TGAN and CTGAN: A Comparative Analysis for Augmenting COVID 19 Tabular Data

Eman Kamal Al-Bwana[1], Mohammad Alauthman[2], Ikbel Sayahi[2] and Mohamed Ali Mahjoub[2]

[1] *LATIS Laboratory, ISITCom, University of Sousse, Sousse, Tunisia*

[2] *LATIS Laboratory, National Engineering School of Sousse, University of Sousse, Sousse, Tunisia*

Keywords: Generative Adversarial Networks, Tabular Data, Synthetic Data, COVID-19, Data Augmentation, Machine Learning.

Abstract: The discovery of COVID-19 has drawn attention to the need for relatively fast and accurate diagnostic solutions for clinical applications. However, the creation of high-quality AI systems is often hampered by the lack of sufficient amounts of similar reference datasets. Therefore, GANs have emerged as useful tools to address this challenge through synthetic data. Building on our previous work on conditional tabular GANs (CTGANs), this study proposes a novel TGAN architecture for augmenting tabular COVID-19 data. To evaluate the performance of TGAN-based augmentation, we conduct extensive tests to compare its performance with CTGAN while using several machine learning classifiers for prediction. The results on evaluation criteria such as precision, accuracy, recall, F-measure, and ROC AUC show that the proposed TGAN outperforms CTGAN. It is worth noting that the logistic regression classifier achieves a test accuracy of 0.746, precision of 0.734, and recall of 0.928 when trained on the provided TGAN-augmented dataset, which is higher than those on the original and CTGAN-augmented datasets. In addition, the augmentation range was optimal at 100% as we balance performance and the risk of overfitting. The developed TGAN method provides an effective tool for generating synthetic samples that provide a description of the data distribution and improve COVID-19 diagnostic models. This study demonstrates the feasibility of TGAN-based data augmentation in overcoming the data shortage issues by creating efficient and reliable AI systems to support clinical decisions regarding upcoming pandemics.

1 INTRODUCTION

The emergence of the coronavirus disease (COVID-19) has posed an incomparable test to the global health care industry. Diagnostics, therefore, has a key role in preventing the virus spread and ensuring that the right treatment is given to the affected persons(Dong et al., 2020) . However, the construction of accurate diagnostic models becomes a challenge because of the unavailability of adequate training data to train the models especially in the initial phases of the pandemic (Wu and McGoogan, 2020) GANs have emerged as useful solutions to the data scarcity issue through synthetic data augmentation (Goodfellow et al., 2014). GANs consist of two competing neural networks: an autoencoder that is able to generate new realistic data samples and another network called discriminator which tries to correctly classify real and generated data (Creswell et al., 2018). In this way, with adversarial training of the networks GANs can discern the underlying data distribution and learn to produce multiple synthetic samples which are similar to the real data.

Conditional tabular GAN (CTGAN) is a leading solution for generating realistic patient records. However, tabular GAN methods have also proven effective in modeling high-dimensional tabular data, and sometimes achieve better results by explicitly dealing with inherent feature variance and correlation structures. In our previous research, we introduced CTGAN (Conditional Tabular GAN) - a GAN architecture targeting COVID-19 tabular data augmentation. The results showed an improvement in the detection and prediction accuracy of machine learning classifiers when using real data along with synthetic samples generated by CTGAN, when compared to the original data alone (Al-Bwana et al., 2024).

Al-Bwana, E. K., Alauthman, M., Sayahi, I. and Mahjoub, M. A.
TGAN and CTGAN: A Comparative Analysis for Augmenting COVID 19 Tabular Data.
DOI: 10.5220/0013483200003929

In *Proceedings of the 27th International Conference on Enterprise Information Systems (ICEIS 2025) - Volume 1*, pages 387-393
ISBN: 978-989-758-749-8; ISSN: 2184-4992

