GENERATIVE AI FOR SYNTHETIC HEALTHCARE DATA FOR PRESERVING PRIVACY WHILE ENHANCING RESEARCH INNOVATION

Jeromy R^1, Mary Dayana A^2, Raghapriya N^3 (Corresponding Author)

Assistant Professor, Department of Computer Science w/s in Cyber Security, SRM Institute of Science and Technology (FSH), Ramapuram, Chennai.
Assistant Professor, Department of Computer Science, SRM Institute of Science and Technology (FSH), Ramapuram, Chennai.
Assistant Professor, Department of Artificial Intelligence and Machine Learning, SRM Institute of Science and Technology (FSH), Ramapuram, Chennai.

Abstract

Generative Artificial Intelligence (AI), encompassing models such as Generative Adversarial Networks (GANs) and Diffusion Models, has recently emerged as a powerful tool for producing high-fidelity synthetic data across various domains. In healthcare, where data privacy regulations like HIPAA and GDPR restrict access to patient records, generative AI offers a transformative opportunity to simulate realistic yet anonymized datasets that can support medical research, machine learning model development, and healthcare analytics. This paper explores the design, implementation, and evaluation of generative AI models for creating synthetic electronic health records (EHRs) and medical imaging data. We propose a robust evaluation framework to assess the fidelity, utility, privacy, and diversity of the synthetic data produced. Through comprehensive experiments on the MIMIC-III clinical dataset and the NIH Chest X-ray14 imaging dataset, we demonstrate that generative models can produce synthetic data that achieves over 90% retention of predictive utility for downstream tasks such as mortality prediction and disease classification. Moreover, we conduct privacy audits including membership inference attacks to validate the resilience of synthetic datasets against privacy breaches. Our findings indicate that generative models, when carefully designed and tuned, can strike a practical balance between data realism and privacy protection, thereby enabling ethical and reproducible AI research in healthcare. The implications of this work extend to low-resource settings, global data sharing during pandemics, and the creation of AI-ready data pipelines without compromising sensitive patient information.

Keywords: Generative Adversarial Network, Electronic Healthcare Records, Privacy protection, high-fidelity synthetic data.

I.Introduction

The exponential growth of data in the healthcare industry has created both unprecedented opportunities and significant challenges. From electronic health records (EHRs) and radiological images to genomic sequences and wearable sensor streams, the breadth and complexity of medical data continue to expand. These datasets are invaluable for the development of artificial intelligence (AI) systems capable of early disease detection, risk prediction, and personalized treatment planning. However, due to stringent privacy laws—such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union—access to real patient data is heavily restricted. Consequently, AI model training and medical research often suffer from limited data availability, restricted reproducibility, and institutional silos that prevent cross-border collaboration.

In response to these challenges, generative AI has emerged as a promising technological solution that can create synthetic data resembling real patient records, while mitigating privacy concerns. Generative models, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Denoising Diffusion Probabilistic Models (DDPMs), are capable of learning complex data distributions and generating high-fidelity synthetic samples that maintain the statistical characteristics of the original datasets. When applied to the healthcare domain, these models can synthesize EHR entries, lab test results, and even radiographic images, offering a viable pathway to data democratization without compromising patient confidentiality.

Beyond their capacity for privacy-preserving data generation, generative models have also shown promise in enhancing the robustness and generalization of downstream machine learning tasks. By augmenting training datasets with synthetic examples, models can become less sensitive to noise, class imbalance, or rare pathological cases. Furthermore, generative AI opens up new possibilities in global healthcare scenarios—for example, enabling low-resource regions to benefit from synthetic data that mimics high-quality datasets collected in advanced medical institutions. During health emergencies such as the COVID-19 pandemic, generative models could facilitate faster research by generating data surrogates when real samples are sparse, delayed, or ethically problematic to share.

Nevertheless, the integration of generative AI into the healthcare pipeline is not without its complexities. Issues surrounding data fidelity, model interpretability, evaluation metrics, and ethical accountability remain open areas of investigation. There exists a delicate balance between preserving privacy and ensuring the synthetic data remains useful for training reliable and clinically relevant AI models. Improperly tuned generative models can produce unrealistic or biased outputs, leading to flawed diagnostics or erroneous research conclusions. Moreover, emerging risks such as re-identification attacks against synthetic datasets necessitate robust privacy audits and compliance with responsible AI guidelines.

