Utilizing Generative Adversarial Networks for Preserving Privacy in Developing Machine Learning Models for the Healthcare Industry

Shahnawaz Khan[1], Bharavi Mishra[2], Sultan Alamri[3] and Philippe Pringuet[1]
[1]*School of Information & Communications Technology, Bahrain Polytechnic, Isa Town, Bahrain*
[2]*Department of Computer Science & Engineering, The LNM Institute of Information Technology, Jaipur, Rajasthan, India*
[3]*College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia*

Keywords: Generative Adversarial Networks (GANs), Differential Privacy, Machine Learning.

Abstract: Privacy preservation is a critical challenge while developing machine learning models utilizing medical data. This research investigates the application of Generative Adversarial Networks (GANs) for generating synthetic medical dataset while preserving the properties of the real-world dataset. It investigates on both types of medical datasets which are tabular and image-based datasets. This research employs Conditional Tabular GANs (CTGANs) for tabular data (Heart Disease Cleveland dataset) and Deep Convolutional GANs (DCGANs) for image data (chest X-ray dataset). The primary purpose is to synthesize datasets within the healthcare domain that closely mimic the statistical properties and diagnostic relevance of their real-world counterparts while safeguarding patient privacy. The proposed research focuses on training GANs to learn complex patterns and dependencies within the data. Thus, enabling GANs to generate realistic synthetic samples that can be used for training machine learning models. The generated datasets have been evaluated using various classifiers. The results demonstrate that models trained on synthetic data achieve comparable performance to those trained on real data. The results demonstrate the efficacy of our approach in balancing data utility and privacy. Furthermore, this research explores different techniques for privacy enhancement. These techniques include parameter tuning, differential privacy, and layer-wise perturbation, to further strengthen privacy preservation. The findings suggest that GAN-based synthetic data generation offers a robust and versatile solution for privacy-preserving machine learning in medical applications.

1 INTRODUCTION

In the recent years the application of artificial intelligence (AI) has been overwhelming in almost every field and industry. Healthcare industry is no exception. Application of AI has unlocked new avenues in disease diagnosis, personalized healthcare and treatment planning, etc. However, developing AI (or machine learning) models that can accomplish these tasks requires medical data which usually contains sensitive information (Khan et al., 2021; Prowal et al., 2021; Beriwal et al., 2022). Medical data usually consists of sensitive information about the patients. Therefore, while training machine learning models on sensitive medical data, there are critical threats to the privacy of the sensitive information about the patient. The chances of revealing private information in such cases become very high. Therefore, preserving the privacy of the sensitive data pose a critical challenge while developing machine learning models for the healthcare industry. Traditional methods of anonymizing the data do not fully meet the requirements because either they compromise the data utility or do not fully safeguard the sensitive information.

This research addresses these challenges and proposes an approach by investigating the application of Generative Adversarial Networks (GANs). The GANs have been utilized in generating the synthetic data that carries or almost mimics the properties of the real-world medical data. The generated synthetic data have the statistical properties and the diagnostic relevance of the real-world data. Therefore, it preserves the privacy and yet can be used in developing the machine learning models for the healthcare industry. This research investigates two different data modalities which are tabular data and image data. The tabular dataset used in this research is Cleveland heart disease dataset (Janosi et al., 1989), and the image dataset is Chest X-Ray Images by (Kermany et al., 2018). This research employs Conditional Tabular GANs (CTGANs) for tabular data and Deep Convolutional GANs (DCGANs) for image data. The research aims to demonstrate the feasibility and efficacy of generating high-fidelity synthetic datasets.

Khan, S., Mishra, B., Alamri, S., Pringuet and P.
Utilizing Generative Adversarial Networks for Preserving Privacy in Developing Machine Learning Models for the Healthcare Industry.
DOI: 10.5220/0013567600003967
In *Proceedings of the 14th International Conference on Data Science, Technology and Applications (DATA 2025)*, pages 544-551
ISBN: 978-989-758-758-0; ISSN: 2184-285X

Furthermore, the generated datasets can be used to train machine learning models. The developed models will not compromise the privacy of the patients and yet will be able to provide robust AI models in assisting the diagnosis and other healthcare-industry related tasks.

