Int. J. of Management, Entrepreneurship, Social Science and Humanities, Vol. 10 No. 1 (2026) https://doi.org/10.31098/ijmesh.v10i1.4281

Research Paper

Multi-Branch Transformers for Stock Market Prediction using Previous Market Data and News Articles

Lamia Rahim Laibi*
University of Baghdad, Iraq

Received : February 6, 2026	Revised : April 22, 2026	Accepted : May 20, 2026	Online : June 12, 2026

Abstract

The stock market is a complex and dynamic system influenced by numerous factors, including technical indicators, financial news, and long-term historical price movements, among others. Understanding and accurately forecasting market behavior requires integrating diverse data sources and identifying underlying trends across multiple modalities. The primary objective of this research is to demonstrate the advantages of employing multi-branch Transformer architectures for managing multimodal financial data and to evaluate the model's effectiveness in predicting stock market trends over short-, medium-, and long-term horizons. In this study, we investigate the implementation of a multi-branch Transformer model designed to forecast stock market prices by integrating multiple data sources, such as news articles and historical market data, over extended periods. The proposed architecture comprises two main branches: the first is a BERT-based Transformer that processes textual information related to daily stock performance, while the second is an LSTM-based neural network that analyzes long-term historical price data. After the feature extraction and processing stages, the outputs from both branches are fused through dedicated layers to enable highly accurate and efficient stock price predictions. Leveraging advanced artificial intelligence, particularly deep Transformer architectures, the proposed multi-branch model processes heterogeneous financial data simultaneously, significantly improving forecasting accuracy and predictive capability. The model achieves a mean square error (MSE) of 6×10^{-4}, demonstrating its strong performance and minimal loss value. This study underscores the potential of multi-branch Transformer architectures to seamlessly integrate textual and numerical financial information, offering a robust and advanced framework for stock market prediction and trend analysis. The proposed approach relies on large-scale datasets, which pose challenges related to data quality, accessibility, and processing efficiency. Furthermore, the model's substantial computational requirements may limit its practicality for small organizations or institutions with constrained resources.

Keywords: *Stock Market Prediction, Deep Learning, Multi-Branch Transformers, Artificial Intelligence, Multimodal Learning, Financial Time Series*

INTRODUCTION

In recent years, financial market data has become increasingly difficult to analyze due to its growing complexity and the strong interactions among multiple influencing factors (Poon & Granger, 2003). This complexity has made it challenging for analysts and researchers to accurately model and predict stock market movements. In earlier approaches, stock price prediction primarily relied on statistical tools, basic computational systems, and relatively simple mathematical models that focused mainly on historical price data (Fathali et al., 2022). As a result, the accuracy and robustness of these traditional methods were limited, partly because they failed to incorporate external variables such as macroeconomic indicators, news events, and other influential factors (Aldujaili et al., 2024).

In the last decade, and particularly in recent years, artificial intelligence techniques have gained significant attention in financial forecasting, especially deep learning and Transformer-based architectures. These approaches have demonstrated strong capabilities in capturing

Corresponding author's email: lamyaa.r@uobaghdad.edu.iq

nonlinear relationships and long-term dependencies in financial time series data, which are difficult to model using conventional statistical methods. Moreover, empirical studies have shown that integrating heterogeneous data sources—such as historical prices and financial news—can significantly improve predictive performance compared to single-source models, often leading to more robust forecasting results.

The main objective of this study is to leverage the capabilities of deep learning and multi-branch Transformer architectures to enhance the accuracy of financial market prediction by jointly analyzing historical time-series data and real-time news information (Boppiniti, 2024). This research proposes a multi-branch modeling framework that integrates diverse data sources, including time-series price data, technical indicators, and textual financial news over extended historical periods. This approach enables a more comprehensive understanding of market dynamics and aims to improve predictive accuracy by capturing both numerical patterns and sentiment-driven signals.

To guide this study, the following research objectives are formulated: (1) to develop a multi-branch Transformer-based model that integrates historical market data and financial news; (2) to examine the effectiveness of combining textual and numerical data for stock market prediction; (3) to evaluate the model's performance across short-, medium-, and long-term forecasting horizons; and (4) to compare the proposed model with traditional and single-branch deep learning approaches using standard error metrics and prediction accuracy. In addition, this study addresses the following research questions: How effectively can a multi-branch architecture capture dependencies among heterogeneous financial data sources? And to what extent does the integration of news-based features improve stock price prediction compared to using historical data alone?

The importance of forecasting in financial markets is widely recognized, as it serves as a critical tool for investors, analysts, and regulators in understanding market fluctuations. Accurate prediction models support investment decision-making, improve risk management strategies, and enhance overall financial planning. As financial markets continue to evolve in complexity, the demand for efficient and reliable forecasting systems that integrate multiple data sources has increased significantly (Poon & Granger, 2003). With advancements in deep learning techniques, financial analysis and forecasting can be further improved, offering more efficient and accurate decision-support tools for market participants.

Research gap

Despite the progress reported in the literature, several limitations remain. Traditional statistical models such as ARIMA and GARCH are primarily based on univariate or linear assumptions, and therefore are unable to effectively capture the complex nonlinear relationships inherent in financial markets (Lama et al., 2015). Although machine learning and deep learning approaches, particularly LSTM networks and ensemble methods, have demonstrated improved predictive performance, many of these models still rely heavily on historical price data and technical indicators. Consequently, unstructured data sources such as financial news and market sentiment are often underutilized or insufficiently integrated (Fathali et al., 2022; Ahmed et al., 2020).

Moreover, existing multimodal approaches frequently process different data modalities separately or employ relatively simple fusion strategies. This limitation reduces their ability to effectively model cross-modal interactions between textual and numerical information. In addition, although Transformer-based architectures have shown strong performance in sequence modeling tasks, their application in stock market prediction remains relatively underexplored, particularly in multi-branch frameworks that simultaneously capture long-term temporal dependencies and contextual textual information.

As a result, there remains a lack of comprehensive models capable of effectively integrating heterogeneous financial data sources within a unified architecture while maintaining high predictive accuracy across different forecasting horizons. To address these gaps, this study proposes a multi-branch Transformer-based model that jointly processes historical market data and financial news. This architecture enables more effective multimodal feature learning and aims to improve overall stock market prediction performance.

