JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2590 - 2596
https://publishoa.com
ISSN: 1309-3452

IMAGE AUGMENTATION ON ECG IMAGE DATASET USING GAN

Pon Bharathi A[1], Poornima R[2], Sasikala T S[3], Nandhakumar A[4], Dr Parthiban K G[5]
[1] Assistant Professor, Dept of ECE, Amrita College of Engineering and Technology, bharathpon@gmail.com
[2] Assistant Professor, Dept of ECE, JCT College of Engineering and Technology , rpoorani6@gmail.com
[3] Assistant Professor, Dept of CSE, Amrita College of Engineering and Technology, sasikalaselva2012@gmail.com
[4] Assistant Professor, Dept of ECE, Dhaanish Ahmed institute of technology , nandhakumar3107@gmail.com
[5] Principal and Professor, Dept of ECE, Dhaanish Ahmed institute of technology , kgparthiban@gmail.com

ABSTRACT

Customized healthcare should organize an environment for the manipulation of reliable and secure private data. Using Generative Adversarial Networks (GANs), this paper proposes a method for constructing synthetic ECGs (electrocardiograms) (GANs). The objective is to develop data that may be used in educational and research contexts while reducing the risk of sensitive data leakage to the absolute minimum possible. For GANs to work, we recommend converting raw data to an image and then decoding it back to the original data domain so that GANs may operate on photos and video frames. Our transformation and processing theory' viability is basically shown. The primary disadvantages of each stage in the suggested process are then discussed for the specific situation of ECGs. As a result, a new study avenue into the use of GANs to anonymize health data is opened, and simple new advancements are anticipated.

Keywords: Ecosystem, Anonymize, Decoded, Anticipated and Trustworthy

1.INTRODUCTION

Mobile health applications and intelligent wearable sensors are only two examples of modern systems that store, process, and infer information from personal health data [1]. The construction of digital twins for customised healthcare is proposed in Reference [2] as a way to orchestrate an ecosystem of data manipulation that is trustworthy and secure.

Privacy for ethical reasons is a concern in health-related endeavors. Privacy and legal issues when exchanging and exposing patient health information are key problems in the healthcare business when it comes to the sources of medical data that are collected. However, a good grasp of medicine and a richness of patient information are essential for a clear diagnosis. [4] In this case, anonymization may be utilized to mitigate the risks involved with collecting and processing vast volumes of personal data. A GAN-based anonymization of private health data [5] is proposed to create a seedbed from training data that allows not

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2590 - 2596
https://publishoa.com
ISSN: 1309-3452

only the capture of information from the original data but also the production of new information with a similar behavior.

Generative Adversarial Network (GAN) methods were proposed in 2014 and have subsequently been recognized as viable solutions for data augmentation and missing data concerns [8]. As a result of these applications, the possibility of using GAN systems to develop synthetic data that resembles the features of a personal health database was questioned. This produced machine, if it were conceivable, would be a highly helpful tool since it allows for an infinite amount of data that is comparable to the original data without compromising the privacy of the original pieces. Due to the fact that this technology may make sensitive data from any field available without running the danger of personal information being leaked, applications might vary from instructional ones to scientific simulations and investigations.

2.RESEARCH WORK

As a result of the Affordable Care Act, healthcare systems need to be reimagined. In particular, the adoption of cutting-edge technology that allows for both centralized and decentralized administration of patients, particularly those with critical illnesses, is essential. In addition to the COVID-19 pandemic and the aging of the population, there are many more contributory variables.

The Internet of Medical Things (IoMT), data processing and fusion, and telemedicine have all recently seen significant evolution [9]. In especially for older persons with chronic conditions, the area of IoMT has created new options for patient monitoring and remote service delivery [10, 11]. However, the growth of the IoMT has also brought forth new difficulties in areas like processing power, security, and privacy.

There are, however, few multi-sensor systems with native AI integration. Telemedicine systems now on the market have the ability to perceive and interact via a data gateway, but their processing power is frequently restricted [12]. As particular patients may have distinct "desired" ranges for their biomarkers and physicians can only have a limited understanding of the patient condition, these streamlined, one-size-fits-all warning systems may not be able to adequately help physicians in monitoring their patients. There is a pressing need for new, sophisticated devices that can track each patient individually.

