10.48047/jocaaa.2024.33.05.77

Synthetic Financial Transaction Data Generation Using Lightweight GAN Models

Adarsh Naidu
Independent Researcher, USA
Email : adarshnaidu2910@gmail.com

Abstract

The rapid digitalization of financial services has led to an unprecedented growth in transactional data, which serves as a critical foundation for applications such as fraud detection, credit risk assessment, customer behavior modeling, and regulatory compliance. Despite its importance, access to real-world financial transaction data remains highly restricted due to stringent privacy regulations, institutional confidentiality, and the risk of sensitive information leakage. These constraints significantly hinder research progress, reproducibility, and the development of robust machine learning models. Synthetic data generation has emerged as a compelling solution to this challenge, enabling the creation of artificial datasets that preserve the statistical and structural properties of real data without exposing private information. Among various generative approaches, Generative Adversarial Networks (GANs) have demonstrated superior capability in learning complex, high-dimensional data distributions, making them particularly suitable for modeling financial transactions. However, state-of-the-art GAN models are often computationally intensive, difficult to train, and impractical for deployment in resource-constrained environments. This paper proposes a lightweight GAN-based framework tailored for synthetic financial transaction data generation. The proposed model emphasizes architectural simplicity, efficient feature representation, and stable adversarial training while maintaining high data fidelity and privacy preservation. Experimental analysis shows that the generated synthetic data closely aligns with real transaction distributions and retains downstream utility, while significantly reducing model complexity. The results highlight the potential of lightweight GANs as a scalable and practical solution for privacy-preserving financial data synthesis.

I. Introduction

The financial industry has undergone a profound transformation driven by advances in digital payment systems, online banking, and real-time transaction processing. As a result, massive volumes of financial transaction data are generated daily, capturing intricate patterns of consumer behavior, spending habits, and anomalous activities. This data plays a central role in modern financial intelligence systems, particularly in applications such as fraud detection, anti-money laundering (AML), credit scoring, and personalized financial services. Machine learning and deep learning models trained on transaction-level data have demonstrated remarkable effectiveness in identifying subtle patterns that are often imperceptible to rule-based systems.

10.48047/jocaaa.2024.33.05.77

Despite its strategic value, real financial transaction data is rarely accessible for research and development purposes. Regulatory frameworks such as the General Data Protection Regulation (GDPR), along with industry standards like PCI-DSS, impose strict limitations on data sharing and usage. Even anonymized datasets pose re-identification risks due to the highly granular and behavioral nature of transaction records. These constraints create a significant gap between academic research and real-world financial systems, limiting reproducibility, benchmarking, and innovation.

Synthetic data generation has gained increasing attention as a viable approach to addressing these challenges. By generating artificial datasets that statistically resemble real data, synthetic data enables safe data sharing, model training, and system testing without compromising individual privacy. Traditional synthetic data generation methods in finance have largely relied on statistical modeling, rule-based simulations, or heuristic sampling techniques. While such approaches are computationally efficient, they often fail to capture complex nonlinear dependencies, temporal correlations, and rare event patterns that are intrinsic to financial transactions.

In recent years, deep generative models have emerged as powerful alternatives for synthetic data generation. In particular, Generative Adversarial Networks (GANs) have demonstrated exceptional performance in learning rich data distributions across various domains. GANs employ an adversarial learning paradigm in which a generator network produces synthetic samples while a discriminator network attempts to distinguish them from real data. Through this competitive process, GANs are capable of generating highly realistic data that closely approximates the underlying distribution of the training dataset. These properties make GANs especially appealing for financial transaction data, which is characterized by high dimensionality, skewed distributions, mixed feature types, and complex inter-feature dependencies.

However, applying GANs to financial data synthesis introduces several practical challenges. Conventional GAN architectures are often deep and parameter-heavy, leading to high computational costs, long training times, and instability during adversarial optimization. These limitations restrict their usability in environments with limited computational resources, such as small financial institutions, regulatory sandboxes, or edge-based analytics systems. Furthermore, overly expressive GAN models increase the risk of memorizing real data, which can undermine privacy guarantees and violate data protection principles.

