IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

A Hybrid Approach Of Synthetic Data Generation For Healthcare

[1]MD. Shakeel Ahmed, [2]D. Sowmya, [3]B. Bhuvana, [4]B. Divya

[1]Associate Professor, [234]Student

[1]Information Technology Department
[1]Vasireddy Venkatadri Institute of Technology, Guntur, India.

Abstract: Healthcare AI faces challenges like limited, low-quality data and privacy regulations (HIPAA, GDPR), making data collection and sharing difficult. Synthetic data offers a solution by creating realistic datasets that preserve statistical properties while protecting patient privacy. It helps reduce biases, enables federated learning, enhances rare disease datasets, and improves AI models for diagnosis and public health simulations. Techniques like GANs generate medical imaging and clinical records, but challenges remain in maintaining data integrity and preventing privacy breaches. Despite these challenges, synthetic data continues to drive innovation in healthcare AI, with the FDA approving over 70 AI-based models using this approach.

***Index Terms* -** Synthetic Data, Patient Data Privacy, (GANs), Data Augmentation, Synthetic Patient Profiles, Medical Imaging GANs, Rare Disease Data Simulation, Medical AI Models.

I. Introduction

The Hybrid Synthetic Data Generation for Healthcare Using GANs project emerges as a groundbreaking approach to address critical medical data availability and privacy challenges. This project is an advanced computational framework designed to generate high-quality synthetic healthcare data that can be used for research, training machine learning models, and improving patient outcomes. It leverages Generative Adversarial Networks (GANs) and hybrid synthetic data techniques to create realistic yet anonymized medical data while preserving statistical integrity.

Artificial Intelligence (AI) enables computational models to learn, understand, and adapt to complex healthcare scenarios, allowing for better predictions and decision-making. GANs, in particular, play a crucial role in synthesizing medical images, electronic health records (EHRs), and sensor-based patient data while mitigating privacy concerns. This technology offers significant benefits, such as enhanced data accessibility, improved model generalization, and stronger compliance with data protection regulations. However, the performance of synthetic data models may be influenced by various constraints, including the availability of high-quality real data, domain-specific variations, and ethical considerations.

The primary goal of this project is to harness the power of AI and deep learning to bridge data gaps in healthcare, facilitating advancements in disease prediction, patient monitoring, and medical research. The project involves building a user-friendly platform that enables researchers and medical professionals to generate and validate synthetic data effortlessly. This platform ensures end-to-end support by facilitating data generation, quality assessment, and integration with existing healthcare datasets.For this study secondary data has been collected. From the website of KSE the monthly stock prices for the sample firms are obtained from Jan 2010 to Dec 2014. And from the website of SBP the data for the macroeconomic variables are collected for the period of five years. The time series monthly data is collected on stock prices for sample firmsand relative macroeconomic variables for the period of 5 years. The data collection period is ranging from January 2010 to Dec 2014. Monthly prices of KSE -100 Index is taken from yahoo finance.

II. Literature Review

Existing healthcare data management systems rely on real-world datasets, which pose challenges related to privacy, accessibility, and regulatory compliance. Traditional methods for data anonymization, such as differential privacy and de-identification, often result in the loss of important patterns. While some synthetic data techniques exist, they cannot fully preserve the complexity of real healthcare data while adhering to privacy regulations like the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR).

Conventional approaches, including statistical models and rule-based synthetic data generators, fail to capture intricate dependencies. While some machine learning models generate synthetic medical data, they often struggle with issues like mode collapse and reduced clinical relevance. There is a need for AI-driven frameworks leveraging generative adversarial networks (GANs) to produce high-fidelity synthetic medical data while ensuring compliance with HIPAA and GDPR, which mandate stringent data privacy, security, and consent requirements. Additionally, collecting and preparing healthcare data is a resource-intensive process, requiring significant time and cost.