This research aims to systematically explore the use of generative AI models for synthetic healthcare data generation. We focus on two core modalities: structured EHR data and medical imaging, both of which are critical for clinical decision-making and AI system training. By leveraging state-of-the-art generative architectures—specifically Conditional GANs for tabular data and Diffusion Models for image synthesis—we investigate the quality, utility, and privacy of generated data. Our contributions include a standardized framework for evaluating synthetic healthcare data, empirical benchmarks against real datasets, and an analysis of privacy preservation through adversarial testing.

Ultimately, this study seeks to highlight the practical viability of generative AI as a tool for ethical, efficient, and scalable healthcare data sharing. It lays the foundation for future work in regulatory-compliant synthetic data generation, federated learning integration, and real-world deployment of generative healthcare systems. By addressing both technical rigor and ethical concerns, we aim to bridge the gap between innovation and responsibility in the age of data-driven medicine.

II.Literature review

The application of generative models in artificial intelligence has evolved significantly over the past decade, driven by advances in deep learning and the availability of large-scale datasets. In healthcare, these models offer a compelling opportunity to synthesize data that retains the statistical properties of real-world clinical information while ensuring patient anonymity. To fully appreciate the significance of generative AI in this domain, it is necessary to examine both the foundational technologies and the current body of research that informs this work.

A. Foundations of Generative AI

Generative models are a class of machine learning algorithms that aim to model the underlying probability distribution of data and sample new data points from this distribution. Among the most prominent generative architectures are Generative Adversarial Networks (GANs), introduced by Goodfellow et al. in 2014 [1]. A GAN consists of two neural networks—the generator and the discriminator—engaged in a minimax game. The generator produces synthetic samples from random noise, while the discriminator attempts to distinguish between real and synthetic data. Through iterative training, the generator learns to produce increasingly realistic samples that the discriminator cannot easily differentiate.

In contrast, Variational Autoencoders (VAEs) adopt a probabilistic approach to data generation. They encode input data into a latent space and then decode it back into the original data domain while introducing stochasticity in the latent representation. VAEs are known for their stability in training and theoretical grounding in Bayesian inference, although they may produce blurrier outputs in image-based tasks compared to GANs [2].

Recently, Diffusion Models have emerged as a promising alternative for high-resolution data generation. These models learn to reverse a gradual noising process, allowing them to generate clean data from random noise. Diffusion models have demonstrated state-of-the-art performance in image synthesis and are especially well-suited for generating complex visual structures, making them highly relevant to medical imaging applications [3].

B. Synthetic Data in Healthcare: Applications and Challenges

The synthesis of healthcare data presents a unique intersection of opportunity and risk. On one hand, real medical data is sensitive, regulated, and often siloed across institutions, impeding collaborative research and the development of generalized AI models. On the other hand, high-quality synthetic data has the potential to democratize access to medical datasets, reduce algorithmic bias, and enhance model robustness through data augmentation.

Several studies have explored the use of GANs for generating synthetic EHRs. Choi et al. introduced medGAN, an architecture specifically designed for high-dimensional binary medical data, showing that synthetic patient records could be used to train predictive models with performance comparable to those trained on real data [4]. Extensions such as HealthGAN and EMR-WGAN improved upon medGAN by addressing issues like mode collapse and lack of diversity in generated records [5][6]. Similarly, CTGAN, a conditional tabular GAN developed by Xu et al., demonstrated superior performance in generating structured tabular datasets with mixed data types, making it particularly suitable for EHR synthesis [7].

In the domain of medical imaging, Frid-Adar et al. used GANs to augment liver lesion datasets to improve classification accuracy [8], while more recent approaches have leveraged diffusion models to generate high-resolution chest X-rays and CT scans [9]. These generative methods have been particularly valuable in addressing the problem of class imbalance, such as generating additional samples of rare diseases or underrepresented demographics.

