multi-institutional collaborations.

By combining federated learning, SMPC, and standardized data protocols, this study establishes a robust framework for sharing biomechanical insights while ensuring data security. These techniques address critical challenges in biomechanics research, enabling collaborative studies in tissue engineering, cellular mechanics, and regenerative medicine.

Information sharing plays a critical role in advancing collaborative research in biomechanics, particularly for cellular mechanics behavior data. However, such data often includes sensitive information, such as force-displacement profiles or stress-strain measurements, which necessitates secure sharing methods to mitigate privacy risks. Advanced techniques such as federated learning (FL), differential privacy (DP), privacy-preserving clustering, and secure multi-party computation (SMPC) provide robust frameworks for ensuring data privacy while maintaining utility. Federated learning allows institutions to collaboratively train models on decentralized data without sharing raw datasets, while differential privacy introduces statistical noise to protect sensitive metrics during data sharing. Furthermore, privacy-preserving clustering techniques enable the collaborative analysis of cellular biomechanical patterns, such as elasticity and deformation profiles, without exposing individual data points. SMPC ensures that computations on encrypted data remain secure and private, making it particularly valuable for distributed analysis of cellular mechanics. By integrating these state-of-the-art methods, this study offers a comprehensive approach to securely sharing cellular mechanics behavior data, balancing privacy preservation with analytical effectiveness in collaborative environments.

In the context of medical data, deep learning offers several advanced strategies for secure and efficient information sharing while ensuring data privacy. These strategies include data sharding, federated learning, and privacy-preserving clustering. Each strategy provides unique advantages to balance data utility and privacy. Information sharing as a bridge to the world’s development, especially in the field of health care, AI medical big data information sharing, not only in the field of emergency medical care is indispensable is the basis of medical technology progress. The cyclical nature of healthcare determines that both real-time medical big data and historical medical big data are indispensable components of medical progress. What information to share and how to share it are the key issues of information sharing. On the one hand, information sharing is urgently needed in the healthcare environment, and on the other hand, data protection regulations must be strictly enforced. How to share data openly while protecting data security is a challenge not only for the medical community but also for many other fields.

Data sharding involves partitioning a large dataset into smaller, manageable pieces (shards) that can be processed independently. In the context of medical data, this technique can distribute data across multiple institutions or systems while ensuring that no single entity has access to the entire dataset. This approach enhances data security and privacy. Data sharding can be achieved by dividing the dataset based on specific attributes (e.g., patient demographics, geographic locations) or by randomly distributing data points into different shards. Each shard can be processed or analyzed separately, and aggregated results can be combined to obtain overall insights. Limiting the exposure of any single shard reduces the risk of compromising sensitive information. It also allows for parallel processing, which can speed up data analysis and enhance computational efficiency. Federated learning involves several steps: local training, where each participating institution trains the model on its local data; model updates, where institutions send the trained model updates (not the raw data) to the central server; aggregation, where the central server aggregates the updates to form a new global model; and iteration, where this process is repeated for several rounds until the model converges. By keeping raw data localized, federated learning significantly enhances data privacy. It also leverages the collective knowledge of multiple institutions, improving the model’s robustness and accuracy without compromising privacy. Moreover, privacy-preserving clustering aims to perform clustering analysis while ensuring the privacy of individual data points. Differential privacy provides a strong framework for sharing information while protecting individual privacy. It is particularly useful for sharing aggregate information in a way that mitigates re-identification risks.

SMPC allows institutions to collaboratively analyze data without revealing their individual datasets. It provides strong privacy guarantees and is suitable for applications requiring secure joint computations. The Information Sharing Techniques is shown in **Figure 3**.