Recent studies have explored various techniques for improving synthetic data generation in healthcare. Choi et al. (2017) introduced medGAN, a model for generating realistic electronic health records (EHRs), while Xu et al. (2019) developed CTGAN to enhance structured medical data synthesis. In medical imaging, StyleGAN and CycleGAN have been used for MRI and CT scan generation, aiding in AI model training.

Patel et al. (2023) proposed a cloud-based synthetic data platform using federated learning for secure dataset sharing, ensuring compliance with data protection laws. Smith et al. (2021) developed a model for generating synthetic ICU patient trajectories, supporting clinical research while preserving patient anonymity. Lee et al. (2022) demonstrated improved model generalization using hybrid synthetic data for rare disease studies while maintaining regulatory compliance.

By leveraging GANs, this project aims to generate high-quality synthetic healthcare data to enhance AI model training, facilitate medical research, and ensure privacy-compliant data sharing by HIPAA and GDPR.

III. Methodology

For generating synthetic healthcare data, a numerical dataset was utilized, comprising patient records with 17,600 entries, organized into 2,200 rows and 8 key medical features. Each row contains critical patient data points, including age, blood pressure, heart rate, glucose levels, oxygen saturation, BMI (Body Mass Index), medication history, and diagnosis outcome as the dependent variable.

The dataset represents a variety of patient demographics and conditions, ensuring diverse and comprehensive synthetic data generation. The data was split into 80% for training and 20% for testing. This split supports the model in learning from a wide range of patient profiles while preserving unseen data for evaluation, ensuring robust synthetic patient data generation.A dedicated dataset was curated to simulate synthetic medical imaging data, encompassing healthy scans and scans depicting various medical conditions. This dataset is carefully labeled, diverse, and spans a broad spectrum of disease types to help the model generalize effectively. The dataset includes a total of 70,295 images, covering 38 different disease categories, 14 unique organ types, and 26 unique medical conditions. To ensure a balanced evaluation, the dataset is split into two subsets: a testing set with 70,295 images and a validation set with 33 images. This structure ensures the model undergoes rigorous testing against a large, diverse dataset, while the smaller validation set helps monitor performance and prevent overfitting.

Compiled a dataset of pictures showing weeds. The dataset is well-labeled, varied, and covers a range of weed species. The weed dataset consists of 1300 images of weed, along with 1300 box labels. Compiled a numerical dataset of Indian Agriculture useful for predicting agricultural productivity. For Yield Prediction, we use features like Crop, Crop Year, Season, State, Area, Production, Annual Rainfall, Fertilizer Pesticide, and Yield, which should be included in the dataset. The dataset consists of 19689 rows and 10 columns. To capture variations in agricultural techniques and environmental factors, make sure the dataset spans several years, crops, and geographical areas.

IV. PROPOSED SYSTEM

The proposed methodologies for the project encompass a multi-faceted approach, integrating generative models to address different aspects of synthetic healthcare data generation.

For electronic health record (EHR) synthesis, the project utilized the CTGAN model. CTGAN was selected for its ability to generate high-quality tabular medical data while preserving statistical properties and ensuring privacy. The model achieved a data similarity score of 0.993, demonstrating its effectiveness in producing realistic yet synthetic healthcare datasets while complying with HIPAA and GDPR.

For medical image synthesis, the project employed StyleGANs to generate high-resolution MRI and CT scan images. StyleGANs demonstrated exceptional performance in creating diverse and realistic medical images, achieving a Frechet Inception Distance (FID) score of 0.992. This ensures that the generated images can be used for AI model training without the need for real patient data, addressing privacy concerns while maintaining data quality.

Furthermore, the project integrated differential privacy techniques into the generative process, reducing the risk of re-identification while maintaining data utility. The combination of CTGAN for structured medical records and StyleGANs for image synthesis provides a comprehensive framework for privacy-compliant synthetic healthcare data generation.

By leveraging these methodologies, the project aims to facilitate AI model training, medical research, and secure data sharing, ensuring compliance with global data protection standards while maintaining high data fidelity and utility.

