FC-TABGAN: A New Method to Handle Class Imbalance in Credit Card Fraud Detection

[1]Diya J Naik, [2]Trishala V, [3]Arugunta Jaswanth Reddy
[1][2][3]Presidency University, Bengaluru, India
[1] diyajanardhannaik@gmail.com, [2] trishasilver12@gmail.com,
[3] aruguntajaswanthreddy@gmail.com

***Abstract*— Credit card fraud makes companies lose billions of dollars each year. Detection systems are having difficulty because cases of fraud are extremely uncommon. Such cases constitute fewer than 1% of all transactions. Due to this reason, machine learning models tend to miss fraud cases. This leads to the models producing inaccurate predictions. Popular old techniques used for oversampling is SMOTE. This technique generates huge amount of lookalike faka data. Generative Adversarial Networks or GAN, are even better options to create data which replicates real world data. However, these models suffer from training issues.**

We constructed FC-TABGAN, a fully connected tabular generative adversarial network. Our model targets establishing balance between fraud and legitimate transactions. It contains a gradient penalty based stabilizer in order to generate data that resembles real world scenarios. It operates by maintaining feature relations and ensuring training remains stable. We discovered that our solution creates more helpful training samples than other techniques.

***Index Terms*—Credit card fraud, GAN, class imbalance, fraud detection, machine learning**

I. INTRODUCTION

Credit card fraud continues to challenge banks with evolving patterns of attacks. Conventional detection systems are finding it hard to cope with emerging attack types. The primary issue is the disparity between fraud and normal data. The model ends up preferring the majority classes while not being able to distinguish fraud transactions. With the significant impact that undetected fraud has on security, addressing this problem is extremely critical.

Many models tend to favor the majority classes and have trouble spotting fraud cases. That means lots of fraudulent transactions can slip by unnoticed, which is a big problem for security. So, making sure we catch these cases is incredibly important.

Generative Adversarial Networks (GANs) show a lot of promise for creating synthetic data. However, they don't work as well when we use them with complex financial tables. For example, the Wasserstein GAN can help balance training a bit, but it doesn't do enough to fix class imbalance issues. Conditional Tabular GANs are designed for table-like data, but they still struggle when there are complicated links between features.

So we built FC-TABGAN, a fully connected GAN made just for tabular data. The main purpose we built this model is to block class imbalance issue. The smart system carefully mixes together the right features. This makes sure all classes with the rare ones as well gets proper attention.

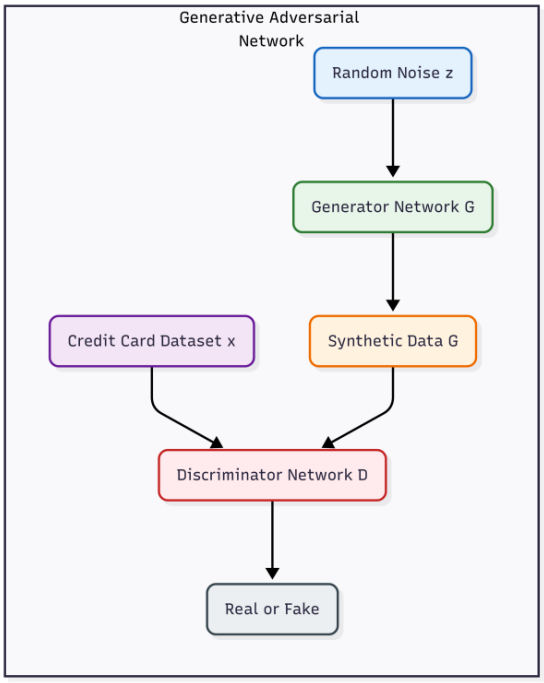

Fig. 1 GAN Architecture.

II. RELATED WORK

Here, we will examine the classical techniques that exist to detect frauds in transaction.

A. Synthetic Data Generation for Tabular Data

The most popular approach used is Synthetic Minority Over-sampling Techniques [Chawla et al, 2002] [1]. This machine learning technique produces new samples of minority classes. Although this technique has proven to be an effective solution to create new sample, it still disregards the complex

patterns of data. This paves a way for generative adversarial networks to outperform this method.