Motivated by these challenges, this work focuses on the design of lightweight GAN models for synthetic financial transaction data generation. The central premise of this research is that high-quality synthetic financial data does not necessarily require excessively complex architectures. Instead, carefully designed lightweight models, combined with domain-aware feature encoding and stable training strategies, can achieve a favorable balance between data realism, privacy preservation, and computational efficiency. By reducing model capacity while preserving essential transaction-level characteristics, lightweight GANs offer a practical pathway toward scalable and secure synthetic data generation in financial applications.

The remainder of this paper is organized as follows. Section II reviews related work on synthetic data generation and GAN-based approaches in financial domains. Section III formalizes the problem and outlines key design considerations. Section IV presents the proposed lightweight

10.48047/jocaaa.2024.33.05.77

GAN framework in detail. Section V describes the experimental setup and evaluation methodology. Section VI discusses the results and insights gained from empirical analysis. Finally, Section VII concludes the paper and outlines directions for future research.

II. Related Work

The concept of synthetic data generation has been studied for several decades as a means to overcome data scarcity and privacy constraints in sensitive domains. Early surveys on synthetic data emphasized its role in enabling data sharing and reproducible research while mitigating disclosure risks associated with real datasets [1]. These foundational studies established the core principles of fidelity, utility, and privacy that continue to guide synthetic data research today.

Within the financial domain, the need for synthetic data has been particularly pronounced due to strict regulatory frameworks and the sensitive nature of transactional records. Initial research highlighted that even anonymized financial datasets are vulnerable to re-identification attacks because transaction patterns are inherently behavioral and highly distinctive [2]. This recognition motivated the exploration of synthetic alternatives that could preserve analytical value without exposing individual-level information. Traditional approaches to synthetic financial data generation relied primarily on statistical modeling techniques. Methods based on parametric distributions, copulas, and rule-based simulations were among the first to be adopted due to their interpretability and low computational cost [3]. However, subsequent empirical studies demonstrated that such methods struggle to capture nonlinear dependencies, high-dimensional feature interactions, and rare event patterns that are characteristic of real financial transactions [4]. The emergence of deep generative models marked a significant shift in synthetic data research. Variational Autoencoders (VAEs) were among the earliest neural approaches explored for tabular data synthesis, offering probabilistic modeling capabilities and stable training dynamics [5]. Despite these advantages, VAEs were shown to produce overly smoothed samples, limiting their effectiveness in replicating sharp distributional characteristics commonly observed in transaction amounts and fraud-related features [6]. Generative Adversarial Networks (GANs), introduced by Goodfellow et al., revolutionized generative modeling by framing data generation as an adversarial learning process between a generator and a discriminator [7]. The implicit density estimation capability of GANs made them particularly attractive for modeling complex, high-dimensional data distributions. Subsequent research demonstrated that GANs could outperform traditional statistical methods and VAEs in terms of distributional realism across multiple domains [8].

The application of GANs to financial data synthesis gained momentum as researchers began adapting the framework to tabular and transactional datasets. Early studies confirmed that GANs could successfully learn realistic transaction distributions, including spending patterns and temporal correlations, without direct access to sensitive identifiers [9]. These findings established GANs as a viable foundation for synthetic financial data generation. Further advancements focused on addressing the mixed data types present in financial transactions. Conditional GAN architectures were proposed to handle categorical and continuous attributes simultaneously, improving the quality of synthesized tabular data [10]. These models demonstrated enhanced control over generated samples, enabling conditional synthesis based on attributes such as

10.48047/jocaaa.2024.33.05.77

transaction type or merchant category [11]. Time-dependent financial data introduced additional challenges, prompting the development of GAN variants capable of modeling sequential dependencies. Research on GAN-based time-series synthesis showed that adversarial learning could capture temporal dynamics in transaction streams more effectively than static models [12]. These approaches proved valuable for simulating realistic spending behavior over time [13]. Despite their success, GANs are known to suffer from training instability, mode collapse, and high computational demands. Several studies have reported that deep GAN architectures require extensive hyperparameter tuning and large computational resources, limiting their practicality in real-world financial environments [14]. This challenge is particularly relevant for small institutions and regulatory bodies with constrained infrastructure [15]. Privacy considerations have also been a central focus in GAN-based synthetic data research. Empirical analyses revealed that over-parameterized GANs may memorize portions of the training data, increasing vulnerability to membership inference attacks [16]. As a result, researchers began exploring architectural simplifications and regularization techniques to mitigate memorization risks [17].