2 LITERATURE REVIEW

AI in healthcare has revolutionized disease diagnosis, treatment, and patient personalized healthcare services. However, training machine learning algorithms on sensitive medical data raises grave privacy concerns (Venugopal et al., 2022; Chen and Esmaeilzadeh, 2024; Cai et al., 2021). This section on related works provides a summary of approaches utilized today and the trends in privacy-preserving schema, including the specific use of Generative Adversarial Neural Networks (GANs) to generate synthetic medical datasets.

In the area of creating synthetic data, Generative Adversarial Networks (GANs) have become a ground-breaking method that shows promise in addressing privacy issues in the medical industry (Cai et al., 2021; Shafik, 2025). Two neural networks—a discriminator and a generator—combine to create synthetic data that closely resembles real-world datasets in GANs. The discriminator assesses the validity of the new data instances produced by the generator. GANs are especially useful for medical datasets that contain sensitive patient data because of this confrontational process, which lets them to identify intricate patterns and networks in the data (Grassucci et al., 2022). Many medical applications, for instance the making of artificial medical imageries and electronic health data, have shown that GANs can create high-fidelity synthetic data while maintaining anonymity.

A major use case of GAN in the healthcare sector is generating synthetic tabular data using Conditional Tabular GANs (CTGANs). This is because tabular datasets are often heterogeneous, containing both continuous and categorical variables, and so CTGANs are particularly designed to handle such complexities. Because they can learn the statistical dependencies and characteristics in the data by conditioning the generation process on specific parameters, CTGANs are suitable for application in privacy-preserving data synthesis. As an example of good data utility protection together with patient privacy protection, CTGANs have been successfully applied to generate synthetic datasets for heart disease prediction (Kermany et al., 2018). This is also how the second half of common anonymity approaches suffer from just lowering the data utility or not covering sensitive data sufficiently.

Deep Convolutional GANs (DCGANs) have established potential in the area of medical imagery by creating synthetic images that reduce privacy concerns while preserving diagnostic relevance. Deep convolutional neural networks are used by DCGANs to extract complex features and spatial hierarchies from medical pictures, including chest X-rays. DCGANs have the potential to make machine learning models without revealing private patient information by producing artificial images that closely mimic actual data of medical images. The creation of artificial chest X-ray image data for the resolution of training diagnostic models is an important target of the medical imaging applications where this method has been confirmed (Venugopal et al., 2022). The potential of DCGANs to improve privacy preservation while preserving the diagnostic value of the generated data is demonstrated by its application in medical imaging.

3 METHODOLOGY

3.1 Proposed Framework

This research utilizes the Generative Adversarial Networks (GANs) (Grassucci et al., 2022) to generate the synthetic data that mirrors the statistical properties of the real-world data while preserving individual privacy in healthcare data while developing the machine learning models for the healthcare industry. The proposed generative model automatically identifies and learns from regularities or patterns in the input data to produce or generate new examples that look like as they may have been realistically extracted from the original dataset. The proposed methodology is shown in Figure 1.

GAN architecture can be considered similar to a two-player game. One player can be referred to as a Generator, and the other can be considered a Discriminator. The discriminator's goal is to differentiate between produced data and factual input data, while the generator's goal is to produce new data (numbers, photos, etc.) that is as near to or comparable to the dataset that is supplied as input.

The discriminator works as a binary classifier and takes real dataset and synthetic dataset samples as input. The output probability values are between 0 and 1. In which a value close to zero indicates that the data is most likely is the fake which a value close to one indicates that the data is real. The discriminator extracts the relevant features from the data and these features

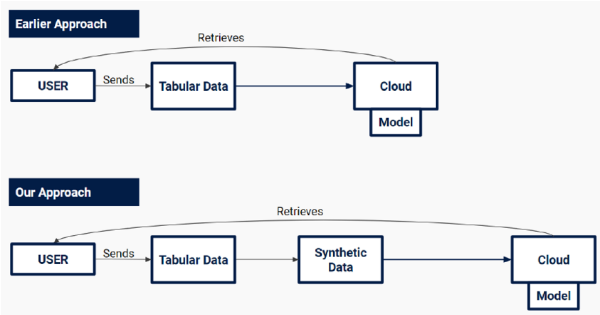

Figure 1: Proposed Methodology.

aids the discriminator network to classify the data. During the training, any misclassification by the discriminator network is penalized and weights (parameters) are updated to improve the accuracy. The discriminator parameters are subsequently revised and backpropagated by this discriminator loss, enhancing discriminator's prediction. The Discriminator Training Process is shown in Figure 2.

