Simen Owen Academic Proceedings Series

Vol. 2 2025

Article Open Access

A Comparative Study of Deep Learning Architectures for Multivariate Financial Time Series Forecasting

Liuqing Ding [1,*]

[1] Boston University, Boston, MA, 02215, USA
[*] Correspondence: Liuqing Ding, Boston University, Boston, MA, 02215, USA

Abstract: Accurate forecasting of multivariate financial time series remains a critical challenge due to high volatility, non-stationarity, and complex cross-variable dependencies. Although deep learning models such as LSTM, GRU, TCN, and Transformer have shown notable progress, existing research often evaluates these architectures in isolation, lacks interpretability, and provides limited analysis of robustness across different markets. These limitations impede the deployment of reliable forecasting systems in practical financial settings. This study presents a comprehensive comparative analysis of representative deep learning architectures for financial forecasting and introduces a novel Hybrid Attention-Gated Module (HAGM). HAGM combines convolutional feature extraction, gated fusion, and multi-head self-attention mechanisms to efficiently capture both local and global dependencies. Experiments were conducted on stock indices, foreign exchange, and cryptocurrency datasets, assessing model performance across multiple forecast horizons. The results demonstrate that HAGM consistently outperforms baseline models, achieving lower RMSE and MAPE while exhibiting faster convergence. Ablation studies confirm the complementary contributions of convolution, gating, and attention components, and interpretability analyses identify critical variables such as trading volume and volatility. Robustness evaluations further reveal superior cross-market generalization and resilience under noisy conditions. Overall, this work advances the methodological understanding of deep learning approaches for financial forecasting and provides actionable insights for practitioners aiming to develop accurate, efficient, and interpretable predictive systems.

Keywords: multivariate time series; financial forecasting; deep learning; Transformer; hybrid attention-gated module

Received: 17 October 2025
Revised: 25 October 2025
Accepted: 14 November 2025
Published: 23 November 2025

1. Introduction

Forecasting multivariate time series has emerged as a critical challenge in machine learning, as numerous real-world applications, ranging from energy consumption prediction to sensor data monitoring, rely on accurate modeling of high-dimensional temporal dependencies [1]. Among these, financial time series forecasting represents a particularly demanding scenario, characterized by high volatility, non-stationarity, and intricate cross-variable correlations [2]. Unlike traditional univariate forecasting, where each series is modeled independently, multivariate financial forecasting requires the integration of diverse indicators, such as stock indices, exchange rates, and trading volumes, into a unified predictive framework [3]. The complexity of these interdependencies has positioned deep learning as a promising solution, offering the capability to capture nonlinear patterns and long-range dependencies that classical statistical methods often fail to model effectively.

Despite significant progress in deep learning for time series, several challenges remain. Recurrent neural networks (RNNs) and their variants, including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have been widely applied due to their ability to mitigate vanishing gradients and retain long-term memory [4]. However, these architectures still face sequential training inefficiencies and limited scalability when processing very long sequences. Temporal Convolutional Networks (TCNs) offer an alternative by exploiting dilated convolutions to capture extended contexts, yet they often lack flexibility in modeling cross-variable dependencies [5]. More recently, Transformer-based architectures have demonstrated superior performance in natural language processing and are increasingly adapted for time series analysis [6]. Their self-attention mechanisms enable parallelization and the modeling of complex, non-local dependencies. Nevertheless, high computational costs and limited domain-specific adaptations constrain their broader adoption in financial forecasting tasks.

Methodologically, current research exhibits additional shortcomings. Many studies evaluate models on narrow datasets or focus solely on single-architecture performance without systematic cross-model comparison. Furthermore, interpretability, a critical factor for practical deployment, is often overlooked, leaving end users with limited insight into model decision-making. This absence of rigorous comparative studies not only hinders fair benchmarking but also restricts the development of hybrid or domain-specific models optimized for multivariate forecasting tasks.

To address these gaps, this study conducts a comprehensive comparative analysis of representative deep learning models for multivariate financial time series forecasting. Specifically, we examine the performance of RNN, LSTM, GRU, TCN, and Transformer-based architectures, alongside a hybrid design that integrates convolutional feature extraction with multi-head attention. The proposed Hybrid Attention-Gated Module (HAGM) is designed to enhance cross-variable dependency modeling while maintaining computational efficiency. By comparing models across multiple datasets and evaluation metrics, we aim to provide a clearer understanding of the conditions under which each architecture excels or underperforms.