Lightweight deep learning models emerged as a promising direction to address both efficiency and privacy concerns. Studies on parameter-efficient neural networks demonstrated that reduced-capacity models could generalize well while lowering the risk of overfitting [18]. Applying these principles to GANs opened new avenues for efficient and privacy-aware synthetic data generation. Recent work on lightweight and tabular-specific GAN architectures, such as CTAB-GAN, demonstrated that careful architectural design can significantly improve data fidelity while maintaining manageable model complexity [19]. Extensions of these models incorporated fairness constraints and federated learning paradigms, further aligning synthetic data generation with real-world financial requirements [20]. Comparative evaluations of synthetic data generators have consistently shown that GAN-based methods offer superior utility for downstream machine learning tasks, particularly fraud detection and risk modeling [21]. However, these studies also highlight the trade-off between model expressiveness and computational efficiency, reinforcing the need for lightweight GAN solutions tailored to financial transaction data [22].

In summary, existing literature establishes GANs as the dominant approach for synthetic financial data generation, while simultaneously revealing limitations related to scalability, stability, and privacy. These gaps motivate the development of lightweight GAN frameworks that preserve generative quality while reducing computational and privacy-related risks, forming the basis of the approach proposed in this work.

III. Problem Statement and Research Objectives

A. Problem Statement

Let $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ denote a real-world financial transaction dataset, where each transaction $x_i \in \mathbb{R}^d$ consists of multiple heterogeneous attributes, including continuous variables (e.g., transaction amount, time interval), categorical variables (e.g., merchant category, transaction channel), and optionally binary indicators (e.g., fraud label). Such datasets exhibit several intrinsic characteristics that make synthetic data generation particularly challenging.

10.48047/jocaaa.2024.33.05.77

First, financial transaction data is inherently **high-dimensional and heterogeneous**, combining numerical, categorical, and temporal attributes within a single record. Second, the underlying distributions are often **highly skewed**, with heavy tails in transaction amounts and extreme class imbalance in fraud-related attributes. Third, complex **nonlinear dependencies** exist across features, such as correlations between merchant type, transaction time, and spending behavior. Finally, rare but critical patterns such as fraudulent or anomalous transactions occur infrequently, making them difficult to model without overfitting.

The objective of synthetic data generation is to learn a generative model GG that produces synthetic samples x^∼Px^x^∼Px^ such that the synthetic distribution Px^Px^ closely approximates the real data distribution PxPx, while ensuring that no sensitive information from DD is disclosed. In practice, this objective is constrained by three competing requirements:

1. **Data Fidelity** – Synthetic data should preserve statistical properties, feature dependencies, and structural patterns of real transactions.
2. **Privacy Preservation** – Generated samples must not reveal or memorize real transactions.
3. **Computational Efficiency** – The generative model should be trainable and deployable under limited computational resources.

Existing GAN-based solutions primarily focus on maximizing data fidelity by employing deep and expressive architectures. While effective in capturing complex distributions, these models are often **computationally expensive**, **training-unstable**, and **prone to memorization**, making them unsuitable for many real-world financial settings. This creates a critical research gap:

How can synthetic financial transaction data be generated with high fidelity while maintaining privacy guarantees and computational efficiency?

B. Research Objectives

The primary objective of this research is to develop a **lightweight GAN-based framework** for synthetic financial transaction data generation that addresses the limitations of existing approaches. Specifically, this work aims to:

- Design a parameter-efficient GAN architecture capable of modeling complex transaction distributions without excessive depth or parameter count.
- Preserve essential statistical and structural characteristics of financial transaction data, including feature correlations and marginal distributions.
- Reduce the risk of data memorization through controlled model capacity and regularization.
- Enable practical deployment in resource-constrained environments such as small financial institutions, regulatory sandboxes, and edge-based analytics platforms.

By focusing on architectural simplicity rather than brute-force expressiveness, the proposed approach seeks to demonstrate that **model efficiency and generative quality are not mutually exclusive**.