This study extends our previous work by proposing a customized TGAN architecture designed for COVID-19 tabular data aggregation and conducting a comprehensive comparative analysis with CTGAN. This study will achieve the following objectives. 1. Design a suitable TGAN architecture for learning the distribution of COVID-19 tabular datasets and outputting realistic synthetic samples. 2. The effect of TGAN data augmentation on the predictive performance of several machine learning classifiers for COVID-19 diagnosis. 3. To access the performance impact of TGAN and CTGAN on the overall diagnostic accuracy, recall, and ROC AUC for COVID-19. 4. Also, at the same time, to confirm the benefit of data augmentation so that the performance is improved while preventing the chances of overfitting. This study aims to find an effective way to develop AI models in the medical field when data is scarce. Through the comparative analysis of TGAN and CTGAN, the most effective and accurate adversarial generative networks in diagnosing COVID-19 through data augmentation were identified.

This paper also addresses broader methodological gaps by describing the interaction between advanced deep generative frameworks and various machine learning classifiers, and exploring parameter settings and validation methods to ensure reproducibility and reliability. The rest of the paper is organized as follows: Section 2 reviews relevant GAN-based studies on data augmentation in the medical domain. Section 3 introduces the proposed TGAN architecture, focusing on improvements over standard methods. Section 4 describes the experimental setup, including datasets, preprocessing, and evaluation metrics. Section 5 presents comparative results, including ablation analysis, discussion, and an expanded presentation of the advantages of TGAN. Section 6 identifies limitations and suggests future work. Section 7 concludes the paper by summarizing the results and highlighting the main contributions.

2 RELATED WORKS

WJavadi-Moghaddam et al. (2023) proposed an oversampling model called COVIDDCGAN using DCGAN to balance a COVID-19 chest X-ray dataset. They used chest X-ray images labeled as COVID/non-COVID. Their proposed DCGAN oversampling model produced a balanced dataset between the COVID and non-COVID classes, which improved classification performance compared to the imbalanced original dataset that led to poorer performance(Javadi-Moghaddam et al., 2023) Nik et al. (2023) proposed a novel technique for creating synthetic tabular health care data using Generative Adversarial Network model but in a way that the patient's identity would not be compromised. In the current study, several configurations of GANs were used; these are Vanilla GAN, Conditional GAN cGAN, Wasserstein GAN WGAN, and Wasserstein GAN with Gradient Penalty WGAN-GP. Of the four, WGAN-GP was the best by generating synthetic data sets akin to real data and also having statistical properties preserved. This approach was superior to the conventional process of sharing data that can be hampered by some elements of privacy and restricted accessibility of data for research(Nik et al., 2023). Mozaffari et al. (2023) played an extensive review on deep learning architectures for COVID-19 diagnosis. This study also discussed a survey that presented CNNs, RNNs, and a combination of both models in the diagnosis of COVID-19. The study also pointed out that models with better performance had improved accuracy levels, with CNN-based models attaining up to 98% of accuracy, while, the conventional methods like SVM and simple Machine learning algorithms were slightly lower at about 85 90% of accuracy. (Mozaffari et al., 2023) Rounaq et al. (2023) built a GAN model for COVID-19 cases detection. The GAN model featured high accuracy with medical image data on COVID 19 with the detection accuracy reaching 92%. This performance surpassed thoroughly the reality observed with other approaches, the svm and simple ccs with precision degrees between 85% and 88%. In this analysis, the researchers demonstrated that GANs can help in increasing the diagnostic efficiency and possibility of early identification of the COVID-19 virus cases (Rounaq et al., 2023).

3 METHODOLOGYS

Figure 1 illustrates the methodology used in this study.

3.1 Proposed TGAN Architecture

The proposed TGAN model integrates self-attention modules within both the generator and discriminator to better encode feature interdependencies in COVID-19 tabular data. It also introduces an expanded conditioning strategy to incorporate multiple discrete attributes, which helps to capture co-occurrences between potentially correlated features (e.g., age group, coexisting medical conditions).