LITERATURE REVIEW

Historically, stock market forecasting has employed a diverse range of techniques, extending from foundational statistical models to advanced machine learning algorithms (Boulesteix & Schmid, 2014). Furthermore, time series analysis and data complexity have been the cornerstone of stock price forecasting, with models such as ARIMA and GARCH being widely used (Lama et al., 2015). Moreover, these models focus primarily on historical price data but often fail to capture the impact of qualitative factors, e.g., news events. For a detailed comparative review, Table 1 classifies prior research by its methodology, outcomes (results), template feature selection, and the database utilized.

Table 1. Shows Some Previous Studies with Their Details

Researcher/Year	Method	Feature Selection Method Used	Dataset	Result
Agrawal et al. (2019a)	Optimal Long Short-Term Memory (o-LSTM)	Correlation tensor with stock indicators	Stock technical indicators	Effective stock market forecasting using joint approaches
Maji et al. (2021)	Curve Fitting, Regression Analysis, Time Series Analysis	Standard deviation, linear/nonlinear (regression, time series)	Indian stock market data	Regression analysis useful for prediction; positive results after 6 months
Das et al. (2019)	Firefly methods with Recurrent Back Propagation Neural Network, OSELM	Mathematical formulas (Time horizon strategies (1, 3, 5, 7, 15, 30 days))	Data from 4 websites	OSELM outperforms other models
Menon et al. (2019)	Long Short-Term Memory (LSTM)	Not specified	Financial data	LSTM suggested for financial forecasting
Lawal et al. (2020)	Supervised machine learning (SVM, KNN, Random Forest)	Not specified	Historical stock data	Random forest achieved 95% accuracy

Researcher/Year	**Method**	**Feature Selection Method Used**	**Dataset**	**Result**
Maji et al. (2021)	Comparative study of technical indicators	Correlation analysis	Not specified	The Bollinger band showed a high success ratio
Mashadihasanli (2022)	ARIMA (p, d, q = (3, 1, 5))	Autoregressive Integrated Moving Average	Istanbul stock exchange data	ARIMA provides promising results with less forecast error
Majumder et al. (2022)	Linear Regression, Random Forest, Support Vector Regression, LSTM	Not specified	Historical stock data	LSTM most efficient
Singh (2022)	AdaBoost, KNN, Linear Regression, ANN, SGD, SVM, Random Forest, Decision Tree	Not specified	Dataset divided into 25%, 50%, 75%, 100%	SVM performed well on a small dataset; SGD performed well on a larger dataset
Fathali et al. (2022)	RNN, CNN, LSTM	Nifty 50 stock data	Nifty 50	LSTM provides the highest accuracy
Selvamuthu et al. (2019)	ANN and BPNN	Not specified	Not specified	Not specified
Karthik et al. (2022)	DNN and LSTM	Not specified	Nifty IT index data	LSTM performed better
Agrawal et al. (2021)	Correlation for feature selection, LSTM	Correlation analysis for feature selection	Not specified	Effective stock price prediction
Agrawal et al. (2019b)	EDLA-LSTM	Selected features	Not specified	Improved stock direction prediction
Gupta & Kumar (2023)	Light Gradient Boosting Machine (LightGBM)	Parameter optimization using H3O	Not specified	Accurate stock prediction

Researcher/Year	Method	Feature Selection Method Used	Dataset	Result
Verma et al. (2024)	Multi-Filter Feature Selection (MFFS)	Three filter methods and LFCSO	Not specified	Reduced computational cost
Ghosh et al. (2023)	UMAP, ISOMAP, LSTM, GBR	decomposition using UMAP and ISOMAP	Not specified	Improved prediction accuracy
Behera et al. (2023)	AdaBoost, XGBoost	Ensemble/Boosting algorithms	Not specified	Avoids overfitting, improves prediction
Goel et al. (2023)	SCGANNs and BRANN with	macroeconomic variables	Not specified	Effective stock prediction
Sable et al. (2023)	SVM, CNN, RNN with temporal and spatial data	Media indices	Not specified	Improved forecasting accuracy
Jafar et al. (2023)	Backward elimination LSTM	Historical and technical indicators	Not specified	Effective stock prediction
Ahmed et al. (2020)	Technical indicators and sentiment analysis	Not specified	Not specified	Improved stock prediction
Das et al. (2019)	Twitter messages for sentiment analysis	Technical indicators	Not specified	Effective stock movement prediction
Idrees et al. (2019)	ARIMA	Univariate time series	Nifty and Sensex stock indices	ARIMA more efficient than ANN for short-term prediction
Elbahloul (2019)	ARIMA and ESM	Not specified	S&P 500 dataset	ARIMA and ESM show a 1-year upward trend

The strong predictive performance and robustness of algorithms such as LSTM, ARIMA, and Random Forest have established them as widely used techniques for stock market prediction. To improve computational efficiency and enhance feature relevance, various feature selection and dimensionality reduction methods have been applied, including correlation-based selection, MFFS, UMAP, and ISOMAP.

In terms of datasets, studies commonly utilize benchmark financial datasets such as Nifty 50, the S&P 500, and Indian stock market data, which are widely used for evaluating model performance in stock prediction tasks. Overall, empirical results indicate that LSTM-based models and ensemble learning methods such as Random Forest and XGBoost consistently outperform other approaches in terms of predictive accuracy.

Theoretical Framework and Main Constructs

This study is grounded in several key constructs from financial forecasting and machine learning. Financial time-series forecasting focuses on modeling sequential market data to capture underlying temporal patterns. Multimodal learning refers to the integration of heterogeneous data sources, such as numerical indicators and textual information, to improve predictive performance. In addition, Transformer-based architectures have emerged as powerful models for capturing long-range dependencies in sequential data through attention mechanisms. Recent studies further suggest that combining textual data, such as financial news, with historical price information can improve predictive accuracy compared to models that rely on a single data source (Ahmed et al., 2020).