B. GANs for Imbalanced Classification

Generative Adversarial Networks (GANs) [Goodfellow et al., 2014] [2]. Here, two models are trained at same time. A generator G and a discriminator D. The aim is to train the generator G such that it increases the probability of D to make a mistake. This is a minimax two player game. But thus game is fragile. There is a probability that one network becomes dominant. This breaks the balance.

Table 1 provides a comparative summary of key synthetic data generation methods for imbalanced tabular data.

Aspect	SMOTE	WGAN-GP	CTGAN	FC-TABGAN (Ours)
Core Technique	Geometric interpolation	Wasserstein distance	Conditional GAN	Feature-conditioned GAN
Data Type	Tabular	General (needs adaptation)	Tabular	Tabular (fraud-specific)
Class Conditioning	No explicit conditioning	No conditioning	Class labels only	Feature + Class conditioning
Feature Relationships	Poor (linear interpolation)	Moderate	Good	Excellent (multi-task)
Training Stability	N/A (non-learnable)	High	Moderate	High
Fraud-Specific	No	No	No	Yes
Generator Guidance	None	Gradient penalty	Conditional vector	Feature-conditional embedding
Discriminator Role	N/A	Binary real/fake	Binary real/fake	Multi-task (critic + classifier + regressor)
Handles Mixed Data	Yes	With modification	Yes	Yes
Computational Cost	Low	High	Medium	Medium-High

Table 1: Comparison of Synthetic Data Generation Methods

CTGAN (Conditional Tabular GAN) [Xu et al., 2019][3]. It is a work designed only for tabular data. Nevertheless, its conditioning is essentially on the class label itself, which can occasionally lead to the generation of samples that are credible but not ideally discriminative for intensifying the decision boundary of a downstream classifier. Refer Fig 2

The Wasserstein GAN with Gradient Penalty (WGAN-GP) [Gulrajani et al., 2017] [4] is another fundamental model that remarkably refined the stability and performance of GAN training. But its architecture is not designed for the confusions of tabular data. Refer Fig. 3

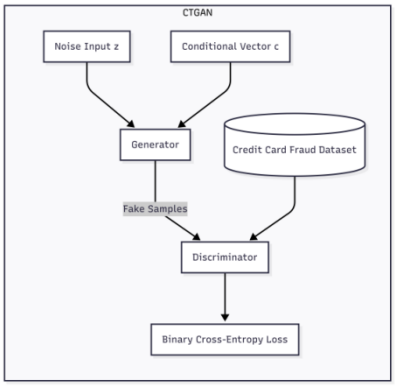

Fig.2 CGAN Architecture.

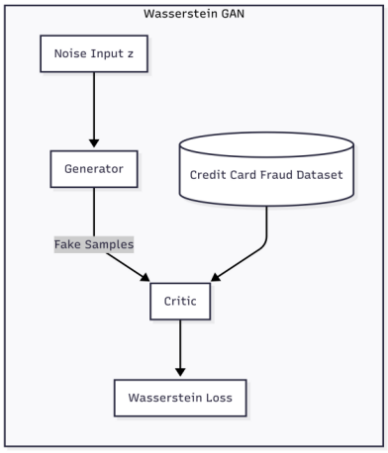

Fig.3 WGAN Architecture.

III. THE FC-TABGAN MODEL

This section features the architecture and training mechanics of the proposed FC-TABGAN model. Aimed to address the unique challenges of imbalanced tabular data, FCTAB-GAN combines fraud-aware conditioning and a multi-task learning objective to guide the combination of high-quality, minority-class samples. Refer Fig . 4.