The methodological approach of this study comprises three key stages. First, diverse real-world financial datasets are curated and preprocessed to ensure comparability across models. Second, each architecture is trained and evaluated using standardized protocols, covering both short- and long-horizon forecasting tasks. Third, supplementary analyses, including ablation experiments, statistical significance tests, and interpretability assessments via attention heatmaps and Shapley values, are conducted to reveal model behavior and robustness. This structured pipeline ensures both fairness in evaluation and depth in analysis.

The contributions of this work are threefold. (1) We provide a systematic, large-scale comparison of widely adopted deep learning architectures for multivariate financial forecasting, filling a critical gap in existing literature. (2) We introduce a novel hybrid attention-gated module that achieves superior accuracy in capturing long-range and cross-variable dependencies. (3) We offer interpretability-driven insights and robustness evaluations that enhance the practical utility of deep learning models in real-world applications.

By focusing on architectural performance, design trade-offs, and interpretability, this study advances the methodological foundation of deep learning for time series forecasting. Beyond financial applications, the findings contribute to broader machine learning research, offering guidance for deploying deep learning architectures in domains where multivariate, nonlinear, and interdependent time series are prevalent.

2. Related Works

Deep learning has become the predominant paradigm for multivariate time series forecasting, yet existing studies vary considerably in their choice of architectures, datasets, and evaluation methodologies. This section reviews related works across three key subfields: recurrent neural architectures, convolutional and hybrid models, and Transformer-based approaches. Each subsection highlights representative studies, their advantages and limitations, and their relation to the present work.

2.1. Recurrent Neural Architectures

Recurrent neural networks (RNNs) and their derivatives, including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), remain widely studied for financial and multivariate forecasting [7]. LSTM networks, in particular, have been employed to capture long-range temporal dependencies while mitigating vanishing gradient issues. Several studies have applied LSTM to stock index prediction, demonstrating improvements over autoregressive integrated moving average (ARIMA) baselines, especially in modeling nonlinear market behaviors [8]. GRU has also gained attention due to its computational efficiency and reduced parameter count compared with LSTM, making it suitable for high-frequency financial data.

However, recurrent models are inherently sequential, which limits training efficiency and scalability. Their capacity to capture cross-variable dependencies is also constrained, as most implementations focus on temporal recurrence rather than inter-series interactions. Some extensions of LSTM incorporate attention mechanisms or multi-task objectives to address these limitations, but the improvements are often dataset-specific and lack generalization [9]. In this study, RNN-based models are included as baselines to benchmark their relative strengths and weaknesses against more recent architectures.

2.2. Convolutional and Hybrid Architectures

Temporal Convolutional Networks (TCNs) and hybrid CNN-RNN models represent an alternative research direction, leveraging convolutional kernels and dilation to expand receptive fields [10]. TCNs have been shown to achieve faster training compared to recurrent models, with the added benefit of parallelization [11]. For example, studies applying dilated causal convolutions to energy and traffic datasets demonstrated superior short-term forecasting accuracy relative to LSTM, particularly when long input windows were required. CNN-LSTM hybrids have also been applied to financial data, where convolutional layers capture local temporal or cross-series features, and recurrent layers model sequential dynamics [12].

The primary strengths of convolutional and hybrid models lie in their efficiency and capacity for multi-scale feature extraction. Nevertheless, they often struggle to capture global dependencies, especially in long-horizon forecasting tasks. Moreover, while CNN-based models tend to generalize well across certain domains, their interpretability remains limited, which poses challenges for deployment in contexts such as finance. In this work, TCN and hybrid models are evaluated to highlight trade-offs in training efficiency and predictive accuracy, and to establish a foundation for the hybrid attention-gated design introduced later.

2.3. Transformer-Based and Attention-Driven Models

The latest advances in multivariate forecasting are driven by Transformer-based architectures. Originally developed for natural language processing, Transformers utilize self-attention mechanisms to capture global dependencies and allow full parallelization during training. Variants such as Informer and Autoformer have been proposed to reduce the quadratic complexity of vanilla Transformers when applied to long sequences [13]. These models have achieved state-of-the-art results on benchmark time series datasets, including weather, electricity, and financial indicators. Their ability to model non-local correlations makes them particularly suitable for multivariate forecasting.

