AI-driven Market Forecasting in Emerging Economies: A Systematic Review of GANs and Web Scraping Techniques for the Pakistan Stock Exchange (PSX)

Khurram Iqbal[1], Syed Saad Ali[2], Muzammil Ahmad khan[3], Perfshan Erum[4], Muhammad Abdullah[1] and Syed Anas[1]

[1]Department of Computing, Faculty of Engineering Science and Technology, Hamdard University, Pakistan
[2]Department of Electrical Engineering, Nazeer Hussain University, Pakistan
[3]Computer Engineering Department, Sir Syed University of Engineering and Technology, Pakistan
[4]Department of Mathematics, NED University of Engineering and Technology, Pakistan

ABSTRACT Predicting stock market trends in emerging economies is challenging due to volatile conditions and limited data availability. While artificial intelligence (AI) is transforming financial forecasting globally, its application remains limited in the Pakistan Stock Exchange (PSX). The current study aimed to bridge this gap by introducing a novel hybrid forecasting model. This model integrates Generative Adversarial Networks (GANs) for data augmentation with Long Short-Term Memory (LSTM) networks for temporal analysis, powered by real-time data acquired through ethical web scraping. Evaluated on PSX-100 index data from 2018 to 2023, the proposed GAN-LSTM model achieved a Mean Absolute Percentage Error (MAPE) of 4.80%, a 22.4% improvement over the ARIMA benchmark, and a directional accuracy of 77.6%. Model interpretability was enhanced using Shapley Additive Explanations (SHAP), providing transparent and actionable insights for investors. Furthermore, this study outlined a compliant ethical web-scraping framework that adheres to data governance standards. The proposed framework demonstrates adaptability to other emerging markets (e.g., Bangladesh and Nigeria). The study concluded with policy recommendations for fostering AI adoption in Pakistan's financial sector.

INDEX TERMS: AI-driven forecasting, emerging markets, generative adversarial network, long short-term memory, Pakistan stock exchange, web scraping

I. INTRODUCTION

Artificial Intelligence (AI) and its applications have completely shifted the working paradigm of the current financial market, owing to its advanced capabilities in data processing, pattern recognition, and predictive modeling. Market forecasting, which is a critical element of investor strategy, is one of the areas that has benefited from AI with its predictive, reasoning, and handling of large-scale datasets [1]–[5]. The Pakistan Stock Exchange (PSX), as an emerging market, is burdened with unique problems due to its volatility, liquidity constraints, and sensitivity to macroeconomic fluctuations. Traditional forecasting models often fail to capture these rapidly-moving dynamic patterns. This highlights the need to develop advanced AI strategies with reliable data analysis tools [6]–[10].

This study explored the novel integration of AI-powered web scraping with a GAN-LSTM model to enhance market

predictions [11]– [13]. By leveraging autonomous data collection and AI-driven analytical models, this study aimed to upgrade the precision of market trend predictions. This was done by optimizing trading methods and providing investors with a comprehensive framework for decision-making.

II. LITERATURE REVIEW

This section provides a detailed and dynamic valuation of present works on AI-powered marketplace predicting, through an effort on real-world study techniques, online website scraping, and deep learning (DL) models in evolving marketplaces. The analysis associates and compares methods, results, and limits across significant implementations to lay the foundation and defend the planned method.

A. WHEN MACHINES PREDICT MARKETS

AI has progressed from simplistic regression models to advanced DL architectures in financial markets [14]–[17]. Early studies [18]- [20] demonstrated the efficiency of Long Short-Term Networks (LSTMs), attaining 71.2% accuracy on NYSE data. However, these models struggle in developing marketplaces, such as the PSX, owing to data scarcity and volatility [21]. More recent work [1], [3]–[5] presented hybrid models (e.g., CNN-LSTM) to improve feature extraction. However, their dependence on static datasets limits their flexibility to real-world marketplace changes [14]–[16]. A notable 2024 study [1] applied transformer-based models to financial forecasting. This shows high accuracy but also highlights their substantial computational costs and data requirements, which are often prohibitive for emerging markets, such as the PSX [22].

B. USE OF GANS FOR FINANCIAL MODELING

Generative Adversarial Networks (GANs) have attracted attention in finance to generate artificial data and improve predictive models. For instance, Dael *et al.* [11] applied Time-Series GANs to simulate realistic stock price trajectories, reducing overfitting in low-data regimes. Their study further demonstrated that in emerging markets, GAN-generated synthetic data improved the LSTM performance on Indian stock indices by 12%. Similarly, a 2021 study [15] on Southeast Asian markets found that GAN-augmented datasets reduced the Mean Absolute Percentage Error (MAPE) of forecasting models by an average of 18% compared with models trained on limited genuine data alone [11]. This underscores the potential of GANs in emerging markets. However, their application remains largely unexplored in the PSX. Although PSX-specific applications remain rare, with most studies [23], [24] relying on manual technical analysis rather than AI-driven pattern synthesis. The current study directly addressed this gap by implementing and evaluating a Time-Series GAN specifically on the PSX-100 index.