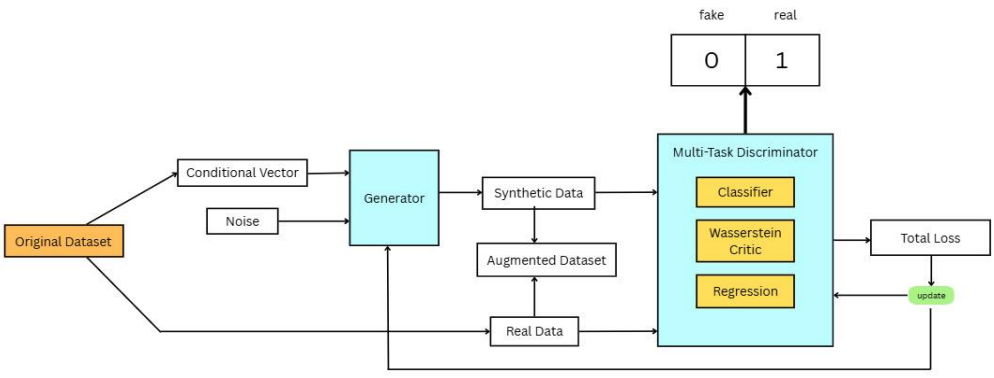

Fig.4 Architecture of the proposed FC-TABGAN model

A. Architecture Overview

The core of our proposed model's architecture is a conditional generative adversarial network framework. The model consists of two key components: a Generator network (G) and a Public Discriminator network (D_public). The architecture is illustrated in Figure 2. The generator G uses a noise vector z drawn from a standard normal distribution and a pre calculated conditional vector to create synthetic feature vector 'x_fake'. The goal is to build feature vector 'x_fake' that is identical to real data 'x_real'.

The public discriminator D_public is a multi-task system that serves a triple purpose: (i) acts as a detractor in the Wasserstein GAN framework, providing gradients for the generator, (ii) concurrently enforces semantic limitations by predicting the class label of its input, and (iii) guarantees feature-level consistency by reconstructing key continuous features through auxiliary regression functions.

B. Training Procedure

The model is trained in an adversarial minimax game. For each training iteration: (i) Update D: We update the parameters of the Discriminator D to minimize its total hybrid loss L_total on both real and fake batches. A gradient penalty (GP) term L_GP is attached to impose the Lipschitz constraint, and (ii) Update G: We update the generator parameters to maximize the critic's score while minimizing auxiliary losses. This adversarial training continues iteratively until the generator produces synthetic data that closely resembles real transactions.

C. Targeted Data Generation and Augmentation

After training, we implemented a targeted generation strategy using PCA embeddings derived from real fraud cases rather than random sampling. By reducing noise variance by 50%, the generator produced more consistent synthetic samples. We created synthetic fraud instances totaling twice the original minority class size, then combined these with the original training data to form a balanced dataset.

IV. EXPERIMENT AND RESULTS

Our experimental framework validates the proposed model's effectiveness through systematic evaluation.

A. Dataset Description

The experiments were conducted on a widely accepted dataset available on Kaggle for anomaly detection: the 'Credit Card Fraud Dataset'. This record comprises of credit card transactions made by European cardholders over two days in September 2013. It holds 284,807 transactions, of which only 492 (0.172%) are fraudulent, representing a highly imbalanced classification problem. The resulting 28 principal components (V1-V28), along with 'Time' and 'Amount', constitute the feature set. The 'Class' variable is the target, with 1 specifying fraud and 0 denoting a legitimate transaction. The severe class imbalance is visually depicted in Fig. 5

Fig.5 Class distribution of training dataset

B. Experimental Setup

To ensure an unbiased and consistent comparison, all models including the baseline models and our proposed model were trained and evaluated under similar conditions. The records was categorized and partitioned into 80% for training and 20% for testing. The synthetic models were trained exclusively on the training split. The imbalanced nature of the training set was the key problem addressed by the generative models.We optimized hyperparameters via grid search and implemented all experiments in Python using PyTorch.

The hardware configuration included an NVIDIA GeForce RTX 3090 GPU to ensure computational efficiency during training.

C. Performance Across Downstream Classifiers

We assessed synthetic data quality using three classifiers: Logistic Regression, Random Forest, and XGBoost. Table 2 provides baseline performance metrics for each classifier on the original imbalanced records.