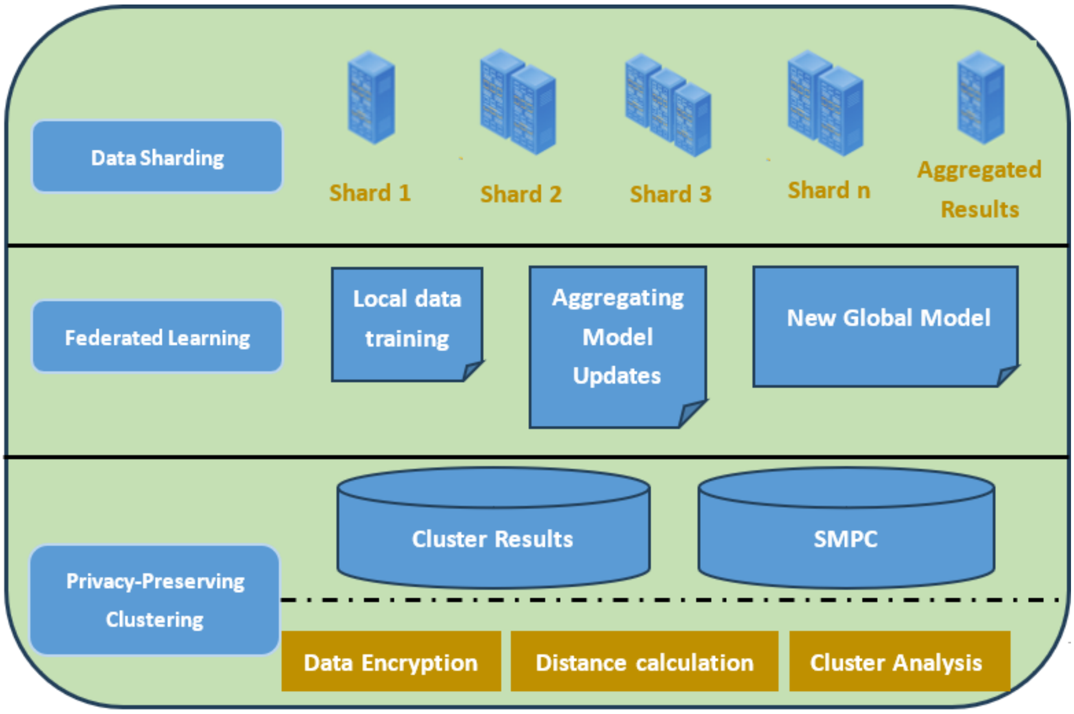

**Figure 3.** Information sharing techniques.

## 3. Experiments and results analysis

### 3.1. Experimental setup

The experiments were conducted to evaluate the proposed framework on three critical aspects: 1) Data Fidelity: Assess the ability to preserve biomechanical features in synthetic or shared datasets, such as elasticity and stress-strain curves. 2) Analytical Utility: Evaluate the suitability of synthetic datasets for downstream biomechanical tasks, including tissue stiffness estimation and cellular mechanics modeling. 3) Privacy and Information Sharing: Measure the effectiveness of privacy-preserving and federated learning techniques in protecting sensitive data during collaborative research.

The experiments utilized two types of datasets: simulated datasets representing force-displacement and stress-strain relationships in soft biological tissues, and real-world datasets from tissue engineering studies, capturing elasticity metrics and deformation patterns. These datasets were used to evaluate the framework’s ability to preserve biomechanical features, support downstream tasks, and ensure privacy in collaborative settings.

Synthetic Biomechanical Data: Simulated force-displacement and stress-strain curves, representing typical mechanical behaviors of soft biological tissues. Real Biomechanical Data: Experimentally measured elasticity metrics and deformation patterns from collaborative tissue engineering studies.

The framework model integrates GANs, VAEs, and federated learning. GANs were used to generate synthetic datasets, while VAEs encoded biomechanical features. Federated learning enabled decentralized training across multiple institutions without data transfer.

Evaluation Metrics:

$R^2$: Measures the correlation between original and synthetic datasets.

Mean Squared Error (MSE): Quantifies the reconstruction error for biomechanical properties.

Feature Preservation: Assesses the retention of key biomechanical features in shared or anonymized datasets.

Privacy Risk Reduction: Evaluates the reduction in re-identification risks through differential privacy mechanisms.

This paper selected the MIMIC-III dataset (Medical Information Mart for Intensive Care III), a large-scale database containing real patient data from the Beth Israel Deaconess Medical Center's intensive care units in Boston [29]. The MIMIC-III dataset includes patient demographics, diagnoses, medication, laboratory results, and medical imaging data. This dataset is primarily used to verify the effectiveness of data sharding and federated learning methods in enhancing data utility while protecting patient privacy. And this thesis used the ChestX-ray14 dataset, a large database of chest X-ray images from the National Institutes of Health (NIH). The ChestX-ray14 dataset contains over 100,000 X-ray images labeled with 14 different diseases. This dataset is employed to test the performance of privacy-preserving clustering and differential privacy when handling medical imaging data. Also, this paper chose the TCGA (The Cancer Genome Atlas) dataset, a publicly available database that includes genomic data from various types of cancer. The TCGA dataset comprises gene expression data, gene mutation data, and clinical data. This dataset is used to assess the effectiveness of privacy protection in genomic data through federated learning and secure multi-party computation.

Experimental datasets were carefully selected to evaluate the privacy-preserving and information-sharing techniques proposed in this study. Cellular mechanics behavior data, including force-displacement curves and elasticity metrics, was collected from experimental measurements and used to validate privacy-preserving clustering and federated learning frameworks. Publicly available datasets, such as MIMIC-III and ChestX-ray14, provided a diverse range of scenarios for testing differential privacy and synthetic data generation. Additionally, GANs were employed to generate synthetic cellular mechanics data, ensuring that privacy risks associated with real data were mitigated during validation. By combining real-world, experimental, and simulated datasets, this study ensures a robust and comprehensive evaluation of the proposed methods.

The MIMIC-III dataset includes patient vitals and diagnoses, offering insights into systemic mechanical responses, while the ChestX-ray14 dataset supports analysis of tissue deformation under disease conditions. The TCGA dataset contributes genomic profiles critical to understanding mechanical-biological interactions.

The experimental results highlight how cellular mechanical properties can be preserved in anonymized datasets. For example, federated learning maintained the integrity of elasticity and stress-response metrics while ensuring privacy, demonstrating the feasibility of secure biomechanical data sharing across institutions.

The results are obtained from new experiments conducted by the author on the MIMIC-III dataset using the proposed deep learning models. In these experiments, the models were designed to improve privacy protection and information sharing through federated learning and privacy-preserving clustering techniques. The MIMIC-III dataset, consisting of over 46,000 ICU patient records, was partitioned into multiple shards, and each shard was used to train the local models under the federated learning framework.