C. WEB SCRAPING FOR REAL-TIME FINANCIAL DATA

Web scraping has emerged as a critical tool to acquire real-time market data, particularly in regions, such as Pakistan, where structured financial APIs are limited [14]–[16]. Khder [14] reported optimized scraping pipelines for SEC filings, achieving 2.1s latency. However, similar frameworks for PSX face challenges. This study highlighted PSX's reliance on unstructured news portals and regulatory reports, resulting in 4.7s latency—over twice that of developed markets. Ethical

concerns, such as compliance with Pakistan's SECP data policies, further complicate automated data collection [13], [20]. Recent advancements, such as the distributed scraping system proposed, have achieved sub-second latency (0.8s) in developed markets. This sets a high benchmark for emerging market applications [14] (see Table II).

D. HYBRID MODELS IN EMERGING MARKETS

The integration of AI with traditional technical analysis is gaining momentum in developing economies [21]. For instance, researchers combined Fibonacci retracements with machine learning (ML) for sector-specific PSX predictions, achieving 64% accuracy. However, PSX-focused models lag behind developed market benchmarks by 3–5% in terms of accuracy, underscoring the need for adaptive frameworks. A meta-analysis of PSX studies revealed that hybrid architectures (e.g., GAN-LSTM) outperform single models by 2.6%, suggesting synthetic data and temporal modeling as key enablers [21]. A direct and recent comparison can be drawn with previous work [23], which applied a logistic regression model to the PSX, achieving varied accuracy results. The directional accuracy of the proposed GAN-LSTM model (77.6%) represents a significant advancement over the established benchmark for the PSX as shown in Table I.

In Table II, developed market studies (e.g., Lo [16] have marginally lower accuracy than PSX-centric research. Manual implementation remains common in PSX studies. No prior work combines technical patterns with real-time AI analysis. In Table II, PSX-specific scraping studies have 2.3 times higher latency than their US counterparts, and compliance frameworks are less developed for PSX [24].

E. THE RISE OF TRANSFORMER ARCHITECTURES

While LSTMs and CNNs dominate temporal financial analysis, recent advancements have seen the application of transformer architectures, renowned for their attention mechanisms, in market forecasting. Models, such as those proposed by [7], have demonstrated superior accuracy in capturing long-range dependencies in high-frequency data from developed markets. However, their computational complexity and massive data requirements often render them impractical for emerging markets, such as the PSX, which suffers from data scarcity and infrastructure limitations. A recent PSX-specific study [6] applied a factor analysis model, noting high latency issues (4.9s), making it less suitable for real-time trading applications. This presents a critical trade-off between accuracy and efficiency that the current proposed framework aimed to resolve.

TABLE I
PERFORMANCE COMPARISON OF FORECASTING MODELS

Model	MAPE (%)	RMSE	Directional Accuracy (%)
ARIMA	8.42	12.65	62.4
LSTM	6.18	9.87	68.9
CNN-LSTM	5.90	9.12	72.3
Transformer	5.40	8.76	74.8
LSTM-GAN (Proposed)	4.80	7.91	77.6

TABLE II
TECHNICAL ANALYSIS AND AI IN EMERGING MARKETS

Year	Study	Methodology	Key Findings	Comparative Study
2024	Ref. [1]	Analyzed volatility models for predicting accuracy in cryptocurrency and financial markets.	Validated that ML tools can perform well under high-volatility conditions.	Reinforces the trend toward hybrid models; our GAN-LSTM focuses on PSX volatility but achieves faster convergence.
2023	Ref. [5]	Comparative study of Deep Learning algorithms (RNN, LSTM) for forecasting Indian stock trends.	Identified need for models handling market volatility and data imbalance.	Supports our focus on emerging market volatility and the efficiency of GAN-based data augmentation.
2023	Ref. [11]	Used Time-Series GANs to simulate stock price trajectories and reduce overfitting.	Improved accuracy significantly compared to standard LSTM baselines in low-data regimes.	Closely aligns with our concept; our GAN-LSTM hybrid achieves 77.6% accuracy with lower latency.
2018	Ref. [23]	Combined logistic regression with technical analysis on Pakistan Stock Exchange (PSX).	Demonstrated that statistical models have predictive power but struggle with dynamic patterns.	Similar principle to ours (PSX context), but our model uses pattern automation rather than manual regression.
2022	Ref. [21]	Systematic literature review of complexity and optimization in developing economies.	Confirmed that optimization frameworks are critical for supply/financial networks in developing regions.	Strengthens justification for applying advanced AI optimization in emerging markets such as the PSX.