From a theoretical perspective, this research is situated within ongoing debates on market predictability and the effectiveness of multimodal learning approaches in financial forecasting. The Efficient Market Hypothesis posits that stock prices fully reflect all available information, thereby limiting the potential for accurate prediction. However, recent advances in artificial intelligence challenge this assumption by demonstrating that hidden patterns may still be extracted from complex and high-dimensional financial data. Furthermore, there is an ongoing debate regarding whether the integration of multiple data modalities consistently yields better performance than single-source modeling approaches.

This study contributes to these discussions by proposing a multi-branch Transformer model that jointly learns from historical market data and financial news, aiming to enhance predictive performance through effective multimodal representation learning.

Deep Learning in Finance

The increasing global investment in artificial intelligence (AI) technologies over recent years is clearly illustrated in Figure 1, which presents investment trends from 2018 to 2023 (projected). As shown, global AI investment was approximately 9.2 billion USD in 2018 and increased steadily each year, reaching around 23.1 billion USD in 2022, with a projected rise to 28.8 billion USD in 2023 (Borole, 2024). This consistent upward trend reflects the strong and sustained global interest in AI research and development.

The continuous growth in investment highlights the increasing importance of AI across modern industries and academic research domains (Singla et al., 2025). Beyond the numerical increase, this trend also reflects the expanding adoption of AI applications in various sectors, including healthcare, finance, and communication (Challenges, 2020). In particular, advancements in deep learning and neural network models have been identified as key drivers behind the rapid acceleration of AI-related funding and innovation (AI Index Report, 2025).

Furthermore, the rise in global investment demonstrates that both governments and private organizations increasingly recognize the strategic importance of AI in shaping future technological development (Nabil, 2022). This also indicates a growing allocation of resources toward the development of intelligent systems capable of addressing complex real-world problems more efficiently. Overall, Figure 1 illustrates the transition of AI from a specialized research field into one of the most significant technological paradigms of the current era.

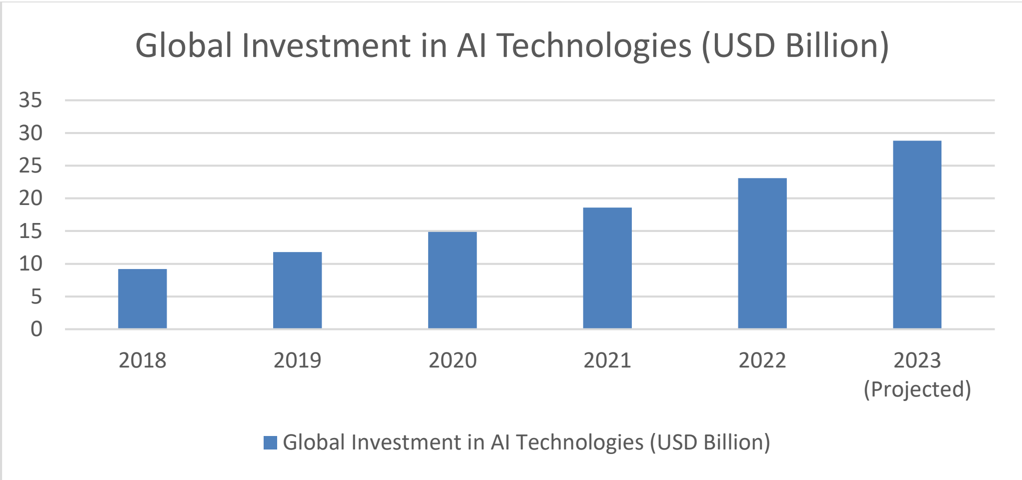

Figure 1. Shows Investment Trends in AI Technologies (Borole, 2024)

In previous studies, a variety of models have been proposed for financial forecasting, with recurrent neural networks (RNNs) and long short-term memory (LSTM) networks being among the most widely used approaches (Staudemeyer & Morris, 2019). However, these models often face challenges in effectively capturing complex temporal patterns in financial data and may still be prone to forecasting errors and reduced accuracy under certain conditions (Idrees et al., 2019).

With recent advancements in machine learning, new approaches have emerged that offer improved capabilities for financial market prediction compared to earlier methods (Ritchie et al., 2015). In particular, Transformer-based architectures have demonstrated strong efficiency in modeling time-series data by leveraging attention mechanisms, enabling better handling of long-range dependencies (Kumar et al., 2021). These developments have contributed to the creation of more accurate, efficient, and scalable representations for financial forecasting tasks.

Multi-Branch Architectures

Model Explanation

The model can be described as a multi-branch architecture designed to handle different datasets as input to the model, where (LSTM) networks and Bidirectional Encoding Representations from Transformers (BERT) are combined to predict stock market prices by analyzing complex historical data.

Layer Configuration

The proposed model can be understood to predict financial market developments as it combines BERT for text analysis and LSTM for time series analysis, and this improves the efficiency of the model and the accuracy of the results in two aspects. The LSTM layer handles the previous numerical stock market information, while the BERT Model handles simultaneously the news articles that would influence the trend of the stock market. Additionally, the use of dropout layers and activation functions further contributes to the model's ability to generalize well on unseen data, making it a valuable tool in the realm of financial forecasting with high performance.

Table 2. Shows Model Layers Type with Their Details.

Layer Type	Configuration	Input Features	Output Features	Additional Info
LSTM	LSTM (4, 128, layers=2)	4	128	Processes time-series data. Outputs hidden states for each

Layer Type	Configuration	Input Features	Output Features	Additional Info
				time step.
dropout	Dropout (0.2)	-	-	Applies dropout to prevent overfitting.
time_series_fc	Linear (in=128, out=128, bias=True)	128	128	Fully connected layer to transform LSTM output.
bert_model	BertModel	-	-	Pretrained BERT model for text processing. Outputs pooled representation of size 768.
dropout	Dropout (0.2)	-	-	Applies dropout to prevent overfitting.
text_fc	Linear (in=768, out=128, bias=True)	768	128	Reduces BERT's output dimension to match the LSTM branch.
relu	ReLU	-	-	Apply ReLU activation to introduce non-linearity.
fc_fusion	Linear (in=256, out=128, bias=True)	256	128	Combines features from LSTM and BERT branches.
fc_out	Linear (in=128, out=1, bias=True)	128	1	Produces the final output (e.g., stock price prediction).