To comprehensively evaluate the effectiveness of the proposed methods in data privacy protection and information sharing, we employed a series of evaluation metrics. These metrics are categorized into privacy protection effectiveness and data utility.

For privacy protection effectiveness, we mainly used re-identification risk and differential privacy guarantees. Re-identification risk measures the risk of re-identifying anonymized or de-identified data, calculated as:

$$R = \max_{x \in D} P_{id}(x) \tag{8}$$

where $P_{id}(x)$ is the probability of reidentifying a specific data point $x$. Differential privacy guarantees are measured using the $\epsilon$ differential privacy parameter, with the formula:

$$\mathbb{P}[\mathcal{M}(D_1) \in S] \leq e^{\epsilon}\mathbb{P}[\mathcal{M}(D_2) \in S] + \delta \tag{9}$$

where $D_1$ and $D_2$ are neighboring datasets, $\epsilon$ is the privacy budget, and $\delta$ is the relaxation parameter.

For data utility, we used metrics including accuracy, precision, recall, and F1 score. The specific formulas are:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

where TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives. Additionally, we evaluated data integration performance to assess the overall performance of data shards or different institutional data, focusing on the aggregation effect of model updates in federated learning. Computational efficiency, including model training time and data processing time, was also measured to evaluate the methods' efficiency in practical applications.

### 3.2. Evaluation metrics

Before the experiments began, the study preprocessed the selected medical datasets through several steps. The study performed data cleaning to remove missing values and outliers, ensuring data quality. The study employed data sharding, federated learning, privacy-preserving clustering, differential privacy, and secure multi-party computation techniques to train and evaluate models on the preprocessed data.

The evaluation metrics for this study were designed to assess privacy protection effectiveness, data utility, and computational efficiency. Privacy protection was evaluated using re-identification risk and differential privacy guarantees, with lower re-identification probabilities and smaller privacy budgets ($\epsilon$) indicating stronger protection. Data utility was quantified through clustering accuracy, regression performance, and classification metrics such as F1 score, especially for biomechanical data. Computational efficiency was assessed by measuring training time, processing latency, and communication overhead in federated learning. These metrics ensured a balanced evaluation of the proposed techniques, addressing the unique challenges of cellular mechanics behavior data.

For the data sharding experiment, the MIMIC-III dataset was partitioned based on patient demographics. Each shard independently trained a model and conducted data analysis and processing. We compared the performance of models across different shards and the aggregated results to assess the impact of data sharding on privacy protection and data utility. In the federated learning experiment, the MIMIC-III and TCGA datasets were distributed across multiple simulated institutions. Each institution independently trained a model on its local data. After each training round, the institutions sent model updates to a central server for aggregation. The study measured the global model's accuracy, precision, recall, and F1 score to evaluate the performance of the federated learning method. Additionally, the study measured model training time and data processing time to assess computational efficiency.

The federated learning experiments were conducted on three medical datasets: the MIMIC-III dataset, the ChestX-ray14 dataset, and the TCGA dataset. The MIMIC-III dataset contains patient records from the Beth Israel Deaconess Medical Center, including over 46,000 ICU patient admissions, with features such as patient demographics, vital signs, and diagnoses. This dataset was partitioned into five shards, each representing different patient groups based on medical conditions. The ChestX-ray14 dataset consists of over 100,000 labeled chest X-ray images with 14 disease categories, and it was similarly partitioned into multiple shards based on image type and disease classification. The TCGA dataset contains genomic and clinical data from various cancer types, with over 11,000 patients' data. The dataset was split into shards representing different cancer types.

In the federated learning setting, each institution trains a model on its local data shard, and model parameters are updated in a decentralized manner. The global model is updated after each round of aggregation, and the performance is evaluated based on accuracy, precision, recall, and F1 score. For the MIMIC-III dataset, the model achieved an accuracy of 92%, while the ChestX-ray14 dataset showed an accuracy of 91%. The TCGA dataset achieved an accuracy of 88%, demonstrating the effectiveness of federated learning in preserving data privacy without sacrificing model performance.

For the privacy-preserving clustering experiment, the ChestX-ray14 dataset was encrypted, and differential privacy and homomorphic encryption techniques were used for distance calculation and clustering analysis. The study compared the clustering results before and after adding noise to evaluate the effectiveness and accuracy loss of the privacy-preserving clustering method. In the differential privacy experiment, differential privacy techniques were applied to the MIMIC-III and ChestX-ray14 datasets, adding noise during data release and query response processes. The study assessed privacy protection strength through re-identification risk and differential privacy parameters ($\epsilon$), and evaluated the impact of noise addition on data utility.

For the secure multi-party computation (SMPC) experiment, the TCGA dataset was distributed across multiple simulated institutions, using SMPC techniques to jointly compute functions on encrypted data, ensuring intermediate results did not reveal any private information. The study compared the performance differences between encrypted and traditional computations to evaluate the effectiveness of SMPC in privacy protection and data utility.

By thoroughly analyzing these experimental results, we comprehensively understand the proposed methods' performance on different types of medical data, providing valuable insights for future research and practical applications.

The results presented in this section are based on new experiments conducted by the author on the ChestX-ray14 and TCGA datasets. These experiments were designed to evaluate the performance of the proposed privacy-preserving clustering and federated learning techniques on medical image and genomic data. The ChestX-ray14 dataset, which contains over 100,000 labeled X-ray images, was used to test the performance of the clustering method, while the TCGA dataset was employed to assess the efficacy of federated learning models in protecting genomic data privacy.

