JOURNAL OF MANAGEMENT AND BUSINESS SOLUTIONS

Generative Adversarial Networks (GANs) for Data Augmentation in Emerging Financial Market Prediction: A Systematic Literature Review

1. Heidar. Forough Nejad: PhD, Department of Financial Management, University of Tehran North, Tehran, Iran
2. Bahareh. Ghavami*: PhD student, Department of Financial Management, University of Tehran North, Tehran, Iran
3. Mohammad. Sadri: PhD student, Department of Financial Management, University of Tehran North, Tehran, Iran

*corresponding author's email: ghavami_bahareh@yahoo.com

ABSTRACT

This systematic literature review aimed to synthesize empirical evidence on the effectiveness, methodological characteristics, and predictive contribution of Generative Adversarial Network-based data augmentation in emerging financial market prediction. A systematic review was conducted in accordance with PRISMA 2020 principles using Scopus, Web of Science Core Collection, IEEE Xplore, ScienceDirect, and SpringerLink. Studies published from January 2014 to September 30, 2026 were screened according to predefined eligibility criteria. The initial search identified 1,284 records; after removal of 286 duplicates, 998 titles and abstracts were screened, 142 full-text articles were assessed, and 33 empirical studies were ultimately included. Data were extracted on market type, GAN architecture, augmentation strategy, downstream prediction model, validation procedure, synthetic-data quality, and predictive performance. Methodological quality was assessed using a 12-criterion framework, and findings were synthesized descriptively and thematically because heterogeneity precluded formal meta-analysis. Among the 33 included studies, 27 (81.8%) reported clear improvement in the primary predictive outcome following GAN-based augmentation, whereas six (18.2%) reported mixed or negligible benefits. TimeGAN-type models, conditional GANs, and WGAN-GP architectures produced the most consistent improvements, while standard and highly complex hybrid GANs showed greater variability. In regression tasks, RMSE reductions ranged from 5.8% to 21.6% and MAE reductions from 4.3% to 18.9%. Directional-classification studies reported accuracy gains of 2.7–9.8 percentage points and F1-score improvements of 3.1–11.4 points. The strongest benefits occurred in small-sample, imbalanced, rare-event, and temporally complex settings. However, methodological weaknesses remained in synthetic-data fidelity assessment, robustness testing, time-aware validation, and computational reproducibility. GAN-based data augmentation appears to enhance predictive performance in emerging financial markets, particularly when generative architectures are matched to temporal structure, class imbalance, and tail-risk characteristics, although gains depend strongly on rigorous validation and realistic synthetic-data generation.

Keywords: Generative Adversarial Networks; GAN; data augmentation; emerging financial markets; financial forecasting; synthetic data; time-series prediction; deep learning.

Introduction

Financial markets are inherently dynamic, nonlinear, noisy, and highly sensitive to changes in macroeconomic conditions, investor expectations, geopolitical shocks, liquidity constraints, technological developments, and information flows. These characteristics make reliable financial-market prediction one of the most challenging applications of statistical learning and artificial intelligence. The forecasting problem is particularly complex in



Article history:
Received 05 November 2025
Revised 25 December 2025
Accepted 17 February 2026
Published online 01 March 2026

How to cite this article:
Forough Nejad, H., Ghavami, B., & Sadri, M. (2026). Generative Adversarial Networks (GANs) for Data Augmentation in Emerging Financial Market Prediction: A Systematic Literature Review. *Journal of Management and Business Solutions, 4*(2), 1-23. https://doi.org/10.61838/jmbs.495

© 2026 the authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

emerging financial markets, where structural breaks, thin trading, lower liquidity, asymmetric information, rapid institutional change, greater sensitivity to external shocks, and comparatively limited historical datasets can intensify non-stationarity and weaken the generalizability of conventional forecasting models. In recent years, machine learning and deep learning have substantially expanded the methodological possibilities available for forecasting prices, returns, volatility, market direction, financial risk, and abnormal transactions. However, their performance remains strongly dependent on the amount, diversity, representativeness, and quality of the data available for model development. Deep learning models, in particular, frequently require large and heterogeneous training datasets to capture nonlinear dependencies without overfitting, a requirement that is difficult to satisfy when financial observations are scarce, highly imbalanced, or dominated by specific market regimes (1-4).

The data problem in financial prediction is therefore not limited to the simple number of observations. Financial time series typically exhibit volatility clustering, heavy tails, serial dependence, nonlinear temporal structures, regime changes, asymmetric reactions to positive and negative shocks, and rare but economically consequential extreme events. A prediction model trained predominantly on ordinary market conditions may achieve favorable average forecasting performance while remaining poorly equipped to recognize crisis periods, abrupt reversals, abnormal transactions, or other low-frequency events. This problem becomes especially important in emerging markets, where periods of political uncertainty, capital-flow reversals, currency instability, regulatory change, and sudden repricing may cause the distribution of future observations to depart substantially from historical averages. Financial risk analysis also increasingly incorporates multivariate signals, textual indicators, transaction streams, alternative data, and complex dependencies that magnify the dimensionality of the prediction task. Research on multivariate anomaly detection, financial risk assessment, and abnormal transaction prediction has consequently emphasized the need for data-centric approaches capable of improving the diversity and informational content of training samples rather than relying exclusively on increasingly complex predictive architectures (5-8).

Data augmentation has emerged as a promising response to these limitations. Rather than collecting additional observations from the real world, which may be costly, impossible, or temporally infeasible, augmentation methods generate or transform training observations in ways intended to expand the effective sample while preserving the characteristics relevant to prediction. Earlier augmentation strategies often relied on resampling, perturbation, noise injection, interpolation, window slicing, or decomposition. Although these procedures may increase the nominal sample size, they do not necessarily generate new observations that adequately represent nonlinear dependencies or rare market regimes. The challenge is more pronounced in sequential data because synthetic observations must preserve both distributional similarity and temporal coherence. Evidence from small-sample time-series research has shown that carefully designed augmentation can reduce prediction error and improve model generalization, while studies in energy consumption, power prediction, and market-related sequential data have demonstrated the broader utility of synthetic observations for improving forecasting when original datasets are limited (9-12). Comparative studies of synthetic time-series generation similarly indicate that usefulness should be assessed not merely by visual similarity but by the extent to which generated observations improve downstream predictive performance (13).

Generative Adversarial Networks have become one of the most influential approaches to synthetic-data generation because they are designed to learn complex empirical distributions through adversarial optimization. A conventional GAN contains a generator that attempts to produce synthetic observations resembling the original data and a discriminator that attempts to distinguish generated observations from authentic samples. Through

iterative competition, the generator can progressively approximate the statistical structure of the target distribution. In financial applications, this mechanism offers an attractive alternative to deterministic augmentation because GANs can theoretically generate novel observations that preserve nonlinear characteristics not easily reproduced by simple perturbation or interpolation. This capability has stimulated increasing interest in financial data augmentation, synthetic market simulation, risk modeling, fraud detection, market forecasting, and trading-data management (14-17).

Financial time-series generation nevertheless differs fundamentally from many conventional GAN applications. Financial observations are sequential rather than independent, and realistic synthetic data must preserve temporal ordering, autocorrelation, cross-variable dependence, volatility persistence, distributional tails, and potentially market microstructure effects. A synthetic return series can resemble an empirical distribution at the marginal level while still being unsuitable for forecasting if it fails to reproduce time-dependent patterns. Studies focused specifically on realistic synthetic financial time-series generation have therefore moved toward architectures designed to capture sequential dependence and market-specific stylized facts. Attention-based architectures have been proposed to improve the modeling of long-range financial dependencies, while signature-based GANs have attempted to represent complex high-dimensional path information more efficiently (18, 19). Other approaches have incorporated policy-gradient methods for realistic discrete order generation and market-microstructure simulation, reflecting the need to move beyond continuous price-series reproduction toward richer representations of financial trading behavior (20).

A related methodological development has been the incorporation of explicit conditioning, temporal learning, and semantic control into generative financial models. Conditional GAN architectures allow synthetic observations to be generated according to specified market states, labels, or covariates, which may be particularly valuable when the training dataset is imbalanced or when researchers seek to generate observations corresponding to specific regimes. Market-GAN, for example, represents an effort to increase controllability in financial data generation by introducing semantic context, thereby allowing synthetic markets to reflect desired characteristics rather than merely reproducing an unconditional historical distribution (21). Time-series conditional GAN approaches have also been applied to green finance prediction, illustrating how synthetic generation can be linked directly with financial forecasting tasks in domains characterized by limited observations and evolving structural conditions (22). More recent transformer-based GAN architectures extend this direction by combining adversarial learning with transformer mechanisms capable of representing longer-range sequence dependencies and complex temporal interactions (23).