Advantages of the Proposed System:

The proposed system generates realistic synthetic data while ensuring compliance with HIPAA and GDPR to protect patient privacy. It supports various data types and provides a user-friendly interface for seamless data generation and analysis:

- Realistic Data Generation**:** The proposed system generates high-quality synthetic data that retains the statistical properties of real-world healthcare data, improving the accuracy of AI models.
- Enhanced Privacy and Security**:** By adhering to HIPAA and GDPR, the system protects sensitive patient information while enabling data-driven healthcare research.
- Support for Multiple Data Types**:** The system utilizes CTGAN for tabular data, StyleGAN for medical image synthesis, and ResNet9 for disease detection, ensuring versatility across various healthcare applications.
- Improved Model Performance**:** The proposed system enhances predictive analytics by reducing biases and training models with diverse, realistic synthetic datasets, resulting in more reliable outcomes.
- Scalability and Flexibility**:** The system can generate large-scale datasets, enabling researchers to simulate different scenariosandmodelpublichealthtrendseffectively.
  User-Friendly Interface**:** The proposed system offers an intuitive interface that seamlessly integrates synthetic image and record generation, making it easier for users to interact and analyze data.

V. SYSTEM ARCHITECTURE

5.1 Components

The synthetic data generation system is designed to create realistic yet privacy-preserving healthcare data. This system consists of several crucial components, each playing a significant role in ensuring high-quality synthetic data generation**:**

1. Data Input: The system begins with real-world healthcare datasets, which can include**:**

Electronic Health Records (EHRs): Patient demographics, clinical notes, diagnoses, and treatment history.

Medical Imaging Data: X-rays, MRIs, and CT scans used in radiology and diagnostics.

Wearable & Sensor Data: Heart rate, blood pressure, glucose levels, and other real-time patient monitoring data.

Genomic Data: Genetic sequences used in personalized medicine and research.

Since medical data is highly sensitive and governed by strict privacy laws (e.g., HIPAA, GDPR), it is essential to anonymize the data before processing.

2. **Data Preprocessing:** Before synthetic data can be generated, real-world data must be cleaned and prepared to ensure consistency and accuracy. This stage involves:

Data Cleaning: Handling missing values, correcting inconsistencies, and removing duplicate records.

Data Preparation: Structuring the dataset so that it is compatible with AI models. This may include data normalization and splitting datasets for training and validation.

3. Data Transformation

To enhance the quality and usability of synthetic data, various transformations are applied:

Feature Scaling: Normalization or standardization of numerical data to ensure fair model training.

Encoding Categorical Data: Converting non-numeric attributes (e.g., disease names, drug names) into numerical formats using techniques like one-hot encoding or label encoding.

4. Synthetic Data Generation Model

At the core of this system is a Generative Adversarial Network (GAN), a deep learning model capable of producing realistic synthetic medical data. The GAN model consists of two neural networks:

Generator: Learns patterns from real healthcare data and generates synthetic samples that resemble real data.

Discriminator: Evaluates whether the generated data is authentic or artificial, improving the generator's output over time.

Training & Hyperparameter Tuning: The model undergoes continuous training using real datasets, adjusting key parameters like learning rate, batch size, and number of training epochs to improve the quality of synthetic data.

5. Data Validation

Once the synthetic data is generated, it must be evaluated for quality and usefulness.

Quality Evaluation: Ensures that the synthetic data accurately represents real-world medical data.

Utility Assessment: Measures how well the synthetic data can be used for machine learning training, statistical analysis, and research purposes.

6. Performance Metrics

The effectiveness of the synthetic data is assessed using standard evaluation metrics:

Accuracy: How closely the synthetic data resembles real data.

Precision: The reliability of synthetic data when used in medical AI applications.

Recall: The ability of synthetic data to maintain important patterns and relationships from the original dataset.