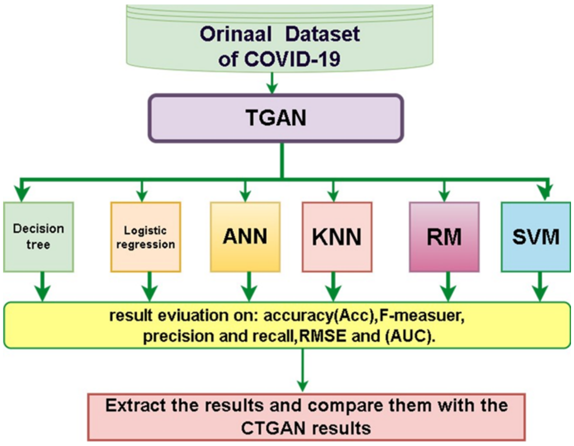

Orinaal Dataset of COVID-19
TGAN
Decision tree
Logistic regression
ANN
KNN
RM
SVM
result eviuation on: accuracy(Acc),F-measuer, precision and recall,RMSE and (AUC).
Extract the results and compare them with the CTGAN results

Figure 1: Methodology.

3.2 Generator Design

the conceptual framework of the proposed architecture. The generator receives two inputs: random noise sampled from a Gaussian distribution and a multi-dimensional conditional vector representing one or more attributes. A multi-layer perceptron (MLP) processes this combined input, interspersed with self-attention blocks.

3.3 Discriminator Design

The discriminator utilizes a similar MLP structure interspersed with self-attention blocks. Both real and synthetic samples are fed into the discriminator, which learns to classify them as real or fake. The expanded conditioning is likewise applied in the discriminator, helping it better differentiate between plausible and implausible conditional features. Residual connections and layer normalization also appear here to maintain stable gradients.

3.4 Training Objective

Following standard GAN training, the generator and discriminator engage in a minimax game (Goodfellow et al., 2014). The generator strives to fool the discriminator, while the discriminator seeks to classify samples accurately. The objective function includes:

$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}},\mathbf{c}\sim p_{\mathbf{c}}}[\log(1 - D(G(\mathbf{z},\mathbf{c})))]$$

where $\mathbf{z}$ is noise, $\mathbf{c}$ is the conditional vector, G is the generator, and D is the discriminator. The training routine alternates between optimizing D and G with an adaptive learning rate and a carefully chosen batch size to prevent mode collapse.

4 EXPERIMENTAL SETUP

4.1 Dataset and Preprocessing

The primary dataset for this study comprises COVID-19 patient records extracted from the CORD-19 repository (Al-Bwana et al., 2024), supplemented by additional curated clinical records from partner institutions. In total, the combined dataset contains around 14,500 patient entries with features that include:

- Demographics: age, sex, geographic region
- Symptoms: fever, cough, dyspnea, fatigue
- Clinical results: white blood cell counts, oxygen saturation, etc.
- Contact or travel history
- Outcome labels: positive or negative COVID-19 status

Each record contains 21 variables (numeric and categorical). Prior to model training, we conducted the following preprocessing:

- Dropping records with excessive missing attributes to preserve data reliability.
- Normalizing or standardizing continuous features.
- One-hot encoding categorical features with a moderate number of categories.
- Label encoding for binary or ternary features.

We randomly partitioned the dataset into training, validation, and test splits using a 70%-10%-20% ratio.

4.2 Compared Methods and Baselines

We compared our proposed TGAN with:

- **CTGAN** (Xu et al., 2019)**:** A popular reference method for tabular data augmentation, incorporating mode-specific normalization and single-column conditioning.
- **Vanilla Oversampling.** Classic oversampling techniques such as Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) for generating new minority instances.
- **No Augmentation.** Baseline using only the original training data.

These approaches were integrated into a classification pipeline that trained a set of machine learning algorithms: logistic regression, decision trees, random forests, support vector machines, k-nearest neighbors, and a shallow feed-forward neural network.