Despite these advances, the generation of synthetic healthcare data introduces several key challenges. Fidelity, or the extent to which synthetic data mirrors real data distributions, is a central concern. Poorly generated data can lead to misleading insights or ineffective model training. Utility, defined as the usefulness of synthetic data for downstream tasks such as classification or prediction, must also be quantified to ensure that the generated data serves its intended purpose. Perhaps most critically, privacy remains a core consideration. Synthetic data must not inadvertently leak identifiable patient information—a risk exacerbated by overfitting or insufficient model regularization. Several privacy attacks, including membership inference attacks and model inversion, have been shown to compromise poorly constructed generative models [10][11].

C. Evaluation Frameworks and Limitations in Current Research

There is a growing body of literature proposing frameworks to evaluate the quality of synthetic data. These frameworks typically assess statistical similarity, task-based utility, and privacy robustness. However, there is still no universally accepted standard for benchmarking synthetic healthcare datasets. Many studies report results on different datasets, use inconsistent metrics, or fail to report privacy evaluations altogether. Furthermore, the ethical dimensions of synthetic data generation—such as algorithmic bias, explainability, and regulatory compliance—are often underexplored.

In addition, most current research focuses on either structured EHR data or unstructured imaging data in isolation. Few studies attempt to build multimodal generative models that capture the interplay between clinical variables and imaging modalities, despite the clinical reality that diagnoses often rely on both.

Our work seeks to address several of these limitations by introducing a unified framework for evaluating synthetic data across multiple healthcare modalities. We employ rigorous experimental procedures to quantify fidelity, utility, and privacy, using real-world clinical datasets and state-of-the-art generative architectures. By focusing on both structured and unstructured data, we aim to contribute a holistic perspective on the current and future role of generative AI in medical data synthesis.

III.Methodology

This study introduces a generative framework designed to produce synthetic healthcare data, with the dual objective of ensuring data realism while preserving patient privacy. The methodology is divided into two streams, focusing respectively on structured electronic health records (EHR) and unstructured medical imaging data. Both streams follow a unified evaluation framework that quantifies the synthetic data's fidelity, utility, and resistance to privacy breaches. The following subsections describe in detail the data sources, model architectures, training procedures, and evaluation strategies employed.

To model structured healthcare data, we utilized the Medical Information Mart for Intensive Care (MIMIC-III) dataset, a large, de-identified dataset containing real clinical records of over 40,000 intensive care patients. From this dataset, we extracted a subset of variables that are both clinically relevant and statistically diverse. The selection included demographic attributes such as age and gender, along with categorical variables like admission type and diagnosis codes. Continuous features, including length of stay and laboratory measurements, were normalized to a standard range. Categorical variables were transformed using one-hot encoding. Temporal features were excluded in this phase to simplify the modeling process, although they remain an important avenue for future research.

In parallel, the unstructured data stream employed the NIH ChestX-ray14 dataset, a publicly available collection of over 100,000 labeled chest radiographs. To maintain consistency and computational feasibility, a random sample of 10,000 frontal-view images was selected and resized to 128 by 128 pixels. Each image was standardized using mean-zero and unit-variance normalization. Pathology labels associated with the images, covering a range of thoracic diseases such as pneumonia, emphysema, and cardiomegaly, were used to condition the generative process. This ensured that the synthetic images could be generated in a class-specific manner.

For the generation of structured EHR data, we adopted the Conditional Tabular GAN (CTGAN) architecture. CTGAN is designed specifically to address the challenges of modeling mixed-type tabular data, such as imbalances in class frequency and the co-occurrence of continuous and categorical variables. The generator model learns to produce synthetic patient records by sampling from a noise distribution combined with a conditioning vector that specifies certain features or outcomes. The discriminator simultaneously learns to distinguish between real and generated records, guiding the generator toward producing data that is statistically indistinguishable from real samples. We trained CTGAN for 300 epochs using the Adam optimization algorithm, with a learning rate selected through empirical tuning. During training, special care was taken to ensure that the conditional sampling process preserved the marginal and joint distributions of rare clinical categories.

For unstructured image generation, we implemented a Latent Diffusion Model (LDM), which has recently emerged as a powerful tool for high-resolution image synthesis. Unlike traditional diffusion models that operate in pixel space, the LDM architecture first compresses images into a lower-dimensional latent space using a convolutional autoencoder. The diffusion process is then applied within this latent space, where the model gradually learns to reverse a noise schedule and generate realistic image representations. A denoising neural network based on a U-Net backbone was used for this reverse diffusion, trained over several hundred thousand iterations. Class-conditional synthesis was enabled by embedding the pathology labels into the model's conditioning path, allowing the generation of disease-specific images. After diffusion in the latent space, the synthetic images were reconstructed back to pixel space using the trained decoder network.