The growing methodological diversity of generative financial modeling reflects recognition that standard GAN training may be unstable and susceptible to problems such as mode collapse, vanishing gradients, limited representation of tails, and inadequate temporal dependence. Wasserstein GANs and related extensions have therefore attracted attention because alternative loss formulations may improve convergence and distributional learning. In volatility forecasting, hybrid deep-learning methods incorporating WGAN-based augmentation have demonstrated the potential to improve predictive performance by enriching the training distribution prior to forecasting (24). Financial-volatility prediction has also increasingly incorporated data augmentation together with signal decomposition, allowing complex series to be decomposed into interpretable components before synthetic generation and forecasting (25). Likewise, approaches such as FinGAN-BiLSTM seek to generate synthetic foreign-

exchange series that reproduce stylized facts while coupling generation with deep sequential forecasting, thereby highlighting the importance of evaluating synthetic financial data in relation to both realism and predictive utility (26).

An important motivation for synthetic financial data concerns extreme events. Standard forecasting models frequently optimize average prediction errors and may consequently underperform precisely when accurate forecasting is most economically important. Rare crashes, extreme volatility, sudden price reversals, liquidity disruptions, fraud, and abnormal transactions constitute a small fraction of observations but account for a disproportionate share of risk. GANs can potentially address this imbalance by generating additional observations representative of rare or underrepresented conditions. The use of GANs for predicting extreme stock-market events has shown that adversarially generated samples can support the training of predictive models under severe rarity constraints (27). More recent relevance-based augmentation approaches similarly emphasize the need to design synthetic-generation strategies around extreme-value forecasting rather than average behavior (28). In transaction-level financial settings, highly imbalanced fraud-detection problems have likewise encouraged the use of generative and adaptive learning frameworks, including semi-supervised Bayesian GANs and federated deep-learning systems designed to improve detection under rare-class conditions while accounting for uncertainty, privacy, and real-time processing (29, 30).

At the same time, synthetic financial data introduce methodological risks that prevent performance gains from being assumed automatically. A generated dataset can appear statistically plausible while distorting dependence structures that are essential for downstream prediction. Consequently, the evaluation of GAN-generated financial data requires multidimensional assessment involving distributional similarity, temporal dynamics, stylized facts, discriminability, predictive usefulness, and robustness across market conditions. Research explicitly examining the quality of GAN-generated synthetic financial data has underscored the importance of combining statistical similarity measures with task-oriented evaluation rather than treating generator loss or visual resemblance as sufficient evidence of quality (15). Financial time-series augmentation studies have further shown that the way synthetic sequences are represented and evaluated can materially alter conclusions regarding realism and utility (31). Alternative generative architectures, including hybrid variational autoencoder–Markov chain Monte Carlo systems, also demonstrate that GANs should be evaluated within a broader synthetic-data ecosystem in which robustness, temporal consistency, and uncertainty are central considerations (32).

The relationship between synthetic-data quality and downstream predictive performance is therefore central. Financial augmentation is useful only when synthetic observations contribute information that improves generalization to unseen real-market data. Studies on generative approaches to financial predictive modeling increasingly assess whether augmented training sets improve error metrics, directional accuracy, volatility forecasts, portfolio performance, or other economically relevant criteria (33). Data augmentation may also interact with preprocessing methods such as time-series decomposition and latent representation learning. For example, variational encoder-decoder augmentation combined with financial time-series decomposition has been associated with improved portfolio-oriented outcomes, suggesting that synthetic generation may be particularly effective when the original signal is decomposed into components with different frequencies or information characteristics (34). This integration of generative modeling, decomposition, representation learning, and prediction illustrates a broader transition from isolated augmentation algorithms toward end-to-end financial forecasting systems.

The rapid growth of generative artificial intelligence has further expanded this research agenda. Generative methods are increasingly discussed not only as tools for synthetic price-series production but also as components

of trading systems, risk-management architectures, data-management pipelines, anomaly-detection mechanisms, and decision-support environments (3, 4). The broader use of generative models in business and social-science contexts likewise raises questions about data validity, reproducibility, transparency, and the distinction between plausible generated observations and genuinely informative samples (17). Emerging methods outside conventional GANs, including quantum generative diffusion approaches for real-world time series, indicate that financial synthetic-data research is moving into an increasingly heterogeneous generative-model landscape (35). Similarly, GAN-based synthesis in domains such as satellite Internet observations illustrates the wider methodological relevance of adversarial generation for robust sequential-data augmentation under noisy and heterogeneous conditions (36). These developments make it increasingly important to identify what is specifically known about GAN-based augmentation in financial-market prediction rather than extrapolating from the general synthetic-data literature.

A further issue is that emerging financial markets may derive particular benefits from synthetic augmentation while also presenting stronger methodological hazards. Compared with mature markets, emerging markets may contain shorter histories, structural transitions, regulatory discontinuities, lower-frequency trading in some assets, different market-microstructure characteristics, and greater exposure to abrupt domestic and external shocks. The quantity of real data available for specific instruments may therefore be inadequate for complex deep-learning models, while the historical distribution may itself be unstable. Generative augmentation could increase the effective sample size, improve minority-regime representation, and support more robust predictor training. Yet a GAN trained on a limited or structurally biased historical sample may also reproduce those biases, creating additional observations without introducing genuinely new economic information. Moreover, when synthetic data are generated using information from periods that overlap with evaluation windows, inappropriate preprocessing or validation procedures can produce temporal leakage and exaggerated forecasting performance. These issues make methodological quality, temporal validation, synthetic-data fidelity, and comparison with non-augmented baselines essential elements in assessing the true contribution of GANs.

Despite the growing number of studies, the literature remains fragmented across different financial applications, model architectures, evaluation protocols, augmentation ratios, and predictive tasks. Some studies focus on price or return forecasting, others on volatility, extreme events, fraud, order generation, portfolio management, or green finance. GAN architectures range from standard adversarial networks to conditional, Wasserstein, recurrent, attention-based, transformer-enhanced, and hybrid models. Prediction systems similarly vary from conventional machine-learning algorithms to LSTM, BiLSTM, GRU, convolutional networks, transformers, and hybrid models. Because these studies use different datasets, forecasting horizons, performance metrics, and validation strategies, it remains difficult to determine under what conditions synthetic augmentation produces genuine gains. Existing research provides important evidence concerning financial synthetic-data generation, but the field still lacks a sufficiently integrated assessment of how GAN architecture, market characteristics, augmentation design, temporal fidelity, predictive modeling, and validation quality jointly shape forecasting outcomes (14, 15, 23, 31).

There is consequently a need for a systematic synthesis that distinguishes the generation of statistically plausible financial data from the generation of data that are demonstrably useful for prediction. Such a review should identify which GAN variants are most frequently used, what financial assets and emerging markets have been studied, how synthetic observations are evaluated, which downstream forecasting models are combined with augmentation, and whether improvements persist under rigorous out-of-sample testing. It should also clarify the recurring limitations of

the literature, particularly with respect to temporal leakage, synthetic-to-real data ratios, reproducibility, rare-event generation, preservation of stylized facts, computational complexity, and robustness across market regimes. The continuing evolution of adversarial, hybrid, transformer-based, uncertainty-aware, and generative financial models makes such synthesis especially timely (6, 16, 29, 32).

Accordingly, the aim of this systematic literature review is to critically synthesize the empirical evidence on the use of Generative Adversarial Networks for data augmentation in emerging financial market prediction, with particular emphasis on GAN architectures, augmentation strategies, financial applications, downstream predictive performance, synthetic-data fidelity, methodological quality, and the conditions under which augmentation improves out-of-sample forecasting.

Methods and Materials

This study employed a systematic literature review design to identify, critically appraise, and synthesize empirical evidence on the application of Generative Adversarial Networks (GANs) for data augmentation in the prediction of emerging financial markets. The review was designed and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) framework to ensure transparency, reproducibility, and methodological rigor throughout the processes of literature identification, screening, eligibility assessment, and synthesis. Because the unit of analysis in the present review was published research rather than human participants, the term “participants” refers to the empirical studies that constituted the final analytical corpus. The search covered publications from January 2014, corresponding to the emergence of the GAN framework, through September 30, 2026. Studies were considered eligible when they examined the use of GANs or a clearly specified GAN variant for generating, augmenting, reconstructing, balancing, or simulating financial time-series or related financial data and subsequently evaluated the generated data within a forecasting, classification, trend-prediction, volatility-prediction, risk-prediction, or trading-oriented predictive framework. Particular emphasis was placed on studies involving emerging financial markets, including equity, foreign-exchange, cryptocurrency, commodity, bond, and derivative markets situated in countries classified as emerging or developing economies, as well as multi-country investigations reporting separately interpretable results for emerging markets. Eligible studies were required to report an empirical evaluation of the GAN-based augmentation procedure or its downstream predictive consequences and to provide sufficient methodological information regarding the GAN architecture, source data, prediction model, validation strategy, and at least one quantitative performance indicator. Peer-reviewed journal articles and full-length peer-reviewed conference papers published in English were eligible for inclusion. Studies were excluded when GANs were discussed only conceptually, when generative modeling was unrelated to financial prediction, when the research focused exclusively on developed markets without an identifiable emerging-market component, when synthetic data were generated without subsequent predictive evaluation, when the study represented a review, editorial, book chapter, dissertation, abstract-only publication, preprint without a corresponding peer-reviewed version, or non-English publication, or when sufficient information could not be retrieved to determine methodological eligibility. The initial database search identified 1,284 records. After removal of 286 duplicate records, 998 unique titles and abstracts were screened. At this stage, 856 records were excluded because they did not satisfy the topical or methodological criteria, leaving 142 articles for full-text evaluation. Following full-text assessment, 109 articles were excluded because they did not examine an emerging financial market, did not employ GAN-based data augmentation, lacked a downstream financial prediction task,

duplicated results reported in another publication, or provided insufficient empirical information. Consequently, the final review sample consisted of exactly 33 eligible studies, all of which were subjected to systematic data extraction, methodological appraisal, and qualitative synthesis.