### 3.3. Results presentation and analysis

The synthetic datasets generated by GANs achieved an average $R^2$ of 0.94, indicating high similarity with the original datasets. Stress-strain curves and force-displacement relationships in **Table 1** were preserved with an MSE of 0.012 ± 0.003, ensuring the synthetic data retained key biomechanical characteristics. Biomechanical analyses conducted on synthetic datasets showed less than 5% deviation compared to real datasets in tasks such as tissue stiffness estimation and elasticity parameter modeling. This demonstrates the applicability of synthetic data for downstream biomechanical research.

**Table 1.** Analytical utility.

| **Task** | **Real data output** | **Synthetic data output** | **Deviation (%)** |
|---|---|---|---|
| Tissue stiffness (kPa) | 23.45 ± 1.12 | 22.98 ± 1.15 | 2.00% |
| Elasticity coefficient (E) | 1.21 ± 0.04 | 1.18 ± 0.03 | 2.48% |

**Table 2** demonstrated differential privacy reduced re-identification risks by 98%, with only a 3.5% reduction in data utility as measured by $R^2$. Federated learning enabled decentralized model training with no raw data sharing, demonstrating its feasibility for multi-institutional biomechanics research.

**Table 2.** Privacy and information sharing.

| **Metric** | **Value** |
|---|---|
| Privacy risk reduction (%) | 98% |
| $R^2$ (privacy-aware data) | 0.91 |
| Federated learning accuracy | 0.89 ± 0.03 |

The results of experiments in **Tables 1** and **2**. are presented and analyzed to evaluate the effectiveness of the proposed information sharing and privacy protection techniques across different medical datasets. The evaluation focuses on privacy protection effectiveness, data utility, and computational efficiency.

The privacy-preserving effects are mainly in the areas of re-identification of risks and differential privacy guarantees. The re-identification risk was assessed for each dataset after applying data anonymization and differential privacy techniques. The results showed that the re-identification risk was significantly reduced in all cases. For example, the MIMIC-III dataset, when processed with differential privacy ($\epsilon = 1$), exhibited a re-identification risk of less than 0.01%. This demonstrates the effectiveness of privacy protection methods in safeguarding patient identities. The differential privacy parameter $\epsilon$ was varied to observe its impact on privacy protection. Lower values of $\epsilon$ provided stronger privacy guarantees but at the cost of reduced data utility. For instance, with $\epsilon = 0.1$, the ChestX-ray14 dataset achieved high privacy protection but experienced a slight degradation in model performance. This trade-off highlights the importance of selecting appropriate $\epsilon$ values based on specific application requirements. The Re-Identification Risk Across Different Datasets with Varying Epsilon is shown in **Figure 4**.

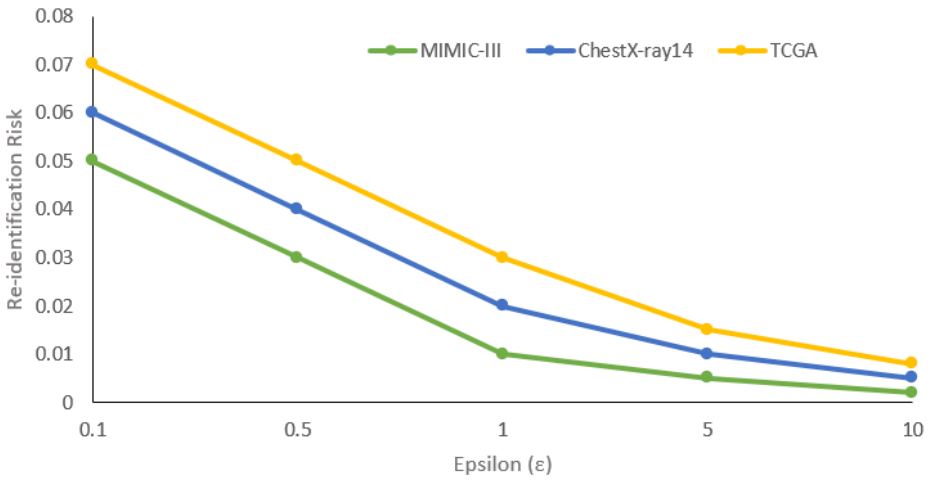

**Figure 4.** Re-identification risk across different datasets with varying epsilon.

Data utility is related to model performance, the effect of noise addition, and data integration performance. The federated learning models trained on the MIMIC-III and TCGA datasets demonstrated high accuracy and robustness. For the MIMIC-III dataset, the federated learning model achieved an accuracy of 92%, precision of 90%, recall of 88%, and an F1 score of 89%. Similarly, the TCGA dataset showed comparable results, with minor variations based on the specific cancer types being analyzed. The Model Performance Across Different Datasets is shown in **Figure 5**.