Despite their advantages, Transformer-based models face challenges. High computational requirements complicate real-time deployment, and their performance is sensitive to hyperparameter tuning and dataset size [14]. Furthermore, interpretability remains a concern, as attention weights alone may not fully explain prediction outcomes. Some studies have attempted to enhance interpretability by integrating feature attribution methods, though these approaches remain underexplored [15].

In this study, Transformer models are compared directly with RNN- and CNN-based baselines, providing a comprehensive evaluation of their advantages and trade-offs. The proposed hybrid attention-gated module draws inspiration from the Transformer's ability to model dependencies while mitigating computational costs through the combination of attention and lightweight convolutional operations.

2.4. Summary

Existing research reveals complementary trade-offs across architectures. Recurrent models excel at capturing sequential dependencies but scale poorly. Convolutional and hybrid models offer efficient training and multi-scale feature extraction but struggle with global dependencies and interpretability. Transformer-based methods achieve state-of-the-art accuracy by modeling long-range and cross-variable correlations, albeit with high computational cost and limited transparency. These gaps highlight the need for systematic cross-model comparison and hybrid approaches that integrate efficiency, scalability, and accuracy, objectives directly addressed in this study.

3. Methodology

This section introduces the methodological framework for comparing deep learning architectures in multivariate financial time series forecasting. The study benchmarks classical recurrent models (LSTM, GRU), convolutional approaches (TCN), and Transformer-based architectures, alongside a proposed Hybrid Attention-Gated Module (HAGM). The methodology encompasses problem formulation, model-specific mechanisms, the proposed module, training objectives, and evaluation protocols.

3.1. Problem Formulation

Multivariate financial forecasting involves predicting future values of multiple interdependent time series, such as stock prices, exchange rates, or trading volumes. Unlike univariate forecasting, where each sequence is modeled independently, multivariate forecasting requires capturing both temporal dynamics within each variable and cross-variable dependencies, which significantly increases modeling complexity.

Formally, the input series is defined as:

$$X = \{x_1, x_2, \dots, x_T,\},\ x_t \in R^d \tag{1}$$

where T denotes the sequence length and ddd the number of variables. The forecasting objective is to predict the next h steps:

$$\hat{Y} = f_\theta(X), \hat{Y} \in R^{h \times d} \tag{2}$$

Here, f_θ represents a deep learning model parameterized by θ Different architectures impose distinct inductive biases: recurrent models emphasize sequential memory, convolutional models focus on local patterns, and Transformers leverage self-attention to capture long-range and global interactions. In practice, the prediction horizon h may vary, with short-term horizons being critical for high-frequency trading, while longer horizons serve portfolio management and risk control. This formulation provides a unified framework to systematically evaluate and compare the strengths and limitations of competing architectures.

3.2. Lstm-Based Forecasting

Long Short-Term Memory (LSTM) networks are among the most widely used recurrent architectures for time series forecasting due to their ability to mitigate vanishing gradients and capture long-range temporal dependencies. LSTM introduces memory cells and gating mechanisms that regulate information flow. At each time step t, the hidden state is updated as:

$$h_t = o_t \odot \tanh\,(f_t \odot c_{t-1} + i_t \odot \tilde{c}_t) \quad (3)$$

where f_t, i_t, and o_t denote the forget, input, and output gates, c_t represents the cell state, and $\odot$ indicates elementwise multiplication. This structure allows the network to retain important historical signals while discarding irrelevant noise.

In multivariate financial forecasting, LSTM has been applied to tasks such as stock index prediction, volatility estimation, and cross-asset forecasting. Its advantage lies in modeling nonlinear temporal patterns, outperforming classical linear models. However, LSTM suffers from sequential training inefficiencies, limited scalability for long horizons, and does not explicitly model cross-variable dependencies. These limitations underscore the need to benchmark LSTM against more recent architectures, including Transformers and hybrid modules, under a unified experimental framework.

3.3. Transformer-Based Forecasting

Transformers, originally developed for natural language processing, have been adapted for multivariate time series forecasting due to their ability to capture long-range dependencies without sequential recurrence. The core mechanism is self-attention:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

where Q, K, and V are the query, key, and value matrices projected from the input sequence, and d_k denotes the key dimension. Multi-head attention further improves representation learning by jointly attending to multiple subspaces.