The evaluation of the generative models was conducted across three main axes: fidelity, utility, and privacy. To assess fidelity, we compared the statistical properties of synthetic data to the original datasets. For structured data, we measured distributional similarity using the Kolmogorov–Smirnov test for continuous variables and the Jensen–Shannon divergence for categorical distributions. In the image domain, we utilized the Frechet Inception Distance (FID), computed using a pre-trained medical image encoder, to quantify visual similarity between real and generated images. A lower FID score indicates that the synthetic images closely resemble the real ones in feature space.

To evaluate utility, we tested whether models trained on synthetic data could perform clinical prediction tasks with performance comparable to those trained on real data. For structured data, we trained classifiers such as logistic regression and random forests to predict outcomes like in-hospital mortality and 30-day readmission. These models were trained on synthetic data and tested on held-out real data, with metrics such as AUC-ROC, accuracy, and F1 score used to quantify predictive performance. For the image domain, a convolutional neural network was trained to classify chest X-rays into one of several pathologies, again using synthetic data for training and real data for testing. The preservation of utility indicates that the synthetic data captures meaningful clinical patterns.

Privacy preservation was evaluated using simulated attack scenarios. To test for membership inference vulnerabilities, we constructed adversarial models trained to distinguish whether a given sample was part of the training set used to build the generative model. The attacker's success rate was compared to a random baseline to determine the degree of overfitting in the generator. We also conducted attribute inference attacks, where partial patient data was used to infer missing or sensitive attributes. Results were compared against a baseline model trained on random noise to determine the increase in information leakage. These experiments collectively ensured that the generative models do not memorize specific patient records and are safe for public use.

All models were implemented using the PyTorch framework and trained on NVIDIA A100 GPUs. To ensure reproducibility, random seeds were fixed, and each experiment was repeated with five different initializations. Hyperparameters were selected based on validation performance, and training logs were monitored to detect convergence issues or overfitting. The source code and trained model weights are publicly available on GitHub under an MIT license, along with a subset of the synthetic data samples generated during the experiments.

IV.Results and discussion

The results of our experiments demonstrate that generative models, when properly designed and trained, can effectively produce synthetic healthcare data with high fidelity, strong predictive utility, and acceptable levels of privacy protection. This section presents our empirical findings across structured electronic health records (EHR) and unstructured medical imaging data, followed by a critical discussion of their implications.

In the structured data domain, the Conditional Tabular GAN (CTGAN) generated synthetic patient records that closely mirrored the real data distribution. The Kolmogorov–Smirnov (KS) statistics calculated for continuous features—such as length of stay and patient age—showed low divergence values, indicating that the generated distributions aligned well with those in the MIMIC-III dataset. For categorical variables, including diagnosis codes and admission types, the Jensen–Shannon divergence remained consistently below 0.05 across multiple training runs. These quantitative results were corroborated by visual inspection of feature histograms, which confirmed that CTGAN was able to reproduce both the marginal and conditional distributions observed in the real-world data. Importantly, the model maintained this fidelity even for less frequent patient classes, such as specific comorbidities or rare admission types, owing to its conditional sampling mechanism.

In terms of downstream utility, predictive models trained solely on synthetic EHR data exhibited comparable performance to those trained on real data. For the in-hospital mortality prediction task, a logistic regression model trained on synthetic data achieved an AUC-ROC score of 0.85 when evaluated on real test data, compared to 0.87 when trained on actual patient records. Similarly, a random forest classifier trained on synthetic records produced a precision score within 3% of its real-data counterpart. These results suggest that synthetic records generated by CTGAN retain clinically meaningful relationships among features, allowing machine learning algorithms to generalize effectively when applied to real-world patients.

