Dean&Francis
ISSN 2959-6157

Research on Stock Price Prediction Based on a Hybrid LSTM-Transformer Model

Bingzhang Wang

College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China
202211081016@sdust.edu.cn

Abstract:
The goal of this paper is to establish a hybrid Long Short-Term Memory (LSTM)-Transformer model to achieve stock price prediction. By comparing the baseline LSTM model to the hybrid model, it is evident that the hybrid model outperforms the baseline in stock price prediction. It is vital for achieving optimal capital allocation, effective risk management, and successful execution. This model is based on next-day log return regression with return to price reconstruction, trained on daily OHLCV (data of stock about OPEN, HIGH, LOW, CLOSE, and VOLUME) features from 2021 to 2023 and evaluated in the first quarter of 2024 (Q1 2024). During the testing period, the mean absolute error (MAE) drops from 10.43 to 1.58, the root mean square error (RMSE) from 12.34 to 2.33, and the Price R^2 rises from 0.25 to 0.97. Directional Accuracy improves from 60.66% to 65.57%, and on strong move days, it improves from 59.57% to 65.96%. The above results demonstrate that the hybrid LSTM-Transformer model outperforms the LSTM baseline model in terms of stock price prediction. Residual diagnostics reveal a mean close to zero and heavy tails with occasional spikes, which is consistent with event-driven jumps that cannot be anticipated using end-of-day data alone. These results support the use of the hybrid model for daily equity prediction and point out directions for future improvement, such as adding the process of volatility normalization, considering explicit event or gap features, and adopting tail robust or quantile objectives.

Keywords: Long short-term memory (LSTM); Transformer model; Adjusted Close Price.

1. Introduction

As a core component of the financial sector, the stock market has always attracted widespread attention [1]. In recent years, rapid advances in artificial intelligence and deep learning have expanded financial forecasting, especially stock prediction. Deep models, from Convolutional Neural Networks (CNNs)

ISSN 2959-6157

to Recurrent Neural Networks (RNNs), capture market nonlinearities and achieve strong index prediction, underscoring the role of AI and offering exceptional innovations in the world of commerce in trade [2,3]. Within this context, this study focuses on NVIDIA, which is a large, liquid, AI-linked mega-cap that is widely followed and exhibits event-driven volatility. To evaluate generalization and avoid overfitting or temporal leakage, this model was trained on 2021–2023. The training span covers distinct regimes, including post-pandemic recovery, tightening-led drawdowns, and the 2023 AI rally, enabling the model to learn across heterogeneous conditions. The 2024 Q1 out-of-sample period then assesses the model's robustness to regime shifts and validates its performance in a recent market environment, using identical preprocessing and evaluation methods.

Recently, a great deal of interesting work has been done in applying machine learning to analyze price patterns and predict stock and index changes. For instance, one study applied and compared an Autoregressive Integrated Moving Average Model (ARIMA model) to obtain an accurate stock forecasting model, highlighting the potential of ARIMA. Another study used CNN models to identify and predict stock price trends and offered valuable guidance for forecasting. Further research implemented single and multilayer LSTM architectures and evaluated them with various metrics to determine the optimal design. Subsequent work showed that a neural network combining convolutional units and long short-term memory outperformed statistical methods and traditional CNNs and LSTMs in prediction tasks. More recent work developed a Graph Convolutional Networks (GCN)-based stock prediction model to address limitations in inter-stock relationship modeling and examined how correlations in trading data affect predictions [4-8].