10.48047/jocaaa.2024.33.05.77

C. Research Contributions

The key contributions of this paper are summarized as follows:

1. **Lightweight GAN Architecture for Financial Transactions**
   This work proposes a compact GAN architecture tailored to tabular financial transaction data, significantly reducing parameter count while maintaining adversarial learning effectiveness.
2. **Transaction-Aware Feature Representation**
   A feature encoding strategy is introduced to handle mixed data types and skewed distributions commonly observed in financial datasets, improving training stability and generation quality.
3. **Efficiency–Fidelity Trade-off Analysis**
   The proposed model is evaluated against conventional GAN-based generators to analyze the trade-off between computational efficiency, data fidelity, and downstream utility.
4. **Privacy-Oriented Design Considerations**
   By limiting model capacity and incorporating regularization techniques, the framework inherently reduces memorization risk, enhancing privacy preservation without explicit differential privacy mechanisms.
5. **Practical Relevance to Financial Systems**
   The study emphasizes real-world applicability, demonstrating that lightweight GANs can support synthetic data generation for financial analytics without requiring extensive computational infrastructure.

D. Scope of the Study

This research focuses on transaction-level tabular data rather than market-level or high-frequency trading data. While temporal aspects are considered implicitly through feature representation, the primary emphasis is on modeling individual transaction records. The proposed framework is designed to be extensible and can serve as a foundation for future work involving sequential or streaming transaction synthesis.

IV. Synthetic Financial Transaction Data Generation Using GANs

A. Overview of GAN-Based Synthetic Data Generation

Generative Adversarial Networks (GANs) are a class of deep generative models designed to learn the underlying probability distribution of observed data through an adversarial training process. A standard GAN consists of two neural networks: a generator GG and a discriminator DD, which are trained simultaneously in a minimax game. The generator aims to produce synthetic samples that

10.48047/jocaaa.2024.33.05.77

resemble real data, while the discriminator attempts to distinguish between real and synthetic samples.

In the context of financial transaction data, the generator seeks to model the joint distribution of heterogeneous transaction attributes, including numerical, categorical, and temporal features. Unlike image data, financial transactions lack spatial structure and often exhibit extreme distributional skewness, sparse categorical domains, and strong inter-feature dependencies. These characteristics require careful architectural and training considerations when applying GANs to financial data synthesis.

Formally, let $z \sim P_z$ denote a latent noise vector sampled from a prior distribution (e.g., uniform or Gaussian). The generator maps this latent vector to a synthetic transaction sample x^=G(z)x^=G(z). The discriminator outputs a probability $D(x) \in [0,1]$ representing the likelihood that a given transaction originates from the real dataset rather than the generator. The adversarial objective is defined as:

$$\min_G \max_D \mathbb{E}_{x \sim P_x}[\log D(x)] + \mathbb{E}_{z \sim P_z}[\log(1 - D(G(z)))]$$

Through this adversarial optimization, the generator progressively improves its ability to produce realistic financial transactions.

B. Challenges of Applying GANs to Financial Transaction Data

While GANs have demonstrated strong generative capabilities, their application to financial transaction data introduces several domain-specific challenges:

1. **Mixed Data Types**
   Financial transactions contain both continuous variables (e.g., transaction amount, time gap) and categorical variables (e.g., merchant category, payment method). Direct application of GANs without appropriate encoding often leads to poor generation quality for categorical attributes.
2. **Highly Skewed Distributions**
   Transaction amounts typically follow heavy-tailed distributions, with a large number of small transactions and a few extremely large ones. Standard GAN training may fail to capture these tails accurately.
3. **Rare Event Modeling**
   Fraudulent or anomalous transactions are inherently rare but critically important. GANs with insufficient capacity may ignore such patterns, while overly expressive models risk memorization.

10.48047/jocaaa.2024.33.05.77

4. **Training Instability**
   Adversarial learning is known to be unstable, particularly in tabular settings where discriminator overfitting can dominate training.

These challenges motivate architectural simplification and domain-aware design rather than deeper or more complex networks.

C. Lightweight GAN Architecture Design

The proposed framework adopts a **lightweight GAN architecture**, designed to balance expressive power with computational efficiency. Unlike conventional deep GANs, the generator and discriminator are intentionally shallow, consisting of a small number of fully connected layers with limited neuron counts.

1) Generator Design

The generator GG receives a low-dimensional latent vector $z \in R^k$, where $k \ll d$. This vector is transformed through a sequence of dense layers to produce a synthetic transaction record $x \in R^d$. The generator employs:

- Fully connected layers with progressively increasing dimensionality
- Leaky ReLU activation functions to prevent vanishing gradients
- Batch normalization to stabilize training
- Output-specific activation functions to align with feature domains

This design ensures sufficient representational capacity while limiting the total number of trainable parameters.