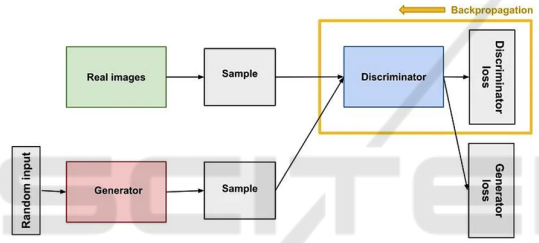

Figure 2: Discriminator Training Process.

Generator network creates or generate synthetic data. It receives random noise (a latent vector) as input which acts as the starting point for generating output. The noise is then transformed into the synthetic data which carries the properties of the real-dataset. The goal is of course to generate the data which is indistinguishable from the real data. The generator network learns by receiving feedbacks from its counterpart (discriminator). Based on the feedbacks from the discriminator, generator adjusts its parameters. The parameters (weights and biases) are updated through backpropagation. The Generator Training Process is shown in Figure 3. The Generating process along with the real-dataset sample is shown in Figure 4.

3.2 Proposed Approach for Privacy Preservation in Tabular Datasets

Traditional methods of data anonymization are not sufficient sometimes while work with tabular datasets for developing machine learning models. The probability of unauthorized access to personal information are very high in sensitive tabular datasets and

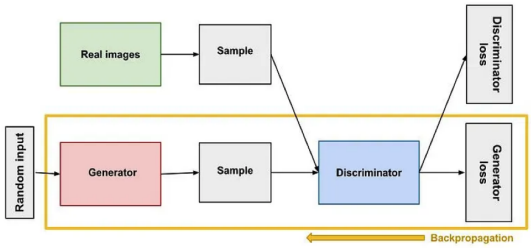

Figure 3: Generator Training Process.

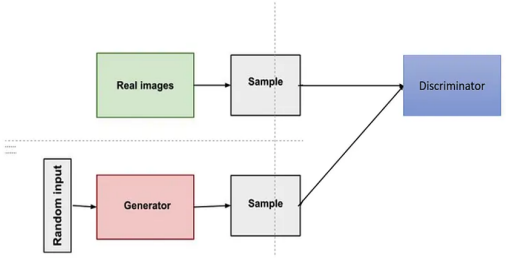

Figure 4: Generating process along with the real-dataset sample.

may lead to severe consequences. The proposed approach presents a method using GANs to preserve the privacy of the healthcare personal datasets. The proposed approach is based on a Conditional Tabular Generative Adversarial Network (CTGAN) (as shown in Figure 5). CTGANs have the capability of generating synthetic tabular data that closely mimics real-world dataset complexities. The model works on the principle of competition between a usually generator-grilled generator that comes up with realistic data and a discriminator trained to tell apart real data samples from synthetically generated ones. What is unique to CTGAN is that it possesses a novel ability to preserve, with as much detail as possible, relationships and dependencies in the data while keeping the complex relationships between different columns intact. This works really well for tasks that require structure and context, such as sensitive health information, which may be found in datasets like Heart Disease Cleveland. The practicality and effectiveness of CTGAN in generating realistic synthetic tabular data have made it useful across a wide range of domains, availing promising solutions to privacy-aware data synthesis applications.

This research utilizes the heart disease Cleveland dataset. The dataset contains sensitive health-related information and can be an ideal candidate for testing any privacy preservation technique. Conventional methods, such as masking or generalization, usually reduce data utility in machine learning tasks. This research offers a solution using GANs, which is one of the deep learning techniques in the state of the art,

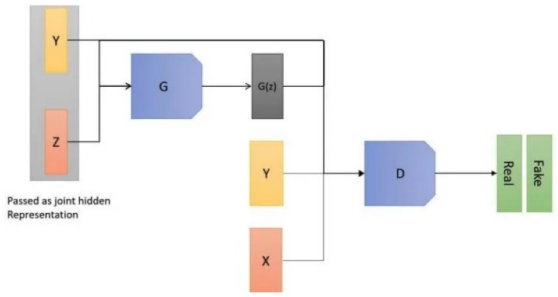

Figure 5: Architecture of CTGAN.

generating synthetic data by closely mimicking the original distribution while preserving privacy. The research follows the following steps in implementing the proposed method:

•Dataset Preparation:
The dataset has been processed for missing values, features normalization and encoding categorical variables. Preprocessing the dataset and cleaning it is a crucial step in ensuring that GAN will learn the underlying pattern effectively.