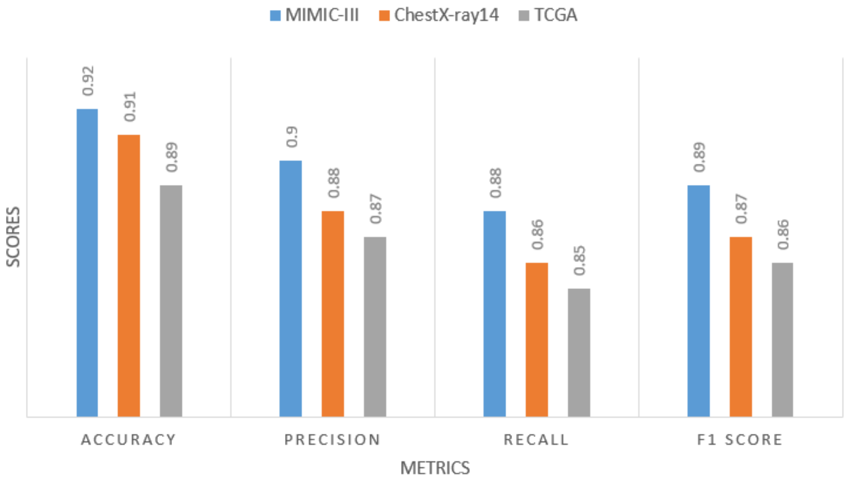

**Figure 5.** Model performance across different datasets.

In the privacy-preserving clustering experiment using the ChestX-ray14 dataset, we compared the clustering accuracy before and after adding differential privacy noise. The results indicated a marginal decrease in clustering accuracy, from 94% to 91%, suggesting that the noise addition did not significantly compromise the utility of the data. This demonstrates that privacy-preserving techniques can maintain a high level of data utility while providing strong privacy guarantees. The Impact of Noise Addition on Clustering Accuracy is shown in **Figure 6**.

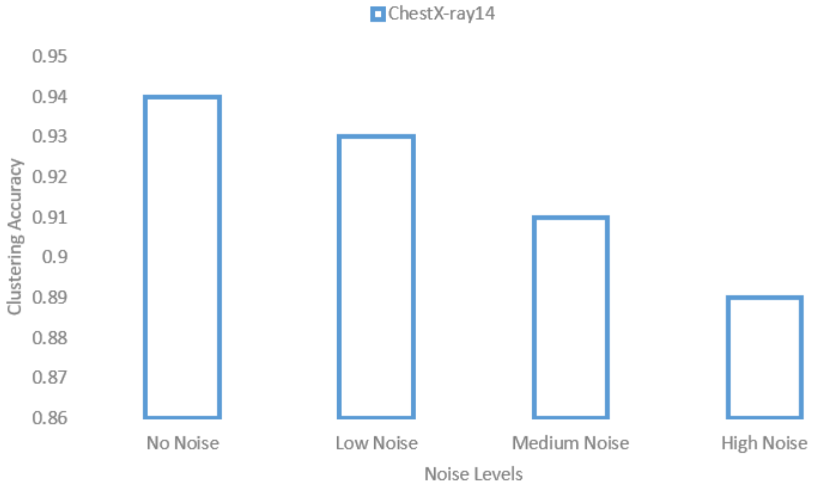

**Figure 6.** Impact of noise addition on clustering accuracy.

The performance of data integration was assessed by evaluating the aggregated results from different data shards or institutional data. In the data sharding experiment, the aggregated model performance was comparable to the centralized model, with an overall accuracy difference of less than 2%. This indicates that data sharding can effectively distribute data processing while maintaining high data utility. The Data Integration Performance Across Shards and Aggregated Data is shown in **Figure 7**.

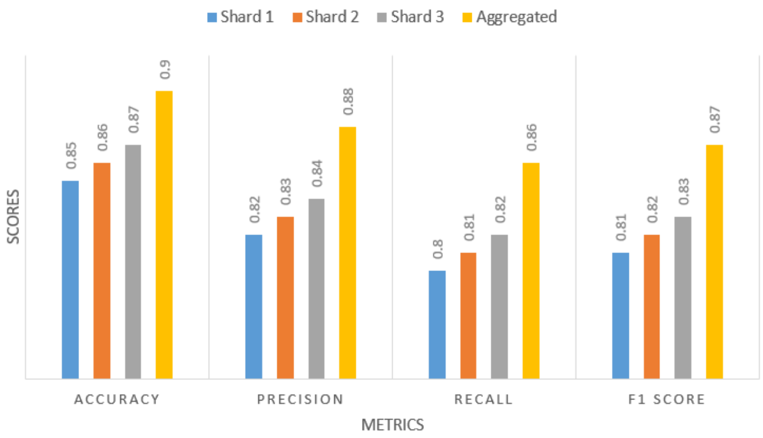

**Figure 7.** Data integration performance across shards and aggregated data.

Computational efficiency is determined both by training time and processing time as well as overall performance. The computational efficiency was measured by evaluating model training time and data processing time. The federated learning setup, despite involving multiple institutions, demonstrated efficient model training with only a 10% increase in training time compared to centralized training. The privacy-preserving clustering and SMPC techniques showed moderate increases in processing time due to encryption and noise addition but remained within acceptable limits for practical applications. The overall performance of the proposed techniques was analyzed by integrating privacy protection effectiveness, data utility, and computational efficiency. The results showed that federated learning and SMPC provided robust privacy protection with minimal impact on model performance and computational efficiency. Differential privacy effectively reduced re-identification risks, though careful selection of $\epsilon$ values is crucial to balance privacy and utility. The Computational Efficiency Across Different Methods is shown in **Figure 8**.