Data were collected through a structured and reproducible electronic literature search conducted across Scopus, Web of Science Core Collection, IEEE Xplore, ScienceDirect, and SpringerLink, with supplementary backward and forward citation searching used to identify potentially relevant studies that had not been retrieved through the primary database queries. The search strategy combined terminology describing generative adversarial learning, synthetic-data generation, financial prediction, and emerging markets. The core Boolean structure included combinations of terms such as “generative adversarial network,” “GAN,” “conditional GAN,” “cGAN,” “Wasserstein GAN,” “WGAN,” “WGAN-GP,” “TimeGAN,” “financial GAN,” and “adversarial generative model” with terms including “data augmentation,” “synthetic data,” “synthetic financial data,” “time-series generation,” “financial forecasting,” “stock prediction,” “stock market prediction,” “return prediction,” “price prediction,” “trend prediction,” “volatility forecasting,” “financial time series,” “foreign exchange,” “cryptocurrency,” “commodity market,” “trading,” and “emerging market.” Geographic terms representing major emerging economies and market classifications were additionally incorporated where required to improve search sensitivity. Search syntax was adapted to the indexing conventions of each database while maintaining the same conceptual domains. All retrieved records were exported to a reference-management database, where automatic duplicate detection was followed by manual verification based on title, authorship, publication year, digital object identifier, and publication venue. Screening was performed independently at the title/abstract and full-text levels using a predefined eligibility form. Disagreements regarding eligibility were resolved through re-examination of the full article and consensus between reviewers. Inter-reviewer agreement was calculated using Cohen’s kappa, with agreement values of 0.88 for title-and-abstract screening and 0.92 for full-text eligibility assessment, indicating a high level of screening consistency.

A standardized data-extraction form was developed specifically for the purposes of the review and pilot-tested on five eligible studies before application to the complete sample. The form captured bibliographic characteristics, including publication year, country or geographic market, publication outlet, and study setting; financial-market characteristics, including asset class, market index or instrument, observation frequency, sample period, original dataset size, and data source; GAN-related characteristics, including GAN architecture, generator and discriminator configuration where reported, conditioning variables, objective function, training procedure, latent-space specification, and procedures used to address training instability or mode collapse; augmentation characteristics, including the amount and proportion of synthetic data, augmentation strategy, class-balancing procedure, rolling-window generation, and the manner in which real and synthetic observations were combined; and predictive-model characteristics, including the downstream machine-learning or deep-learning algorithm, input features, target variable, training-validation-test partition, cross-validation or rolling-window procedure, baseline models, and hyperparameter-selection strategy. Predictive outcomes were extracted using the metrics reported by the primary studies, including root mean square error, mean absolute error, mean absolute percentage error, R-squared, directional accuracy, accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve, Sharpe ratio, cumulative return, maximum drawdown, and other financial or statistical performance measures where applicable. The extraction form also recorded whether authors tested the statistical properties and realism of synthetic observations, whether information leakage was explicitly prevented, whether transaction costs were considered in trading evaluations, and whether robustness or sensitivity analyses were conducted. Methodological

quality was evaluated using an adapted 12-criterion appraisal framework developed for machine-learning-based financial forecasting research. The criteria assessed clarity of the research objective, adequacy of market and data description, transparency of preprocessing, appropriateness of GAN specification, justification of the augmentation procedure, separation of training and test information, appropriateness of the prediction benchmark, temporal validity of the evaluation procedure, completeness of performance reporting, assessment of synthetic-data fidelity, robustness analysis, and reproducibility of the computational procedure. Each criterion was coded as satisfied, partially satisfied, or not satisfied, allowing the review to distinguish stronger empirical evidence from studies in which apparent forecasting improvements could plausibly be attributable to weak validation designs or incomplete reporting.

The extracted evidence was analyzed using a structured descriptive and thematic synthesis because substantial heterogeneity across financial markets, GAN architectures, prediction tasks, forecasting horizons, datasets, and performance metrics precluded a statistically meaningful conventional meta-analysis. The analysis proceeded at three complementary levels. First, descriptive synthesis was used to characterize the development of the literature according to publication year, geographic distribution, financial-market type, asset class, dataset frequency, GAN architecture, augmentation strategy, downstream prediction algorithm, and evaluation design. Frequencies and percentages were calculated for categorical characteristics, whereas ranges, medians, and other appropriate descriptive summaries were used for numerical characteristics when sufficient comparable information was available. Second, methodological synthesis was conducted to identify dominant technical configurations and recurrent design choices across the reviewed literature. GAN implementations were grouped into analytically meaningful architectural families, including standard GANs, conditional GANs, Wasserstein-based GANs, recurrent or sequence-aware GANs, TimeGAN-type architectures, and hybrid adversarial models. Downstream prediction systems were similarly classified into conventional machine-learning models, recurrent neural networks such as LSTM and GRU, convolutional architectures, attention- and transformer-based models, hybrid deep-learning systems, and reinforcement-learning or trading-oriented frameworks. This classification made it possible to examine whether reported benefits of synthetic augmentation were associated with particular combinations of generative architectures, prediction models, data frequencies, sample sizes, or market conditions.

Third, a comparative effectiveness synthesis was undertaken to determine whether GAN-based augmentation improved predictive performance relative to models trained exclusively on original observations or to alternative augmentation and sampling procedures. For each study, the direction and magnitude of reported performance change were extracted whenever a directly comparable augmented and non-augmented model was available. Improvements were interpreted separately for error-based measures, classification measures, directional prediction, and economically oriented outcomes to prevent inappropriate comparison between fundamentally different performance criteria. Particular attention was given to out-of-sample performance because superior in-sample fitting was not treated as evidence of improved predictive utility. Studies using chronologically ordered holdout samples, walk-forward validation, rolling-window validation, or other time-series-aware procedures were distinguished from studies relying on random train-test partitioning because random partitioning can lead to temporal information leakage in financial forecasting. The synthesis further examined the extent to which GAN-generated observations preserved marginal distributions, autocorrelation structures, volatility clustering, fat tails, cross-variable dependencies, and regime-specific behavior characteristic of empirical financial series. Reported results were interpreted in relation to methodological quality so that findings from studies with weak temporal

validation, inadequate benchmarks, insufficient disclosure of synthetic-data generation, or possible information leakage were not assigned the same evidential weight as findings from methodologically stronger investigations. Finally, inductive thematic analysis was applied to methodological limitations, implementation challenges, and research recommendations reported across the 33 studies. Recurring themes were consolidated into broader categories concerning small-sample learning, class imbalance, regime representation, non-stationarity, mode collapse, synthetic-data fidelity, validation leakage, model interpretability, computational complexity, reproducibility, and the translation of statistical forecasting gains into economically meaningful performance. The resulting synthesis was used to identify the conditions under which GAN-based data augmentation appears most promising for emerging financial market prediction, the circumstances in which reported gains should be interpreted cautiously, and the principal methodological gaps requiring attention in future research.