AI algorithms, particularly explainable AI (xAI) approaches that may give predictive models of the patient's health, may bring about a shift in this domain [13,14]. For example, changing particular biomarkers in accordance with the model criteria might help doctors uncover countermeasures to reduce deterioration risk. The best way to achieve this is to create models in the form of comprehensible guidelines (if-then-else). It is possible to have a greater understanding of the factors that influence how a patient's health changes and to identify personalized prevention measures and treatment approaches using this method of strategy-building.

[15] Supplemented their dataset with the artificial beats produced by a 14-layer ACGAN. [16] Investigated the creation of realistic synthetic signals using a variety of GAN designs. They evaluated the produced beats quantitatively using the metrics Maximum Mean Discrepancy (MMD) and Dynamic Time Warping (DTW). [17] produced synthetic ECG signals using the same architecture (two-layer BiLSTM). To assess the produced beats, they employed MMD as well as two cutting-edge measures. [18] Produced 3 classes of ECG using a multiclass DCGAN model. [14] Developed a brand-new BiLSTM-CNN GAN

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2590 - 2596
https://publishoa.com
ISSN: 1309-3452

to produce fake ECG signals. [19] used a cutting-edge GAN model to the typical 12 lead ECG readings. Only [20] employed AC-GAN in the works mentioned above; neither WGAN-GP nor AC-WGAN-GP were implemented.

3.PROPOSED SYSTEM

This section defines the issue setting and briefly introduces the initial results in [9,10]. This work's initial outcome is the definition of an unique five-step technique for GAN-based global data anonymization. As seen in Figure 1, we focus on the white blocks in this part in order to investigate the broad issue of health data anonymization for both pictures and raw static data. The blue blocks, which represent the raw dynamic data in the form of ECGs, and four of the five phases in the process formulation will be examined.

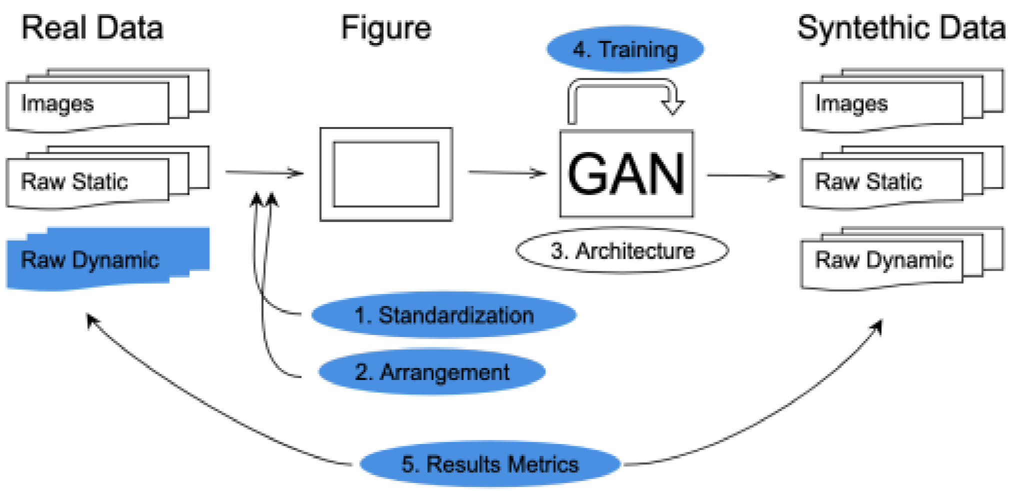

Figure:1 Block Diagram

Therefore, it is suggested to uniformize the data and represent them visually to the GAN system in order to apply the same tools as for images samples. There are two processes in this pre-processing phase.

- **Data Standardization:**reducing all samples to the same number of values if necessary and bringing all dataset features to the same range value, which is from 0 to 255.
- **Data Augmentation:**arranging characteristics in a shaped image—typically a square—to determine how the data is laid down.

Data Standardization

To equalise the weight of each variable range, all units are standardised to a shared range. The typical feature normalisation or scaling would be

$$\frac{x-\min(x)}{\max(x)-\min(x)} \times (255-a)+a$$

An is set to 1 for the Thyroid database. Continuous features in this form have values between 1 and 255, which is the typical range for picture files. In order to be utilized just as a NULL value, the value

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2590 - 2596
https://publishoa.com
ISSN: 1309-3452

0 is omitted from the range. The normalization formula might be made simpler by putting a = 0 when there should be no NULL value taken into account, as is the case with the Cardiogram database.