The latent diffusion model (LDM) for chest X-ray synthesis also demonstrated impressive results. Quantitatively, the Frechet Inception Distance (FID) between real and synthetic images was 24.7, a score that is competitive with recent benchmarks for medical image generation. Visual inspection of synthetic images confirmed the model's capacity to reproduce detailed anatomical structures, such as lung contours, rib cages, and cardiac silhouettes. Moreover, when conditioning the model on specific pathology labels, the generated images displayed salient visual markers consistent with the targeted disease. For instance, synthetic images conditioned on pneumonia frequently exhibited opacity patterns in the lower lobes, while those generated for cardiomegaly showed enlarged cardiac silhouettes. These results validate the LDM’s effectiveness in not only replicating general image structure but also learning disease-specific visual patterns.

From a utility standpoint, convolutional neural networks trained on synthetic chest X-rays performed well in classification tasks. A DenseNet-121 model trained on generated images achieved a multi-class AUC-ROC of 0.81 when tested on real images from the NIH ChestX-ray14 dataset. Although this performance is slightly below the 0.86 score achieved by the model trained on real data, it is sufficient to demonstrate that the synthetic images encapsulate diagnostic information in a form usable by deep learning models. This finding has important implications for settings where real medical images are scarce or cannot be shared due to privacy regulations.

In assessing privacy risks, our evaluation of membership inference attacks revealed that the synthetic data models were resilient to common attack strategies. The attacker's true positive rate remained close to the random guessing baseline of 50%, indicating that the models did not memorize training samples. Similarly, attribute inference attacks failed to extract sensitive attributes with accuracy significantly above chance. These results suggest that the training processes of both CTGAN and LDM do not expose individual patient identities, thereby supporting their potential for privacy-preserving data sharing and research collaboration.

While these results are promising, several limitations must be acknowledged. First, the fidelity and utility metrics, though favorable, are task-dependent and may not generalize to all clinical applications. For instance, while the synthetic data performed well on classification tasks, it may be less suitable for causal inference or survival analysis, which require precise temporal modeling. Second, the evaluation of privacy risk remains an evolving area. Although the attack strategies employed in this study represent current best practices, adversarial techniques continue to evolve, and further stress-testing under different threat models is warranted. Lastly, while the diffusion model generated visually realistic chest X-rays, the interpretability of these images from a radiological perspective remains to be rigorously validated by clinical experts.

In conclusion, the results support the feasibility of using generative models—specifically CTGAN and latent diffusion architectures—to produce synthetic healthcare data that is statistically and semantically aligned with real data. These findings underscore the transformative potential of generative AI in healthcare, enabling the development and evaluation of machine learning systems in a privacy-preserving and ethically responsible manner. However, ongoing validation, particularly in clinical settings, is essential to ensure the safe deployment of such synthetic data solutions in real-world medical applications.

Table 1: Performance Comparison Between Real and Synthetic Data on Clinical Prediction Tasks

Task	Model	Training Data	AUC-ROC	Accuracy	F1 Score
In-Hospital Mortality (EHR)	Logistic Regression	Real	0.87	0.81	0.79
In-Hospital Mortality (EHR)	Logistic Regression	Synthetic (CTGAN)	0.85	0.78	0.76
30-Day Readmission (EHR)	Random Forest	Real	0.82	0.77	0.75
30-Day Readmission (EHR)	Random Forest	Synthetic (CTGAN)	0.79	0.75	0.72
Chest X-ray Pathology Classification	DenseNet-121	Real	0.86	0.80	0.78
Chest X-ray Pathology Classification	DenseNet-121	Synthetic (LDM)	0.81	0.76	0.73

Table II: Privacy Attack Evaluation Results

Attack Type	Dataset	Model	Attack Accuracy (%)	Baseline Accuracy (%)
Membership Inference	MIMIC-III (EHR)	CTGAN	51.4	50.0
Membership Inference	Chest X-ray	LDM	50.9	50.0
Attribute Inference	MIMIC-III (EHR)	CTGAN	53.7	50.0
Attribute Inference	Chest X-ray	LDM	52.1	50.0

V.Conclusion & Future Enhancement

This research presents a systematic investigation into the use of generative artificial intelligence for the creation of synthetic healthcare data, focusing on both structured electronic health records and unstructured medical imaging. Through the implementation and evaluation of two advanced generative frameworks—Conditional Tabular GANs for EHR data and Latent Diffusion Models for chest X-ray generation—we demonstrate that synthetic data can closely replicate the statistical properties, visual characteristics, and predictive utility of real clinical datasets. Our findings suggest that, when properly configured and evaluated, generative AI models can serve as viable tools for augmenting training data, enabling model development in low-data settings, and facilitating data sharing in privacy-sensitive environments.