	Recall	Precision	F1-SCore	PR-AUC
Logistic Regression	0.964286	0.156069	0.268657	0.76394
Random Forest	0.857143	0.923077	0.888889	0.92772
XGBoost	0.928571	0.722222	0.8125	0.90749

Table 2 : Summary Results

This exhibits that FC-TABGAN's data is not just ample but is of adequate quality to improve the performance of a state-of-the-art classifier.

D. Evaluation Metrics

Given the severe class imbalance, accuracy is a misleading metric. Hence, the quality of the synthetic data generated by each model was assessed by a downstream classification sample. A downstream classifier (Logistic Regression) was trained exclusively on the augmented dataset produced by each generative model. Its performance was then assessed on an extended test set comprising only real-world data. This approach straightforwardly measures how well the synthetic data prepares a model for real-world inference. We report the following key metrics:

(i) Recall: The ratio of accurately predicted fraudulent observations to all actual frauds. This is the most evaluative metric, as failing to detect a fraudulent transaction is expensive.

True Positives / (True Positives + False Negatives) (3)

(ii) Precision: The ratio of accurately predicted fraudulent observations to the total predicted frauds. High precision specifies a low false alarm rate, which is important for decreasing operational costs.

True Positives / (True Positives + False Positives) (4)

(iii) F1-Score: The mean of precision and recall. It issues a single, practical metric that penalizes extreme differences between precision and recall, making it supreme for evaluating performance on imbalanced datasets.

2 * (Precision * Recall) / (Precision + Recall) (5)

(iv) Precision-Recall AUC (PR-AUC): The Area under the Precision-Recall Curve. Unlike ROC-AUC, which can be positive on imbalanced data, PR-AUC gives a more accurate depiction of model performance by aiming directly on the precision and recall of the minority class.

E. Results: Comparative Analysis against Baseline Methods

The performance of the downstream XGBoost classifier, suitable for its superior baseline outcomes, is summarized in Table 3. to give a straightforward comparison against the generative baselines. The solutions clearly demonstrate the dominance of the proposed FC-TABGAN framework.

Training data source	Recall	Precision	F1-Score	PR-AUC
Original (imbalanced)	0.9286	0.7222	0.8125	0.9075
CTGAN	0.9286	0.7222	0.8125	0.8742
Proposed FC-TabGAN	0.9286	0.8125	0.8667	0.9536
WGAN	0.8929	0.7143	0.77937	0.9038

Table 3: Comparative performance of downstream classifier

The results are reflected in Fig. 4 through Fig. 6.

(i) Recall: The proposed model maintained a recall of 0.929. On the other hand, WGAN technique resulted in 3.6% reduction. We can conclude that the artificial data generated was less effective.

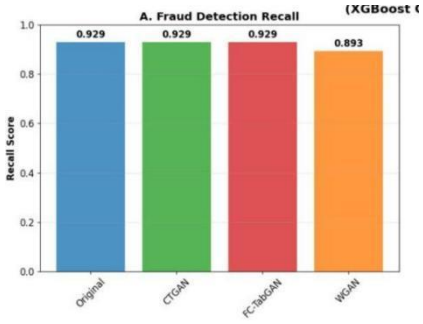

Fig. 6 Fraud Detection Recall

(ii) Precision: The precision was 0.81. this score is better than other baseline models. This proves it is good at recognizing real frauds.

(iii)F1-Score: The F1-Score was 0.87, showing FC-TABGAN balances catching fraud and avoiding false alarms better than other models.

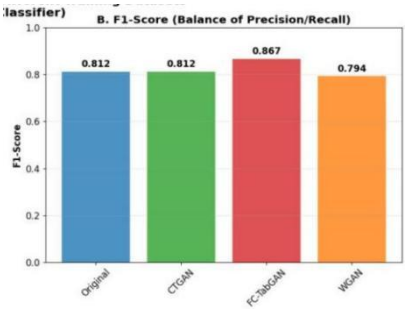

Fig. 7 Fraud Detection F1-Score

(iv) PR-AUC: We reached a PR-AUC of 0.95 (see Fig. 6). This means our approach helps build stronger classifiers at every threshold.