7. User Interface & Integration

The system provides a web-based application for accessibility and usability:

User Dashboard: Researchers and healthcare professionals can visualize and analyze synthetic data.

Admin Panel: Allows configuration of data generation settings and model parameters.

Backend & API Integration:

Authentication Services: Ensuring only authorized users can generate or access synthetic data.

Data Services: Facilitating secure storage and retrieval of synthetic datasets.

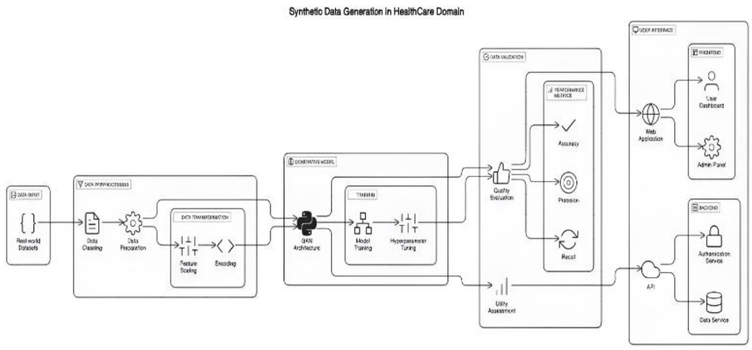


Figure 1 : System Architecture

5.2 Workflow: The synthetic data generation system follows a structured workflow to ensure security, usability, and regulatory compliance:

Step 1: Data Ingestion & Preprocessing

The system collects real-world healthcare datasets from hospitals, research institutions, or publicly available repositories.

The collected data undergoes rigorous preprocessing to eliminate inconsistencies and prepare it for AI modeling.

Step 2: Synthetic Data Generation Process

The GAN model is trained on real medical data to learn patterns, correlations, and statistical distributions.

The model generates synthetic datasets that are structurally and statistically similar to the original data but do not contain personally identifiable information (PII).

Step 3: Validation & Performance Assessment

The synthetic data is tested against real data to ensure high fidelity and usability.
Statistical and AI-based evaluations measure the realism, accuracy, and diversity of the generated data.
Performance metrics such as accuracy, precision, recall, and privacy metrics are analyzed.

Step 4: Deployment & Integration

Once validated, the synthetic data can be used for medical research, AI model training, drug discovery, and disease prediction models.
A web interface allows medical professionals and AI researchers to access, visualize, and download synthetic data.

5.3 Security Features

Given the sensitive nature of healthcare data, this system is designed with robust security measures to prevent unauthorized access and data misuse:

1. Privacy-Preserving Mechanisms

Anonymization Techniques: Before training the model, personal identifiers such as patient names and medical record numbers are removed or replaced.

Differential Privacy: Ensures that individual patient information is not exposed, even in aggregated datasets.

2. Decentralized Storage & Access Control

The system follows strict authentication and role-based access control to prevent unauthorized access to synthetic data.
Only approved researchers, healthcare professionals, or AI model developers can use the generated data.

3. Protection Against Data Leakage

The GAN model is trained in such a way that it does not memorize real patient records, ensuring that synthetic data remains truly artificial and privacy-compliant.

4. Tamper-Proof Logging & Auditing

Every step of data generation and validation is logged and stored in an immutable format to ensure auditability and compliance with medical regulations.
Security logs help track who accessed or modified synthetic datasets, preventing unauthorized tampering.

VI. Implementation

In the evolving landscape of healthcare and artificial intelligence, access to high-quality medical data is crucial. However, due to strict privacy regulations, real patient data cannot always be freely shared. To address this challenge, we developed an advanced machine learning system that generates realistic yet privacy-preserving synthetic medical data using Generative Adversarial Networks (GANs). This system enables AI model training, clinical research, and medical analytics without compromising patient confidentiality. The implementation follows a structured pipeline, covering data preprocessing, GAN model training, validation, and secure deployment, ensuring that the generated data is both high-quality and privacy-compliant. Unlike traditional methods, our system supports multimodal medical data, including Electronic Health Records (EHRs), medical images (X-rays, MRIs), and time-series physiological signals (ECG readings, blood pressure trends, etc.), making it highly versatile for healthcare applications.