Synthesizing prior work, Classical statistical models such as auto ARIMA emphasize parsimony and interpretability, yet struggle with nonlinearity and regime shifts. CNN-based learners identify local price shapes and micro patterns, but have limited access to long-range temporal dependencies. Pure recurrent approaches with single and multilayer LSTM capture sequential memory and nonlinear dynamics, though they may lag at turning points and under volatile conditions. Hybrid CNN-LSTM frameworks combine spatial pattern extraction with temporal memory and often outperform single-family baselines, while their receptive fields remain fixed and may miss distant interactions. Graph-based GCN models encode cross-asset relations and market structure, but rely on stable graph construction and can propagate noise if correlations are transient [4-8]. This study proposes a hybrid LSTM-Transformer model to complement these methods by combining local order modelling with attention-based long-range aggregation within a unified, leakage-free protocol. Thereby addressing the main gaps identified across categories and achieving more accurate predictions.

This study develops a combined model for stock price prediction. The model combines the sequence learning power of LSTM networks with the global attention mechanism of the Transformer model. This approach demonstrates higher predictive Accuracy in comparison to single model methodology while investigating the interpretability aspect of this hybrid framework and its capability to capture complex market dynamics. As they can capture long-term dependencies and preserve the temporal structure of the sequential data, LSTM networks are extensively applied to time-series modelling. However, the single LSTM model has difficulty learning global dependencies efficiently, especially for such large and heterogeneous datasets. On the contrary, the Transformer model is good at capturing long-range dependencies with its self-attention mechanism. Therefore, the hybrid LSTM-Transformer model combines the advantages of both methods and provides a more comprehensive modeling framework for stock price prediction.

Simultaneously, this research employs deep learning to propose some new insights into financial forecasting. It shows that the hybrid neural architectures can capture long-term dependencies and complex nonlinear relationships in stock market data. By combining LSTM and Transformer models, this work demonstrates that using sequential memory in combination with the concept of attention could obtain an enhanced accuracy in predicting future observations. As a result, it has theoretical contributions for financial data modeling. Not only does it have practical implications for investors, analysts, and regulators, but it can also provide a tool for decision makers in a volatile and uncertain economic environment. Future research may also focus on cross-market applicability, explainability methods for hybrid deep learning models, and real-time financial applications. This would further extend the scope and influence of hybrid deep learning in financial technology.

2. Methodology

2.1 Data Sources and Description

The data for this study were obtained from Kaggle, a prominent online platform for data science resources. NVIDIA's historical stock prices and related data were selected as the Dataset used in this article. Stock price data from January 1, 2021, to December 31, 2023, was chosen as the training set for modeling and used to predict and verify stock price trends for the first quarter of 2024 [9].

2.2 Indicator Selection and Rationale

This study used daily NVIDIA OHLCV data after 2021-01-01, with specific indicators detailed in Table 1, yielding 870 trading records and seven raw fields with zero missing values and no duplicate dates. After constructing technical factors and aligning the next-day log-return target, early warm-up rows required by rolling windows are removed, resulting in 850 effective observations, which implies a 2.30% reduction. The indicator set is designed to balance stationarity, trend, volatility, and microstructure signals. The core target is the next-day log return to improve stationarity.

Table 1. Description of Indicator Symbols

Name of the Indicator	Symbols of the Indicator	Explanation of indicators
Adj Close	P_t	Adjusted Closing Price on day t
Open	O_t	Opening Price on day t
High	H_t	High Price on day t
Low	L_t	Low Price on day t
Volume	V_t	Trading Volume on day t
Moving Average	MA_n	n-day simple moving average
Exponential Moving Average	EMA_n	n-day exponential moving average
Rolling Standard Deviation	$\sigma_{n,t}$	n-day rolling standard deviation at t
Positive Constant	ϵ	a small positive constant (e.g., 10^{-8})

2.3 Method Introduction

This study formulates one-step forecasting as a regression on the next-day log return. The baseline model is an LSTM that captures local sequential dependence and short-horizon momentum. To improve the performance of the basic model, an LSTM-Transformer hybrid model was established.

The LSTM–Transformer hybrid improves upon the LSTM representation by incorporating multi-head self-attention, which enables the adaptive integration of long-range lags and cross-temporal interactions. This enhances the model's ability to detect trend continuation.