•CTGAN Implementation:
The architecture of CTGAN has been developed explicitly to preserve the conditional relationships in the source dataset. CTGAN has been conditioned on selected attributes to ensure that the generated samples resemble real ones with regard to statistical properties. Thus, the implementation of the CTGAN algorithm will generate synthetic tabular data.

•Privacy Parameter Tuning:
In the proposed approach, parameters of GAN are tuned in a way which will help in enhancing the privacy preservation. In addition, it allows to control the balance between privacy and utility of data. In this stage, the level of privacy preservation can be personalized based on the application needs.

•Evaluation Metrics:
The metrics used to evaluate our approach to privacy preservation include anonymity, diversity, and closeness quantifying anonymity, diversity, and closeness to the original distribution of the synthetic data, respectively.

3.3 Proposed Approach for Privacy Preservation in Image Datasets

Machine learning models trained on sensitive medical image datasets can pose a serious challenge for preserving the privacy of the data. This research utilizes the medical images of chest X-rays to demonstrate the privacy-preservation in image datasets. It leverages the use of Deep Convolutional Generative Adversarial Network (DCGAN) in generating synthesized chest X-ray images, thus avoiding the privacy concern of the original medical data. DCGAN Architecture is shown in Figure 6. DCGAN forms one of the significant developments in generative deep learning subject to the domain of image generation. DCGAN is one of the GANs variants, using deep convolutional neural networks for learning complex visual patterns and synthesizing them. The new network embeds convolutional layers, adding a new capability to the network in capturing spatial hierarchies and intricate details present within the data for high-resolution realistic images. Originally designed for the goal of image synthesis, DCGAN turned out to be really effective in many domains, including medical imaging. DCGAN supplies a very powerful tool in solving the task of privacy preservation for machine learning through generating synthetic images with the main features of the original dataset, thus mitigating privacy risks. Success here lies in its capability to learn and reproduce these complex structures, which finally makes it an indispensable workhorse for applications that demand visual fidelity with data privacy.

There has been an increasing demand for robust privacy preservation approaches to deal with sensitive patient information in medical image datasets. The existing methods failed in striking a good trade-off between retaining the data utility in machine learning tasks and preserving the privacy of individuals. This research proposes a solution that incorporates DCGAN for generating synthetic images that capture the essence of the original distribution with reduced privacy risks. The research follows the following steps in implementing the proposed method:

3.4 Preprocessing the Dataset Before DCGAN Application

Similar to the tabular dataset, preprocessing is required for the image dataset as well. A good dataset will ensure that a DCGAN learns most of the features with minimum chances of compromising privacy. In this case, the chest X-ray dataset has to be extensively preprocessed. Preprocessing in this experiment includes normalizing image sizes, cleaning noise or artifacts, and normalizing pixel values.

•DCGAN Implementation:
DCGAN is a variant of GANs designed especially for the generation of images. Hence, for synthesizing

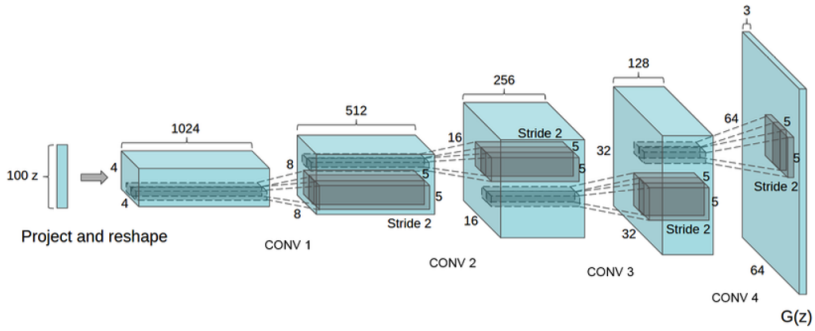

Figure 6: DCGAN Architecture.

chest X-ray images, this research has applied the same technique. The network has been trained on the comprehension of complex patterns and structure in the original images to make generated images look similar but private due to the generation process.