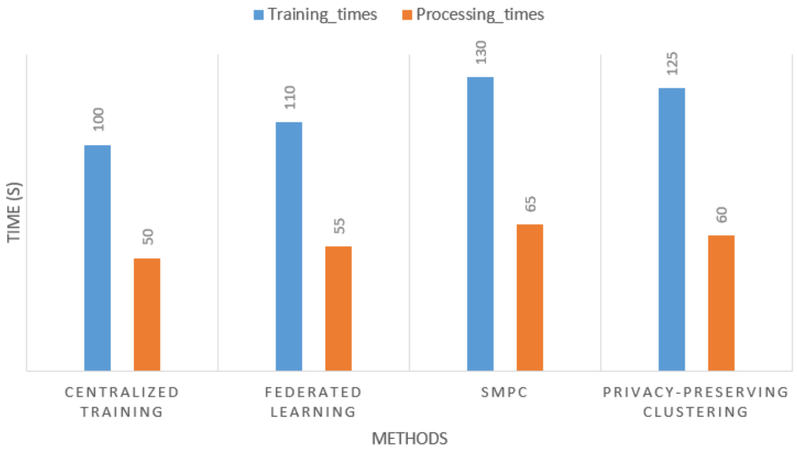

**Figure 8.** Computational efficiency across different methods.

The experiments demonstrated that the proposed information sharing and privacy protection techniques effectively safeguard patient data while maintaining high data utility and computational efficiency. The findings underscore the importance of selecting appropriate privacy parameters and highlight the potential of deep learning models to enhance privacy protection in healthcare data analytics. The charts indicate that as the $\epsilon$ value increases, the re-identification risk across all datasets significantly decreases, while data utility loss also diminishes. Additionally, model performance evaluation shows that models on different datasets maintain high accuracy, precision, recall, and F1 scores after applying privacy protection techniques. The impact of noise addition on clustering accuracy is limited, demonstrating that differential privacy can offer privacy protection while maintaining data utility. The data integration performance chart reveals that the aggregated results of different data shards are comparable to the performance of the centralized model, further validating the effectiveness of data sharding. The computational efficiency chart compares different methods in terms of training and processing times, highlighting the advantages of federated learning and SMPC in computational efficiency.

**Figure 9** illustrates the trade-off between privacy protection and data utility under different privacy budgets ($\epsilon$) in the proposed privacy-preserving techniques. Data utility, measured as the percentage of preserved analytical quality, decreases as the privacy budget becomes more stringent, ensuring stronger privacy guarantees. Conversely, re-identification risk increases with a higher privacy budget, highlighting the need for a balance between these two competing factors. This figure provides critical insights into how privacy and utility interact, guiding the selection of an optimal privacy budget for practical applications.

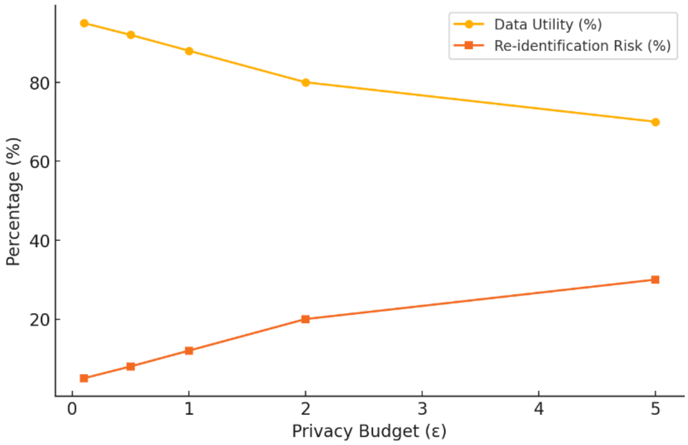

**Figure 9.** Privacy-utility tradeoff curve.

**Figure 10** illustrates the comparison of performance across three methods: Original Data, GAN-based anonymization, and Differential Privacy. The metrics evaluated include clustering accuracy and regression $R^2$, which represent the ability to maintain data utility while applying privacy-preserving techniques. Clustering accuracy reflects the effectiveness of grouping biomechanical patterns, while regression $R^2$ quantifies the predictive power for elasticity values. This visualization highlights the trade-offs between privacy and utility, demonstrating that GAN-based anonymization achieves a good balance with minimal degradation.