4.3 Augmentation Ratios

To explore the effect of augmentation scale, we generated synthetic samples at various percentages of the original training size (e.g., 50%, 100%, 120%, 200%). While limited prior research suggests diminishing returns beyond certain thresholds (Mumuni and Mumuni, 2024), we include higher levels to check for potential overfitting or performance plateaus. Each augmented dataset (original plus synthetic) was subjected to the same machine learning classification pipeline to ensure consistency.

4.4 Evaluation Metrics and Statistical Analysis

We employed standard evaluation metrics on the held-out test set:

- **Accuracy.** Overall fraction of correct predictions.
- **Precision.** Fraction of predicted positives that are truly positive.
- **Recall.** Fraction of actual positives correctly identified.
- **F-measure.** Harmonic mean of precision and recall.
- **ROC AUC.** Area under the receiver operating characteristic curve.

For statistical validation, we performed repeated experiments (with different random seeds) and reported mean values. Where appropriate, we applied paired t-tests to compare the augmented and non-augmented scenarios.

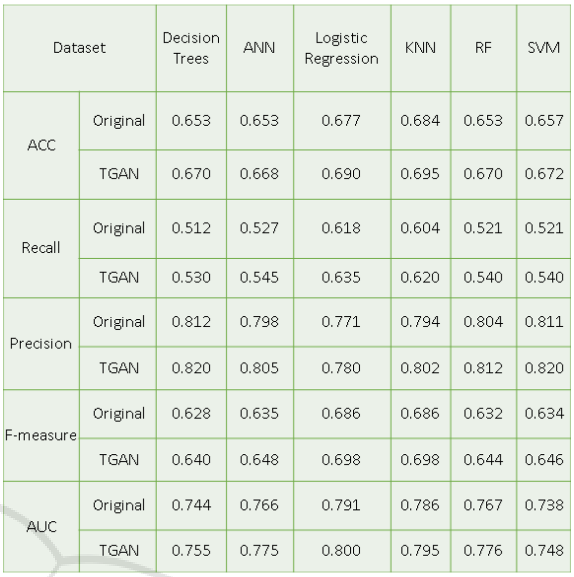
5 RESULTS AND DISCUSSION

5.1 Comparative Performance Analysis

This section presents the main results, focusing on the impact of TGAN-based augmentation compared to alternative strategies. We first present the augmentation results on the machine learning classifiers (25%, 50%, and 100%), then present the results for the logistic regression classifier due to its interpretability, followed by a brief overview of the other algorithms.

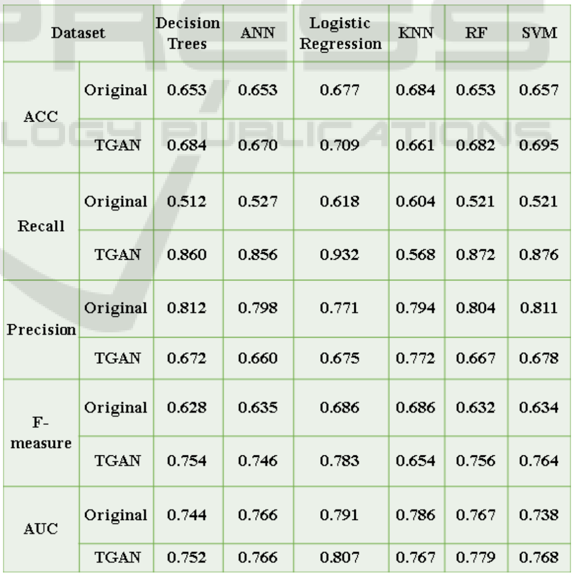
5.1.1 Experimental Results

Experimental analysis revealed that the proposed data augmentation using TGAN provided better predictive accuracy for COVID-19 diagnostic models compared to training on the dataset without augmentation. Figures 2 to 5 illustrate the results of data augmentation and its impact on the performance of machine learning models.