•Privacy Enhancement Techniques:
In this work, the privacy preservation has been enhanced by incorporating additional techniques, such as differential privacy and layer-wise perturbation during the DCGAN training process. These are additional layers of protection against any potential leakage of privacy in the generated synthetic images.

•Evaluation Metrics:
The privacy preservation will be quantitatively assessed with the Structural Similarity Index, SSI, and Peak Signal-to-Noise Ratio, PSNR, metrics. These give insight into visual similarity between the original and synthetic images with regard to maintaining diagnostic relevance while ensuring privacy.

4 SIMULATION AND RESULTS

4.1 Tabular Dataset

As discussed earlier, this research utilizes the Heart Disease Cleveland dataset. The analysis of the privacy-preservation in the tabular dataset has been conducted by generating synthetic data by utilizing the heart disease dataset and then training multiple classification models. The performance of these models has been compared. The classifiers have been trained on the original dataset and then on the generated dataset. The following Figures (7,8) demonstrate the snippets of the original dataset and the generated dataset.

The classifiers which have been trained on both datasets are logistic regression, naïve-Bayes, k-nearest neighbors, support vector classifiers (linear, radial basis and polynomial) and random forest. The generated dataset has been utilized in training and then to check the accuracies of various machine learning classifiers. While comparing accuracies of various ML models on synthetic dataset and original datasets, it has been observed that self-generated dataset performed better in certain cases, and overall, it can be used in training the machine learning models while retaining the core features and preserving the privacy. The following table 1 and Figure 9 demonstrate the comparison between the accuracies obtained from original dataset and the generated dataset.

Table 1: Comparison between accuracies of original and synthetic datasets.

Classifier	Orig. (%)	Synth. (%)	Diff. (%)
Log Reg.	77.5	77.5	0.00
Naïve Bayes	63.0	72.0	9.00
KNN	82.22	75.56	6.67
SVC Lin.	82.22	77.78	4.45
SVC RBF	77.78	75.56	2.22
SVC Poly	82.22	73.33	8.89
RF	80.00	68.89	11.11

4.2 Image Dataset

This research uses deep convolutional generative adversarial networks to generate synthetic images for testing privacy preservation in medical chest X-rays image datasets. The generated synthetic images retain diagnostic relevance with minimal risks regarding privacy. The experiments have been carried out focusing on the assessment of visual fidelity and the determination of how well the generated images preserve privacy. Figures 10 shows the Chest X-Ray Original Dataset while Figure 11 shows the Chest X-Ray Synthetic Dataset.

The creation of a synthetic dataset of chest X-ray images using Deep Convolutional Generative Adver-

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	condition
69	1	0	160	234	1	2	131	0	0.1	1	1	0	0
69	0	0	140	239	0	0	151	0	1.8	0	2	0	0
66	0	0	150	226	0	0	114	0	2.6	2	0	0	0
65	1	0	138	282	1	2	174	0	1.4	1	1	0	1
64	1	0	110	211	0	2	144	1	1.8	1	0	0	0
64	1	0	170	227	0	2	155	0	0.6	1	0	2	0
63	1	0	145	233	1	2	150	0	2.3	2	0	1	0
61	1	0	134	234	0	0	145	0	2.6	1	2	0	1
60	0	0	150	240	0	0	171	0	0.9	0	0	0	0
59	1	0	178	270	0	2	145	0	4.2	2	0	2	0
59	1	0	170	288	0	2	159	0	0.2	1	0	2	1
59	1	0	160	273	0	2	125	0	0	0	0	0	1
59	1	0	134	204	0	0	162	0	0.8	0	2	0	1
58	0	0	150	283	1	2	162	0	1	0	0	0	0
56	1	0	120	193	0	2	162	0	1.9	1	0	2	0
52	1	0	118	186	0	2	190	0	0	1	0	1	0
52	1	0	152	298	1	0	178	0	1.2	1	0	2	0
51	1	0	125	213	0	2	125	1	1.4	0	1	0	0
45	1	0	110	264	0	0	132	0	1.2	1	0	2	1
42	1	0	148	244	0	2	178	0	0.8	0	2	0	0
40	1	0	140	199	0	0	178	1	1.4	0	0	2	0
38	1	0	120	231	0	0	182	1	3.8	1	0	2	1
34	1	0	118	182	0	2	174	0	0	0	0	0	0
74	0	1	120	269	0	2	121	1	0.2	0	1	0	0