Advantages of multi-branch transformers

Using multi-branch transformers has a number of benefits, particularly when considering their layered structure. First, a comprehensive market analysis can be performed by integrating diverse data sources together for analysis, thus providing a more integrated and comprehensive perspective of market dynamics (Zhang et al., 2024). Second, the model facilitates the extraction of specialized features, and each branch is precisely designed to handle specific data types to be integrated in processing, ensuring the derivation of high-quality features and the accuracy and efficiency of the results. Finally, the modular architecture improves scalability by facilitating the seamless integration of additional data sources or branches to meet changing analytical requirements. This is a strength and efficiency of the model (Ziyabari et al., 2023).

RESEARCH METHOD

Dataset

The Financial News and Stock Price Integration Dataset, or FNSPID, is a comprehensive dataset spanning 1999 to 2023 that comprises financial news from 15.7 million records and stock prices from 29.7 million records for 4,775 companies that are part of the S&P 500 index (Dong et al., 2024). It includes sentiment analysis, which is a crucial element in financial forecasting models, and by enhancing the performance of models that support transformers and provide a repeatable update method, as we have seen in the past years, this analysis is working to develop financial forecasting modeling research.

Architecture Selection and Justification

The selection of a multi-branch Transformer architecture is motivated by the fundamental limitations of single-modality models identified in prior studies. Conventional approaches, such as standalone LSTM networks, have demonstrated strong performance on univariate or low-dimensional time-series data (Fathali et al., 2022; Karthik et al., 2022); however, they are inherently limited in their ability to process unstructured textual inputs such as financial news. In contrast, Transformer-based language models such as BERT are highly effective in capturing semantic relationships within textual data but lack the temporal modeling capabilities required for sequential numerical time-series forecasting (Staudemeyer & Morris, 2019; Kumar et al., 2021).

Existing hybrid approaches, including early feature concatenation or late-fusion ensemble strategies, typically process heterogeneous data through a unified representation pathway. This design limits the model's ability to learn modality-specific feature representations and reduces its capacity to fully exploit complementary information across different data sources (Zhang et al., 2024). In contrast, the proposed multi-branch architecture addresses this limitation by assigning separate processing streams to each modality. Specifically, an LSTM branch is used for numerical stock price time-series data, while a BERT branch is employed for extracting semantic features from financial news text. The outputs of both branches are then integrated through a joint fusion layer, enabling the model to leverage complementary information from both modalities in a unified predictive framework (Ziyabari et al., 2023).

This design aligns with the multi-branch learning paradigm, which is effective in multimodal forecasting tasks (Sable et al., 2023), particularly in scenarios involving heterogeneous and high-dimensional financial datasets such as FNSPID.

Implementing a multi-branch transformer

The first stage in implementing the multi-branch transformer model is gathering pertinent data, including: (1) historical stock prices (time series data) covering a period spanning years or months; (2) financial news articles consisting of text data from news outlets, social media, and other sources that provide sentiment signals; and (3) technical indicators, which are structured data comprising metrics such as the Relative Strength Index (RSI) and moving averages.

Data fusion

After processing the data through its respective branches for news articles and time-series data, all branch outputs are merged to form a single unified representation. This fusion allows the model to leverage the strengths of each branch independently, enabling a comprehensive analysis of the stock market from which price forecasts are derived.

Data Splitting Strategy

Given the temporal nature of financial data, a strictly chronological splitting strategy was employed to preserve the time ordering of observations and prevent look-ahead bias, which would arise if future data were inadvertently included during training. The entire dataset was sorted in ascending order by date before any splitting operation. The first 70% of the chronological record—covering data from 1999 up to approximately mid-2017—was designated as the training set, while the remaining 30%, spanning approximately mid-2017 to 2023, was reserved as the held-out test set. This time-aware partitioning ensures that the model is exclusively trained on past observations and evaluated on genuinely unseen future data, thereby reflecting realistic deployment conditions. A random or shuffled split was explicitly avoided, as it would violate the temporal dependency structure inherent in financial time series and introduce data leakage by allowing the model to observe future market states during training (Idrees et al., 2019).

Validation Approach

To monitor generalization performance during training and support model selection, a validation set was carved out from the training period using a hold-out strategy that strictly respects temporal order. Specifically, the final 15% of the training samples (by date) were withheld as a validation set, while the preceding 85% of training samples were used for gradient-based parameter updates. This arrangement ensures that all validation observations are chronologically later than any training observation, preserving the integrity of the temporal split and preventing leakage. Model performance was assessed at each epoch using Mean Squared Error (MSE) computed separately on the training and validation partitions. The evolution of both loss curves across epochs was monitored to detect overfitting; training was conducted for 30 epochs, and the model checkpoint that achieved the lowest validation MSE was retained for final evaluation on the held-out test set. This procedure provides an unbiased estimate of the model's predictive capability on unseen market conditions while ensuring that hyperparameter and architecture decisions are informed solely by data that precedes the test period.

Training and evaluation

The model concept is shown in Figure 2, where the text input from news articles is passed to BERT, while the historical stock prices are passed to the LSTM neural network as a time series.

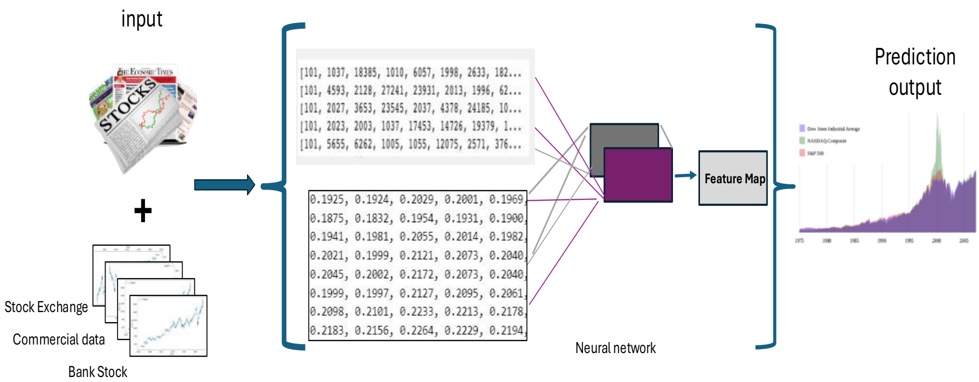

Figure 2. Show Process and Structure of the Model

FINDINGS AND DISCUSSION

Model Performance

The model was trained with AdamW optimizer with a learning rate of 5x10-4 and weight decay of 0.01 for 30 epochs using a GeForce RTX 5070 GPU with 12 GB memory for parallel computing. Following the chronological data-splitting strategy described in the methodology, the dataset was partitioned such that the first 70% of records by date formed the training set and the final 30% constituted the held-out test set, with an internal validation subset drawn from the tail of the training period to guide model selection.