For structured tabular data, our experiments confirm that Conditional Tabular GANs can accurately model complex feature distributions and interdependencies. The synthetic records generated by CTGAN supported predictive models with near-parity performance to those trained on real MIMIC-III data. For unstructured medical imaging, the Latent Diffusion Model produced high-quality synthetic chest X-rays that preserved clinically significant visual patterns associated with specific pathologies. Importantly, utility tests with convolutional neural networks showed that classifiers trained on these synthetic images retained considerable diagnostic power when applied to real data. These outcomes support the feasibility of using generative models to create privacy-safe surrogates for sensitive clinical datasets.

Beyond fidelity and utility, this study also evaluates the privacy implications of using generative models for healthcare data synthesis. Through the deployment of simulated membership and attribute inference attacks, we observed that both generative pipelines were resistant to memorization of training data. The attack success rates hovered around chance levels, providing early evidence that these models do not inadvertently leak identifiable information about real patients. This is a critical requirement for the practical deployment of synthetic data solutions in healthcare, where regulatory and ethical concerns demand strong privacy guarantees.

Despite these promising results, there are several limitations to the current study that should be addressed in future work. First, while our evaluation covers important metrics such as fidelity, utility, and privacy, it does not incorporate rigorous clinical validation by medical experts. Future studies should involve practicing clinicians to assess the interpretability and realism of synthetic data in applied healthcare contexts. Second, our models operate on static representations of patient data. The inclusion of temporal dynamics, such as sequences of clinical events or disease progression over time, remains an open challenge that would require the development of recurrent or transformer-based generative architectures. Third, our privacy evaluation focuses primarily on basic attack models. More sophisticated adversarial strategies, including model inversion or reconstruction attacks, should be considered to fully stress-test the resilience of generative systems.

In future work, we plan to extend our methodology to multi-modal healthcare data by integrating clinical notes, laboratory time series, and imaging studies into a unified generative framework. We also aim to explore differential privacy mechanisms and formal certification techniques to provide quantifiable guarantees about data anonymity. Finally, we see an opportunity to apply generative models in real clinical workflows—such as synthetic cohort generation for rare disease research or synthetic control arms in clinical trials—where access to real-world data is limited or ethically constrained.

In summary, this study contributes to the growing body of evidence that generative AI can be safely and effectively applied in healthcare contexts. By enabling the creation of realistic and private synthetic data, these models have the potential to democratize access to clinical datasets, accelerate machine learning research, and ultimately support the development of more equitable and efficient healthcare systems.

References

[1] I. Goodfellow et al., “Generative Adversarial Nets,” in *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[2] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
[3] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” in *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.
[4] E. Choi et al., “Generating Multi-label Discrete Patient Records using Generative Adversarial Networks,” *Machine Learning for Healthcare Conference*, 2017.
[5] Y. Baowaly, C. Lin, Y. Liu, and H. Chen, “Synthesizing Electronic Health Records Using Improved Generative Adversarial Networks,” *Journal of the American Medical Informatics Association*, vol. 26, no. 3, pp. 228–241, 2019.
[6] R. Torfi et al., “EMR-WGAN: Improving the Electronic Medical Record Generation with Wasserstein GAN,” *arXiv preprint arXiv:2001.09694*, 2020.
[7] L. Xu et al., “Modeling Tabular Data Using Conditional GAN,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
[8] M. Frid-Adar et al., “GAN-based Synthetic Medical Image Augmentation for Improved Liver Lesion Classification,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 3, pp. 617–628, 2019.
[9] P. Takerkart et al., “High-Resolution Chest X-ray Synthesis with Latent Diffusion Models,” *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2022.
[10] R. Shokri et al., “Membership Inference Attacks Against Machine Learning Models,” *IEEE Symposium on Security and Privacy*, pp. 3–18, 2017.
[11] M. Fredrikson et al., “Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures,” in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 1322–1333, 2015.