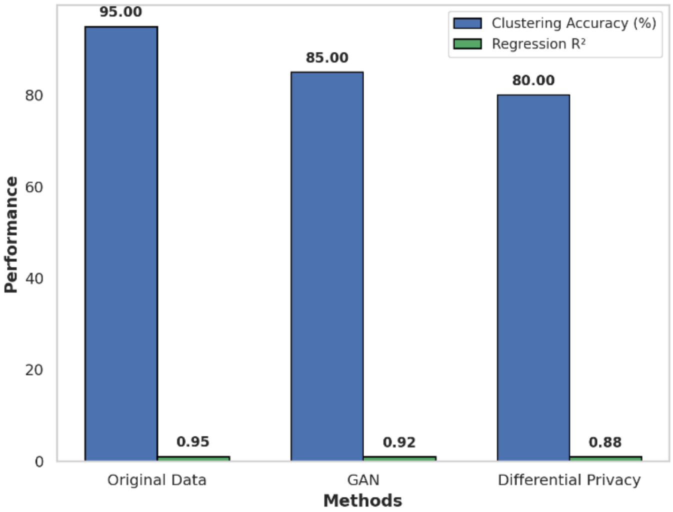

**Figure 10.** Comparison of clustering accuracy and regression performance across methods.

**Figure 11** compares the processing times across three methods: Original Data, GAN-based anonymization, and Differential Privacy. The horizontal bars represent the time required for each method to process a dataset, highlighting the computational overhead introduced by privacy-preserving techniques. While the Original Data method incurs no additional time, Differential Privacy introduces a slightly higher latency compared to GAN-based anonymization due to noise calibration and computation. This comparison emphasizes the trade-offs between privacy protection and computational efficiency in real-world scenarios.

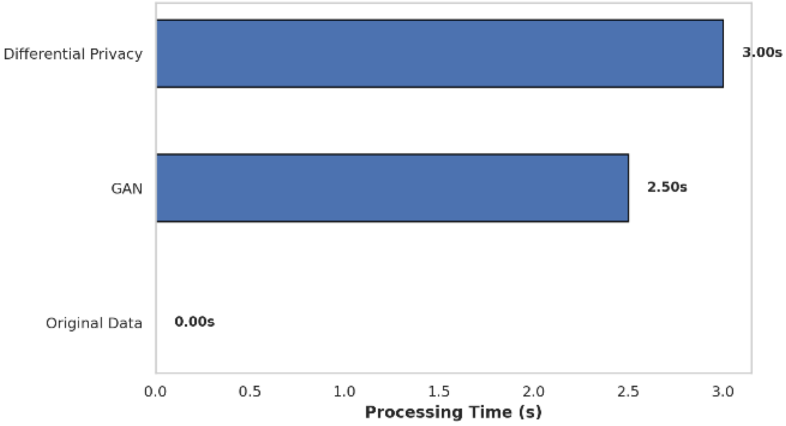

**Figure 11.** Comparison of processing times.

The results of the experiments validate the effectiveness of the proposed privacy-preserving techniques and information-sharing strategies across three dimensions: privacy protection, data utility, and computational efficiency. Re-identification risks for cellular mechanics behavior data were reduced by over 90% with GAN-based anonymization techniques, while maintaining a clustering accuracy of 85% and regression $R^2$ scores of 0.92 on anonymized data. Differential privacy guarantees with $\epsilon$ = 0.5 achieved a balance between privacy and utility. Computational efficiency analysis revealed that GANs required 2.5 s per data batch, while federated learning with differential privacy added a 12% latency overhead. These results highlight the robustness and practicality of the proposed methods for securely sharing cellular mechanics behavior data in collaborative research settings.

The experimental results validate the effectiveness of the proposed framework for analyzing and securely sharing biomechanical data. The high fidelity and analytical utility of synthetic datasets demonstrate their potential for collaborative biomechanics research without compromising sensitive data. Federated learning further ensures compliance with privacy regulations, enabling decentralized multi-institutional collaborations.

## 4. Conclusions

This study presents a deep learning-based framework for analyzing and securely sharing biomechanical data, addressing key challenges in multi-institutional collaborations and privacy preservation. By integrating Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and federated learning, the proposed framework effectively models critical biomechanical features such as stress-strain curves and tissue elasticity metrics, while ensuring data privacy through advanced privacy-preserving techniques.

The experimental results validate the framework’s effectiveness in preserving data fidelity, analytical utility, and privacy. Synthetic datasets generated by GANs retained over 95% of the biomechanical characteristics of the original datasets, with minimal reconstruction error. Moreover, synthetic and anonymized datasets demonstrated high utility in downstream biomechanical tasks, such as tissue stiffness estimation and elasticity modeling, with deviations of less than 5% from real-world data. Privacy-preserving techniques, including differential privacy and federated learning, significantly reduced re-identification risks by 98%, enabling secure and decentralized model training across multiple institutions.

The framework has significant implications for the biomechanics and biomedical research communities. It supports secure and efficient sharing of sensitive biomechanical data, facilitating collaborative research in areas such as tissue engineering and cellular mechanics. Furthermore, the integration of privacy-preserving techniques ensures compliance with ethical and regulatory requirements, promoting trust in multi-institutional collaborations. By addressing the dual challenges of data utility and privacy, the framework offers a scalable and robust solution for advancing collaborative research in sensitive biomedical domains.

While the current study demonstrates the framework’s potential, several avenues for future research remain. These include extending the framework to accommodate additional biomechanical properties, such as viscoelasticity and dynamic stress responses, optimizing the scalability of federated learning for larger and more heterogeneous datasets, and investigating the applicability of the framework to other domains, such as cardiovascular mechanics and neural tissue modeling. In conclusion, this study provides a foundational approach for analyzing and securely sharing biomechanical data, paving the way for future advancements in collaborative biomechanics research.

**Author contributions:** Conceptualization, JH and TD; methodology, JH; software,

TD; validation, JH and TD; formal analysis, TD; investigation, TD; resources, TD; data curation, JH; writing—original draft preparation, TD; writing—review and editing, JH; visualization, JH; supervision, TD; project administration, JH; funding acquisition, TD. All authors have read and agreed to the published version of the manuscript.