Figure 7: Original Tabular Dataset.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	condition
44	0	1	168	314	0	2	201	0	0.081187	1	0	0	0
80	1	3	137	269	1	2	66	1	5.981982	1	2	0	1
61	0	2	208	333	0	2	189	1	-0.15144	0	1	2	1
59	0	2	122	329	0	0	203	0	-0.16115	2	3	1	0
55	1	2	103	254	1	2	174	0	1.413229	0	0	2	0
78	1	3	108	243	0	2	180	0	4.026862	1	0	0	1
67	1	3	106	174	0	2	164	0	2.077912	0	0	0	0
68	1	3	129	317	0	2	178	0	1.553058	0	3	2	0
77	1	1	91	339	0	0	194	1	1.104283	1	0	0	0
78	1	3	138	357	0	0	202	0	0.405732	2	0	0	0
74	1	3	130	323	0	2	105	1	0.678881	1	0	2	1
61	1	3	160	280	0	2	97	1	2.302824	1	0	2	1
52	1	0	119	350	0	2	174	0	1.121329	1	0	2	1
50	1	1	155	326	0	2	209	0	-0.10342	0	0	0	0
76	1	2	170	424	0	1	124	1	0.265089	2	3	0	1
41	1	3	127	334	0	0	189	0	1.583265	0	2	0	1
39	0	1	140	266	0	0	209	1	-0.30532	0	0	0	0
74	0	3	118	237	0	2	83	1	2.826592	1	1	2	1
74	1	3	163	386	0	2	194	0	-0.26446	0	2	0	0
46	1	3	138	243	0	2	187	0	-0.17335	0	0	2	1
60	1	3	161	138	0	0	60	1	1.445313	1	0	0	0
36	1	1	164	216	0	0	152	0	0.042709	0	0	0	0
48	1	3	170	314	0	2	173	0	2.189425	0	2	0	0
64	1	1	107	301	0	2	137	0	2.721544	0	0	2	1
58	0	3	141	220	0	2	125	1	5.675099	2	2	2	1

Figure 8: Generated (synthetic) dataset.

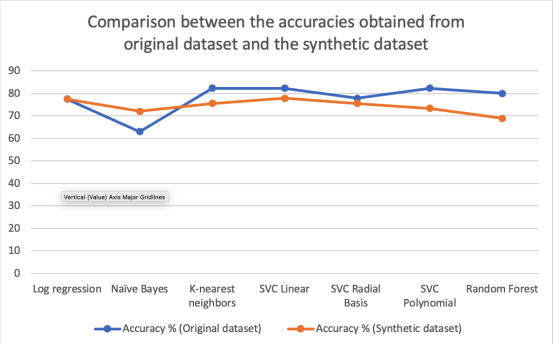

Figure 9: Comparison between the accuracies obtained from original dataset and the synthetic dataset

sarial Networks has been proposed. This research aims the dual goals of preserving diagnostic relevance and mitigating privacy risks. Extensive testing using a wide range of classifiers ranging from traditional methods to deep learning architectures showed remarkable consistency in these synthetically generated images. Synthetic images represented fidelity related to fidelity in both anatomical structures and pathological features regardless of the classifier used.

The performance metrics across various classifiers has been almost identical, demonstrating the robustness and generalizability of the synthetic dataset for ML model training. These results also make the proposed approach practically applicable in privacy-preserving machine learning tasks with regards to medical image analysis. This research strongly emphasizes the feasibility of using synthetic datasets for training of classifiers when real patient data is unavailable.