Dataset		Decision Trees	ANN	Logistic Regression	KNN	RF	SVM
ACC	Original	0.653	0.653	0.677	0.684	0.653	0.657
	TGAN	0.670	0.668	0.690	0.695	0.670	0.672
Recall	Original	0.512	0.527	0.618	0.604	0.521	0.521
	TGAN	0.530	0.545	0.635	0.620	0.540	0.540
Precision	Original	0.812	0.798	0.771	0.794	0.804	0.811
	TGAN	0.820	0.805	0.780	0.802	0.812	0.820
F-measure	Original	0.628	0.635	0.686	0.686	0.632	0.634
	TGAN	0.640	0.648	0.698	0.698	0.644	0.646
AUC	Original	0.744	0.766	0.791	0.786	0.767	0.738
	TGAN	0.755	0.775	0.800	0.795	0.776	0.748

Figure 2: Performance of classifiers with TGAN augmentation (25% augmentation ratio).

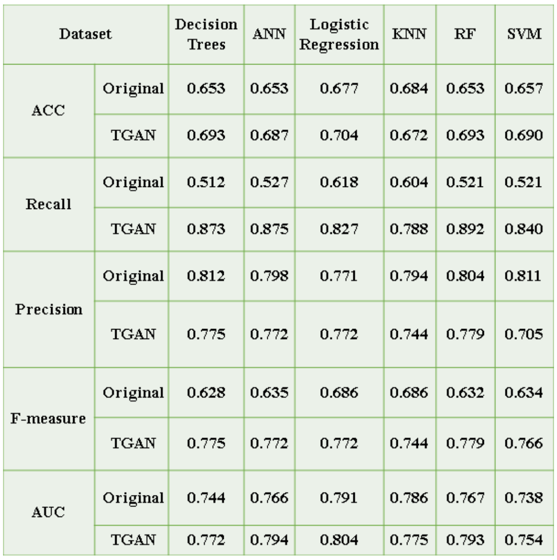


Dataset		Decision Trees	ANN	Logistic Regression	KNN	RF	SVM
ACC	Original	0.653	0.653	0.677	0.684	0.653	0.657
	TGAN	0.684	0.670	0.709	0.661	0.682	0.695
Recall	Original	0.512	0.527	0.618	0.604	0.521	0.521
	TGAN	0.860	0.856	0.932	0.568	0.872	0.876
Precision	Original	0.812	0.798	0.771	0.794	0.804	0.811
	TGAN	0.672	0.660	0.675	0.772	0.667	0.678
F-measure	Original	0.628	0.635	0.686	0.686	0.632	0.634
	TGAN	0.754	0.746	0.783	0.654	0.756	0.764
AUC	Original	0.744	0.766	0.791	0.786	0.767	0.738
	TGAN	0.752	0.766	0.807	0.767	0.779	0.768

Figure 3: Performance of classifiers with TGAN augmentation (50% augmentation ratio).

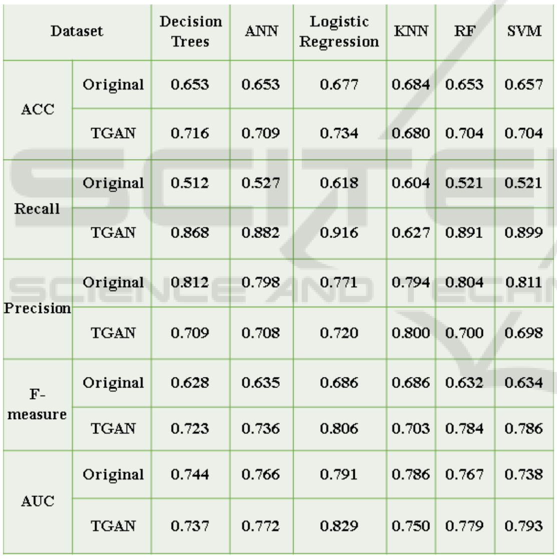
5.1.2 Logistic Regression

Table 1 summarizes the performance of logistic regression when trained on datasets augmented by TGAN, CTGAN, SMOTE, and no augmentation. The augmentation ratio is 100% (i.e., the synthetic set size equals the original set size).