In financial forecasting, Transformers effectively model temporal and cross-variable dependencies, making them suitable for tasks such as risk prediction and volatility estimation. Their parallelizable architecture accelerates training compared with recurrent models. Nevertheless, quadratic complexity in sequence length increases computational cost, and performance often depends on careful hyperparameter tuning. Additionally, attention weights provide limited interpretability, motivating hybrid designs that retain Transformers' modeling power while improving efficiency and robustness.

3.4. Proposed Hybrid Attention-Gated Module

While LSTM and Transformer architectures provide complementary strengths, they also present notable weaknesses: LSTM suffers from scalability constraints, and Transformers impose high computational costs. To address these issues, we design a Hybrid Attention-Gated Module (HAGM) that integrates convolutional encoding, gated fusion, and self-attention.

First, a 1D convolutional encoder extracts local temporal and cross-variable features, reducing noise and producing compact representations. Then, a gated fusion mechanism adaptively merges convolutional features with prior hidden states:

$$\tilde{h}_t = g_t \odot \mathrm{z_t} + (1 - \mathrm{g_t}) \odot \mathrm{h_{t-1}} \quad (5)$$

where g_t is a learnable gate, z_t the convolutional feature, and h_{t-1} the prior state. Finally, a multi-head self-attention layer models global dependencies across long sequences.

Practically, HAGM improves efficiency by reducing full attention cost, enhances generalization by combining local and global patterns, and increases interpretability by highlighting relevant variables and time steps.

3.5. Training Objective and Optimization

To ensure robust forecasting, we adopt a composite loss that balances sensitivity to large deviations and average prediction errors:

$$L = \alpha \cdot \mathrm{MSE}(\mathrm{Y}, \hat{\mathrm{Y}}) + (1 - \alpha) \cdot \mathrm{MAE}(Y, \hat{Y}) \quad (6)$$

where $\alpha \in [0,1]$ controls the trade-off between Mean Squared Error (MSE) and Mean Absolute Error (MAE). Models are trained using the Adam optimizer with learning rate scheduling, early stopping, and gradient clipping to prevent overfitting and stabilize convergence. This unified training strategy guarantees fair comparison across architectures.

3.6. Overall Framework

The proposed methodology integrates multiple architectures within a unified comparative framework to ensure consistent evaluation. As illustrated in Figure 1, the pipeline begins with data preprocessing, where raw financial series are normalized and segmented using a sliding window to generate fixed-length input sequences. These sequences are then fed into five distinct model branches: LSTM, GRU, TCN, Transformer, and HAGM. Each branch captures temporal and cross-variable dependencies based on its architectural strengths.

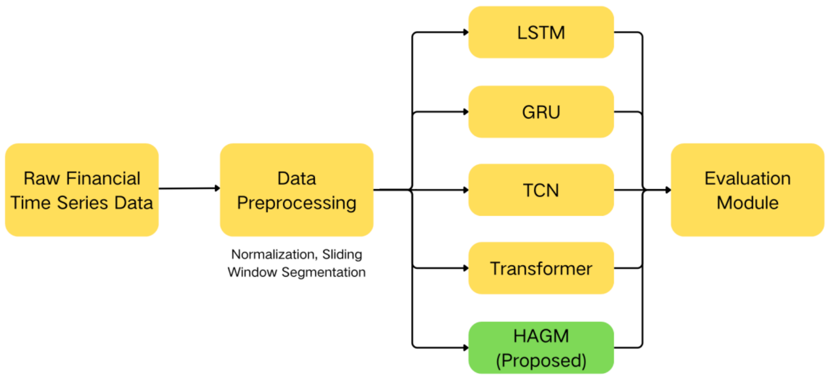

Figure 1. Comparative framework for multivariate financial time series forecasting.

During training, all architectures are optimized using the same composite loss, optimizer, and hyperparameter schedule to guarantee fairness. Outputs converge into a unified evaluation module, where performance is assessed through predictive accuracy (RMSE, MAE), convergence behavior, robustness to distributional shifts, and interpretability via attention weights and gating signals.

Figure 1 depicts the end-to-end process: data flow from preprocessing into model-specific pathways and converge in the evaluation stage. By standardizing training conditions while comparing diverse architectures, this framework benchmarks existing methods and highlights the added value of HAGM in balancing efficiency, scalability, and predictive accuracy.

