International Conference on Social Development and Intelligent Technology (SDIT2024)

Challenges and Opportunities in Stock Price Prediction: An Exploration Using an ANN-LSTM-Transformer Hybrid Model

Shengkai Ma
Xi'an Jiaotong Liverpool University, China

Abstract: This paper explores the effectiveness of a hybrid model combining ANN, LSTM, and Transformer architectures for stock price prediction. The model integrates ANN for feature extraction, LSTM for capturing long-term dependencies, and Transformer for modeling complex relationships. However, its performance on real-world datasets fell short of traditional models like RF, SVM, and XGBoost, primarily due to insufficient hyperparameter tuning, inadequate data preprocessing, and the challenge of managing model complexity with limited training data. The findings emphasize the need for systematic optimization, advanced preprocessing techniques, and the inclusion of diverse data sources to improve predictive accuracy and robustness. While the current results highlight limitations, the hybrid model approach remains a promising avenue for tackling the complexities of stock prediction and enhancing financial decision-making.

Keywords: Stock Price Prediction; Artificial Neural Network (ANN); Long Short-Term Memory (LSTM); Transformer; Hybrid Model

1. Previous Literature

The stock market has long been a vital component of the broader financial system, consistently attracting investors due to its potential for high returns. However, the inherent risks associated with stock market investments make accurate stock prediction essential for informed decision-making. The highly dynamic nature of the stock market, driven by complex interactions among various economic, social, and psychological factors, necessitates the development of predictive models capable of accommodating this complexity. As a result, stock market forecasting has become a prominent focus for both researchers and investors. In recent years, with advancements in artificial intelligence (AI), many researchers, including our team, have sought to apply AI models to stock prediction. Our goal is to harness state-of-the-art AI techniques to enhance the accuracy of market forecasts and address the multifaceted challenges inherent to this domain.

Initially, we explored the use of ARIMA, a traditional time series model, for stock prediction. ARIMA's strength lies in its ability to model linear relationships in time-series data by decomposing it into autoregressive and moving average components, making it an effective tool for short-term forecasting. However, after reviewing several studies, we found that while ARIMA yields favorable results over short time periods [1], it is fundamentally limited in addressing the nonlinear nature of stock market dynamics. The ARMA model, which assumes the future value of a variable is a linear combination of its past values and errors, fails to account for the influence of nonlinear factors such as economic conditions, political events, and investor sentiment [2]. These limitations became particularly evident when attempting to model complex interactions among variables or when forecasting over extended time horizons. Consequently, it became clear that ARIMA alone could not adequately capture the multifaceted and evolving nature of stock markets, prompting the need to explore alternative approaches capable of incorporating these nonlinearities.

With the rapid progress in artificial intelligence theory and technology, Artificial Neural Networks (ANN) emerged as a promising option for stock price prediction. ANN were initially proposed by Rosenblatt in his foundational work in 1958. However, their practical development was hindered by their inability to solve fundamental logical problems, such as the XOR problem. This limitation was addressed in 1986 with the introduction of the backpropagation algorithm, a groundbreaking advancement that significantly improved the training of multilayer ANN models. By overcoming these early challenges, backpropagation catalyzed the rapid evolution of ANN, enabling them to achieve exceptional performance across diverse tasks, ranging from natural language processing to computer vision. Today, ANN has become a versatile and widely adopted approach in financial forecasting [3]. Studies have demonstrated that ANN is capable of identifying intricate patterns and trends in stock prices, establishing them as an essential asset for financial analysis and forecasting [4]. ANN represents a paradigm shift from traditional statistical models by adopting a data-driven and adaptive framework, requiring minimal assumptions about the underlying data distribution [5]. This characteristic makes ANN particularly well-suited for modeling complex and nonlinear relationships between stock performance and its influencing factors. ANN's ability to approximate nonlinear functions with arbitrary precision through its multilayer architecture enables it to uncover intricate patterns that traditional methods often overlook. Moreover, ANN excels in integrating diverse and high-dimensional data sources, such as macroeconomic indicators, market trends, and sentiment analysis, thereby providing a more holistic view of stock market behavior. Empirical studies have consistently demonstrated ANN's superiority in predictive accuracy when compared to traditional models such as Support Vector Machines (SVM) and Logistic Regression (LR), particularly in scenarios requiring the integration of multiple factors [6][7].