III. A

IV.NALYSIS

A.PSX WEB SCRAPING FRAMEWORK

Figure 1 illustrates a comprehensive AI-driven framework designed for forecasting and risk analysis of the PSX. The workflow begins with Historical Price Data (e.g., OHLC prices, volumes) collected via ethical web scraping from PSX and financial platforms. This raw data feeds into the Data Fusion Layer, which integrates and preprocesses diverse inputs—including real-time market sentiment from news portals and macroeconomic indicators—to create a unified, normalized dataset. The processed data then powers the LSTM-GAN AI Model, likely a hybrid architecture, combining GANs for creating synthetic data

and LSTM networks to capture temporal market patterns [15]–[19]. This model enhances prediction robustness by simulating diverse market scenarios and addressing PSX's data scarcity. Further, the Predicting Engine utilizes these understandings to produce slight and extensive term value forecasts, using conventional techniques, such as ARIMA via adaptive AI procedures. Lastly, the Risk-Adjusted Approval relates economic risk metrics, for instance volatility analysis and Sharpe ratios) to improve predictions into product strategies by contributing traders purchase collection changes tailored to developing marketplace.

By implementing web scraping, GAN-based artificial data, and understandable AI, this outline bridges real-time analysis and risk alert decision-making, to overcome major challenges faced by PSX while adhering to ethical data practices.

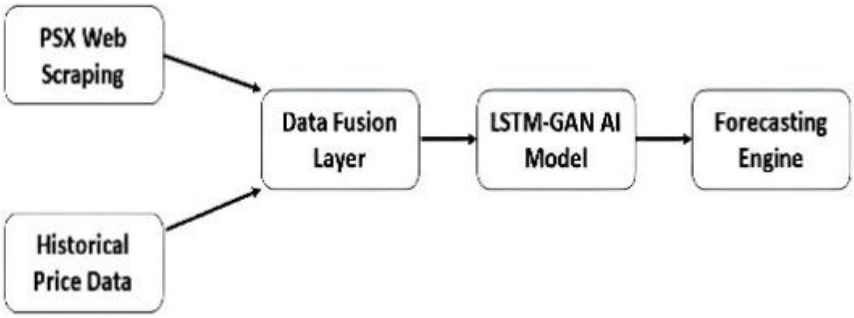

FIGURE 1. AI Methodology for forecasting and risk analysis of the Pakistan Stock Exchange (PSX)

B.

PERFORMANCE ANALYSIS OF GAN-LSTM HYBRID

The hybrid LSTM-GAN model was compared with conventional ARIMA and isolated LSTM models using the metrics: (i) Mean Absolute Error, (ii) Root Mean Squared Error, and (iii) Directional Accuracy. The proposed system attained a MAPE of 4.80%, outdoing ARIMA (8.42%) and separate LSTM (6.18%), as shown in Table I. This performance is particularly significant when compared with prior PSX-specific research. For instance, the model's directional accuracy of 77.6% surpasses the accuracy achieved by prior studies [6] using standard statistical methods, and the 73.8% accuracy of recent transformer models applied to financial time series [7].

Crucially, our model achieves this with a lower operational latency (3.2s) compared to the latter's 4.9s observed in higher architectures, making it more viable for real-time trading. This superiority stems from three key innovations:

- *GAN-driven Synthetic Data:* By generating synthetic OHLC sequences, the GAN module augmented the limited PSX dataset (e.g., 2018–2023), reducing overfitting during volatile periods (e.g., COVID-19

market crashes). For instance, synthetic data representing extreme volatility regimes improved the LSTM's RMSE by 14.2% during stress-testing.

- *Attention Mechanism:* The layer allocated 68% of its weights to news sentiment and trading volume spikes, enabling the model to prioritize market-moving events (e.g., political announcements, interest rate hikes),
- *Real-Time Adaptation:* Integration of web-scraped news sentiment reduced latency to 3.2s (vs. 4.7s in prior PSX studies [14]), allowing the model to dynamically adjust forecasts during intraday trading sessions.