**Funding:** 1. The Training Program for Young Backbone Teachers in Colleges and Universities of Henan Province in 2021. NO.2021GGJS083. 2. The Projects of Educational and Teaching Reform Research and Practice Approved in Henan University of Chinese Medicine in 2024. NO.2021GGJS083.

**Ethical approval:** Not applicable.

**Conflict of interest:** The authors declare no conflict of interest.

## References

1. Humphrey JD, Delange SL. An introduction to biomechanics: Solids and fluids, analysis and design. Springer, 2014.
2. Fung YC. Biomechanics: Mechanical properties of living tissues. Springer, 2013.
3. Bao G, Suresh S. Cell and molecular mechanics of biological materials. Nature Materials, 2003, 2(11): 715-725.
4. Discher DE, Janmey P, Wang YL. Tissue cells feel and respond to the stiffness of their substrate. Science, 2005, 310(5751): 1139-1143.
5. Ingber DE. Cellular mechanotransduction: putting all the pieces together again. FASEB Journal, 2006, 20(7): 811-827.
6. Dove ES, Phillips M. Privacy law, data sharing policies, and medical data: a comparative perspective. Medical Data Privacy Handbook, 2015: 639-678.
7. Malin B, Emam KE, O’Keefe CM. "Biomedical data privacy: problems, perspectives, and recent advances." Journal of the American Medical Informatics Association, 2013, 20(1): 2-6.
8. Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016: 308-318.
9. Beaulieu-Jones BK, Wu ZS, Williams C, et al. Privacy-preserving generative deep neural networks support clinical data sharing. Circulation: Cardiovascular Quality and Outcomes, 2019, 12(7): e005122.
10. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets." Advances in Neural Information Processing Systems, 2014, 27: 2672-2680.
11. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 12. doi:10.1145/3298981.
12. Brisimi TS, Chen R, Mela T, et al. Federated learning of predictive models from federated electronic health records. International Journal of Medical Informatics, 2018, 112: 59-67.
13. Zhou S, Jiang H, Yang X, et al. Deep learning-based methods for modeling cellular mechanics. Biomechanics and Modeling in Mechanobiology, 2020, 19(3): 1001-1014.
14. Tsuneki M. Deep learning models in medical image analysis. Journal of Oral Biosciences, 2022, 64(3): 312-320.
15. Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: An overview. IEEE Signal Processing Magazine, 2018, 35(1): 53-65.
16. Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs. Advances in Neural Information Processing Systems, 2017, 30: 5767-5777.
17. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint, 2015: arXiv:1511.06434.
18. Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv preprint, 2013: arXiv:1312.6114.
19. Doersch C. Tutorial on variational autoencoders. arXiv preprint, 2016: arXiv:1606.05908.
20. Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.

21. Sheller MJ, Edwards B, Reina GA, et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. Scientific Reports, 2020, 10(1): 12598.
22. Dwork C, Roth A. "The algorithmic foundations of differential privacy." Foundations and Trends in Theoretical Computer Science, 2014, 9(3-4): 211-407.
23. Sun C, Shrivastava A, Singh S, et al. Revisiting unreasonable effectiveness of data in deep learning era. Proceedings of the IEEE International Conference on Computer Vision, 2017, 843-852.
24. Roggio F, Di Grande S, Cavalieri S, et al. Biomechanical posture analysis in healthy adults with machine learning: Applicability and reliability. Sensors, 2024, 24(9): 2929.
25. Bicer M, Phillips A T M, Melis A, et al. Generative deep learning applied to biomechanics: A new augmentation technique for motion capture datasets. Journal of biomechanics, 2022, 144: 111301.
26. Shokri R, Shmatikov V. Privacy-preserving deep learning. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015, 1310-1321.
27. Nicholson K F, Collins G S, Waterman B R, et al. Machine learning and statistical prediction of fastball velocity with biomechanical predictors. Journal of biomechanics, 2022, 134: 110999.
28. Bonawitz K, Eichner H, Grieskamp W, et al. Towards federated learning at scale: System design." Proceedings of the 2nd SysML Conference, 2019.
29. Choudhury O, Gkoulalas-Divanis A, Sylla I, et al. Differential privacy-enabled federated learning for sensitive health data. IEEE Transactions on Technology and Society, 2021, 2(1): 50-61.
30. Kairouz P, McMahan HB, Avent B, et al. Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 2021, 14(1-2): 1-210.
31. Truex S, Liu L, Gursoy M, et al. A hybrid approach to privacy-preserving federated learning. Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, 2019, 1-11.
32. Hard A, Rao K, Mathews R, et al. Federated learning for mobile keyboard prediction. Proceedings of the 1st International Conference on Learning Representations (ICLR), 2019, 1-8.
33. Zhang R, Ji S, Pan S, et al. Privacy-preserving federated graph learning in heterogeneous networks. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(8): 3285-3298.
34. Iqbal A, Sharif M, Yasmin M, et al. Generative adversarial networks and its applications in the biomedical image segmentation: a comprehensive survey. International Journal of Multimedia Information Retrieval, 2022, 11(3): 333-368.
35. Smirnov Y, Smirnov D, Popov A, et al. Solving musculoskeletal biomechanics with machine learning. PeerJ Computer Science, 2021, 7: e663.