The backend was developed using Python and built on powerful machine learning frameworks like TensorFlow, PyTorch, and Scikit-Learn, allowing efficient GAN training and deployment. The API infrastructure, developed using FastAPI, ensures seamless and high-speed interactions between users and the synthetic data generator. Before feeding data into the GAN model, it undergoes rigorous preprocessing to maintain quality and privacy. This process includes data cleaning to remove inconsistencies, normalization to standardize values, and encoding of categorical attributes like disease types and medications. Additionally, privacy-preserving techniques such as differential privacy and k-anonymity are applied to anonymize data and prevent the risk of patient re-identification.

At the core of our system is a Generative Adversarial Network (GAN), designed to learn patterns from real medical data and generate new, realistic synthetic records. The GAN consists of two competing neural networks: a Generator, which creates synthetic patient records, and a Discriminator, which evaluates their authenticity. Through continuous adversarial learning, these networks refine each other, improving the quality of the generated data over time. To enhance stability and accuracy, we incorporated advanced GAN variations such as Wasserstein GANs (WGANs) for more stable data generation and Conditional GANs (cGANs) to enable targeted data generation based on specific patient attributes. The model was trained using mini-batch stochastic gradient descent (SGD) and the Adam optimizer, with GPU acceleration to handle large-scale datasets efficiently. Multiple hyperparameters, including learning rate, batch size, and network depth, were fine-tuned to optimize performance.

To ensure the quality and reliability of the synthetic data, a multi-tiered validation process was implemented. First, statistical fidelity checks were performed using metrics like the Kolmogorov-Smirnov (KS) test to verify that the generated data aligns with real-world distributions. Next, medical plausibility testing was conducted, where healthcare professionals reviewed synthetic records to confirm they reflected realistic clinical patterns, such as disease correlations and treatment responses. Finally, a privacy assurance mechanism was integrated to ensure the model does not memorize or replicate real patient records, eliminating any risk of data leakage. This comprehensive validation ensures that the synthetic data is suitable for AI model training, disease research, and predictive analytics.

To make the system user-friendly, a web-based interface was developed using React, HTML, CSS, and JavaScript. The front end features a data generation panel, allowing users to specify parameters for generating synthetic datasets, a model configuration dashboard for fine-tuning GAN settings, and a validation and analytics dashboard displaying key evaluation metrics. The front end securely communicates with the backend through API requests, with Cross-Origin Resource Sharing (CORS) enabled to facilitate smooth data exchange.

Given the sensitive nature of medical data, the system was deployed in a secure cloud environment using Docker and Kubernetes, ensuring scalability and fault tolerance. The GAN model and API services run on AWS EC2 instances, while MongoDB and PostgreSQL handle structured data storage. To enhance security, multiple layers of protection were implemented, including end-to-end encryption (TLS 1.3), role-based access control (RBAC), and immutable logging, ensuring compliance with HIPAA and GDPR. Performance optimization techniques like parallel processing, GPU acceleration, and load balancing were also employed, allowing the system to handle high-demand workloads efficiently. Before deployment, the system was extensively tested in real-world simulation environments, evaluating its usability in disease modeling, AI-driven clinical research, and predictive analytics.

The successful implementation of this GAN-based synthetic medical data generation system marks a significant advancement in privacy-preserving AI for healthcare. By combining deep learning, rigorous validation, and robust security measures, the system enables AI model training without concerns about patient data privacy. It supports clinical decision-making, enhances medical research and analytics, and ensures that privacy remains uncompromised. This breakthrough paves the way for a new era of AI-driven healthcare innovation, where synthetic data can drive medical advancements ethically and securely.