2.3.1 Long Short-Term Memory (LSTM)

An LSTM model maintains two states: a long-term memory and a short-term hidden state. The core structure of LSTM includes three main gates: the forget gate, the output gate, and the input gate. This gated design prevents vanishing gradients and enables the network to learn patterns over multiple horizons. For financial data, LSTMs capture local sequential dependence, short-term momentum, mean reversion, and mild seasonality, while filtering out noise. They are easy to train and provide a robust foundation for predicting one-step-ahead next-day returns. Through these mechanisms, LSTM can not only remember important information but also forget irrelevant details when necessary, thereby improving the model‘s performance in complex tasks.

2.3.2 Transformer Model

Compared to LSTM, Transformer allows all time features to be learned in parallel, reducing performance degradation caused by long-term dependencies and accommodating feature memory for the entire sequence [10]. It computes pairwise interactions across time steps to model long-range dependencies. With queries, keys, and values (Q, K, V), scaled dot-product attention is shown in equation (1). The Transformer model could use the self-attention mechanism to effectively capture long-range dependencies in sequences. Compared with purely recurrent propagation, attention adaptively aggregates cross-temporal information and better exploits distant yet informative signals commonly present in financial time series.

$$Atteneion(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \qquad (1)$$

2.3.3 Model assumptions

(1) This study does not consider macroeconomic factors, news, and intraday data.

(2) The ultimate prediction goal of this study is the Adjusted Close Price.

(3) Assuming that the Adjusted Close Price properly ac-

ISSN 2959-6157

counts for splits or dividends;

(4) Assuming that most predictive information resides in the past 60 trading days of Price and volume.

3. Result

3.1 Data Preprocessing

This study utilizes NVIDIA's daily stock price data from January 2021 to March 2024. The hybrid model's direct prediction target is the next day's log return of the adjusted closing price. This design focuses the model on learning the relatively stationary and more model-friendly return series. For the final evaluation, the predicted returns r_t are inverse-transformed into concrete price values P_t by using equation (2). This allows for a comprehensive assessment of the model's predictive performance on a more intuitive and practically significant price scale.

$$P_t = P_{t-1} \bullet e^{r_t}. \tag{2}$$

Feature engineering in this study constructs a set of indicators from open-source data to capture market dynamics, as shown in Table 2.

Table 2. Indicator Formula and Description

Symbols of Indicators	Equation of the Indicator	Explanation of indicators
r_t	$r_t = lnP_t - lnP_{t-1}$	Daily log return
HL_t	$HL_t = \frac{H_t - L_t}{P_t + \epsilon}$	High--Low range
$OCGap_t$	$OCGap_t = \frac{P_t - O_t}{O_t + \epsilon}$	Open--Close gap
$\Delta\ell V_t$	$\Delta\ell V_t = ln(V_t + 1) - ln(V_{t-1} + 1)$	Volume change
$AG_{14,t}$	$AG_{14,t} = EMA_{14}\left(max\left(P_t - P_{t-1}, 0\right)\right)$	intermediate indicator
$AL_{14,t}$	$AL_{14,t} = EMA_{14}\left(max\left(P_{t-1} - P_t, 0\right)\right)$	intermediate indicator
$RSI_{14,t}$	$RSI_{14,t} = 100 - \frac{100}{1 + \frac{AG_{14,t}}{(AL_{14,t} + \epsilon)}}$	14 Days Relative Strength Index
$MACD_t$	$MACD_t = EMA_{12}\left(P_t\right) - EMA_{26}\left(P_t\right)$	Moving average convergence/divergence
$Signal_t$	$Signal_t = EMA_9\left(MACD_t\right)$	MACD signal line
$Hist_t$	$Hist_t = MACD_t - Signal_t$	MACD histogram
TR_t	$TR_t = max\{H_t - L_t, \lvert H_t - P_{t-1}\rvert, \lvert L_t - P_{t-1}\rvert\}$	intermediate indicator
$nATR_{14,t}$	$nATR_{14,t} = \frac{MA_{14}\left(TR_t\right)}{P_t + \epsilon}$	Average True Range
$BW_{20,t}$	$BW_{20,t} = \frac{\sigma_{20,t}}{MA_{20}\left(P_t\right) + \epsilon}$	Bollinger Bands

After missing values have been handled, a Standard Scaler is fitted exclusively to the pre-2024 training set to standardize the features and the target. Finally, sequential samples for model input are generated using sliding windows (lookback = 60).