Data Augmentation

The normalized data may now be processed into an image format. When considering the six binary and fifteen continuous characteristics in the Thyroid database, a 7 7 grid may be used for a variety of setups in the Thyroid database. A few of them are shown. To prevent any further issues, whether models like CNN are employed, all features are arranged in the shape of a square region. The initial layout was chosen (left, top scheme). Tocreated from a genuine thyroid sample. All values are set to 0 in the NULL region.

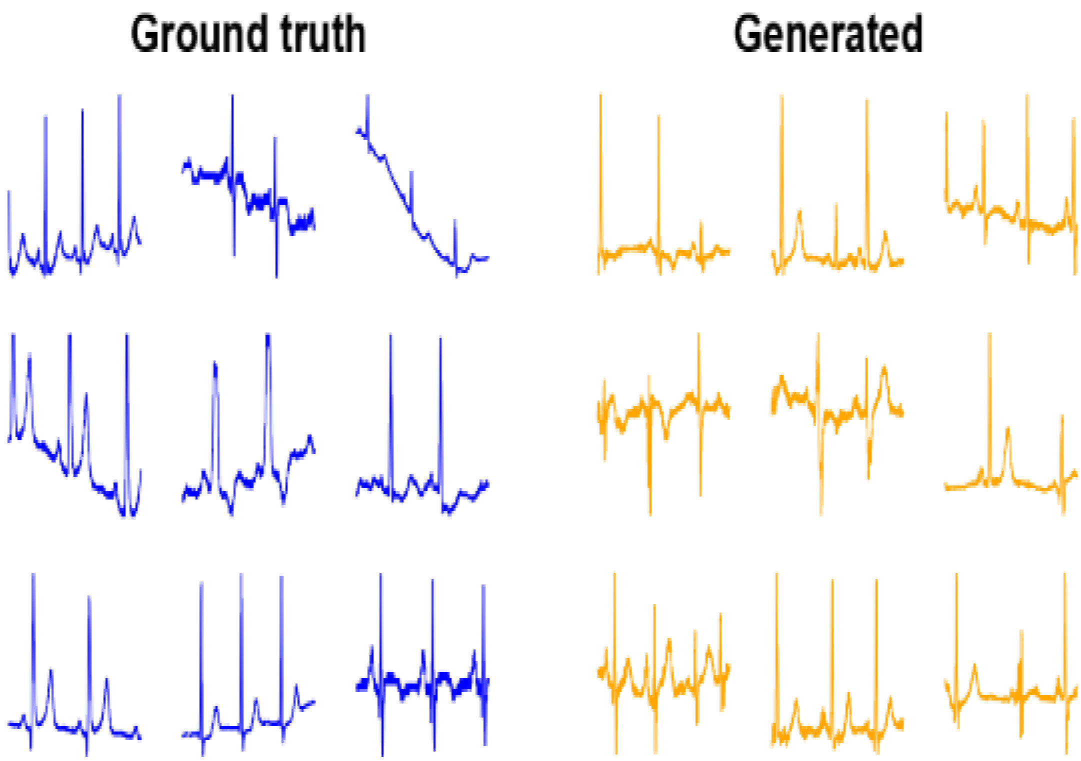

Figure:2Real-world data Cardiogram signals vs created time series from samples after 200 iterations of CNN model training.

Due to the fact that findings are not static, results are compared visually with the original data as a doctor would do, rather than looking at the distribution of a certain feature. Figure 2 shows a subset of random actual and artificial cardiogram data. It is crucial to emphasise that the current assessment is solely based on the visual distinctions between artificial and actual data, necessitating the advice of a physician. It is clear that more investigation into the metrics for evaluating GAN-based solutions is required.

3.1 GAN-Generated Artificial ECGs

A GAN system may be fed with general health data to produce good results on anonymization for personal data, as was confirmed in the preceding. The steps for the suggested process are as follows:

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2590 - 2596
https://publishoa.com
ISSN: 1309-3452

- Standardization of data.
- Consolidation of data into a picture.
- Identify the architecture of GANs.
- Address standard deep learning training issues.
- Choose relevant metrics for results.

Because it is a multidimensional time-series with a dual raw/image nature, the Cardiogram dataset of ECGs from Physionet was chosen. Researchers, algorithm benchmarkers, and educators can all benefit from this collection of digitized ECGs on PhysioNet. It was conferred by the German National Metrology Institute.