However, more experiments are being conducted to take the proposed research in direction of practical applicability for future research on image dataset. Therefore, only limited results have been demonstrated in this section. Further studies could be tailored to concrete clinical scenarios and could also as-

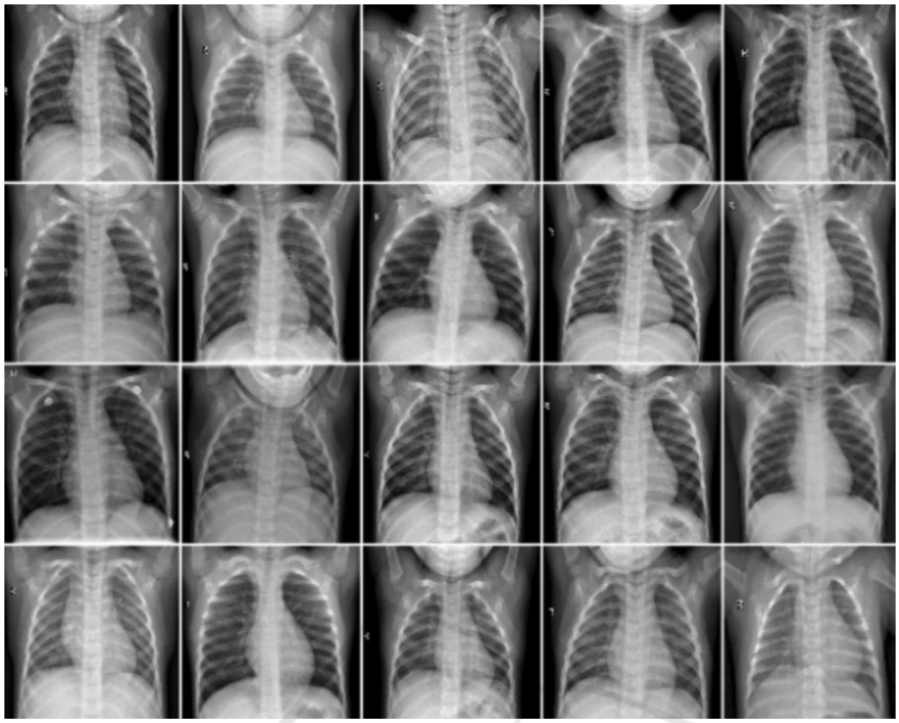

Figure 10: Chest X-Ray Original Dataset.

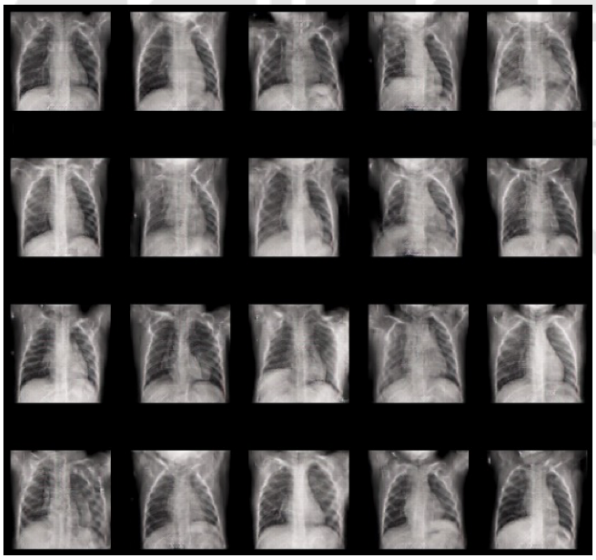

Figure 11: Chest X-Ray Synthetic Dataset.

sess the efficacy of the synthetic dataset against real-world data. Figure 12 presents the snippet of the generated synthetic data and generator and discriminator loss during the training.

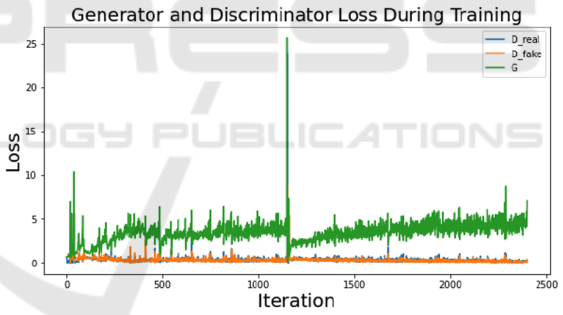

Figure 12: Loss Curve.