3.2 LSTM-Transformer hybrid model

The architecture of the model was implemented using TensorFlow and Keras.

3.2.1 Model Design

This study formed each sample as a 60×14 window, which contained sixty trading days and fourteen features. First, a Dense layer projects features to 64 dimensions and adds a positional code so that the model knows the order. An initial LSTM extracts short-term patterns; the sequence then goes through a Transformer encoder (multi-head self-attention with residuals, normalization, and a feed-forward block) to gather longer-range signals. A second LSTM then compresses the sequence into a single context vector. This passes through a Dropout layer and a 32-unit dense layer before being converted into the final output: the standardized next-day log return. Then convert this back to Price using equation (2). Training uses Huber loss for robustness, Adam optimization, and automatic learning-rate reduction on plateaus, as well as early stopping. Scalers are fitted on training data only; the last 10% of training windows form the validation set, and the test period is Q1 2024. This allows for a fair error comparison with the baseline LSTM.

3.2.2 Evaluation metrics

This study evaluated these two models in both the price domain and the return domain. Price metrics include MAE, RMSE, and R^2. And this study reported two directional measures for returns. These were Directional Accuracy and Directional Accuracy on strong moves. A lower value is better for MSE, RMSE, and MAE, while a higher value is better for R, directional Accuracy, and the 1% tolerance rate. By comparing and analysing the above errors, the differences can be identified between the mixed and baseline models. Based on this analysis, a higher-performing model for the stock price prediction could be selected.

3.3 Result

3.3.1 Baseline vs Hybrid

According to the content of Table 3 and Fig. 1, the LSTM–Transformer hybrid clearly outperforms the baseline LSTM on the 2024 Q1 test set. To be more specific about the part of the price domain, the hybrid model attains RMSE ≈ 2.33, MAE ≈ 1.58, whereas the baseline LSTM model yields RMSE ≈ 12.34, MAE ≈ 10.43. And the Price R^2 improves from 0.25 (baseline) to 0.97 (hybrid), indicating that the hybrid explains nearly all price variation.

Then, with regard to the aspect of return domain, it shows that the directional Accuracy of the hybrid model improved by approximately 4.9% compared to the baseline model. And on strong-move days with $r_t \geqslant 0.5\%$ also improved by approximately 6.4% compared to the baseline model.

All in all, the overlaid plot in Fig. 1 shows the stock price trends predicted by these two models. It provides a more intuitive illustration of these two models' performance. And the overlaid plot shows that the baseline underestimates levels and lags turning points, whereas the hybrid closely tracks the sharp rally in February–March 2024.

Table 3. Value of Evaluation Metrics

Evaluation metrics	LSTM	LSTM+Transformer
MAE	10.43	1.58
RMSE	12.34	2.33
R^2	0.25	0.97
Directional Accuracy	60.66%	65.57%
Directional Accuracy ($\lvert r_t \rvert > 0.5\%$)	59.57%	65.96%