4.RESULT AND DISCUSSION

With varying numbers of epochs, many separate runs of data augmentation are offered (the rest of the neural structures are left untouched). There were produced a thousand synthetic samples. Figure 3 displays the inherent outcomes.

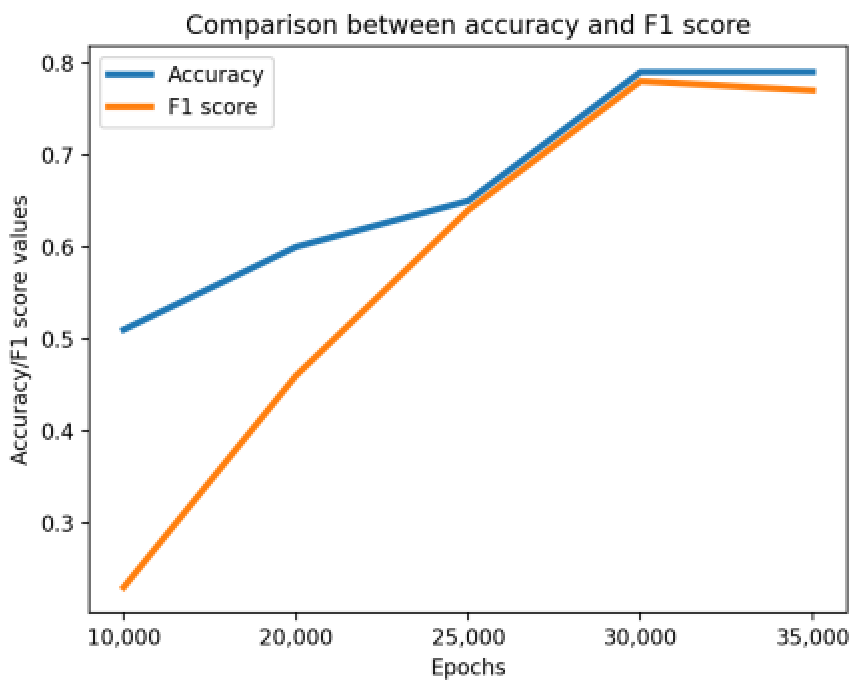

Figure:3Comparison of F1 score and accuracy

JS does not offer consistent indicators of the calibre of XAI, but the FID does. Three thousand epochs is the configuration that produces the greatest results in terms of least FID, most validated rules, best accuracy, and F1 score. As a result, FID may be able to predict how well XAI would perform, reducing the requirement for continuous testing over all potential data augmentation runs.

When using GAN architectures, the evaluation of outcomes and the stopping point for training heavily rely on visual inspection methods, the majority of which are imported from the field of computer

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2590 - 2596
https://publishoa.com
ISSN: 1309-3452

vision research. Several indirect metrics (scores and distances) have been constructed on the convolutional layers when using Inception as the pretrained model, however this is not the norm and neither is our testing setting.

5.CONCLUSIONS

In this work, we looked at how generative adversarial networks (GANs) are applied to data that is retrieved by IoMT devices. Additionally, the technique is used to evaluate the synthetic dataset produced by the GAN in comparison to the actual data. The algorithm's criteria assist the alignment of the synthetic dataset with the real dataset, as shown by the findings. A performance metric has been established that is independent of visual examination. It enables one to define a stopping criterion and indirectly assess the effectiveness of the GAN system when producing synthetic ECGs based on assessing the accuracy for a parallel classification job. It can only be used when the data is labelled and shares the same drawbacks as other indirect performance metrics established for the use of the pre-trained Inception layer: it assures GAN architecture convergence but does not imply that the produced pictures are suitable for visual inspection.For testing GAN performance when creating new pictures, further work has to be done. But in this study, we present a novel approach that holds true in a wide range of situations, particularly supervised categorization.

References

[1] Ramanujam, E., ThinagaranPerumal, and S. Padmavathi. "Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review." *IEEE Sensors Journal* 21, no. 12 (2021): 13029-13040.

[2] Elayan, Haya, MoayadAloqaily, and Mohsen Guizani. "Digital twin for intelligent context-aware IoT healthcare systems." *IEEE Internet of Things Journal* 8, no. 23 (2021): 16749 16757.

[3] Hajihassani, Omid, OmidArdakanian, and HamzehKhazaei. "Anonymizing sensor data on the edge: a representation learning and transformation approach." *ACM Transactions on Internet of Things* 3, no. 1 (2021): 1-26.