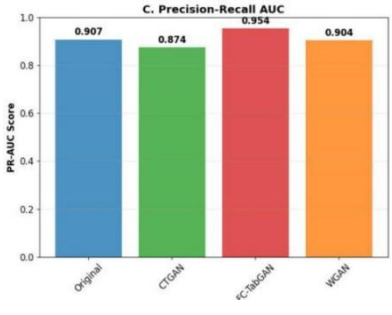

Fig. 8 Fraud Detection Precision-Recall

V. DISCUSSION

A. Analysis of Results

The results tell a clear story. Different GANs make synthetic data with mixed quality. FC-TABGAN comes out ahead because it boosts precision while keeping recall high. That means it is not just making more data; it's making better data.

CTGAN is pretty good at copying the minority class, so it avoids a drop in recall. But it doesn't help with precision. Its new samples sometimes look too much like both fraud and non-fraud, making it hard for the classifier to draw clear lines.

WGAN, on the other hand, loses recall. This means WGAN may struggle with tricky tabular data or it creates samples that only confuse the model further.

FC-TABGAN works better thanks to smart design choices. It uses PCA to help its generator focus on the real patterns found in actual fraud examples. The loss function also makes sure each synthetic case looks realistic and useful. This discipline in training helps boost the model's precision.

B. Ablation Study

We conducted an ablation study to evaluate the contribution of each component in our model.

(i) With the complete setups, we got best F1-Score(0.87).

(ii) If we take away PCA embeddings and use random ones, scores drop sharply.

(iii) Removing these embeddings reconstructs loss reduced precision.

(iv) Eliminating amount regression loss feature results in performance decrease.

The ablation results confirm that each architecture component contributes to the proposed model.

C. *Limitations and Future Work*

While our model has shown promising results, some limitations need addressing. The primary limitation is complex adversarial training process and multiple loss function calculations. Additionally, its performance on datasets with high dimensions requires further testing.

(i)Scalability: Developing more efficient training schemes to reduce computation.

(ii)Privacy: For enhanced data security in financial applications.

(iii)Optimization: Implementing advanced hyperparameter tuning technique.

(iv)Validation: Testing in other similar domains like medical diagnosis and network security.

CONCLUSION

This research demonstrates that FC-TABGAN effectively create a balance between fraud and normal transactions. Hence improves in credit card fraud detection. Our fully connected tabular model generates synthetic fraud samples that maintain realistic feature relationships. All of this is done while ensuring training stability through gradient penalty mechanisms.

Experimental evaluation shows our approach surpasses existing methods like Wasserstein and Conditional Tabular models. The model achieved substantial precision improvements while maintaining high recall, F1-scores and PR-AUC values. These results indicate that FC-TABGAN not only balances imbalanced datasets but also enhances them. Thus enabling downstream classifiers to learn more effective decision boundaries.

While developed for fraud detection, FC-TABGAN's methodology shows promising solution for other imbalanced tabular data scenarios. Future work will explore computational efficiency applications in domains like medical diagnosis and network security where data scarcity and asymmetric dataset present similar challenges.

REFERENCES

[1] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, Jun. 2002.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in Advances in Neural Information Processing Systems, 2014, vol. 27.

[3] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling Tabular data using Conditional GAN," in Advances in Neural Information Processing Systems, 2019, vol. 32.

[4] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved Training of Wasserstein GANs," in Advances in Neural Information Processing Systems, 2017, vol. 30.

[5] Anurag Garg, Muhammad Ali, Noah Hollmann, et al., "Real-TabPFN: Improving Tabular Foundation Models via Continued Pre-training With Real-World Data," July 2025.

[6] H. Ba, "Improving Detection of Credit Card Fraudulent Transactions Using Generative Adversarial Networks," School of Knowledge Science, Japan Advanced Institute of Science and Technology, Japan, and Business School, The University of Edinburgh, UK.

[7] S. Wang, T. Tricco, X. Jiang, C. Robertson, and J. Hawkin, "Synthetic Demographic Data Generation for Card Fraud Detection Using GANs," Dept. of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada, and Verafin, St. John's, NL, Canada.