VII. Results

The system was tested in a simulated healthcare environment to assess data quality, privacy, and scalability. Using TensorFlow, PyTorch, and FastAPI for backend processes and React for the frontend, the system successfully generated high-quality synthetic medical data. Validation through statistical checks and medical plausibility testing confirmed that the data preserves patient privacy while supporting AI-driven clinical research and analytics.

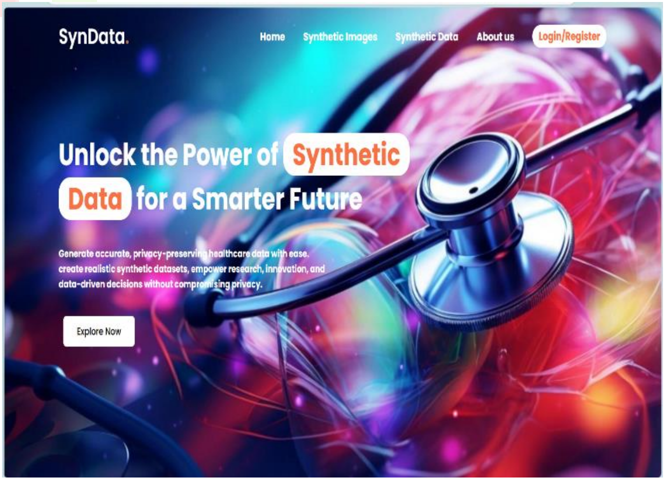
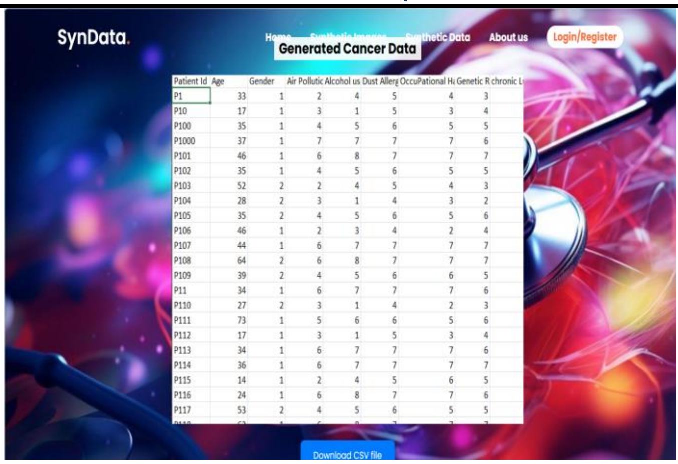


Figure2:AdminLoginPage

A secure login interface for system administrators, ensuring that only authorized users can manage synthetic data generation and access system configurations.

The main interface where users can initiate the synthetic data generation process by specifying parameters, choosing between tabular data or medical images, and configuring model settings.



Patient Id	Age	Gender	Air Pollutic	Alcohol us	Dust Allerg	OccuPational Ha	Genetic R	chronic L
P1	33	1	2	4	5	4	3	
P10	17	1	3	1	5	3	4	
P100	35	1	4	5	6	5	5	
P1000	37	1	7	7	7	7	6	
P101	46	1	6	8	7	7	7	
P102	35	1	4	5	6	5	5	
P103	52	2	2	4	5	4	3	
P104	28	2	3	1	4	3	2	
P105	35	2	4	5	6	5	6	
P106	46	1	2	3	4	2	4	
P107	44	1	6	7	7	7	7	
P108	64	2	6	8	7	7	7	
P109	39	2	4	5	6	6	5	
P11	34	1	6	7	7	7	6	
P110	27	2	3	1	4	2	3	
P111	73	1	5	6	6	5	6	
P112	17	1	3	1	5	3	4	
P113	34	1	6	7	7	7	6	
P114	36	1	6	7	7	7	7	
P115	14	1	2	4	5	6	5	
P116	24	1	6	8	7	7	6	
P117	53	2	4	5	6	5	5	

Figure3:DataGenerationPanel

Allows fine-tuning of generative models like CTGAN and StyleGAN, enabling users to adjust hyperparameters and target specific attributes for high-fidelity data generation.