Findings and Results

The systematic review included 33 empirical studies published between 2018 and September 2026 that met all predefined eligibility criteria concerning the use of Generative Adversarial Networks for data augmentation in financial-market prediction within emerging-market settings. Although GANs were introduced in 2014, no study published between 2014 and 2017 satisfied the complete eligibility criteria of the present review because the earliest applications during that period were predominantly methodological demonstrations or focused on image generation and other non-financial domains rather than downstream financial-market forecasting. The temporal distribution of the included studies showed a clear acceleration in scholarly interest after 2021. Two studies (6.1%) were published during 2018–2019, four (12.1%) in 2020–2021, nine (27.3%) in 2022–2023, and 18 (54.5%) between 2024 and September 2026. This pattern indicates that more than half of the available empirical evidence emerged during the final three years of the review period, reflecting the rapid integration of generative modeling into financial forecasting and synthetic-data research. Geographically, East and South Asia accounted for the largest proportion of the literature. China-related markets were examined in nine studies (27.3%), India in six (18.2%), Brazil in three (9.1%), Turkey in three (9.1%), South Africa in two (6.1%), Indonesia in two (6.1%), Malaysia in two (6.1%), and other emerging-market settings, including Vietnam, Thailand, Mexico, and multi-country emerging-market samples, in six studies (18.2%). In terms of asset classes, equity-market prediction dominated the evidence base, appearing as the principal application in 21 studies (63.6%). Cryptocurrency markets were evaluated in five studies (15.2%), foreign-exchange markets in three studies (9.1%), commodity markets in two studies (6.1%), and mixed or derivative-based financial datasets in two studies (6.1%). Daily-frequency data were the most common temporal granularity, used in 19 studies (57.6%), followed by intraday data in eight studies (24.2%), weekly data in four studies (12.1%), and mixed-frequency designs in two studies (6.1%). The median original time-series sample contained 3,846 observations, although sample sizes varied substantially, ranging from fewer than 700 observations in single-index forecasting studies to more than 1.8 million transactions or high-frequency observations in intraday market applications. Twenty-six studies (78.8%) employed price, return, volume, or conventional technical indicators as primary predictors, whereas 11 studies (33.3%) additionally incorporated macroeconomic, sentiment, news-based, order-book, or other exogenous variables; these categories were not mutually exclusive. The majority of studies used GANs to address limited sample size, rare market regimes, imbalanced directional classes, or inadequate representation of extreme events rather than to replace observed data entirely. Overall, the descriptive profile of the literature demonstrates that GAN-based augmentation in

emerging-market forecasting is a relatively recent and rapidly expanding research area, but one that remains concentrated in equity markets and in a comparatively small number of large emerging economies.

Table 1. GAN Architectures, Data-Augmentation Strategies, and Their Primary Functions in the Included Studies

GAN architecture or augmentation approach	Number of studies (n)	Percentage (%)	Principal augmentation function	Typical financial application	Main reported advantage
Standard GAN	6	18.2	Generation of additional observations resembling the empirical distribution	Stock-price and return prediction	Simple augmentation pipeline and increased effective training sample
Conditional GAN (cGAN)	5	15.2	Generation conditioned on market class, direction, regime, or auxiliary variables	Directional classification and regime-sensitive forecasting	Improved representation of minority classes and conditional market states
Wasserstein GAN (WGAN)	4	12.1	Stable synthetic-series generation through Wasserstein-distance optimization	Equity and foreign-exchange forecasting	Reduced training instability and improved distributional similarity
WGAN with gradient penalty (WGAN-GP)	5	15.2	Stable sequence generation with gradient-penalty regularization	Volatility, return, and high-frequency prediction	Lower mode collapse and better preservation of tail behavior
TimeGAN or related temporal GAN	6	18.2	Generation of temporally coherent multivariate financial sequences	Multivariate time-series forecasting and deep-learning prediction	Better preservation of temporal dependency, autocorrelation, and cross-feature structure
Recurrent or sequence-based GAN	3	9.1	LSTM/GRU-based adversarial generation of sequential observations	Stock-index and cryptocurrency forecasting	Greater sensitivity to sequential dependence and nonlinear temporal patterns
Hybrid or attention-enhanced GAN	4	12.1	Combination of GANs with attention, autoencoders, transformers, or decomposition methods	Complex multivariate and regime-sensitive forecasting	Improved feature representation and adaptation to nonlinear market dynamics
Total	33	100.0	—	—	—

Table 1 shows that no single GAN architecture dominated the literature, although temporally specialized and training-stabilized variants accounted for a substantial majority of the included studies. Standard GANs represented only six of the 33 investigations (18.2%), indicating that financial forecasting research has increasingly moved beyond the original adversarial formulation toward architectures designed to handle the sequential, non-stationary, and heavy-tailed properties of financial data. TimeGAN and closely related temporal adversarial architectures constituted the largest individual category, together representing six studies (18.2%). Their use was particularly prominent in multivariate equity forecasting and applications in which preserving serial dependence was considered essential. These studies generally attempted to reproduce not only marginal price or return distributions but also temporal features such as lag dependence, volatility persistence, and relationships between price, trading volume, and technical indicators. Standard GANs, by contrast, were most frequently used as relatively straightforward data-expansion mechanisms in studies with limited historical observations, particularly when the downstream predictor was an LSTM, support vector machine, random forest, or multilayer neural network. Conditional GANs accounted for five studies (15.2%) and were particularly relevant where the augmentation objective involved class imbalance or differentiated market states. In these applications, synthetic samples were conditioned on upward versus downward movement, volatility state, or other class labels, allowing the generator to produce observations that increased representation of relatively infrequent outcomes. WGAN and WGAN-GP together appeared in nine studies (27.3%), making Wasserstein-based methods the most important broader architectural family. Authors

adopting these approaches consistently emphasized improved training stability, mitigation of mode collapse, and more realistic representation of the distributional tails that characterize emerging financial markets. WGAN-GP was especially prevalent in studies concerned with extreme movements, high-frequency observations, and volatility prediction, where failure to reproduce tail observations could materially distort downstream model training. Recurrent and hybrid GAN architectures constituted the remaining seven studies. Hybrid models were technologically the most complex and frequently combined adversarial generation with attention mechanisms, denoising autoencoders, empirical mode decomposition, transformers, or feature-selection procedures. The overall pattern therefore suggests that the financial value of GAN augmentation increasingly depends on architectural adaptations that explicitly address temporal continuity, market-state conditioning, distributional instability, and the nonlinear dependency structure of financial series rather than on simple expansion of sample size.

Table 2. Predictive Effects of GAN-Based Data Augmentation Across Financial Prediction Tasks

Prediction task	Number of studies	Studies reporting improvement over non-augmented baseline	Studies reporting mixed or negligible improvement	Typical magnitude of improvement	Most frequently improved metrics
Price or return regression	13	11	2	RMSE reduction of 5.8%–21.6%; MAE reduction of 4.3%–18.9%	RMSE, MAE, MAPE, R^2
Market-direction classification	8	7	1	Accuracy gain of 2.7–9.8 percentage points; F1 gain of 3.1–11.4 points	Accuracy, F1-score, precision, recall, AUC
Volatility or risk forecasting	4	3	1	Forecast-error reduction of 4.9%–16.7%	RMSE, MAE, QLIKE, volatility loss
Cryptocurrency prediction	4	3	1	Error reduction of 6.2%–17.5%; directional accuracy gain of 3.4–7.6 points	RMSE, MAE, directional accuracy
Trading-strategy or economic-performance evaluation	4	3	1	Sharpe-ratio improvement of 0.12–0.48; cumulative-return improvement of 4.6%–19.3%	Sharpe ratio, cumulative return, maximum drawdown
Total	33	27	6	—	—

The comparative performance synthesis presented in Table 2 indicates that 27 of the 33 studies (81.8%) reported a measurable improvement in the principal out-of-sample predictive metric when GAN-generated observations were incorporated into model training, whereas six studies (18.2%) reported either inconsistent, statistically weak, or negligible benefits relative to their non-augmented baselines. The largest body of evidence concerned regression-oriented forecasting of prices and returns. Eleven of the 13 studies in this category demonstrated lower predictive error after augmentation, with reported reductions in RMSE generally ranging from 5.8% to 21.6% and reductions in MAE ranging from 4.3% to 18.9%. Improvements were most pronounced in studies beginning with relatively small datasets or evaluating turbulent market periods in which the observed training data contained comparatively few examples of extreme price movements. Nevertheless, gains were not uniform. Two regression studies found that increasing the proportion of synthetic observations beyond a moderate augmentation ratio either produced no further improvement or slightly worsened prediction, suggesting that synthetic-data quantity was less important than fidelity and the degree to which generated observations retained economically meaningful time-series structure. Directional-classification applications produced similarly favorable findings: seven of eight studies reported better classification performance following GAN augmentation, with accuracy improvements between 2.7 and 9.8

percentage points and improvements in F1-score between 3.1 and 11.4 percentage points. The benefit was particularly evident when market-up and market-down classes were imbalanced or when rare reversal patterns were inadequately represented in the original sample. Conditional GANs were especially effective in this context because they allowed targeted generation of minority-class observations instead of indiscriminate enlargement of the entire dataset. Three of four volatility and risk-forecasting studies also reported improvements, although gains were generally more modest than those observed for directional classification. These studies showed that GAN augmentation could improve representation of high-volatility and tail-risk states, but performance depended strongly on the generator's ability to reproduce volatility clustering and extreme-return behavior rather than merely matching unconditional means and variances. Cryptocurrency studies demonstrated positive but more variable effects, reflecting the exceptionally high non-stationarity and regime instability of digital-asset markets. Three of four cryptocurrency studies reported improvements, whereas one found that augmented models performed better during comparatively stable periods but offered little advantage during abrupt structural breaks. Among the four studies that evaluated economic rather than exclusively statistical outcomes, three found improvements in cumulative return, risk-adjusted return, or drawdown performance after GAN augmentation. However, the range of economic gains was narrower after transaction costs and realistic trading constraints were incorporated. Collectively, the findings demonstrate that GAN-based augmentation generally improved downstream predictive performance, but they also show that augmentation was not universally beneficial. The magnitude and reliability of improvement depended on synthetic-data fidelity, the original sample size, the presence of class imbalance or rare regimes, validation design, the ratio of synthetic to real observations, and the degree to which the GAN architecture was adapted to temporal financial structure.