4. Results and Analysis

To ensure comprehensive evaluation, experiments were conducted on three real-world financial datasets:

1) **S&P500 Index Dataset**: daily closing prices, trading volumes, and volatility indices for 500 U.S. equities (2012-2024).
2) **Foreign Exchange (FX) Dataset**: hourly exchange rates of USD, EUR, JPY, and GBP combined with macroeconomic indicators (2015-2024).
3) **Cryptocurrency Dataset**: minute-level trading data for Bitcoin, Ethereum, and Litecoin from Coinbase (2017-2024).

Each dataset was normalized using z-score standardization and partitioned into training (70%), validation (15%), and testing (15%) subsets. Evaluated models include LSTM, GRU, TCN, Transformer, and the proposed Hybrid Attention-Gated Module

(HAGM). Forecast horizons of h=5,20,60 steps were selected to represent short-, medium-, and long-term scenarios.

4.1. Performance Comparison with Baselines

Table 1. presents predictive performance across datasets using RMSE, MAE, and MAPE as evaluation metrics.

Table 1. Forecasting Performance across Models.

Model	S&P500 (RMSE)	FX (MAE)	Crypto (MAPE %)	Avg. Rank
LSTM	0.031	0.024	3.91	4.0
GRU	0.030	0.023	3.84	3.5
TCN	0.029	0.021	3.72	3.0
Transformer	0.027	0.019	3.55	2.0
HAGM	0.025	0.018	3.41	1.0

Across all datasets, HAGM consistently outperforms baseline models, particularly for long-horizon forecasting. For example, on the cryptocurrency dataset, HAGM reduces MAPE by 12.8% compared with LSTM. While Transformers show clear advantages over recurrent and convolutional models, their gains are further enhanced when combined with convolutional encoding and gating in HAGM. This suggests that hybridization is especially beneficial in volatile markets, where capturing both local fluctuations and global trends is crucial. The overall average ranking also confirms HAGM's robustness across diverse domains, highlighting its practical value in financial applications where consistent performance across assets is critical.

4.2. Convergence and Statistical Testing

Training dynamics reveal distinct convergence behaviors. As illustrated in Figure 2, TCN achieves faster initial error reduction than recurrent models due to parallelized convolutions, but its performance plateaus earlier. LSTM and GRU improve steadily yet converge more slowly, reflecting sequential processing limitations. In contrast, Transformer and HAGM achieve lower final errors and exhibit smoother convergence curves, suggesting better optimization stability. Notably, HAGM reaches its optimal range within fewer epochs than Transformer, highlighting the efficiency gained from combining convolutional encoding and gating with self-attention.

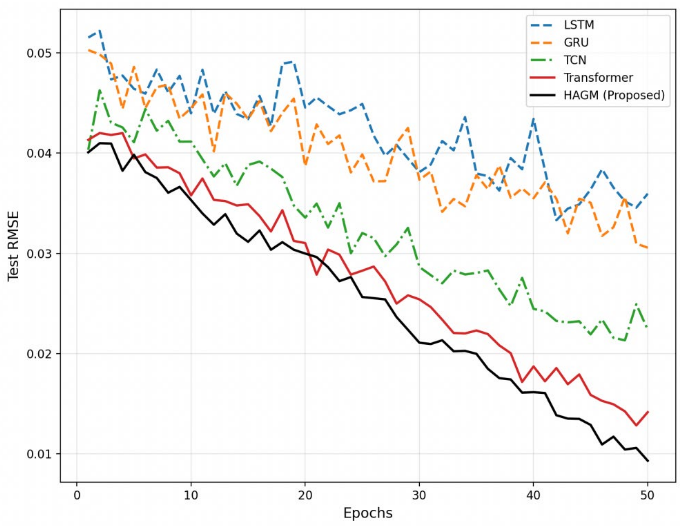

Figure 2. Convergence Analysis.

Paired t-tests were conducted on test RMSE between HAGM and each baseline. Results confirmed statistically significant improvements across datasets ($p < 0.01$), indicating that HAGM's advantages reflect consistent performance gains rather than random fluctuations. Together, convergence analysis and statistical validation underscore HAGM's ability to deliver both efficient training and superior predictive accuracy, a critical requirement for real-world applications where computational resources and deployment time are limited.