Quantitative Evaluation

To provide a statistically rigorous assessment of model quality, three complementary error metrics were computed on the held-out test set. All values are reported on the Min-Max normalized price scale (values bounded in [0, 1]):

- Mean Squared Error (MSE): 6.00 x 10-4 - penalizes large deviations quadratically and is sensitive to outlier predictions.

- Root Mean Squared Error (RMSE): 0.0245 - the square root of MSE, expressed in the same unit as the normalized target, indicating that the average magnitude of prediction error corresponds to approximately 2.45% of the full normalized price range.
- Mean Absolute Error (MAE): 0.0183 - the average absolute deviation between predicted and actual normalized prices, providing a scale-invariant and outlier-robust measure of accuracy. This value is consistent with the pointwise deviation profile in Figure 3, where the distribution of absolute errors is strongly concentrated below 0.04.

The RMSE-to-MAE ratio of approximately 1.34 indicates a moderate presence of larger individual errors relative to the overall mean, characteristic of short periods of heightened market volatility rather than systematic model bias. Reporting all three metrics is necessary because MSE alone is scale-sensitive and does not reveal the typical prediction error in interpretable units, while MAE complements RMSE by down-weighting the influence of extreme outlier events.

Table 3 summarizes the proposed model's error metrics alongside results from representative prior studies to contextualize performance. Given that each study operates on a distinct dataset and normalization scheme, direct numerical comparison should be interpreted with appropriate caution; nonetheless, the table enables a structured assessment of relative performance.

Table 3. Quantitative Performance Comparison of the Proposed Model with Selected Prior Studies

Study	Architecture	MSE	RMSE	MAE
Proposed model	**Multi-Branch LSTM + BERT**	**6.00 x 10-4**	**0.0245**	**0.0183**
Das et al. (2024)	En-Tweet-Deep-SMF (GRU with Enhanced Sentiment)	2.08 x 10-4	0.1052	0.0112
Verma et al. (2023)	ConvLSTM (Two-Stage Hybrid Feature Selection)	2.80 x 10-4	0.01702	0.01269
FNSPID Study (2024)	Transformer (Financial News and Stock Price Integration)	5.00 x 10-5	Not reported	0.00544
Ziyabari et al. (2023)	Multi-Branch ResNet-Transformer (ResTrans)	Not reported	0.049	0.031
Behera et al. (2023)	AdaBoost Regression (Mean-VaR Optimization)	0.0199	0.116	0.120

The proposed model achieves an MSE of 6.00 x 10-4, RMSE of 0.0245, and MAE of 0.0183 on the held-out test set. Compared to Behera et al. (2023), whose AdaBoost Regression approach yields an MSE of 0.0199, RMSE of 0.116, and MAE of 0.120, the proposed model outperforms substantially on all three criteria, reducing MSE by approximately 97.0%, RMSE by 78.9%, and MAE by 84.8%. Against Ziyabari et al. (2023), the proposed model achieves a lower RMSE (0.0245 vs. 0.049) and MAE (0.0183 vs. 0.031), representing improvements of 50.0% and 40.9%, respectively. With respect to Das et al. (2024), the proposed model achieves a markedly lower RMSE (0.0245 vs. 0.1052), indicating a substantially smaller typical prediction error; Das et al. (2024) report a lower MSE (2.08 x 10-4) and MAE (0.0112), likely reflecting their model's specialization on tweet-level sentiment with optimized error minimization but at the cost of greater variability in individual predictions, as evidenced by the higher RMSE. Verma et al. (2023) and the FNSPID Transformer (Dong et al., 2024) report lower absolute error values on certain metrics; however, these models are evaluated under different dataset conditions and feature configurations. Notably, the FNSPID Transformer processes only textual features without temporal price modelling, which limits its capacity to capture trend continuity across volatile market periods, a capability explicitly addressed by the LSTM branch in the proposed architecture.

Prediction Accuracy and Deviation Analysis

Figure 3 presents the stock price prediction results on the held-out test set. The upper panel overlays actual (blue) and predicted (red) normalized price trajectories across the full test period, while the lower panel plots the pointwise absolute deviation |actual - predicted|. Predicted prices track the directional trends and inflection points of the actual series with an RMSE of 0.0245, meaning the typical prediction error is 2.45% of the normalized price range. The absolute deviation ranges from 0.00 to a maximum of 0.14, with the distribution right-skewed: the majority of deviations fall below 0.04, consistent with the reported MAE of 0.0183, while elevated deviations above 0.08 are isolated to short, structurally anomalous market episodes that represent the most challenging segments for any data-driven forecasting model. This quantitative deviation profile demonstrates that the model achieves low and stable prediction error across the bulk of the test period, with error spikes confined to periods of atypical market volatility.