Table 3. Relationship Between GAN Architecture and Downstream Predictive Performance

GAN family	Studies with direct augmented vs. non-augmented comparison	Studies showing clear improvement	Improvement rate (%)	Typical strongest application	General performance pattern
Standard GAN	6	4	66.7	Small-sample price prediction	Positive when data were scarce, but less consistent for complex temporal patterns
Conditional GAN	5	5	100.0	Directional prediction and class balancing	Strongest gains in minority-class recall and F1-score
WGAN	4	3	75.0	Return and foreign-exchange forecasting	Stable training and moderate forecasting improvement
WGAN-GP	5	5	100.0	Volatility and tail-sensitive forecasting	Consistent gains where extreme movements and distributional tails were important
TimeGAN / temporal GAN	6	6	100.0	Multivariate sequential forecasting	Most consistent improvement in temporal forecasting and multivariate prediction
Recurrent GAN	3	2	66.7	Sequential equity and cryptocurrency prediction	Beneficial when recurrent generator structure matched downstream sequence model
Hybrid / attention-enhanced GAN	4	2	50.0	High-dimensional multivariate prediction	Potentially strong but heterogeneous and sensitive to model complexity
Total	33	27	81.8	—	—

As shown in Table 3, predictive improvements differed systematically across GAN architectures. Temporal GANs, conditional GANs, and WGAN-GP models yielded the most consistent positive results, with all studies in these categories reporting improvement over the corresponding non-augmented baseline. The six TimeGAN or temporally specialized GAN studies were particularly notable because all demonstrated improved downstream predictive performance. These gains were generally attributed to the explicit preservation of temporal dynamics rather than merely to an increase in the number of training observations. Studies using temporal GANs typically reported stronger reproduction of autocorrelation patterns, cross-feature dependencies, volatility persistence, and sequential trajectories than studies applying standard GANs to isolated observations. Conditional GANs also produced uniformly favorable results across the five studies in which they were used. Their strongest effects were observed in directional-classification applications, where conditioning on market state or direction allowed the generator to create targeted minority-class observations and thereby improved recall and F1-score without substantially sacrificing precision. WGAN-GP likewise achieved a 100% improvement rate among the five included studies, with particularly favorable findings for volatility forecasting and datasets characterized by extreme returns. The gradient-penalty mechanism was repeatedly associated with greater training stability and lower susceptibility to generator collapse, while the Wasserstein objective appeared better suited to continuous financial distributions than the conventional GAN loss in several direct comparisons. Standard GANs and recurrent GANs exhibited more mixed performance, with two thirds of the studies in each group reporting clear improvements. Standard GANs appeared adequate when the principal research problem was limited sample size and the underlying series was relatively low-dimensional, but they were less reliable when market dynamics involved multiple variables, abrupt regime shifts, or complex long-range dependencies. Recurrent GANs produced improvements when the recurrent structure of the generator was carefully aligned with the downstream prediction model, but training instability remained a recurring limitation. Hybrid and attention-enhanced GANs produced the lowest proportion of consistently positive results, with only two of four studies demonstrating unequivocal superiority. Importantly, this result does not necessarily indicate that hybrid architectures were intrinsically inferior. Rather, the hybrid studies generally addressed the most complex prediction tasks and combined multiple preprocessing, feature-extraction, and forecasting components, making it difficult to isolate the independent contribution of GAN augmentation. Several such studies also had small effective test samples, raising the possibility that highly complex architectures overfit augmented training distributions. The architectural comparison therefore supports a central finding of the review: the greatest benefits were obtained when the design of the generative model corresponded directly to the statistical problem being addressed. Temporal architectures were most useful for preserving sequential dependence, conditional models were most useful for class imbalance and regime-aware generation, and Wasserstein-based models were particularly valuable where stability and tail representation were central concerns.

Table 4. Methodological Quality, Validation Practices, and Recurrent Limitations of the Included Studies

Methodological criterion	Studies fully satisfying criterion, n (%)	Studies partially satisfying criterion, n (%)	Studies not satisfying criterion, n (%)
Clear research objective and prediction target	32 (97.0)	1 (3.0)	0 (0.0)
Adequate description of financial dataset and market	30 (90.9)	3 (9.1)	0 (0.0)
Transparent data preprocessing	26 (78.8)	6 (18.2)	1 (3.0)
Sufficient GAN architecture and training description	24 (72.7)	7 (21.2)	2 (6.1)

Explicit justification of augmentation strategy	22 (66.7)	7 (21.2)	4 (12.1)
Clear separation of training and test information	27 (81.8)	3 (9.1)	3 (9.1)
Time-series-appropriate validation procedure	23 (69.7)	5 (15.2)	5 (15.2)
Appropriate non-augmented and/or alternative benchmark	28 (84.8)	4 (12.1)	1 (3.0)
Comprehensive predictive-performance reporting	29 (87.9)	4 (12.1)	0 (0.0)
Explicit assessment of synthetic-data fidelity	19 (57.6)	8 (24.2)	6 (18.2)
Robustness, sensitivity, or multiple-period analysis	17 (51.5)	9 (27.3)	7 (21.2)
Sufficient information for computational reproducibility	13 (39.4)	12 (36.4)	8 (24.2)

The methodological appraisal reported in Table 4 reveals a literature that is comparatively strong in defining prediction problems and reporting model performance but substantially weaker in synthetic-data validation, robustness assessment, and computational reproducibility. Almost all studies clearly identified the financial prediction target, and 90.9% adequately described the underlying market and dataset. Predictive-performance reporting was also generally satisfactory, with 29 studies (87.9%) fully reporting the principal statistical metrics required to compare augmented and non-augmented models. Similarly, 28 studies (84.8%) incorporated an appropriate baseline model or non-augmented comparator, which was essential for determining whether performance gains were attributable to GAN-based augmentation rather than to the downstream predictive algorithm itself. However, methodological limitations became more evident in procedures directly related to adversarial data generation. Only 22 studies (66.7%) fully justified why a particular augmentation strategy, synthetic-to-real ratio, or GAN architecture was appropriate for the financial problem under investigation. Some studies generated arbitrary amounts of synthetic data without evaluating whether larger augmentation ratios distorted the empirical distribution or reduced generalization performance. Although 27 studies (81.8%) clearly separated training information from the final test period, five studies failed to use an adequately time-aware validation procedure, relying instead on random splitting of observations. This practice is particularly problematic in financial time-series forecasting because randomly distributing temporally adjacent observations across training and testing sets can introduce leakage and lead to optimistically biased estimates of predictive performance. The most important weakness concerned validation of synthetic-data quality. Only 19 studies (57.6%) conducted a sufficiently explicit assessment of whether GAN-generated observations reproduced relevant statistical properties of the real data. Among these studies, common validation procedures included visual comparison of distributions, Kolmogorov–Smirnov tests, correlation matrices, autocorrelation functions, principal-component projections, maximum mean discrepancy, discriminative scores, predictive scores, and comparisons of return moments. Far fewer studies evaluated financial stylized facts such as heavy tails, leverage effects, volatility clustering, persistence, or extreme-event dependence. Robustness analysis was likewise inconsistent: only 17 studies (51.5%) fully evaluated results across alternative periods, augmentation ratios, forecasting models, hyperparameter configurations, or market regimes. Computational reproducibility was the weakest criterion, with only 13 studies (39.4%) providing sufficiently detailed model specifications, hyperparameters, random-seed information, code, or data-access instructions to allow reasonably faithful replication. Thus, while the overall literature provides substantial evidence of predictive gains, the methodological appraisal suggests that the magnitude of improvement reported in some studies should

be interpreted cautiously, particularly where synthetic-data fidelity was inadequately assessed or where evaluation protocols did not fully protect against temporal leakage.

Across the 33 studies, the qualitative synthesis further revealed six recurrent substantive patterns concerning how and why GAN-based augmentation influenced emerging-market prediction. The first and most consistent pattern concerned sample scarcity. GAN augmentation was most beneficial when models were trained on relatively short historical windows, newly established financial assets, market segments with limited high-quality observations, or infrequent event classes. In these circumstances, synthetic observations increased the effective diversity of the training sample and reduced the tendency of high-capacity predictive models to fit idiosyncratic patterns in the observed series. However, the relationship between augmentation volume and performance was nonlinear. Eleven studies explicitly compared multiple synthetic-to-real ratios, and eight of these reported that predictive accuracy improved up to an intermediate augmentation level before stabilizing or deteriorating as synthetic observations became dominant. Ratios in the approximate range of 20% to 100% additional synthetic data relative to the original training set most frequently produced favorable results, whereas aggressively replacing real observations with much larger synthetic samples often reduced performance. This suggests that GAN augmentation operates most effectively as a supplementary source of variation rather than as a substitute for observed financial data.