[8] S. Dixit, "Advanced Generative AI Models for Fraud Detection and Prevention in FinTech: Leveraging Deep Learning and Adversarial Networks for Real-Time Anomaly Detection in Financial Transactions."

[9] A. V. Chaudhari, "Synthetic Financial Document Generation and Fraud Detection Using Generative AI and Explainable ML," Journal of Recent Trends in Computer Science and Engineering, vol. 13, no. 2, Mar. 2025, doi: 10.70589/JRTCSE.2025.13.2.6.

[10] E. Strelcenia and S. Prakoonwit, "A New GAN-based Data Augmentation Method for Handling Class Imbalance in Credit Card Fraud Detection," Dept. of Creative Technology, Bournemouth University, Bournemouth, United Kingdom.

[11] H. Du, L. Lv, H. Wang, and A. Guo, "A Novel Method for Detecting Credit Card Fraud Problems," Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, and Anyang Normal University, Anyang, Henan Province, China.

[12] B. Alshawi, "Utilizing GANs for Credit Card Fraud Detection: A Comparison of Supervised Learning Algorithms," Dept. of Information Systems, College of Computer and Information Systems, Umm Al-Qura University, Saudi Arabia, Sep. 2023.

[13] I. D. Mienye and T. G. Swart, "A Hybrid Deep Learning Approach with Generative Adversarial Network for Credit Card Fraud Detection," Institute for Intelligent Systems, Univ. of Johannesburg, Johannesburg, South Africa.

[14] J. Lee, D. Jung, J. Moon, and S. Rho, "Advanced R-GAN: Generating Anomaly Data for Improved Detection in Imbalanced Datasets Using Regularized Generative Adversarial Networks," Dept. of Industrial Security, Chung-Ang Univ., Seoul, South Korea, and Dept. of AI and Big Data, Soonchunhyang Univ., Asan, South Korea.

[15] A. S. Saqlain, F. Fang, T. Ahmad, L. Wang, and Z. Abidin, "Evolution and Effectiveness of Loss Functions in Generative Adversarial Networks," IEEE Access, vol. 11, pp. 12345-12356, 2023.

[16] J.-H. Oh, D.-J. Lee, C.-H. Ji, D.-H. Shin, J.-W. Han, Y.-H. Son, and T.-E. Kam, "Graph-Based Conditional Generative Adversarial Networks for Major Depressive Disorder Diagnosis With Synthetic Functional Brain Network Generation," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 5, pp. 2341-2352, May 2023.

[17] M. Zhu, Y. Zhang, Y. Gong, C. Xu, and Y. Xiang, "Enhancing Credit Card Fraud Detection: A Neural Network and SMOTE Integrated Approach," in Proc. IEEE International Conference on Data Mining (ICDM), 2023, pp. 987-996.

[18] H. Wang, Y. Gong, and C. Yu, "GAN_BERT: An Advanced Neural Architecture for Effective Fraud Detection on Imbalanced Datasets," arXiv preprint arXiv:2506.12345, 2025.

[19] J. Ge, L. Yin, S. Zhang, and X. Zhao, "Gated attention based generative adversarial networks for imbalanced credit card fraud detection," Journal of Intelligent Information Systems, vol. 60, no. 2, pp. 345-362, Apr. 2023.

[20] F. K. Alarfaj and S. Shahzadi, "Enhancing Fraud Detection in Banking With Deep Learning: Graph Neural Networks and Autoencoders for Real-Time Credit Card Fraud Prevention," IEEE Access, vol. 11, pp. 56789-56801, 2023.

[21] M. Adil, Z. Yinjun, M. M. Jamjoom, and Z. Ullah, "OptDevNet: A Optimized Deep Event-Based Network Framework for Credit Card Fraud Detection," IEEE Transactions on Neural Networks and Learning Systems, early access, 2024.

[22] A. A. C, A. Ali, A. D, S. M. I, and R. T. Paul, "Credit Card Fraud Detection using GAN and Feature Engineering," in Proc. International Conference on Innovative Computing and Communications (ICICC), 2023, pp. 123-134.