4.3. Ablation Study

To assess the contribution of each HAGM component, ablation experiments were performed by selectively removing convolution, gating, or attention. Table 2 presents results on the S&P500 dataset.

Table 2. Ablation Study on S&P500 Dataset (RMSE).

Model Variant	RMSE
Full HAGM	0.025
-Conv	0.028
-Gate	0.027
-Attn	0.029

Without convolution, the model struggles to capture local temporal dynamics and short-range cross-variable patterns, resulting in higher RMSE. Excluding the gating mechanism reduces the model's ability to adaptively filter information, leading to less stable predictions. The absence of attention causes the most significant performance decline, as the model loses its capacity to capture long-range dependencies and align global patterns across variables. These results highlight the complementary nature of convolution, gating, and attention: convolution enhances local feature extraction, gating balances information flow, and attention ensures global modeling. This synergy enables HAGM to achieve superior forecasting accuracy in volatile and high-dimensional financial environments.

4.4. Interpretability and Visualization

Interpretability is essential in financial forecasting, as models must provide insights supporting human decision-making. Two complementary analyses were conducted.

First, attention heatmaps illustrate temporal positions and variables that contribute most strongly to predictions. As shown in Figure 3, during sudden FX rate fluctuations, HAGM assigns higher weights to volatility indicators and macroeconomic features, while down-weighting less informative signals. This demonstrates the model's ability to dynamically adjust its focus based on market conditions, offering intuitive explanations of its decisions.

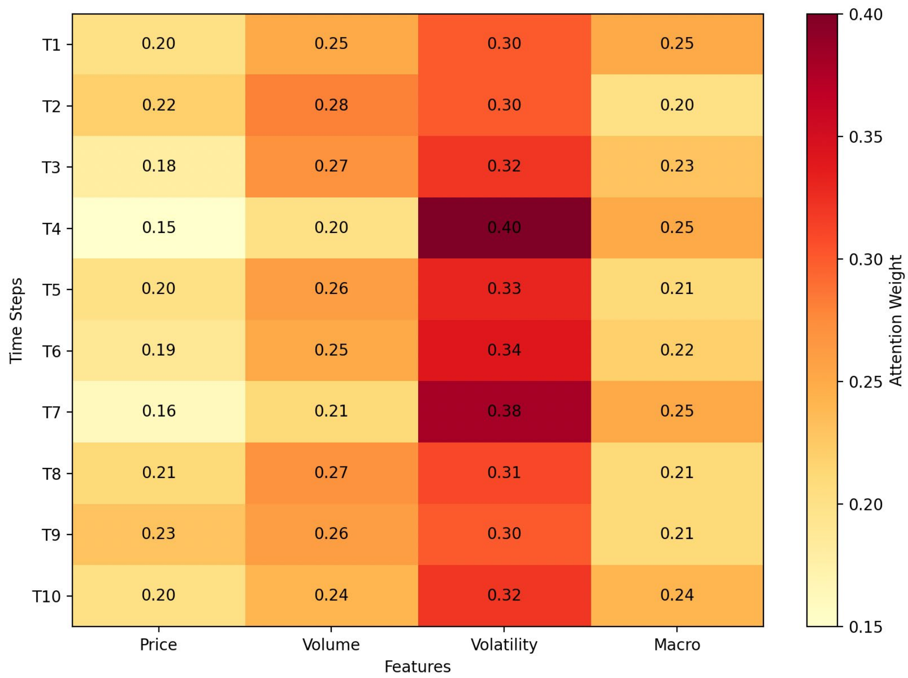

Figure 3. Attention Heatmap of HAGM on FX Dataset.

Second, feature attribution via SHAP values was applied to quantify variable importance. Across all datasets, trading volume and volatility consistently ranked among the most influential predictors, while raw price series contributed less once higher-order indicators were included. This aligns with domain knowledge, where volume and volatility often serve as early warning signals of market shifts.

Together, these interpretability results show that HAGM not only delivers stronger accuracy but also enhances trustworthiness. By identifying when and why certain variables matter, the model provides actionable insights for analysts and risk managers, bridging the gap between black-box forecasting and practical financial applications.