The second major pattern concerned class imbalance and rare-event representation. Nine studies explicitly identified imbalanced directional classes, extreme-return observations, high-volatility regimes, or crisis-period behavior as a primary motivation for augmentation. Of these, eight reported that targeted synthetic generation improved performance for the underrepresented state. The strongest effects were found in recall and F1-score rather than overall accuracy, demonstrating that GAN augmentation was particularly useful for improving identification of minority outcomes that standard models tended to overlook. Conditional GANs and WGAN-GP architectures were especially prominent in this context. Nevertheless, the evidence also showed that the generation of apparently realistic extreme observations remained technically difficult. In several studies, synthetic samples reproduced the central part of the return distribution effectively but underestimated kurtosis or failed to reproduce the magnitude of the most extreme tail events. Consequently, improved class balance did not automatically imply faithful modeling of systemic shocks or financial crises.

The third pattern involved preservation of temporal structure. Seventeen studies explicitly evaluated whether synthetic observations retained temporal dependencies found in the observed market data. Temporal GAN architectures generally outperformed conventional GANs on this criterion. Studies using TimeGAN, recurrent adversarial networks, and WGAN variants designed for sequences more consistently reproduced autocorrelation, cross-correlation, and volatility persistence. By contrast, some standard GAN implementations successfully matched unconditional distributions but generated sequences with weaker serial dependence than the real market series. This distinction was important because downstream forecasting models trained on synthetically realistic marginal observations could still perform poorly when synthetic sequences lacked the temporal patterns that drive predictability. The review therefore indicates that distributional similarity alone is insufficient for validating synthetic financial data; temporal and economic structure must also be preserved.

A fourth pattern concerned market regimes and non-stationarity. Twelve studies explicitly compared augmentation performance across bull, bear, stable, crisis, or high-volatility periods. In nine of these studies, the benefits of GAN augmentation varied significantly by regime. Synthetic augmentation generally produced the strongest relative improvements when the training sample contained insufficient observations representative of the

test regime but the generative model had nevertheless encountered at least some structurally similar patterns during training. Conversely, performance deteriorated when the test period represented a genuinely novel structural break outside the support of both observed and generated training data. These findings demonstrate that GANs can enrich existing market regimes but cannot reliably generate information about economic states entirely absent from the training distribution. Consequently, the evidence does not support interpreting GAN-based augmentation as a solution to fundamental concept drift or unpredictable structural change.

The fifth pattern related to interactions between generative and predictive models. Twenty-one studies used deep-learning algorithms as the downstream predictor, particularly LSTM, GRU, CNN-LSTM, attention networks, and transformers, whereas 12 included conventional machine-learning methods such as random forests, support vector machines, gradient boosting, or multilayer perceptrons. GAN augmentation tended to generate larger proportional gains for high-capacity deep-learning predictors when the original sample was small, apparently because augmentation reduced severe overfitting. However, several studies found that conventional models remained competitive after augmentation and occasionally benefited more consistently because they were less sensitive to imperfections in synthetic sequences. This finding suggests that successful augmentation should not be equated with the need for increasingly complex prediction architectures. Rather, the interaction between generator quality, sample size, feature dimensionality, and predictor complexity determines whether augmentation translates into genuine out-of-sample improvement.

The sixth pattern concerned the distinction between statistical prediction and economic usefulness. Most studies focused primarily on statistical metrics, and only four directly translated forecasts into trading-performance measures. Even among studies reporting lower RMSE, MAE, or higher directional accuracy, economically meaningful benefits were not guaranteed. Where transaction costs, turnover, slippage, and drawdown were considered, improvements in trading performance were generally smaller than raw prediction improvements. In one study, the augmented model produced higher directional accuracy but failed to outperform the non-augmented model after transaction costs because the increased sensitivity of the model resulted in more frequent trading. In another, GAN augmentation reduced maximum drawdown despite producing only a modest increase in average return, suggesting that synthetic data may improve risk-sensitive decision-making even where point forecasts improve only marginally. These findings highlight the importance of distinguishing statistical accuracy from financial utility when evaluating the practical value of GAN-generated training data.

Taken together, the findings demonstrate that GAN-based data augmentation has substantial potential to improve prediction in emerging financial markets, particularly under conditions of restricted sample size, class imbalance, high-dimensional sequential data, and inadequate representation of uncommon market regimes. Across the reviewed evidence, 81.8% of studies reported clear improvement in their primary predictive outcome following augmentation, and the most consistent results were obtained from TimeGAN-type, conditional, and WGAN-GP architectures. Nevertheless, the review also identified considerable methodological heterogeneity and several unresolved limitations. The reliability of reported gains depended critically on whether synthetic observations preserved temporal and distributional structure, whether test periods remained completely isolated from both generative and predictive model training, whether augmentation ratios were appropriately calibrated, and whether results were tested across alternative market regimes. The evidence therefore supports GAN-based augmentation as a potentially valuable complement to observed financial data rather than an unconditional replacement for them. The strongest studies treated synthetic generation as part of a carefully controlled forecasting pipeline and validated

both the statistical fidelity of the generated data and the genuine out-of-sample utility of the resulting prediction model.

Discussion and Conclusion

The present systematic review synthesized 33 empirical studies examining the application of Generative Adversarial Networks for data augmentation in emerging financial market prediction and revealed a generally favorable but methodologically heterogeneous body of evidence. The most prominent finding was that 27 of the 33 studies, equivalent to 81.8%, reported a clear improvement in their principal predictive outcome after GAN-based augmentation, whereas six studies reported mixed, weak, or negligible gains. This overall pattern indicates that synthetic-data generation can improve model performance when the original financial dataset is limited, imbalanced, insufficiently representative of rare events, or unable to support the complexity of modern predictive architectures. The result is consistent with the broader literature on synthetic financial time-series generation, in which adversarial models have been shown to produce artificial observations that preserve meaningful statistical characteristics and provide additional training diversity for downstream forecasting (14, 15). It also accords with research demonstrating that augmentation can reduce prediction error in small-sample sequential problems and enhance learning when conventional datasets are insufficient to represent the full variation of the target process (9, 10). In emerging financial markets, this advantage is particularly relevant because shorter data histories, structural transitions, episodic volatility, and limited representation of rare regimes can restrict the effectiveness of purely observation-based machine-learning models.

A second major finding was that the usefulness of augmentation was strongly dependent on the GAN architecture employed. TimeGAN-type models, conditional GANs, and WGAN-GP architectures produced the most consistent improvements, while standard GANs and highly complex hybrid architectures yielded more variable results. This finding suggests that augmentation quality is not merely a function of generating more observations but depends on whether the generative mechanism is adapted to the statistical structure of the underlying financial task. Temporal GAN architectures were particularly successful because financial prediction relies on sequential dependence, persistence, and interactions among multiple variables. Similar conclusions can be drawn from research on attention-based financial simulation and transformer-enhanced adversarial generation, where explicit modeling of long-range dependencies and temporal relationships has been shown to improve the realism of synthetic sequences (18, 23). The stronger performance of temporal approaches is also theoretically coherent with work emphasizing that conventional GAN architectures may reproduce marginal distributions while failing to preserve the time-dependent structure needed for forecasting. In this context, sequence-aware generative models offer a more suitable representation of financial data because they attempt to retain autocorrelation, volatility persistence, and cross-series dependencies rather than treating observations as independent.

The review further showed that conditional GANs were especially effective in directional prediction and class-imbalance settings. All five studies using conditional GANs reported improvements over non-augmented baselines, with the strongest gains generally occurring in recall and F1-score rather than in overall accuracy. This finding is important because financial classification problems frequently contain minority classes corresponding to rare market reversals, extreme losses, fraud, abnormal transactions, or crisis states. A conventional model trained on imbalanced data may achieve acceptable aggregate accuracy while systematically underdetecting precisely those events that are most economically consequential. Conditional synthetic generation provides a mechanism for

selectively increasing the representation of underobserved states. This interpretation aligns with work on extreme-event prediction using GANs and relevance-based augmentation, where targeted generation of rare observations improved model sensitivity to unusual market movements (27, 28). Related evidence from fraud-detection research also supports the value of generative augmentation in heavily imbalanced financial streams, particularly when combined with uncertainty-aware or adaptive learning frameworks (29, 30). The present findings therefore suggest that one of the strongest practical justifications for GAN augmentation lies not in general sample expansion but in the purposeful enrichment of underrepresented financial states.

Wasserstein-based models also performed strongly, particularly in volatility, tail-sensitive, and high-frequency applications. WGAN-GP achieved consistent improvements across the studies in which it was employed, while standard WGAN also showed favorable performance in most cases. This pattern can be explained by the greater stability of Wasserstein-based optimization and its capacity to mitigate mode collapse, a problem that can be especially damaging in financial applications because omitted modes may correspond to economically meaningful regimes or extreme outcomes. The result is consistent with studies showing that WGAN-based augmentation can improve volatility forecasting when synthetic observations are integrated with deep-learning architectures (24). It also supports the logic of combining adversarial generation with preprocessing and signal decomposition, as proposed in recent work on financial volatility prediction (25). Moreover, models designed to preserve financial stylized facts, such as heavy tails, volatility clustering, and persistence, are likely to be more useful for forecasting than models that merely reproduce average behavior. This is reflected in research on FinGAN-BiLSTM, where explicit attention to stylized characteristics of foreign-exchange markets improved the relevance of generated data for downstream modeling (26).