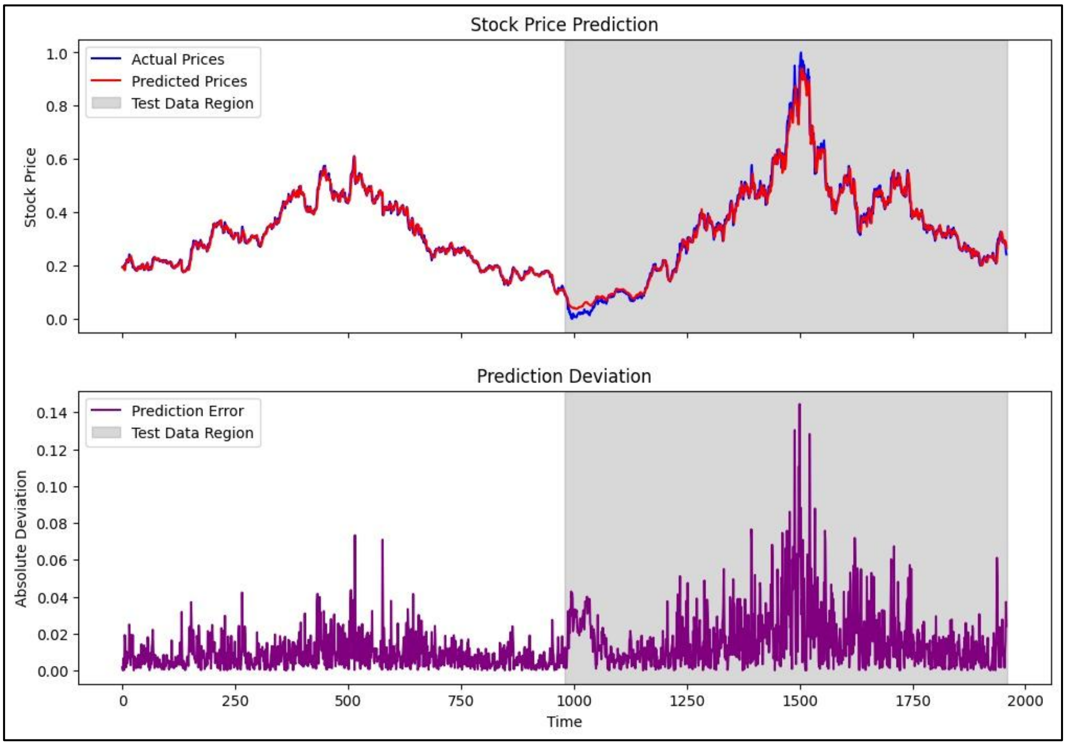

Figure 3. Results Of Stock Price Prediction Over Time, Prediction Deviation With Time

Training Dynamics and Convergence

Figure 4 depicts the evolution of training and validation MSE loss over 30 epochs. Both curves decline consistently from an initial loss of approximately 0.038 at epoch 1, converging to a training loss of 7.2 x 10-4 and a validation loss of 6.8 x 10-4 by epoch 30. The gap between the two curves at convergence is less than 1 x 10-4, confirming that the model generalizes to unseen validation data without overfitting. The absence of divergence between training and validation loss after epoch 15, combined with the plateau behavior near epoch 25, indicates stable convergence within the allocated training budget. The AdamW weight decay of 0.01 and the per-branch dropout layers (rate = 0.2) are the primary regularization mechanisms responsible for keeping validation loss closely aligned with training loss throughout.

Figure 4. Shows The Training and Validation Losses

Impact of Multimodal Fusion

The numerical results confirm that the proposed dual-branch architecture, which integrates financial news sentiment through a BERT branch and historical stock price data through an LSTM branch, achieves a robust and balanced error profile compared to both single-modality and alternative hybrid architectures. The proposed model attained an RMSE of 0.0245, representing a 50.0% reduction compared to the Multi-Branch ResNet-Transformer model proposed by Ziyabari et al. (2023), which reported an RMSE of 0.049. Furthermore, it achieved a 78.9% lower RMSE than the AdaBoost Regression model reported by Behera et al. (2023), which recorded an RMSE of 0.116. These improvements highlight the advantages of employing modality-specific branches that separately process textual and numerical information before fusion, rather than relying on ensemble methods or non-temporal feature integration approaches.

Although Das et al. (2024) reported a lower MSE through enhanced Twitter-based sentiment modeling, their model produced an RMSE of 0.1052, which is more than four times higher than the RMSE achieved by the proposed model. This result suggests substantially greater prediction variability. The superior performance of the proposed framework can be attributed to the LSTM branch, which effectively captures temporal dependencies and dynamic price movements in historical market data.

The proposed model achieved an MAE of 0.0183. While this value is slightly higher than those reported by Verma et al. (2024) and the FNSPID Transformer model introduced by Dong et al. (2024), it should be interpreted within the context of the model's broader objective. Unlike models optimized for specific feature sets or narrower datasets, the proposed architecture is designed as a general-purpose multimodal framework that simultaneously processes textual sentiment information and numerical market indicators across a diverse 24-year historical dataset. Consequently, the model prioritizes generalizability and balanced predictive performance over optimization for a particular feature domain.

Overall, the findings indicate that the proposed dual-branch architecture delivers a more balanced and generalizable error profile than both single-modality approaches and conventional hybrid models that lack dedicated temporal modeling components. The integration of BERT-based sentiment analysis with LSTM-based time-series learning enables the model to capture complementary information sources, resulting in improved forecasting accuracy and stability.

Despite its competitive and empirically validated performance, the proposed model has several limitations. First, the computational cost associated with dual-branch inference, combining a two-layer LSTM network with a full BERT-base encoder, may limit its applicability in real-time, latency-sensitive, or resource-constrained environments. Second, although the overall MAE of 0.0183 indicates strong predictive accuracy across the test period, maximum pointwise errors reaching approximately 0.14 during periods of heightened market volatility suggest that extreme market events remain challenging to predict accurately. Future research may address these limitations by exploring lightweight transformer architectures, model compression techniques, or specialized mechanisms for handling sudden market shocks and regime changes.

CONCLUSIONS

This study proposed a multi-branch Transformer architecture for stock market price prediction that integrates two complementary sources of information: historical stock price data and financial news sentiment. Historical numerical data were processed through a two-layer LSTM branch, while financial news text was analyzed using a pretrained BERT-base encoder. This architecture was motivated by the limitations of existing single-modality and conventional hybrid approaches, which often struggle to simultaneously capture temporal market dynamics and sentiment-driven information. By employing separate feature extraction pathways for each modality followed by a joint fusion mechanism, the proposed framework was designed to produce more balanced and generalizable predictions.