ISSN 2959-6157

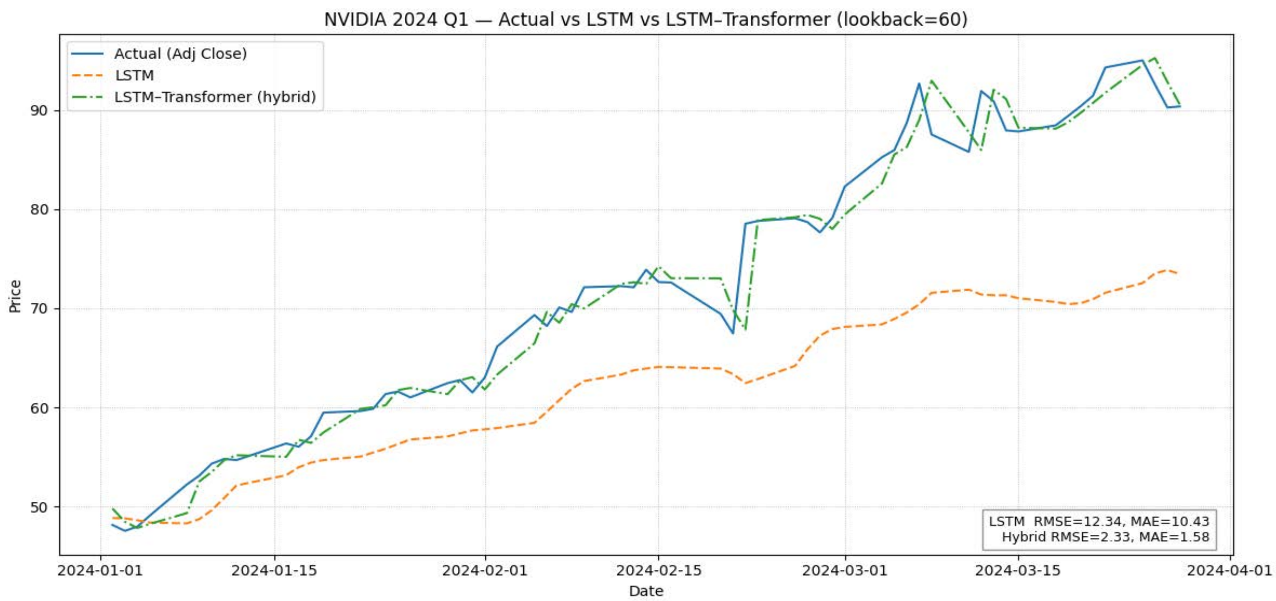

Fig. 1 Actual vs LSTM vs LSTM-Transformer (lookback = 60 days) (Picture credit: Original)

3.3.2 Diagnostic Analysis of the Hybrid Model

For the hybrid model, Fig. 2 has a mean value approximately equal to 0.35, and the Standard Deviation approximately equal to 2.3. It shows a low average relative price level and implies negligible bias relative to the price level. For Fig. 3, the QQ-Plot of Price Residuals, it shows that most points fit on the red line (y=x). However, the QQ-plot curves upwards at the extremes, indicating heavy outliers. This phenomenon also appears in Fig. 4. This deviance usually corresponds to significant one-day price fluctuations that are usually news-related, such as earnings releases, changes in guidance, regulatory or macro announcements, and product news.

For Fig. 4, most residuals cluster around zero without long drifts, so the model adheres well to sustained trends but is challenged by rare, large moves.