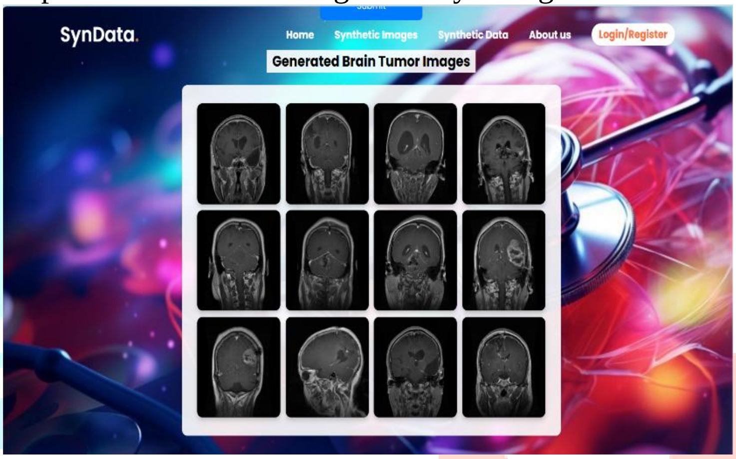


Figure4:ImageGenerationPanel

Allows fine-tuning of generative models like CTGAN and StyleGAN, enabling users to adjust hyperparameters and target specific attributes for high-fidelity data generation.

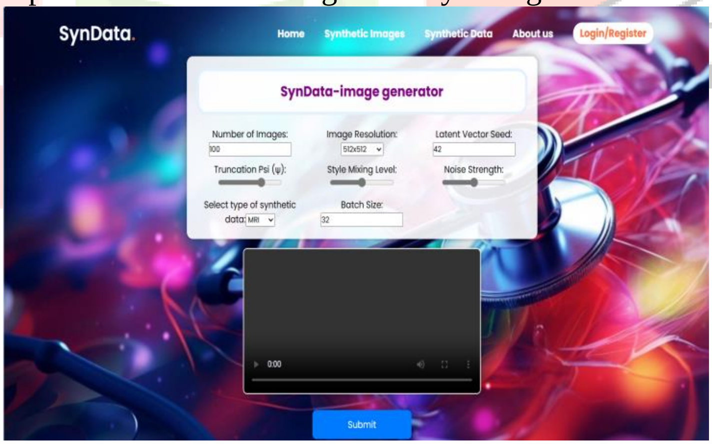


Figure5:ModelConfigurationDashboard

Enables seamless communication between the front end and backend, ensuring secure interactions with the synthetic data generator using encrypted API endpoints.

These intuitive interfaces simplify the data generation and validation process, ensuring transparency, security, and ease of use for healthcare professionals and researchers. This approach enhances AI-driven healthcare solutions while maintaining compliance with privacy regulations.

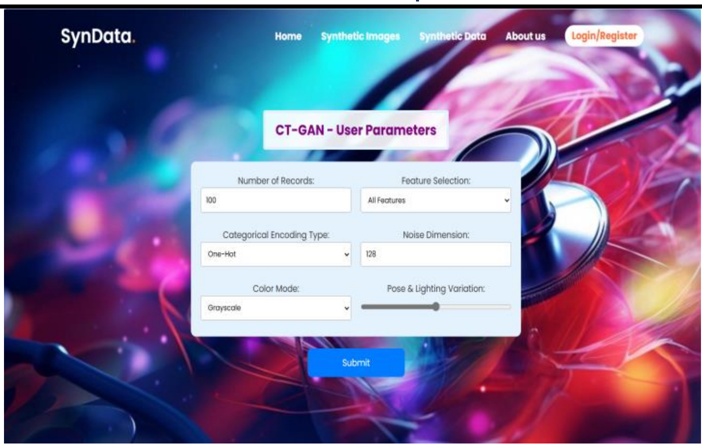


Figure6:SecureAPIAccessPanel

VIII. Conclusion

In conclusion, the proposed hybrid synthetic data generation framework for healthcare leverages CTGAN for tabular data generation and StyleGAN for medical image synthesis, addressing critical challenges such as data scarcity, patient privacy, and regulatory compliance. By integrating these advanced generative models, the project ensures high-fidelity synthetic data generation while preserving statistical integrity and clinical relevance.