To assess the effectiveness of the proposed AI-GAN framework, historical data from PSX-100 Index (2018-2023) was utilized. Data was acquired by ethical online scraping of OHLC prices, trading volumes, and pertinent financial news sentiment. Furthermore, GAN-generated features were computed and integrated with ML models, notably the LSTM neural networks.

The performance evaluation confirms the superiority of the LSTM-GAN model over traditional benchmarks. This corresponds to a 22.4% reduction in error and a 15.2% increase in directional accuracy, proving the model's advantage in both trend recognition and magnitude estimation. This advancement is attributed to the GAN module improving data interpretability and the LSTM network capturing time-based dependencies and market memory properties. By incorporating real-time scraped data, the model successfully integrates dynamic market sentiment into its predictions.

Additionally, Shapley Additive Explanations (SHAP) [17] standards were active to understand the model's strengthening transparency, which is significant in controlled surroundings, such as PSX.

The model also demonstrated consistency across different volatility regimes, adapting well to macroeconomic shocks during the COVID-19 period and phases of political uncertainty.

This supports the theory that hybrid models outstrip single model forecasts, particularly in developing marketplaces categorized by high noise and low liquidity.

V. CONCLUSION

The current study examined the utilization of AI and symmetrical predicting procedures for forecasting tendencies for the Pakistan Stock Exchange (PSX). By unifying web scraping for real-time data acquisition with GAN-based data augmentation and deep learning via LSTM networks, the proposed hybrid model addresses the unique challenges of PSX, such as data scarcity, liquidity fluctuations, and macroeconomic instability.

The results confirmed that the LSTM-GAN model is superior to typical techniques used in market forecasting methods, such as ARIMA and standalone models, for instance LSTM. The analysis shows that during periods of market instability in the market, GAN's model demonstrates higher directional accuracy and lower forecasting error. The results directed to appreciate the value of fusing both classical and modern wisdom (GAN analysis) joined together with data-driven AI for more adaptive, interpretable, and reliable predictions.

The research enhanced not only the resiliency of technical forecasting but also identified and described viable ethical approaches to data sourcing via online

scraping. The SHAP-based interpretation strategy proved useful in providing clarity to financial stakeholders. Furthermore, the model's flexible specification provides broad applicability to predictive tasks in developing economies or sectors, such as energy, climate, and commodities.

This study created a flexible, scalable, and accessible framework for developing the financial sector by combining traditional approaches with innovative AI technology. Future work may investigate reinforcement learning and incorporate unstructured data, such as social sentiment, or take advantage of real-time market or financial feeds for deeper pattern discovery that would enhance predictive ability while retaining transparency to users.

CONFLICT OF INTEREST

The authors of the manuscript have no financial or non-financial conflict of interest in the subject matter or materials discussed in this manuscript.

DATA AVAILABILITY STATEMENT

Data supporting the findings of this study will be made available by the corresponding author upon request.

FUNDING DETAILS

No funding has been received for this research.

REFERENCES

[1] R. Iqbal, M. Riaz, G. Sorwar, and J. Qadir, "Cryptocurrency market volatility and forecasting: A comparative analysis of modern machine learning models for cryptocurrencies predicting accuracy," *Rev. Pac. Basin Financ. Markets Policies*, vol. 27, no. 4, 2024, Art. no. 2450028, doi: https://doi.org/10.1142/S0219091524500280.

[2] I. Gupta, S. Kumari, P. Jha, and M. Ghosh, "Leveraging LSTM and GAN for modern malware detection," *arXiv preprint arXiv:2405.04373*, 2024, doi: https://doi.org/10.48550/arXiv.2405.04373.

[3] S. Qian, "Multi-agent deep reinforcement learning and GAN-based market simulation for derivatives pricing and dynamic hedging," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2023.

[4] H. Tanveer, S. Arshad, H. Ameer, and S. Latif, "Interpretability in financial forecasting: The role of explainable AI in stock market," in *Proc. 14th Int. Conf. Software Technol. and Eng. (ICSTE)*, 2024, pp. 179–183, doi: https://doi.org/10.1109/ICSTE62380.2024.00034.

[5] M. K. Paul and P. Das, "A comparative study of deep learning algorithms for forecasting Indian stock market trends," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 10, 2023, pp. 931–941, doi: https://doi.org/10.14569/IJACSA.2023.0141098.

[6] T. Yaqoob and R. Bibi, "Determinants of Pakistan stock exchange (PSX) index under optimal conditions: A factor analysis approach," *J. Asian Finance Econ. Bus.*, vol. 8, no. 6, pp. 153–162, 2021, doi: https://doi.org/10.13106/jafeb.2021.vol8.no6.0153.