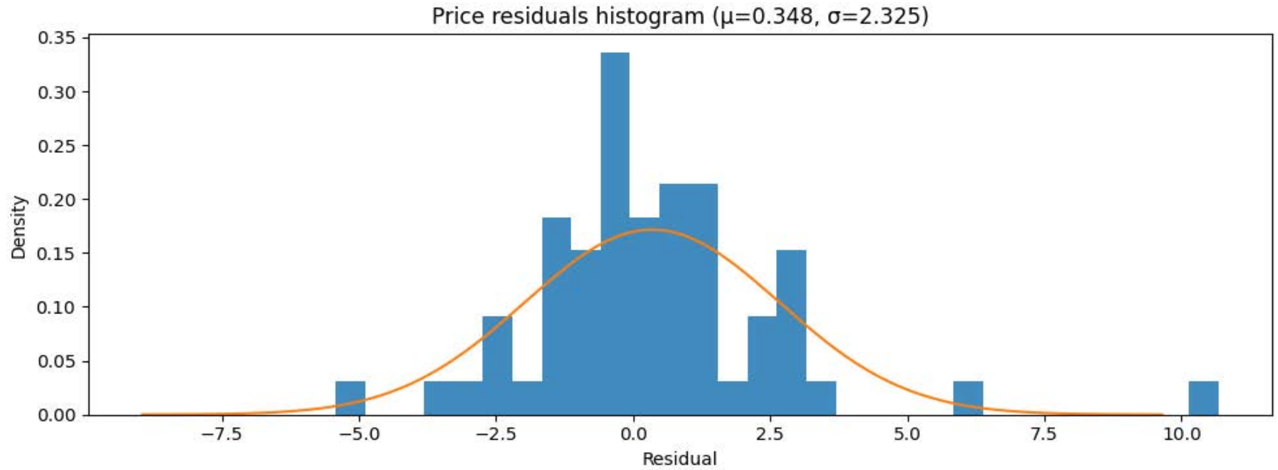

Fig. 2 Price Residuals histogram ($\mu = 0.348, \sigma = 2.325$) (Picture credit: Original)

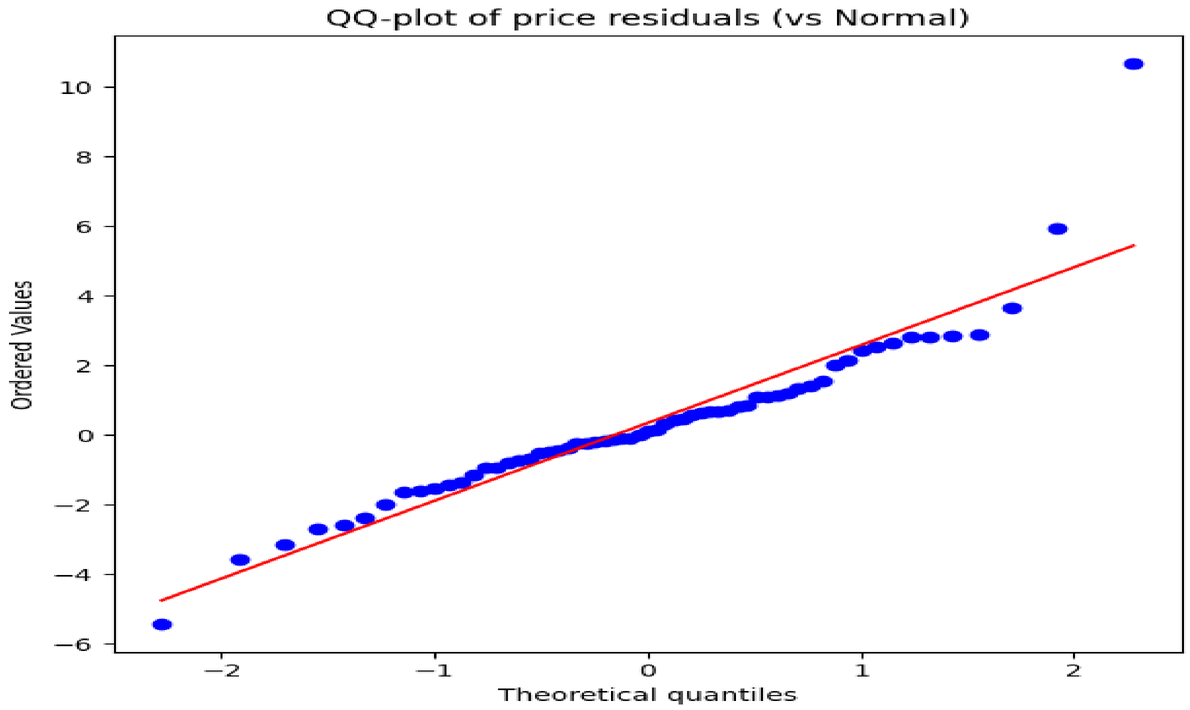

Fig. 3 QQ-Plot of Price Residuals (Picture credit: Original)