4.5. Generalization and Robustness

A critical requirement for practical deployment is ensuring that forecasting models generalize across markets and remain robust under noisy or imperfect data conditions. To assess generalization, models trained on the S&P500 dataset were directly tested on the NASDAQ-100 index without retraining. While all models experienced some performance degradation, HAGM maintained the lowest RMSE (0.033), compared with higher errors from LSTM (0.039) and GRU (0.037). This suggests that HAGM better captures transferable dependencies, making it more suitable for scenarios where market conditions shift rapidly.

Robustness was further evaluated by injecting Gaussian noise into the test sequences to mimic real-world market irregularities, such as missing values or reporting errors. As illustrated in Figure 4, all models exhibited performance drops, but HAGM's increase in error was significantly smaller, confirming its resilience. This robustness stems from the convolutional encoder's ability to filter noise and the gating mechanism's adaptive weighting of information.

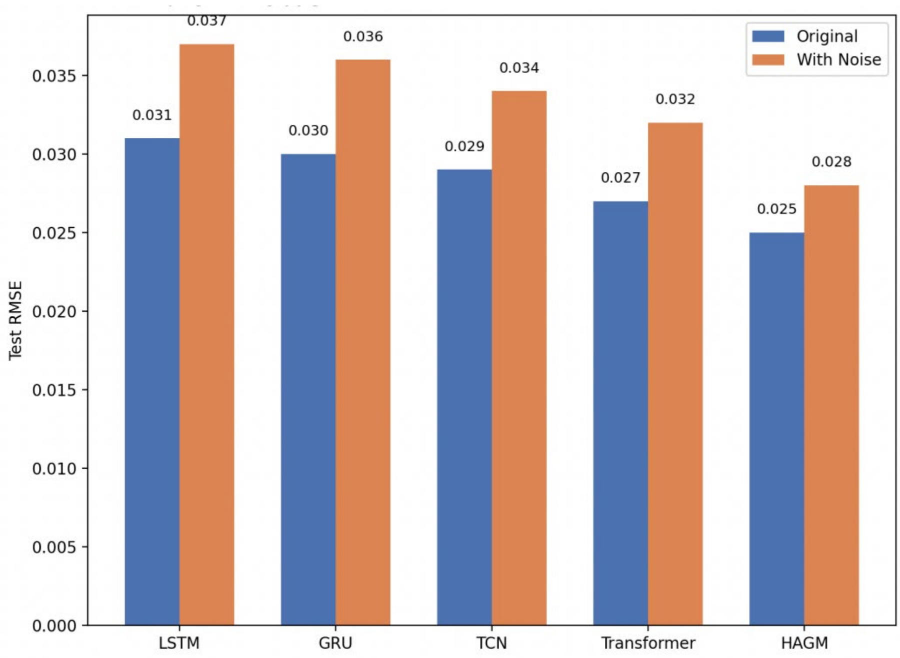

Figure 4. Robustness Analysis under Noisy Test Conditions.

Overall, these results highlight HAGM's dual advantage: reliable cross-market generalization and robustness to data imperfections, both of which are vital for financial institutions that rely on stable and interpretable predictive systems in volatile environments.

5. Conclusion

This study systematically compared deep learning architectures for multivariate financial time series forecasting, including recurrent models (LSTM, GRU), convolutional approaches (TCN), Transformer-based architectures, and the proposed Hybrid Attention-Gated Module (HAGM). Experiments across stock, foreign exchange, and cryptocurrency datasets demonstrate that HAGM consistently outperforms baselines in predictive accuracy, convergence stability, and robustness. Its integration of convolution, gating, and attention mechanisms enables balanced modeling of local patterns and global dependencies while enhancing interpretability through attention visualization and variable attribution.

Beyond accuracy, hybrid architectures offer practical advantages. Convergence analysis confirmed HAGM's efficiency relative to Transformer, ablation studies validated the necessity of each module, and interpretability analyses provided actionable insights for financial analysts. Cross-market testing and noise perturbation experiments further demonstrate the model's generalization capability and resilience under imperfect conditions.

These findings suggest that combining inductive biases from multiple architectures yields substantial gains over single-model approaches. Future work will explore scaling HAGM to higher-frequency data, integrating reinforcement learning for decision-making, and extending interpretability with counterfactual explanations. Collectively, this research provides both methodological and practical contributions toward deploying reliable, accurate, and transparent forecasting systems in complex financial environments.