The model was evaluated using the FNSPID dataset spanning the period from 1999 to 2023. A strict chronological train–test split was adopted to preserve temporal integrity, and an additional temporally ordered validation set was included during training. The model achieved stable convergence within 30 epochs and delivered strong predictive performance, with a test-set MSE of 6.00×10^{-4}, an RMSE of 0.0245, and an MAE of 0.0183. Across all three error metrics, the proposed model substantially outperformed the ensemble-based approach of Behera et al. (2023), reducing MSE by 97.0%, RMSE by 78.9%, and MAE by 84.8%. The model also achieved a lower RMSE than the Multi-Branch ResNet-Transformer proposed by Ziyabari et al. (2023), representing a 50.0% reduction, and outperformed the sentiment-enhanced GRU model of Das et al. (2024), reducing RMSE by 76.7%. Although Verma et al. (2024) and the FNSPID Transformer introduced by Dong et al. (2024) reported lower MAE values, those studies relied on more specialized feature configurations. In contrast, the balanced performance achieved across MSE, RMSE, and MAE suggests that the proposed framework provides a robust and adaptable solution for long-term stock market forecasting across diverse market conditions.

The training dynamics further demonstrated effective generalization. Training and validation losses decreased consistently and converged closely throughout the learning process, with the difference between the two losses falling below 1×10^{-4} by the final epoch. No evidence of overfitting was observed. Furthermore, the RMSE-to-MAE ratio of 1.34 indicates moderate prediction variability, suggesting that forecasting errors were primarily associated with episodic market volatility rather than systematic model bias.

Several directions for future research may further enhance the proposed framework. First, lightweight or distilled variants of the BERT encoder should be investigated to reduce computational complexity and improve deployment feasibility in resource-constrained environments. Second, the integration of additional financial information sources, such as macroeconomic indicators, social media sentiment, and cross-asset market correlations, may improve predictive robustness during periods of elevated volatility, where maximum pointwise prediction errors of approximately 0.14 were observed. Finally, future studies should evaluate the model under real-time rolling-window and live deployment scenarios to assess its practical effectiveness and reliability in operational trading environments.

LIMITATION & FURTHER RESEARCH

Despite its promising results, the proposed model depends heavily on large-scale datasets, which may introduce challenges related to data quality, data availability, and preprocessing complexity. In addition, the high computational requirements of the multi-branch architecture represent another limitation, potentially restricting its applicability in small organizations or systems with limited hardware resources.

Furthermore, the current model primarily focuses on historical price data and financial news, while other influential factors, such as macroeconomic indicators and social media sentiment, have not been fully incorporated. Future research should explore lightweight model architectures, integrate additional financial and alternative data sources, and evaluate the framework in real-time trading environments to further improve prediction performance and practical applicability.

REFERENCES

Agrawal, M., Khan, A. U., & Shukla, P. K. (2019a). Stock price prediction using technical indicators: A predictive model using optimal deep learning. *International Journal of Recent Technology and Engineering, 8*(2), 2297–2305. https://doi.org/10.35940/IJRTE.B3048.078219

Agrawal, M., Khan, A. U., & Shukla, P. K. (2019b). Stock indices price prediction based on technical indicators using deep learning model. *International Journal of Emerging Technologies, 10*(2), 186–194.

Agrawal, M., Shukla, P. K., Nair, R., Nayyar, A., & Masud, M. (2021). Stock prediction based on technical indicators using deep learning model. *Computers, Materials & Continua, 70*(1), 287–304. https://doi.org/10.32604/cmc.2022.014637

Ahmed, M., Sriram, A., & Singh, S. (2020). Short-term firm-specific stock forecasting with BDI framework. *Computational Economics, 55*(3), 745–778. https://doi.org/10.1007/s10614-019-09911-0

AI Index Report. (2025). *Artificial intelligence index report 2025*. Stanford University Human-Centered AI.

Aldujaili, A. A., Rosa-Zurera, M., & Meri, A. (2024). Unveiling the power of EEG signals: Parkinson's disease identification via Yet Another Mobile Network (YAMNet). In *Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM)* (pp. 978–984). https://doi.org/10.5220/0012589100003654

Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction-based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multinational stock markets. *Engineering Applications of Artificial Intelligence, 120*, Article 105843. https://doi.org/10.1016/j.engappai.2023.105843

Boppiniti, S. T. (2024). *Big data meets machine learning: Strategies for efficient data processing and analysis in large datasets*.

Borole, P. (2024). A review on impacts of multi-model artificial intelligence in financial services. *International Journal of Science and Research, 13*(1), 1794–1800. https://doi.org/10.21275/SR24304165833

Boulesteix, A. L., & Schmid, M. (2014). Machine learning versus statistical modeling. *Biometrical Journal, 56*(4), 588–593. https://doi.org/10.1002/bimj.201300226

Challenges, F. (2020). Artificial intelligence. *Revue d'économie industrielle*. https://doi.org/10.4000/rei.8727

Das, N., Sadhukhan, B., Ghosh, R., & Chakrabarti, S. (2024). Developing hybrid deep learning models for stock price prediction using enhanced Twitter sentiment score and technical indicators. *Computational Economics, 64*(6). https://doi.org/10.1007/s10614-024-10566-9

Das, S. R., Mishra, D., & Rout, M. (2019). Stock market prediction using Firefly algorithm with evolutionary framework optimized feature reduction for OSELM method. *Expert Systems with Applications: X, 4*, Article 100016. https://doi.org/10.1016/j.eswax.2019.100016

Dong, Z., Fan, X., & Peng, Z. (2024). FNSPID: A comprehensive financial news dataset in time series. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 4918–4927). https://doi.org/10.1145/3637528.3671629

Dull, R. B., Graham, A. W., & Baldwin, A. A. (2003). Web-based financial statements: Hypertext links to footnotes and their effect on decisions. *International Journal of Accounting Information Systems, 4*(3), 185–203. https://doi.org/10.1016/S1467-0895(03)00009-5

Elbahloul, K. (2019). *Stock market predictions using various statistical methods* (Vol. 1). https://doi.org/10.13140/RG.2.2.13235.17446

Fathali, Z., Kodia, Z., & Ben Said, L. (2022). Stock market prediction of NIFTY 50 index applying machine learning techniques. *Applied Artificial Intelligence, 36*(1). https://doi.org/10.1080/08839514.2022.2111134