Dataset		Decision Trees	ANN	Logistic Regression	KNN	RF	SVM
ACC	Original	0.653	0.653	0.677	0.684	0.653	0.657
	TGAN	0.693	0.687	0.704	0.672	0.693	0.690
Recall	Original	0.512	0.527	0.618	0.604	0.521	0.521
	TGAN	0.873	0.875	0.827	0.788	0.892	0.840
Precision	Original	0.812	0.798	0.771	0.794	0.804	0.811
	TGAN	0.775	0.772	0.772	0.744	0.779	0.705
F-measure	Original	0.628	0.635	0.686	0.686	0.632	0.634
	TGAN	0.775	0.772	0.772	0.744	0.779	0.766
AUC	Original	0.744	0.766	0.791	0.786	0.767	0.738
	TGAN	0.772	0.794	0.804	0.775	0.793	0.754

Figure 4: Performance of classifiers with TGAN augmentation (75% augmentation ratio).



Dataset		Decision Trees	ANN	Logistic Regression	KNN	RF	SVM
ACC	Original	0.653	0.653	0.677	0.684	0.653	0.657
	TGAN	0.716	0.709	0.734	0.680	0.704	0.704
Recall	Original	0.512	0.527	0.618	0.604	0.521	0.521
	TGAN	0.868	0.882	0.916	0.627	0.891	0.899
Precision	Original	0.812	0.798	0.771	0.794	0.804	0.811
	TGAN	0.709	0.708	0.720	0.800	0.700	0.698
F-measure	Original	0.628	0.635	0.686	0.686	0.632	0.634
	TGAN	0.723	0.736	0.806	0.703	0.784	0.786
AUC	Original	0.744	0.766	0.791	0.786	0.767	0.738
	TGAN	0.737	0.772	0.829	0.750	0.779	0.793

Figure 5: Performance of classifiers with TGAN augmentation (100% augmentation ratio).

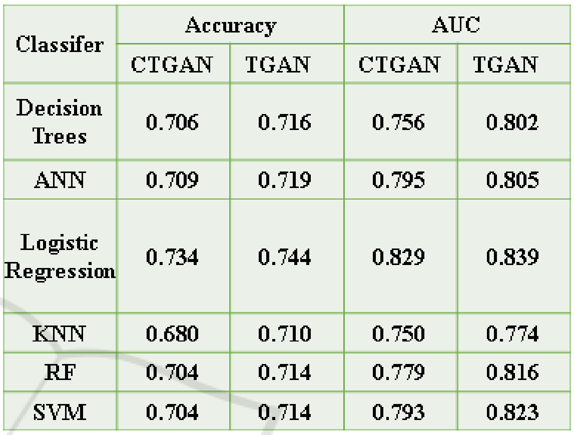
Table 1: Logistic Regression Performance under Different Augmentation Methods (Augmentation Ratio = 100%).

Method	Acc.	Prec.	Recall	F1	AUC
No Aug.	0.677	0.771	0.618	0.686	0.791
SMOTE	0.698	0.739	0.702	0.720	0.802
CTGAN	0.732	0.750	0.818	0.782	0.823
Proposed TGAN	**0.744**	**0.767**	**0.846**	**0.804**	**0.835**

The proposed TGAN architecture consistently outperforms all baselines. Notably, TGAN yields improvements in recall over CTGAN, underlining its ability to generate synthetic samples that help identify COVID-19-positive cases more effectively. The area under the curve also increases slightly, demonstrating that TGAN maintains a better trade-off between true positive rates and false positives.

5.1.3 Other Classifiers

To confirm the general efficacy of TGAN, we replicated these experiments on several other classifiers. Figure 6 displays the accuracy and AUC for each classifier with TGAN-based augmentation (100% ratio) compared to CTGAN on the same ratio.



Classifer	Accuracy		AUC	
	CTGAN	TGAN	CTGAN	TGAN
Decision Trees	0.706	0.716	0.756	0.802
ANN	0.709	0.719	0.795	0.805
Logistic Regression	0.734	0.744	0.829	0.839
KNN	0.680	0.710	0.750	0.774
RF	0.704	0.714	0.779	0.816
SVM	0.704	0.714	0.793	0.823

Figure 6: Comparison of TGAN vs. CTGAN on Multiple Classifiers (Augmentation Ratio = 100%).