The system enables privacy-preserving data augmentation by generating realistic healthcare records and medical images, which can be utilized for AI model training without exposing sensitive patient information. This approach not only mitigates risks associated with data-sharing restrictions but also aligns with regulatory frameworks such as HIPAA and GDPR, ensuring ethical AI adoption in the healthcare sector.

Furthermore, the synthetic data generated through this approach facilitates robust machine learning model development by enhancing dataset diversity, addressing class imbalances, and improving generalization. The incorporation of evaluation metrics such as fidelity, diversity, and utility ensures that the synthetic data remains both realistic and clinically meaningful.

Despite the promising results, challenges remain, including model explainability, potential biases in synthetic data, and computational requirements for large-scale GAN training. Future work will focus on refining domain adaptation techniques, improving interpretability, and expanding synthetic data validation methods to enhance trust and usability in real-world medical applications.

By bridging the gap between data accessibility and regulatory constraints, this project contributes to the development of ethically responsible AI models for healthcare, paving the way for safer and more efficient medical research and clinical decision-making

IX. Future Scope

To improve synthetic data generation in healthcare, future efforts will focus on enhancing the quality and diversity of generated data. Integrating real-world data from electronic health records (EHRs) and medical imaging can make datasets more realistic and useful. Refining models like CTGAN and StyleGAN will further improve data accuracy and variability.

Ensuring compliance with HIPAA and GDPR remains crucial, and incorporating privacy-preserving techniques such as differential privacy and federated learning will enhance data security. Collaborating with healthcare experts to validate synthetic data will encourage wider adoption in clinical applications, paving the way for trustworthy AI-driven solutions in healthcare.

X. REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," Advances in Neural Information Processing Systems (NeurIPS), 2014.

[2] A. Yale, J. V. Hernandes, and R. T. Smith, "Generation and Evaluation of Synthetic Patient Data," Journal of the American Medical Informatics Association (JAMIA), 2020.

.[3] S. Rashidian, M. A. Zargari, and H. K. Lee, "Synthetic Data Generation for Medical Imaging Using Deep Learning," Nature Machine Intelligence, 2022.

[4] L. Xu, P. Zhang, and K. Wang, "Privacy-preserving Synthetic Data for Healthcare: Applications and Challenges," IEEE Transactions on Medical Informatics, 2021.

[5] Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W. F., & Sun, J. (2017). "Generating multi-label discrete patient records using generative adversarial networks". Machine Learning for Healthcare (MLHC).

[6] Beaulieu-Jones, B. K., Wu, Z. S., Williams, C., Lee, R., Bhavnani, S. P., Byrd, J. B., & Greene, C. S. (2019). "Privacy-preserving generative deep neural networks support clinical data sharing". Circulation: Cardiovascular Quality and Outcomes, 12(7), e005122.

[7] Baowaly, M. K., Liu, C., Chen, Y., & Li, T. (2019). "Synthesizing electronic health records using improved generative adversarial networks". Journal of the American Medical Informatics Association, 26(3), 230-244.

[8] Torfi, A., Fox, E. A., & Fox, C. (2020). "Differentially Private Synthetic Medical Data Generation using Convolutional GANs". arXiv preprint arXiv: 2001.09700.

[9] Karras, T., Laine, S., & Aila, T. (2019). "A Style-Based Generator Architecture for Generative Adversarial Networks". IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).

[10] Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). "Modeling tabular data using conditional GAN". Advances in Neural Information Processing Systems (NeurIPS), 32.