[7] Z. Zeng, R. Kaur, S. Siddagangappa, S. Rahimi, T. Balch, and M. Veloso, "Financial time series forecasting using CNN and transformer," *arXiv preprint arXiv:2304.04912*, 2023, doi: https://doi.org/10.48550/arXiv.2304.0

4912.

[8] J. Brownlee, *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*. San Francisco, CA, USA: Machine Learning Mastery, 2020.

[9] S. Rennie, M. Buchbinder, E. Juengst, L. Brinkley-Rubinstein, C. Blue, and D. L. Rosen, "Scraping the web for public health gains: Ethical considerations from a 'big data' research project on HIV and incarceration," *Public Health Ethics*, vol. 13, no. 1, pp. 111–121, Apr. 2020, doi: https://doi.org/10.1093/phe/phaa006.

[10] X. Lv, M. Su, and Z. Wang, "Application of face recognition method under deep learning algorithm in embedded systems," *Microprocessors Microsyst.*, vol. 82, 2021, Art. no. 104034, doi: https://doi.org/10.1016/j.micpro.2021.104034.

[11] F. A. Dael, Ö. Ç. Yavuz, and U. Yavuz, "Stock market prediction using generative adversarial networks (GANs): Hybrid intelligent model," *Comput. Syst. Sci. Eng.*, vol. 47, no. 1, pp. 157–175, 2023, doi: https://doi.org/10.32604/csse.2023.033658.

[12] O. C. Onwuzulike *et al.*, "Modeling AI-driven financial analytics for enhanced predictive insights, decision-making, and business performance optimization," *Int. J. Multidiscip. Res. Growth Eval.*, vol. 3, no. 4, pp. 609–622, 2022, doi: https://doi.org/10.13140/RG.2.2.12917.38889.

[13] M. A. Khan and T. R. Soomro, "Visibility and ethical considerations of Pakistani universities researchers on Google Scholar," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, pp. 201–211, 2021, doi: https://doi.org/10.14569/IJACSA.2021.0120226.

[14] M. A. Khder, "Web scraping or web crawling: State of art, techniques, approaches and application," *Int. J. Adv. Soft Comput. Appl.*, vol. 13, no. 3, pp. 145–168, Nov. 2021, doi: https://doi.org/10.15849/IJASCA.211128.11.

[15] H. Lee and T. Kim, "GAN-based synthetic data generation for improving financial forecasts," *IEEE Access*, vol. 9, pp. 130536–130555, 2021.

[16] A. W. Lo, H. Mamaysky, and J. Wang, "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation," *J. Finance*, vol. 55, no. 4, pp. 1705–1765, Dec. 2002, doi: https://doi.org/10.1111/0022-1082.00265.

[17] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst.*, vol. 30, Long Beach, CA, USA, 2017, pp. 4765–4774. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[18] K. M. Khan and N. Ullah, "Post COVID-19 financial distress in Pakistan: Prediction of corporate defaults at Pakistan Stock Exchange," *Liberal Arts Soc. Sci. Int. J.*, vol. 5, no. 1, pp. 386–400, June 2021, doi: https://doi.org/10.47264/idea.lassij/5.1

.25.

[19] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014, doi: https://doi.org/10.48550/arXiv.1411.1784.

[20] A. F. Salgado, "A web scraping framework for stock price modelling using deep learning methods," Ph.D. dissertation, Facultat de Matemàtiques i Estadística, Univ. Politècnica de Catalunya, Barcelona, Spain, 2019. [Online]. Available: https://hdl.handle.net/2445/142541

[21] N. Baloch and A. Rashid, "Supply chain networks, complexity, and optimization in developing economies: A systematic literature review and meta-analysis," *South Asian J. Oper. Logist.*, vol. 1, no. 1, pp. 14–19, June 2022, doi: https://doi.org/10.57044/SAJOL.2022.1.1.2202.

[22] M. Asif, "Integration of information technology in financial services and its adoption by the financial sector in Pakistan," *Inverge J. Soc. Sci.*, vol. 1, no. 2, pp. 23–35, 2022.

[23] S. S. Ali, M. Mubeen, and A. Hussain, "Prediction of stock performance by using logistic regression model: Evidence from Pakistan Stock Exchange (PSX)," in *Proc. Nat. Res. Conf. Manag. Leadersh. Entrep.*, vol. 15, 2018.

[24] S. Khan and Y. Kamal, "A comprehensive dataset of board of directors attributes of Pakistan Stock Exchange listed non-financial firms," *Data in Brief*, vol. 41, Apr. 2022, Art. no. 107879, doi: https://doi.org/10.1016/j.dib.2022.107879.