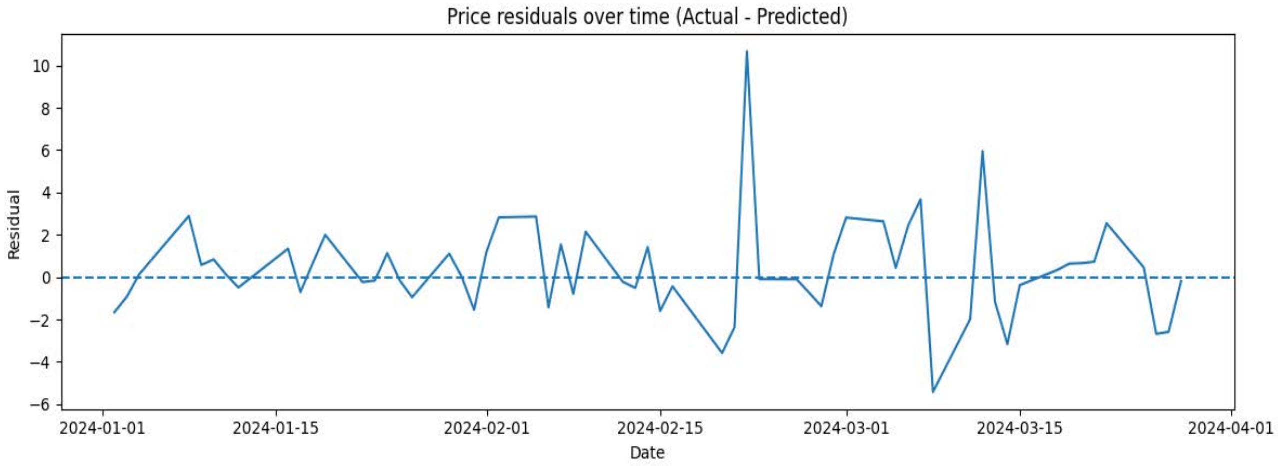

Fig. 4 Price Residuals Over Time (Picture credit: Original)

4. Discussion

Using identical splits, window lengths, and evaluation protocols, the LSTM–Transformer hybrid model consistently outperforms the baseline LSTM in terms of both price errors and directional measures. Also, multivariate technical factors supply momentum and volatility context that a univariate baseline lacks. In the overlaid plots, the hybrid model tracks trend accelerations and slowdowns more closely, while the baseline tends to lag and undershoot around turning points. Compared with yesterday's Price, the hybrid model also shows better adherence in regime changes, indicating that attention helps reweight distant yet informative lags rather than merely smoothing noise.

Diagnostics figures clarify where performance is won and where it remains vulnerable. And the diagnostics figure shows residuals centred near zero without long drifts. It suggests there is no systematic bias within the test window. However, the histogram and QQ-plot reveal heavy tails, and the residual time series shows spikes around dates adjacent to earnings. It shows that the hybrid model is conservative in the event of sharp increases and may

ISSN 2959-6157

overshoot during pullbacks, which is consistent with the use of daily stock data without the modelling of explicit news or earnings.

The hybrid model balances local sequencing and long-range dependence, leverages multivariate technical context, and improves trend tracking and turning-point responsiveness under a unified, fair comparison. Future work can be promoted in some ways. Firstly, to boost the model's performance, it could incorporate exogenous signals, such as earnings windows, event windows, and gap strength. After that, it could be applied to volatility normalization and quantile or heavy-tail losses for more robust tails and interval forecasts. Last but not least, it could be scaled up to include more tickers and longer samples, and feature attribution could be added for interpretability and stress testing.