5 CONCLUSION

This research employs Generative Adversarial Networks (GANs) [6] for generating the synthetic tabular and image datasets. The dataset adopted for tabular dataset is heart disease Cleaveland dataset and synthetic tabular dataset has been generated with the Conditional Tabular GAN (CTGAN). For the image dataset chest X-ray image dataset has been used and the synthetic image dataset has been generated with Deep Convolutional GAN (DCGAN).

The research primarily focuses to improve the preservation of privacy in medical data and also study the performance of synthesized datasets on various machine learning models. It has been observed from

the tests carried out on a very broad range of classifiers that the same pattern has been replicated-regarding robustness and generalizability of the synthetic datasets. The adoption of CTGAN in generating tabular datasets and DCGAN for chest X-ray images has shown promising results. First, with the tabular data, this research has been able to synthesize the data set quite realistically, while keeping the statistical patterns that were inherent in the original data set. Second, the DCGAN managed to generate such chest X-rays that maintained diagnostic relevance, along with a high degree of visual fidelity.

It has also been observed that different machine learning models yield similar performance, hence, underlines the versatility of the synthetic datasets. The consistency of results across various classifiers suggests that synthetic datasets can easily be integrated into a wide variety of machine learning tasks and may become one of the most promising avenues for privacy-preserving applications in both medical imaging and tabular data analysis. It can be inferred that when GANs are developed on specific data types, they can generate synthetic datasets which can perform just as well as real data in training a model without revealing sensitive information. Therefore, there is a potential in applying the proposed approach in other medical domains. Its validation against real-world diverse datasets, and the integration of other measures that may enhance privacy could form new avenues of interest in future studies. The proposed approach contributes to enhance machine learning model development with a view to privacy by providing a reliable and consistent framework for synthetic data generation in different modalities.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Aarya Gard and Harshal Jain for their valuable assistance with simulation and conducting experiments.

REFERENCES

Beriwal, S., Thirunavukkarasu, K., Khan, S., and Abimannan, S. (2022). Detection of infectious diseases in human bodies by using machine learning algorithms. In *Handbook of Research on Machine Learning: Foundations and Applications*, page 229.

Cai, Z., Xiong, Z., Xu, H., Wang, P., Li, W., and Pan, Y. (2021). Generative adversarial networks: A survey toward private and secure applications. *ACM Computing Surveys (CSUR)*, 54(6):1–38.

Chen, Y. and Esmaeilzadeh, P. (2024). Generative ai in medical practice: in-depth exploration of privacy and security challenges. *Journal of Medical Internet Research*, 26:e53008.

Grassucci, E., Cicero, E., and Comminiello, D. (2022). Quaternion generative adversarial networks. In *Generative Adversarial Learning: Architectures and Applications*, pages 57–86. Springer International Publishing.

Janosi, A., Steinbrunn, W., Pfisterer, M., and Detrano, R. (1989). Heart disease [dataset]. UCI Machine Learning Repository.

Kermany, D., Zhang, K., and Goldbaum, M. (2018). Labeled optical coherence tomography (oct) and chest x-ray images for classification. Mendeley Data, V2.

Khan, S., Thirunavukkarasu, K., Hammad, R., Bali, V., and Qader, M. R. (2021). Convolutional neural network based sars-cov-2 patients detection model using ct images. *International Journal of Intelligent Engineering Informatics*, 9(2):211–228.

Prowal, P., Singh, A. S., Thirunavukkarasu, K., Khan, S., and Qader, M. R. (2021). Machine learning algorithms based on feature selection method used for the prediction of breast cancer. In *2021 IEEE International Conference on Data Analytics for Business and Industry (ICDABI)*, pages 100–106.

Shafik, W. (2025). Generative adversarial networks: Security, privacy, and ethical considerations. In *Generative Artificial Intelligence (AI) Approaches for Industrial Applications*, pages 93–117. Springer Nature Switzerland.

Venugopal, R., Shafqat, N., Venugopal, I., Tillbury, B., Stafford, H., and Bourazeri, A. (2022). Privacy preserving generative adversarial networks to model electronic health records. *Neural Networks*, 153:339–348.