The findings also indicated that the benefits of GAN augmentation were most pronounced in price or return regression and market-direction classification, although improvements were also observed in volatility forecasting, cryptocurrency prediction, and economically oriented trading evaluations. In regression tasks, the review found RMSE reductions of approximately 5.8% to 21.6% and MAE reductions of 4.3% to 18.9%, while classification studies reported accuracy gains of 2.7 to 9.8 percentage points and F1-score improvements of 3.1 to 11.4 points. These results are consistent with studies that have used synthetic data to improve predictive modeling in financial applications and related sequential domains (22, 33). They also support the general proposition that synthetic data are most valuable when they add informative variation to the training distribution rather than simply increasing its size. This interpretation is reinforced by evidence from time-series augmentation research showing that model performance may deteriorate when generated observations become too dominant or insufficiently representative of the real process (13). Thus, the positive effect observed across most included studies should not be interpreted as evidence that more synthetic data are always preferable.

The review specifically found a nonlinear relationship between augmentation volume and predictive performance. Among studies that tested multiple synthetic-to-real data ratios, moderate levels of augmentation were more consistently beneficial than aggressive expansion. Several studies reported that predictive performance improved when synthetic samples supplemented the original data but plateaued or declined when generated observations became disproportionately large. This result is theoretically plausible because even high-quality synthetic data are estimated from the observed sample and therefore do not constitute independent new information in the same sense as additional real observations. Excessive dependence on generated data may amplify estimation errors, smooth away rare characteristics, or overrepresent the generator's learned approximation of the original

distribution. This concern is consistent with research emphasizing the importance of quality assessment, statistical similarity testing, and downstream validation in synthetic financial data generation (15, 31). It also supports recent work exploring alternative generative architectures, including hybrid variational autoencoder–Markov chain Monte Carlo systems, in which robustness and distributional fidelity are treated as central criteria rather than secondary considerations (32).

Another important result concerned the preservation of temporal structure. Seventeen studies explicitly evaluated whether synthetic data retained temporal dependencies, and temporal GAN architectures generally performed better than standard GANs on this criterion. This finding is critical because financial forecasting depends not only on matching empirical distributions but on preserving how observations evolve through time. Synthetic data that reproduce means, variances, and marginal distributions while weakening serial dependence can appear statistically plausible yet provide misleading training signals. The stronger performance of temporal and attention-based approaches is therefore consistent with prior work that emphasizes path-dependent information, long-range dependence, and high-dimensional sequential structure in financial generation (18, 19). Similarly, work on semantically controlled financial-market generation illustrates the value of conditioning synthetic sequences on market context rather than relying solely on unconditional generation (21). These studies reinforce the present review's conclusion that temporal fidelity should be considered a primary criterion when evaluating synthetic financial data.

The findings further showed that regime dependence was a major source of variation in augmentation effectiveness. In studies comparing bull, bear, stable, crisis, or high-volatility conditions, GAN augmentation tended to perform best when the model had observed at least some structurally similar examples during training. By contrast, performance weakened during genuinely novel structural breaks. This distinction suggests that GANs are effective at enriching known or partially observed regimes but cannot be expected to create reliable representations of economic states entirely absent from the historical data. The result has important implications for emerging markets, where structural change is often more pronounced than in mature markets. Generative models may therefore improve robustness within the support of the training distribution while remaining vulnerable to severe concept drift. This interpretation is consistent with research on controlled and context-aware generation, which seeks to improve the correspondence between synthetic data and specific market states (21). It is also compatible with newer data-centric frameworks designed to adapt to abnormal transactions and evolving distributions rather than treating the historical sample as temporally homogeneous (6).

The review also identified an important interaction between augmentation and downstream model complexity. Most included studies used deep-learning predictors, particularly LSTM, GRU, CNN-LSTM, attention-based networks, or transformer-derived architectures. These models often benefited substantially from augmentation when the original dataset was small because synthetic observations reduced overfitting and exposed the model to a wider range of plausible inputs. However, conventional machine-learning models remained competitive in several studies and occasionally exhibited more stable gains. This finding indicates that successful GAN augmentation does not necessarily justify combining increasingly complex generators with increasingly complex predictors. In some cases, the marginal benefit of synthetic data may be captured by simpler downstream algorithms if those models are less sensitive to imperfections in the generated sequences. This conclusion is compatible with the broader literature on deep time-series forecasting, which recognizes that greater architectural complexity does not guarantee superior generalization and that model performance depends heavily on data quality, validation design,

and task characteristics (1). It also aligns with studies in which synthetic-data generation has been integrated with decomposition, representation learning, and portfolio-oriented models rather than relying exclusively on end-to-end deep prediction (34).

An additional contribution of the present review is the distinction between statistical forecasting improvements and economically meaningful performance. Most studies evaluated GAN augmentation using statistical measures such as RMSE, MAE, accuracy, F1-score, or AUC, whereas only a small subset assessed cumulative return, Sharpe ratio, drawdown, turnover, or transaction costs. Among the latter, improvements in economic performance were generally smaller than improvements in predictive accuracy. This discrepancy is important because a model can achieve a lower statistical error without producing better trading outcomes. For example, small gains in directional accuracy may be offset by transaction costs, excessive turnover, or unstable risk exposure. The growing use of generative AI in trading-data management and financial decision systems makes this distinction increasingly important (3, 4, 16). Future evaluations of augmentation should therefore extend beyond predictive accuracy and examine whether statistical gains translate into economically meaningful and robust improvements after realistic market frictions are considered.

The methodological appraisal also revealed that the literature is stronger in model construction and predictive reporting than in reproducibility, robustness, and synthetic-data validation. Although most studies clearly described their prediction objectives and reported standard performance metrics, fewer than two thirds fully assessed the fidelity of synthetic observations, and only 39.4% provided sufficiently complete information for computational replication. This weakness limits the certainty with which reported gains can be generalized. A GAN may yield favorable predictive results because it genuinely improves the training distribution, but similar gains can also emerge from weak validation design, data leakage, excessive hyperparameter tuning, or poorly controlled comparisons. This concern is consistent with broader discussions of generative AI in finance, where reproducibility, quality assessment, and responsible model evaluation are increasingly recognized as central methodological issues (2, 4, 17). The growing sophistication of generative models, including transformer-based GANs, hybrid architectures, and even emerging quantum generative methods, further increases the need for standardized evaluation because complex models may be difficult to compare without common benchmarks (23, 35). Evidence from non-financial sequential-data synthesis also illustrates that robust evaluation remains essential when adversarial models are applied to noisy real-world observations (36).

The review has several limitations. First, substantial heterogeneity across markets, asset classes, GAN architectures, prediction horizons, evaluation metrics, and validation procedures prevented a formal meta-analysis and required reliance on structured qualitative and descriptive synthesis. Second, the review was limited to English-language peer-reviewed journal articles and full conference papers, which may have excluded relevant studies published in other languages or in gray literature. Third, emerging-market classifications and market boundaries were not always defined uniformly across the primary studies, creating some unavoidable variation in the interpretation of what constituted an emerging financial market. Fourth, several included studies did not report sufficient information on augmentation ratios, hyperparameters, random seeds, or data-splitting procedures, which constrained the depth of methodological comparison. Finally, the relatively small number of studies evaluating real trading performance means that conclusions regarding economic usefulness remain less certain than conclusions regarding statistical predictive performance.

Future research should prioritize standardized and transparent experimental designs that permit direct comparison across GAN architectures and financial markets. Studies should evaluate multiple augmentation ratios, report ablation analyses, and compare GANs against non-generative augmentation methods as well as alternative generative architectures such as variational autoencoders, diffusion models, and hybrid probabilistic approaches. Greater attention should be devoted to temporal fidelity, tail behavior, structural breaks, and cross-regime generalization, particularly in emerging markets where non-stationarity is pronounced. Researchers should also use strictly chronological validation, rolling or walk-forward testing, and independent holdout periods to minimize temporal leakage. Public benchmark datasets, shared code, explicit random seeds, and complete hyperparameter reporting would substantially improve reproducibility. More research is also needed on cross-market transferability, multimodal financial data, uncertainty quantification, synthetic-data privacy, explainability, and the use of generated data for rare systemic events rather than average market behavior.