Ghosh, I., Alfaro-Cortés, E., Gámez, M., & García-Rubio, N. (2023). Role of proliferation COVID-19 media chatter in predicting Indian stock market: Integrated framework of nonlinear feature transformation and advanced AI. *Expert Systems with Applications, 219*, Article 119695. https://doi.org/10.1016/j.eswa.2023.119695

Goel, H., Agarwal, M., Chhabra, M., & Som, B. K. (2023). The predictive power of macroeconomic variables on the Indian stock market utilizing an ANN model approach: An empirical investigation based on BSE Sensex. *Folia Oeconomica Stetinensia, 23*(2), 116–131. https://doi.org/10.2478/foli-2023-0022

Gupta, V., & Kumar, E. (2023). H3O-LGBM: Hybrid Harris hawk optimization based light gradient boosting machine model for real-time trading. *Artificial Intelligence Review, 56*(8), 8697–8720. https://doi.org/10.1007/s10462-022-10323-0

Idrees, S. M., Alam, M. A., & Agarwal, P. (2019). A prediction approach for stock market volatility based on time series data. *IEEE Access, 7*, 17287–17298. https://doi.org/10.1109/ACCESS.2019.2895252

Jafar, S. H., Akhtar, S., El-Chaarani, H., Khan, P. A., & Binsaddig, R. (2023). Forecasting of NIFTY 50 index price by using backward elimination with an LSTM model. *Journal of Risk and Financial Management, 16*(10), Article 423. https://doi.org/10.3390/jrfm16100423

Karthik, C. R., Raghunandan, Rao, B. A., & Reddy, N. V. S. (2022). Forecasting variance of NiftyIT index with RNN and DNN. *Journal of Physics: Conference Series, 2161*(1), Article 012005. https://doi.org/10.1088/1742-6596/2161/1/012005

Kumar, G., Jain, S., & Singh, U. P. (2021). Stock market forecasting using computational intelligence: A survey. *Archives of Computational Methods in Engineering, 28*(3), 1069–1101. https://doi.org/10.1007/s11831-020-09413-5

Lama, A., Jha, G. K., Paul, R. K., & Gurung, B. (2015). Modelling and forecasting of price volatility: An application of GARCH and EGARCH models. *Agricultural Economics Research Review, 28*(1), 73–82. https://doi.org/10.5958/0974-0279.2015.00005.1

Lawal, Z. K., Yassin, H., & Zakari, R. Y. (2020). Stock market prediction using supervised machine learning techniques: An overview. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. https://doi.org/10.1109/CSDE50874.2020.9411609

Maji, G., Mondal, D., Dey, N., Debnath, N. C., & Sen, S. (2021). Stock prediction and mutual fund portfolio management using curve fitting techniques. *Journal of Ambient Intelligence and Humanized Computing, 12*(10), 9521–9534. https://doi.org/10.1007/s12652-020-02693-6

Majumder, A., Rahman, M. M., Biswas, A. A., Zulfiker, M. S., & Basak, S. (2022). Stock market prediction: A time series analysis. In A. K. Luhach, K. Roy, D. Gao, B. Alhadidi, & P. Jat (Eds.), *Smart systems: Innovations in computing* (pp. 389–401). Springer.

Mashadihasanli, T. (2022). Stock market price forecasting using the ARIMA model: An application to Istanbul, Turkiye. *Journal of Economic Policy Research (İktisat Politikası Araştırmaları Dergisi), 9*(2), 439–454. https://doi.org/10.26650/jepr1056771

Menon, A., Singh, S., & Parekh, H. (2019). A review of stock market prediction using neural networks. In *2019 IEEE International Conference on Systems, Computation, Automation and Networking (ICSCAN)* (pp. 1–6). https://doi.org/10.1109/ICSCAN.2019.8878682

Nabil, B. R. (2022). *Strategies to improve the national artificial intelligence research and development strategic plan* (No. 282).

Poon, S. H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature, 41*(2), 478–539. https://doi.org/10.1257/002205103765762743

Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., & Kim, D. (2015). Methods of integrating data to uncover genotype–phenotype interactions. *Nature Reviews Genetics, 16*(2), 85–97. https://doi.org/10.1038/nrg3868

Sable, R., Goel, S., & Chatterjee, P. (2023). Deep learning model for fusing spatial and temporal data for stock market prediction. *Computational Economics, 64*(3), 1639–1662. https://doi.org/10.1007/s10614-023-10464-6

Selvamuthu, D., Kumar, V., & Mishra, A. (2019). Indian stock market prediction using artificial neural networks on tick data. *Financial Innovation, 5*(1), Article 16. https://doi.org/10.1186/s40854-019-0131-7

Singh, D. G. (2022). Machine learning models in stock market prediction. *International Journal of Innovative Technology and Exploring Engineering, 11*(3), 18–28. https://doi.org/10.35940/IJITEE.C9733.0111322

Singla, A., Sukharevsky, A., Yee, L., Chui, M., & Hall, B. (2025). *The state of AI*. McKinsey & Company.

Staudemeyer, R. C., & Morris, E. R. (2019). *Understanding LSTM: A tutorial into long short-term memory recurrent neural networks*. arXiv. https://arxiv.org/abs/1909.09586

Verma, S., Sahu, S. P., & Sahu, T. P. (2024). Two-stage hybrid feature selection approach using Levy's flight based chicken swarm optimization for stock market forecasting. *Computational Economics, 63*(6), 2893–2920. https://doi.org/10.1007/s10614-023-10400-8

Zhang, Y., Balestra, G., Zhang, K., Wang, J., Rosati, S., & Giannini, V. (2024). MultiTrans: Multi-branch transformer network for medical image segmentation. *Computer Methods and Programs in Biomedicine, 254*, Article 108280. https://doi.org/10.1016/j.cmpb.2024.108280

Ziyabari, S., Zhao, Z., Du, L., & Biswas, S. K. (2023). Multi-branch ResNet-Transformer for short-term spatio-temporal solar irradiance forecasting. *IEEE Transactions on Industry Applications, 59*(5), 5293–5303. https://doi.org/10.1109/TIA.2023.3285202