[4] Cai, Zhipeng, ZuobinXiong, HonghuiXu, Peng Wang, Wei Li, and Yi Pan. "Generative adversarial networks: A survey toward private and secure applications." *ACM Computing Surveys (CSUR)* 54, no. 6 (2021): 1-38.

[5] Bui, Van, Tung Lam Pham, Huy Nguyen, and Yeong Min Jang. "Data Augmentation Using Generative Adversarial Network for Automatic Machine Fault Detection Based on Vibration Signals." *Applied Sciences* 11, no. 5 (2021): 2166.

[6] Vi, Bao Ngoc, Dinh Tan Nguyen, Cao Truong Tran, HuuPhuc Ngo, Chi Cong Nguyen, and Hai-Hong Phan. "Multiple Imputation by Generative Adversarial Networks for Classification with Incomplete Data." In *2021 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pp. 1-6. IEEE, 2021.

[7] Undirwade, Ashish, and Sujit Das. "Image anonymization using deep convolutional generative adversarial network." *Machine Learning Algorithms and Applications* (2021): 305-330.

[8] Piacentino, Esteban, Alvaro Guarner, and Cecilio Angulo. "Generating synthetic ecgs using gans for anonymizing healthcare data." *Electronics* 10, no. 4 (2021): 389.

[9] Parvathy, VelmuruganSubbiah, SivakumarPothiraj, and Jenyfal Sampson. "Automated internet of medical things (iomt) based healthcare monitoring system." In *Cognitive Internet of Medical Things for Smart Healthcare*, pp. 117-128. Springer, Cham, 2021.
[10] Siddiqui, MohdFaizan. "IoMT potential impact in COVID-19: combating a pandemic with innovation." In *Computational intelligence methods in COVID-19: Surveillance, prevention, prediction and diagnosis*, pp. 349-361. Springer, Singapore, 2021.
[11] Jang, Seunghui, and Yanggon Kim. "A Fast Training Method using Bounded Continual Learning in Image Classification." In *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 186-191. IEEE, 2021.
[12] Mongelli, Maurizio. "Design of countermeasure to packet falsification in vehicle platooning by explainable artificial intelligence." *Computer Communications* 179 (2021): 166-174.
[13] Matarese, Maria, Roberta Pendoni, MichelaPiredda, and Maria Grazia De Marinis. "Caregivers' experiences of contributing to patients' self-care in Chronic Obstructive Pulmonary Disease: A thematic synthesis of qualitative studies." *Journal of advanced nursing* 77, no. 10 (2021): 4017-4034.
[14] Han, Daoqi, LingyiXu, Ruohan Cao, HuiGao, and Yueming Lu. "Anti-collision voting based on Bluetooth low energy improvement for the ultra-dense edge." *IEEE Access* 9 (2021): 73271-73285.
[15] Adib, Edmond, FatemehAfghah, and John J. Prevost. "Arrhythmia Classification using CGAN-augmented ECG Signals." *arXiv preprint arXiv:2202.00569* (2022).
[16] Banerjee, Rohan, and AvikGhose. "Synthesis of realistic ecg waveforms using a composite generative adversarial network for classification of atrial fibrillation." In *2021 29th European Signal Processing Conference (EUSIPCO)*, pp. 1145-1149. IEEE, 2021.
[17] Khare, Priyanshi, Rajesh Wadhvani, ManasiGyanchandani, and Banalaxmi Brahma. "Generating Data for Real World Time Series Application with GRU-Based Conditional GAN." In *Proceedings of International Conference on Data Science and Applications*, pp. 335-349. Springer, Singapore, 2022.
[18] Dasgupta, Subhrajyoti, Sudip Das, and Ujjwal Bhattacharya. "CardioGAN: An Attention-based Generative Adversarial Network for Generation of Electrocardiograms." In *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 3193-3200. IEEE, 2021.
[19] Shin, Heean, Sukkyu Sun, Joonnyong Lee, and Hee Chan Kim. "Complementary photoplethysmogram synthesis from electrocardiogram using generative adversarial network." *IEEE Access* 9 (2021): 70639-70649.
[20] Zhang, Yu-He, and Saeed Babaeizadeh. "Synthesis of standard 12-lead electrocardiograms using two-dimensional generative adversarial networks." *Journal of Electrocardiology* 69 (2021): 6-14.