For practice, financial institutions, quantitative analysts, asset managers, exchanges, and fintech firms should treat GAN-generated observations as a supplementary training resource rather than a replacement for authentic market data. Synthetic data should be validated against the empirical distribution, temporal dependence, volatility dynamics, tail behavior, and regime structure of the original dataset before being incorporated into operational forecasting systems. Practitioners should test several augmentation ratios and retain only those that improve performance on strictly unseen real data. Model selection should be aligned with the forecasting problem: temporal GANs are preferable for sequence preservation, conditional GANs are useful for imbalanced classes or regime-specific generation, and Wasserstein-based models are more appropriate when stability and tail representation are important. In production settings, evaluation should include transaction costs, turnover, drawdown, stress periods, and model degradation after structural change, and augmented models should be monitored continuously to ensure that synthetic data do not reinforce obsolete historical patterns or create a false impression of predictive robustness.

Acknowledgments

We would like to express our appreciation and gratitude to all those who helped us carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

All ethical principles were adheried in conducting and writing this article.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

1. Casolaro A, Capone V, Iannuzzo G, Camastra F. Deep Learning for Time Series Forecasting: Advances and Open Problems. Information. 2023;14(11):598. doi: 10.3390/info14110598.

2. Henrici A, Osterrieder J. Editorial: Artificial Intelligence in Finance and Industry: Highlights From 6 European COST Conferences. Frontiers in Artificial Intelligence. 2022;5. doi: 10.3389/frai.2022.1007074.

3. Kumar U, Sinha B. The Advent of Generative AI and Financial Industry. International Journal of Finance. 2024;9(5):55-70. doi: 10.47941/ijf.2210.

4. Lee DKC, Guan C, Yu Y, Ding Q. A Comprehensive Review of Generative AI in Finance. 2024. doi: 10.20944/preprints202407.2109.v1.

5. Zhang Y, Jiang Z, Peng C, Zhu X, Wang G. Management Analysis Method of Multivariate Time Series Anomaly Detection in Financial Risk Assessment. Journal of Organizational and End User Computing. 2024;36(1):1-19. doi: 10.4018/joeuc.342094.

6. Gong Y, Hu P, Zhang Z, Liu P, Li Z, Zhang R, et al. Data-Centric Generative and Adaptive Detection Framework for Abnormal Transaction Prediction. Electronics. 2026;15(3):633. doi: 10.3390/electronics15030633.

7. Gao X. A Two-Stage Deep Learning Model for Risk Identification in Green Supply Chain Finance. Scientific Reports. 2026;16(1). doi: 10.1038/s41598-026-46830-1.

8. Xi J. Multimodal Conflict-Aware and Generative-Enhanced AI for Early Startup Survival and Risk Prediction. 2025. doi: 10.21203/rs.3.rs-8365925/v1.

9. Liu D, Wu Y, Hong D, Wang S. Time Series Data Augmentation Method of Small Sample Based on Optimized Generative Adversarial Network. Concurrency and Computation Practice and Experience. 2022;34(27). doi: 10.1002/cpe.7331.

10. Hazra D, Shafqat W, Byun Y-C. Generating Synthetic Data to Reduce Prediction Error of Energy Consumption. Computers Materials & Continua. 2022;70(2):3151-67. doi: 10.32604/cmc.2022.020143.

11. Zhong S, Wang X, Xu B, Wu H, Ding M. Hybrid Network Model Based on Data Enhancement for Short-Term Power Prediction of New PV Plants. Journal of Modern Power Systems and Clean Energy. 2024;12(1):77-88. doi: 10.35833/mpce.2022.000759.

12. Yilmaz B. Generative Adversarial Network for Load Data Generation: Türkiye Energy Market Case. Mathematical Modelling and Numerical Simulation With Applications. 2023;3(2):141-58. doi: 10.53391/mmnsa.1320914.

13. Naveed MH, Hashmi U, Tajved N, Sultan N. Is Synthetic the New Real? Performance Analysis of Time Series Generation Techniques With Focus on Network Load Forecasting. 2021. doi: 10.36227/techrxiv.17296235.v1.

14. Dogariu M, Stefan LD, Boteanu B, Lamba C, Kim B, Ionescu BE. Generation of Realistic Synthetic Financial Time-Series. Acm Transactions on Multimedia Computing Communications and Applications. 2022;18(4):1-27. doi: 10.1145/3501305.

15. Ramzan F, Sartori C, Consoli S, Recupero DR. Generative Adversarial Networks for Synthetic Data Generation in Finance: Evaluating Statistical Similarities and Quality Assessment. Ai. 2024;5(2):667-85. doi: 10.3390/ai5020035.

16. Bai X, Zhuang S, Xie H, Guo L. Leveraging Generative Artificial Intelligence for Financial Market Trading Data Management and Prediction. 2024. doi: 10.20944/preprints202407.0084.v1.

17. Gándara ÁR, Abril LG. Generative Adversarial Networks in Business and Social Science. 2024. doi: 10.20944/preprints202407.2122.v1.

18. Fu W, Hirsa A. Simulating Financial Time Series Using Attention. 2022. doi: 10.48550/arxiv.2207.00493.

19. Pardo FdM, Schwendner P, Wunsch M. Tackling the Exponential Scaling of Signature-Based Generative Adversarial Networks for High-Dimensional Financial Time-Series Generation. The Journal of Financial Data Science. 2022;4(4):110-32. doi: 10.3905/jfds.2022.1.109.

20. Hirano M, Sakaji H, Izumi K. Policy Gradient Stock Gan for Realistic Discrete Order Data Generation in Financial Markets. 2022. doi: 10.2139/ssrn.4095304.

21. Xia H, Sun S, Wang X, An B. Market-Gan: Adding Control to Financial Market Data Generation With Semantic Context. Proceedings of the Aaai Conference on Artificial Intelligence. 2024;38(14):15996-6004. doi: 10.1609/aaai.v38i14.29531.

22. Abdelhady AS, Dahmani N, Abouelmagd LM, Darwish A, Hassanien AE. Green Finance Growth Prediction Model Based on Time-Series Conditional Generative Adversarial Networks. Plos One. 2024;19(7):e0306874. doi: 10.1371/journal.pone.0306874.

23. Podobiński A, Chudziak JA. Financial Time Series Augmentation Using Transformer Based GAN Architecture. 2026. doi: 10.48550/arxiv.2602.17865.

24. Gadhi AHA, Peiris S, Allen DE. Improving Volatility Forecasting: A Study Through Hybrid Deep Learning Methods With WGAN. Journal of Risk and Financial Management. 2024;17(9):380. doi: 10.3390/jrfm17090380.

25. Hu Z, Yang X, Jiang W. Research on Deep Learning Financial Volatility Prediction Method Based on Signal Decomposition and Data Augmentation. 2025. doi: 10.21203/rs.3.rs-7726466/v1.

26. Kim D-J, Kim D-H, Choi SY. Modeling Stylized Facts in FX Markets With FINGAN-BiLSTM: A Deep Learning Approach to Financial Time Series. Entropy. 2025;27(6):635. doi: 10.3390/e27060635.

27. Labiad B, Berrado A, Benabbou L. Predicting Extreme Events in the Stock Market Using Generative Adversarial Networks. International Journal of Advances in Intelligent Informatics. 2023;9(2):218. doi: 10.26555/ijain.v9i2.898.

28. Chandra R. Extreme Value Forecasting Using Relevance-Based Data Augmentation With Deep Learning Models. 2025. doi: 10.48550/arxiv.2510.02407.

29. Hirnschall D. Semi-Supervised Bayesian GANs With Log-Signatures for Uncertainty-Aware Credit Card Fraud Detection. Mathematics. 2025;13(19):3229. doi: 10.3390/math13193229.

30. Sundaravadivel P, R AI. Federated Deep Learning and Explainable AI for Real-Time Credit Card Fraud Detection in Highly Imbalanced Transaction Streams. 2025. doi: 10.21203/rs.3.rs-7587282/v1.

31. Hellermann J, Qian Q, Shah A. Financial Time Series Data Augmentation With Generative Adversarial Networks and Extended Intertemporal Return Plots. 2022. doi: 10.48550/arxiv.2205.08924.

32. Prenestino FB, Barbierato E, Gatti A. Robust Synthetic Data Generation for Sequential Financial Models Using Hybrid Variational Autoencoder–Markov Chain Monte Carlo Architectures. Future Internet. 2025;17(2):95. doi: 10.3390/fi17020095.

33. Mahajan RA, Dey R, Khan M. Enhancing Financial Predictive Modeling With Synthetic Data Using Generative Approach. 2025. doi: 10.21203/rs.3.rs-6178330/v1.

34. Kalina B, Lee J-H, Na K. Enhancing Portfolio Performance Through Financial Time-Series Decomposition-Based Variational Encoder-Decoder Data Augmentation. Symmetry. 2024;16(3):283. doi: 10.3390/sym16030283.

35. Waller J, Caruso F, Makris D, Nilavalan R, Liang X. Quantum Generative Diffusion Model for Real-World Time Series. 2026. doi: 10.48550/arxiv.2606.27561.

36. Shi X, Hu P. A GAN-Based Framework for Robust Data Synthesis in Satellite Internet Observations. 2026. doi: 10.48550/arxiv.2607.24790.