5. Conclusion

In this research, the LSTM baseline model was compared to a hybrid LSTM-Transformer model in the task of forecasting the NVIDIA Stock Price. These two models maintained data splits, lookback preprocessing, and evaluation unchanged and avoided data leakage. And the hybrid model had complementary roles in predicting the NVIDIA Stock Price compared to the baseline model. Firstly, LSTM extracts short horizon order and denoises the sequence. The Transformer encoder aggregates long-range cues through self-attention. Across the same test window, the hybrid delivered lower price errors and stronger directional performance. The overlaid curves demonstrate more precise tracking during trend accelerations and decelerations. These improvements suggest that attention enhances the signal beyond smoothing and assists in the reweighting of distant yet informative lags. So the hybrid LSTM-Transformer model has a better performance in predicting stock prices.

In reality, the results lead to the fact that the sequential hybrid mode works very well under only end-of-day information. In terms of methodology, the work focuses on fair baselines, tight data alignment, and knowledge return to price reduction. Future work also includes probability prediction, decision use, and volatility, which can be modeled explicitly. Adding prior intraday returns or a range-based estimate of the realised volatility model, like parkinson or Garman-Klass. And then scale targets and features by performing a dynamic rolling of variance or ATR. To address this objective, training can be performed as either quantile loss or pinball loss, and learn systems with calibrated intervals. By using these loss terms, the multi-horizon prediction can be modeled to predict multiple steps. Lastly, add regime detectors to update hyperparameters as market state changes. Report statistical significance with the Diebold-Mariano test, report trading realism with slippage, and cost awareness backtests.

References

[1] Zhou Jingwen. Stock forecasting based on the ARIMA model -- taking the Bank of China as an example [J]. Modern Computer, 2024, 30(14): 89–92.

[2] Wang Chaojie, Chen Yuanyuan, Zhang Shuqi, Zhang Qiuhui. Stock market index prediction using a deep Transformer model. Expert Systems with Applications, 2022, 208, 118128. DOI: 10.1016/j.eswa.2022.118128.

[3] Jaimin Shah, Darsh Vaidya, Manan Shah. A comprehensive review of multiple hybrid deep learning approaches for stock prediction. Intelligent Systems with Applications, 2022, 16, 200111. DOI: 10.1016/j.iswa.2022.200111.

[4] Lounnapha Sayavong, Wu Zhongdong, Sookasame Chalita. Research on a stock price prediction method based on a convolutional neural network. 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Jishou, China, Sept 14–15, 2019. IEEE, 2019: 173–176. DOI: 10.1109/ICVRIS.2019.00050.

[5] Shakir Khan and Hela Alghulaiakh. ARIMA Model for Accurate Time Series Stocks Forecasting. International Journal of Advanced Computer Science and Applications (IJACSA), 2020, 11(7). DOI: 10.14569/IJACSA.2020.0110728.

[6] Hum Nath Bhandari, Binod Rimal, Nawa Raj Pokhrel, et al. Predicting the stock market index using LSTM. Machine Learning with Applications, 2022, 9, 100320. DOI: 10.1016/j.mlwa.2022.100320.

[7] Wu Mingtai, Li Zhongcui, Norbert Herencsar, et al. A graph-based CNN-LSTM stock price prediction algorithm with leading indicators. Multimedia Systems, 2023, 29(3), 1751-1770. DOI: 10.1007/s00530-021-00758-w.

[8] Ma Dongbo. The Research on Stock Price Prediction Method Based on Graph Convolutional Networks. Master's thesis, Shandong Technology and Business University, 2024. Available: CNKI database, Accession No. 10.27903/d.cnki.gsdsg.2024.000081.

[9] Dawood Muhammad. NVIDIA stock data [Dataset]. Kaggle, 2023. Available from: https://www.kaggle.com/datasets/muhammaddawood42/nvidia-stock-data.

[10] Ruan Junhao, Wang Kang, Feng Tianyun, et al. A Transformer-based Stock Forecasting Method [J]. Computer & Digital Engineering, 2025, 53(05),1375-1380.