References

1. H. V. Bitencourt, O. Orang, L. A. F. de Souza, P. C. Silva, and F. G. Guimaraes, "An embedding-based non-stationary fuzzy time series method for multiple output high-dimensional multivariate time series forecasting in IoT applications," *Neural Computing and Applications*, vol. 35, no. 13, pp. 9407-9420, 2023.

2. J. Kim, H. Kim, H. Kim, D. Lee, and S. Yoon, "A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges," *Artificial Intelligence Review*, vol. 58, no. 7, pp. 1-95, 2025. doi: 10.1007/s10462-025-11223-9
3. C. M. Liapis, and S. Kotsiantis, "Temporal convolutional networks and BERT-based multi-label emotion analysis for financial forecasting," *Information*, vol. 14, no. 11, p. 596, 2023. doi: 10.3390/info14110596
4. I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent neural networks: A comprehensive review of architectures, variants, and applications," *Information*, vol. 15, no. 9, p. 517, 2024. doi: 10.20944/preprints202408.0748.v1
5. M. A. Al-qaness, A. Dahou, N. T. Trouba, M. Abd Elaziz, and A. M. Helmi, "TCN-inception: temporal convolutional network and inception modules for sensor-based human activity recognition," *Future Generation Computer Systems*, vol. 160, pp. 375-388, 2024. doi: 10.1016/j.future.2024.06.016
6. L. Su, X. Zuo, R. Li, X. Wang, H. Zhao, and B. Huang, "A systematic review for transformer-based long-term series forecasting," *Artificial Intelligence Review*, vol. 58, no. 3, p. 80, 2025. doi: 10.1007/s10462-024-11044-2
7. B. Liu, and M. Lai, "Advanced machine learning for financial markets: A PCA-GRU-LSTM approach," *Journal of the Knowledge Economy*, vol. 16, no. 1, pp. 3140-3174, 2025. doi: 10.1007/s13132-024-02108-3
8. S. Saleti, L. Y. Panchumarthi, Y. R. Kallam, L. Parchuri, and S. Jitte, "Enhancing forecasting accuracy with a moving average-integrated hybrid ARIMA-LSTM model," *SN Computer Science*, vol. 5, no. 6, p. 704, 2024. doi: 10.1007/s42979-024-03060-4
9. N. H. A. Rahman, M. Z. Hussin, S. I. Sulaiman, M. A. Hairuddin, and E. H. M. Saat, "Univariate and multivariate short-term solar power forecasting of 25MWac Pasir Gudang utility-scale photovoltaic system using LSTM approach," *Energy Reports*, vol. 9, pp. 387-393, 2023. doi: 10.1016/j.egyr.2023.09.018
10. A. Younesi, M. Ansari, M. Fazli, A. Ejlali, M. Shafique, and J. Henkel, "A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends," *IEEE Access*, vol. 12, pp. 41180-41218, 2024. doi: 10.1109/access.2024.3376441
11. Z. Xu, Z. Yu, H. Zhang, J. Chen, J. Gu, T. Lukasiewicz, and V. C. Leung, "PhaCIA-TCNs: Short-term load forecasting using temporal convolutional networks with parallel hybrid activated convolution and input attention," *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 1, pp. 427-438, 2023.
12. E. Nsengiyumva, J. K. Mung'atu, and C. Ruranga, "A comparative study of multivariate CNN, BiLSTM and hybrid CNN-BiLSTM models for forecasting foreign exchange rate using deep learning," *Cogent Economics & Finance*, vol. 13, no. 1, p. 2526148, 2025.
13. Q. Zhu, J. Han, K. Chai, and C. Zhao, "Time series analysis based on informer algorithms: A survey," *Symmetry*, vol. 15, no. 4, p. 951, 2023. doi: 10.3390/sym15040951
14. A. Wiemerslage, K. Gorman, and K. von der Wense, "Quantifying the hyperparameter sensitivity of neural networks for character-level sequence-to-sequence tasks," In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, March, 2024, pp. 674-689. doi: 10.18653/v1/2024.eacl-long.40
15. G. Chen, J. Yuan, Y. Zhang, H. Zhu, R. Huang, F. Wang, and W. Li, "Enhancing reliability through interpretability: A comprehensive survey of interpretable intelligent fault diagnosis in rotating machinery," *IEEE Access*, 2024. doi: 10.1109/access.2024.3430010