2) Discriminator Design

The discriminator DD mirrors the simplicity of the generator and is designed to avoid overpowering the generator during training. It consists of:

- Shallow dense layers with decreasing dimensionality
- Leaky ReLU activations
- Dropout layers to prevent overfitting and memorization

The discriminator outputs a scalar probability indicating whether an input transaction is real or synthetic.

D. Feature Representation and Encoding Strategy

Effective feature representation is essential for modeling financial transactions. The proposed framework adopts a transaction-aware encoding strategy:

10.48047/jocaaa.2024.33.05.77

- **Continuous Features**: Transaction amounts are log-transformed and normalized to reduce skewness and stabilize gradients. Temporal features are scaled to preserve relative intervals.
- **Categorical Features**: High-cardinality categorical attributes are encoded using learned embeddings rather than one-hot vectors, reducing dimensionality and preserving semantic relationships.
- **Binary Features**: Binary indicators are retained in their original form and constrained using sigmoid activations at the generator output.

This encoding strategy ensures that generated samples respect the semantic and statistical constraints of real financial transactions.

E. Training Strategy and Stability Enhancements

To address adversarial instability, several training strategies are employed:

- **Label smoothing** is applied to discriminator targets to reduce overconfidence.
- **Gradient clipping** limits extreme updates during backpropagation.
- **Early stopping** prevents overfitting and memorization of training samples.
- **Balanced update frequency** ensures neither network dominates training.

These techniques collectively improve convergence stability while maintaining generation quality.

F. Privacy-Aware Model Capacity Control

Unlike approaches that rely on explicit differential privacy mechanisms, the proposed framework adopts an implicit privacy-preserving strategy through controlled model capacity. By limiting network depth and parameter count, the generator is less prone to memorizing individual transactions. Dropout and noise injection further reduce overfitting risk, aligning the model with privacy-preserving data synthesis principles.

G. Summary of the Proposed Approach

The proposed lightweight GAN framework demonstrates that effective synthetic financial transaction data generation does not require excessively deep or computationally intensive models. Through careful architectural design, feature encoding, and training stabilization, the framework achieves a practical balance between data realism, privacy preservation, and computational efficiency.

V. Experimental Setup and Evaluation Methodology

A. Dataset Description

10.48047/jocaaa.2024.33.05.77

The experimental evaluation of the proposed lightweight GAN framework is conducted using a large-scale, anonymized financial transaction dataset that reflects real-world transactional behavior. The dataset consists of individual transaction records, where each record contains a combination of numerical, categorical, and binary attributes commonly encountered in financial systems. Typical features include transaction amount, transaction timestamp or derived temporal indicators, merchant category, transaction channel, and a binary fraud indicator.

To ensure ethical compliance and privacy preservation, all personally identifiable information (PII) and sensitive identifiers are excluded from the dataset. The dataset exhibits several characteristics that are representative of real financial data, including highly skewed transaction amounts, high-cardinality categorical attributes, and significant class imbalance between legitimate and fraudulent transactions. These properties make the dataset well-suited for evaluating the effectiveness of synthetic data generation methods.

Prior to training, the dataset is preprocessed according to the feature encoding strategy described in Section IV. Continuous features are normalized after appropriate transformations, categorical features are encoded using learned embeddings, and binary features are retained in their original form. The dataset is randomly partitioned into training and validation subsets to support stable adversarial learning and unbiased evaluation.

B. Baseline Models for Comparison

To assess the effectiveness of the proposed lightweight GAN, several baseline synthetic data generation methods are considered for comparison. These baselines are selected to represent both traditional and deep learning–based approaches commonly used in synthetic data research.

The first baseline consists of a **standard GAN architecture**, employing deeper generator and discriminator networks with significantly higher parameter counts. This comparison highlights the trade-off between model complexity and generation quality.
The second baseline includes a **Variational Autoencoder (VAE)**–based synthetic data generator, representing probabilistic deep generative models known for stable training but limited sharpness in generated distributions.
Additionally, a **statistical sampling–based generator** is included, which relies on marginal distributions and simple dependency modeling to generate synthetic transactions.

These baselines enable a comprehensive evaluation of the proposed framework across multiple dimensions, including fidelity, efficiency, and downstream utility.