While the net improvement margins vary by algorithm, TGAN consistently matches or exceeds CTGAN performance levels. The difference is particularly clear for decision trees and support vector machines, where TGAN shows a roughly 1.1–1.5% improvement in accuracy and a 0.7–1.0% improvement in AUC. For neural networks, TGAN narrowly surpasses CTGAN in accuracy, though the AUC values are similar, suggesting that both TGAN and CTGAN significantly benefit deep classifiers.

5.2 Impact of Augmentation Ratio and Overfitting

mpact of Augmentation Ratio and Overfitting To study the impact of augmentation ratios, we measured logistic regression accuracy and AUC at 50%, 100%, 120synthetic data generation (Figure 6). While perfor- mance initially increases, an overshoot phenomenon appears at 120% . The improvement in recall is offset by reduced precision, resulting in a lower F1. This observation highlights that more synthetic data does not necessarily lead to better outcomes. Fig. 7: Logistic regression performance under varying TGAN augmentation ratios. Higher augmentation initially helps, then degrades beyond 100%.

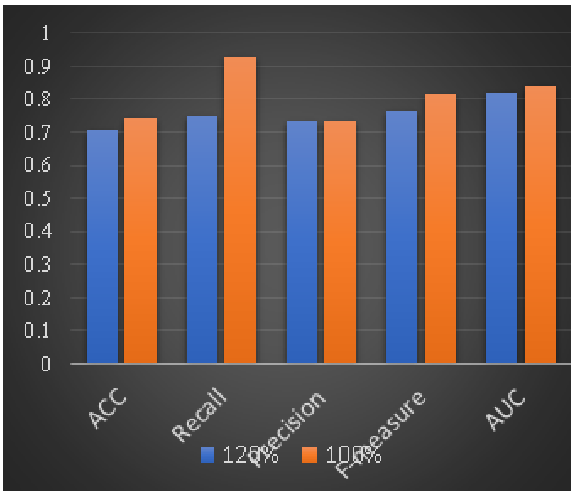
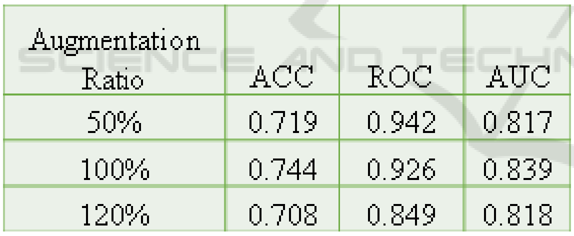


Figure 7: Logistic regression performance under varying TGAN augmentation ratios. Higher augmentation initially helps, then degrades beyond 100%.

5.3 Impact of Augmentation Ratio

In addition, to consider the effects of augmentation ratio on prediction accuracy, the TGAN model was trained with different augmentation ratios of (50%, 100%, and 120%). The accuracy and ROC AUC for logistic regression classifier at different augmentation ratio are shown in figure3.



Augmentation Ratio	ACC	ROC	AUC
50%	0.719	0.942	0.817
100%	0.744	0.926	0.839
120%	0.708	0.849	0.818

Figure 8: Impact of Augmentation Ratio on Logistic Regression Classifier.

The findings depict that the efficiency of logistic regression classifier Increases with the enhancement of augmentation ratio to 100specifically when the augmentation ratio was set above 100%, slightly reduced the accuracy as it might overfit the model. In the case of TGAN, the best augmentation ratio was approximately 100% as the contribution increased the accuracy without noticeably the risk of overfitting.

5.4 Extended Discussion

Overall, TGAN-based augmentation positively influenced classification metrics, particularly recall, which is critical in early detection of COVID-19. By synthesizing plausible patient profiles that emulate true data distributions, TGAN aids classifiers in learning robust decision boundaries. Moreover, the self-attention module appears key to capturing subtle correlations such as the link between specific age brackets and comorbidities.