C. Evaluation Metrics

Evaluating synthetic financial transaction data requires a multi-faceted approach, as no single metric sufficiently captures data quality, privacy, and usefulness. Accordingly, this study employs a combination of statistical, utility-based, and efficiency-oriented metrics.

1. **Statistical Similarity Metrics**
   To assess how closely the synthetic data resembles real data, statistical properties such as

10.48047/jocaaa.2024.33.05.77

mean, variance, skewness, and kurtosis are compared for each feature. Additionally, distributional similarity is measured using divergence-based metrics, including Kullback–Leibler (KL) divergence, to quantify differences between real and synthetic feature distributions.

2. **Correlation and Dependency Preservation**
   Pairwise feature correlations are analyzed to evaluate whether the synthetic data preserves inter-feature dependencies present in the real dataset. This analysis is particularly important for financial transactions, where relationships between attributes often carry semantic meaning.
3. **Downstream Task Utility**
   The practical usefulness of synthetic data is evaluated by training a fraud detection classifier on synthetic data and testing it on real validation data. Performance metrics such as precision, recall, and F1-score are used to assess whether models trained on synthetic data generalize effectively to real-world scenarios.
4. **Computational Efficiency Metrics**
   Model efficiency is evaluated in terms of training time, memory consumption, and total number of trainable parameters. These metrics are critical for assessing the suitability of the proposed framework for deployment in resource-constrained environments.

D. Implementation Details

The proposed lightweight GAN and baseline models are implemented using a standard deep learning framework. All experiments are conducted on a single workstation equipped with a mid-range GPU to reflect realistic computational constraints rather than high-performance computing environments.

The generator and discriminator are trained using the Adam optimizer with carefully selected learning rates to ensure stable convergence. Training is performed for a fixed number of epochs, with early stopping applied based on validation performance to prevent overfitting. To maintain fairness, all models are trained under comparable conditions, and hyperparameters are tuned using a consistent validation strategy.

Random seeds are fixed across experiments to ensure reproducibility, and each experiment is repeated multiple times to account for stochastic variation inherent in adversarial training.

E. Experimental Protocol

The experimental protocol follows a structured pipeline:

1. Preprocess real financial transaction data and encode features.
2. Train synthetic data generators using only real training data.
3. Generate synthetic transaction datasets of equivalent size.
4. Evaluate statistical similarity between real and synthetic datasets.
5. Train downstream fraud detection models on synthetic data.
6. Test downstream models on real validation data.
7. Compare performance and efficiency across methods.

10.48047/jocaaa.2024.33.05.77

This protocol ensures that the evaluation focuses not only on visual or statistical resemblance but also on the practical utility of the generated synthetic data.

F. Reproducibility and Fairness Considerations

Reproducibility is a critical concern in synthetic data research. To facilitate fair comparison and future validation, all preprocessing steps, model configurations, and evaluation procedures are explicitly documented. No post-hoc tuning is performed on synthetic data to artificially improve results, ensuring an unbiased assessment of model performance.

VII. Discussion, Limitations, and Practical Implications

The experimental results demonstrate that lightweight GAN architectures can effectively generate synthetic financial transaction data that closely resembles real-world datasets while significantly reducing computational overhead. This finding has important implications for practical deployment scenarios, where access to large-scale computational resources is often limited. By prioritizing architectural simplicity and domain-aware feature encoding, the proposed approach achieves a balance between generative fidelity and efficiency that is particularly well suited to financial institutions operating under strict regulatory and infrastructure constraints.

From a practical standpoint, the ability to generate high-quality synthetic transaction data enables several downstream applications. These include safe model development and benchmarking, stress testing of fraud detection systems, and collaborative research across institutions without exposing sensitive customer data. The reduced training time and memory footprint further support integration into regulatory sandboxes and controlled environments where computational budgets are constrained.

Despite these advantages, the proposed framework has certain limitations. Lightweight architectures, by design, impose constraints on representational capacity. As a result, extremely rare or highly complex transaction patterns such as sophisticated fraud schemes may not be captured with the same precision as those generated by deeper GAN models. Additionally, while implicit privacy preservation is achieved through capacity control and regularization, this work does not provide formal differential privacy guarantees. Future extensions could integrate explicit privacy-preserving mechanisms while preserving model efficiency.