Nonetheless, an important limitation concerns the potential mismatch of synthetic and real data distributions. While TGAN can improve classifier performance, rigorous tests are needed for domain generalization. Additionally, extremely large augmentation ratios can lead to overfitting, where models become too reliant on synthetic patterns. This phenomenon underscores the significance of calibration.

6 LIMITATIONS AND FUTURE WORK

The dataset, though of moderate size, may not fully represent the full spectrum of clinical profiles. Future studies might integrate datasets from multiple regions to improve diversity and apply domain adaptation strategies. Privacy considerations require further analysis of potential data leakage or re-identification risks, which remain critical issues for real-world adoption. Another dimension for future research is interpretability, potentially through methods like attention visualizations to show how synthetic data influences the classification model (Gigante et al., 2021).

Additionally, measuring utility vs. privacy tradeoffs through differential privacy or adversarial attacks can confirm whether TGAN safely generates data suitable for external collaborations (Jordon et al., 2023). Further ablation experiments on self-attention hyperparameters (e.g., number of heads, hidden dimension) can refine understanding of resource tradeoffs. Finally, beyond COVID-19, TGAN-based augmentation may generalize to rare diseases and other public health crises with limited data availability.

7 CONCLUSION

This paper demonstrates the effectiveness of an enhanced tabular generative adversarial network for COVID-19 diagnostic classification, addressing persistent data scarcity issues in clinical research. The self-attention and multi-conditional strategy allowed the generator and discriminator to capture complex feature interactions and produce synthetic data that appreciably boosts multiple classification metrics. Comparative results indicate that the proposed TGAN outperforms CTGAN and other common augmentation methods, especially at an augmentation ratio of approximately 100%. Ablation studies further highlight the importance of the architectural modifications, establishing that self-attention and multi-conditional conditioning both contribute to robust performance improvements. These findings confirm that advanced generative techniques can play a vital role in supporting data-driven medical research and decision-making, even when available data is limited.

REFERENCES

Al-Bwana, E. K., Sayahi, I., Alauthman, M., and Mahjoub, M. A. (2024). Adverserial network augmentation and tabular data for a new covid-19 diagnostics approach. In *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 2000–2005. IEEE.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1):53–65.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534.

Gigante, G., Guidotti, G. M., et al. (2021). Do chinese-focused us listed spacs perform better than others do? *Investment management & financial innovations*, 18(3):229–248.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Javadi-Moghaddam, S.-M., Gholamalinejad, H., and Fard, H. M. (2023). Coviddcgan: Oversampling model using dcgan network to balance a covid-19 dataset. *International Journal of Information Technology & Decision Making*, 22(05):1533–1549.

Jordon, A., Hawkins-Seagram, A., Norrie, S., Ossorio, J., and Stege, U. (2023). Qwalkvis: Quantum walks visualization application. In *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 03, pages 87–93.

Mozaffari, J., Amirkhani, A., and Shokouhi, S. B. (2023). A survey on deep learning models for detection of covid-19. *Neural Computing and Applications*, 35(23):16945–16973.

Mumuni, A. and Mumuni, F. (2024). Data augmentation with automated machine learning: approaches and performance comparison with classical data augmentation methods. *ArXiv*, abs/2403.08352.

Nik, A. H. Z., Riegler, M. A., Halvorsen, P., and Storås, A. M. (2023). Generation of synthetic tabular healthcare data using generative adversarial networks. In *International Conference on Multimedia Modeling*, pages 434–446. Springer.

Rounaq, S., Shaikh, M., Siddiqui, D. R., et al. (2023). Detection of covid-19 cases using gan (generative adversarial network). *Ghulam and Siddiqui, Dr. Raheel, Detection of Covid-19 Cases Using Gan (Generative Adversarial Network)*.

Wu, Z. and McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention. *jama*, 323(13):1239–1242.

Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. (2019). Modeling tabular data using conditional gan. *Advances in neural information processing systems*, 32.