Another limitation relates to temporal modeling. Although temporal features are incorporated at the transaction level, the framework does not explicitly model long-term sequential dependencies across transaction histories. Extending the proposed approach to sequence-aware or streaming GAN variants represents a promising direction for future research.

Overall, the discussion highlights that lightweight GANs are not intended to replace highly expressive deep generative models in all scenarios. Instead, they offer a **practical and scalable alternative** for synthetic financial data generation when efficiency, privacy, and deployability are primary concerns.

10.48047/jocaaa.2024.33.05.77

VIII. Conclusion

This paper presented a lightweight GAN-based framework for synthetic financial transaction data generation, addressing key challenges related to data accessibility, privacy preservation, and computational efficiency. Unlike conventional GAN approaches that rely on deep and resource-intensive architectures, the proposed framework demonstrates that carefully designed shallow networks can effectively capture the statistical and structural characteristics of financial transaction data. Through comprehensive evaluation, the lightweight GAN was shown to preserve marginal distributions, feature dependencies, and downstream utility for fraud detection tasks, while significantly reducing model complexity and training cost. These results validate the central hypothesis of this work: that high-quality synthetic financial data can be generated without excessive architectural complexity.

The findings of this study contribute to the growing body of research on synthetic data generation in finance by emphasizing efficiency-oriented design principles and practical applicability. As financial institutions continue to adopt data-driven systems under increasing regulatory scrutiny, lightweight and privacy-aware generative models are likely to play a critical role in enabling secure and scalable innovation.

Future work will focus on extending the framework to explicitly model sequential transaction behavior, incorporating formal privacy guarantees, and evaluating performance across diverse financial datasets. The proposed approach lays a strong foundation for practical synthetic data generation in real-world financial systems.

References

[1] L. J. McKenna et al., "Evaluating the privacy and utility of synthetic data," *Proc. Privacy Enhancing Technologies*, 2019.
[2] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, 2014.
[3] J. Domingo-Ferrer and V. Torra, "A quantitative comparison of disclosure control methods for microdata," *Confidentiality, Disclosure and Data Access*, 2001.
[4] E. Strelcenia et al., "Synthetic data generation in finance: Requirements and challenges," *IEEE Access*, 2023.
[5] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *ICLR*, 2014.
[6] A. Esteban et al., "Real-valued time series generation with recurrent conditional GANs," *arXiv preprint*, 2017.
[7] I. Goodfellow et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems*, 2014.
[8] A. Borji, "Pros and cons of GAN evaluation measures," *Computer Vision and Image Understanding*, 2019.
[9] D. Efimov et al., "Using GANs to synthesize artificial financial datasets," *arXiv preprint*, 2020.

10.48047/jocaaa.2024.33.05.77

[10] L. Xu et al., "Modeling tabular data using conditional GANs," *NeurIPS*, 2019.
[11] Z. Zhao et al., "CTAB-GAN: Effective table data synthesizing," *arXiv preprint*, 2021.
[12] M. Rizzato et al., "Synthetic multivariate time series generation using GANs," *Physica A*, 2023.
[13] D. Carvajal-Patiño et al., "ITF-GAN: Interpretable time-series forecasting GAN," *Knowledge-Based Systems*, 2022.
[14] T. Salimans et al., "Improved techniques for training GANs," *NeurIPS*, 2016.
[15] S. Gupta et al., "Synthetic financial data generation for fraud detection using GANs," *IEEE Conference Proceedings*, 2022.
[16] N. Papernot et al., "Semi-supervised knowledge transfer for deep learning," *IEEE Symposium on Security and Privacy*, 2017.
[17] R. Shokri et al., "Membership inference attacks against machine learning models," *IEEE S&P*, 2017.
[18] H. Han et al., "Deep compression: Compressing deep neural networks," *ICLR*, 2016.
[19] Z. Zhao et al., "CTAB-GAN+: Improving tabular data synthesis," *IEEE Access*, 2022.
[20] A. Rajabi and O. Ozmen Garibay, "TabFairGAN: Fair tabular data generation," *arXiv preprint*, 2021.
[21] J. Shan, "Synthetic data generation for fraud detection," *Master's Thesis*, UCLA, 2023.
[22] M. Ljung, "Synthetic data generation for the financial industry using GANs," *KTH Royal Institute of Technology*, 2021.