Building on these advancements, Recurrent Neural Networks (RNNs) introduced a significant innovation in sequence modeling by incorporating feedback connections, allowing them to process temporal dependencies in sequential data. Among the variants of RNNs, Long Short-Term Memory (LSTM) networks have gained prominence due to their ability to overcome the vanishing gradient problem, a critical limitation of standard RNNs. LSTM achieves this by introducing memory cells and gating mechanisms, which enable the selective retention and forgetting of information over extended time horizons. This unique capability makes LSTM particularly advantageous in capturing nonlinear relationships and long-term dependencies in stock price movements [8][9]. Due to its efficiency in financial prediction, LSTM has become highly significant in stock market forecasting [10]. Futher more, Fischer and Krauss (2018) demonstrated the effectiveness of LSTM in financial market predictions, showing that its ability to learn from vast amounts of historical data, capture long-term dependencies in stock prices, and adapt flexibly to different data structures has positioned it as a key architecture in financial time-series analysis [11]. Additionally, researchers have identified that LSTM networks exhibit remarkable predictive capabilities in the financial domain, making them widely adopted for time-series forecasting in capital markets [12].

Recent research [8] has demonstrated the significant potential of hybrid models that combine LSTM with other architectures to address specific challenges in stock market prediction. For instance, integrating LSTM with Transformer models has yielded exceptional results in various applications, particularly in multi-task and real-time prediction scenarios. Hybrid models leverage the complementary strengths of their components, with LSTM excelling in modeling sequential dependencies and Transformers providing robust attention-based mechanisms for parallel computation and long-range dependency modeling. Notably, the Transformer architecture overcomes the constraints of sequential computation inherent in RNNs by utilizing self-attention mechanisms, which enable efficient processing of entire sequences simultaneously. This efficiency is particularly critical in financial applications, where large-scale and high-frequency data require rapid and accurate analysis. Additionally, the integration of LSTM and Transformer architectures has been shown to maintain high performance even when the data retention rate is reduced to 50%, highlighting their potential for efficient time-series processing and accelerated training [13].

Motivated by these findings, we directed our attention to hybrid models that integrate the unique strengths of multiple architectures. Specifically, the ARIMA-LSTM hybrid model combines ARIMA's ability to identify linear trends with LSTM's capacity to recognize nonlinear patterns, resulting in improved accuracy for predicting stock price correlation coefficients. Empirical evidence supports the superior performance of this hybrid approach compared to traditional financial models, demonstrating its robustness and adaptability across diverse asset combinations and time horizons [14]. By synthesizing the complementary capabilities of these models, the ARIMA-LSTM hybrid offers a promising framework for enhancing predictive performance in complex financial contexts.

Meanwhile, the Transformer model has revolutionized sequence modeling by introducing a fully attention-based mechanism that replaces traditional RNN and CNN architectures. The Transformer’s self-attention mechanism allows it to capture intricate dependencies across entire sequences without relying on sequential processing, enabling parallelized training and greater scalability. This innovation has set new benchmarks in a variety of tasks, including machine translation, pre-trained language models, reinforcement learning, and cross-modal learning, due to its simplicity, interpretability, and computational efficiency [15]. Recognizing these advantages, we sought to incorporate the Transformer model into our hybrid framework, leveraging its ability to model long-range dependencies efficiently. By combining the Transformer's strengths with LSTM’s proven capabilities in sequential data analysis, we aimed to develop a robust and scalable model for stock prediction.

In conclusion, we developed an ANN-LSTM-Transformer hybrid model to evaluate its effectiveness in stock prediction. This model leverages the strengths of these advanced AI architectures to address the inherent complexities of financial forecasting, offering a promising approach for achieving more accurate, reliable, and efficient market predictions. Through this work, we seek to discover the growing body of research at the intersection of AI and finance, demonstrating the transformative potential of hybrid models in optimizing investment strategies and enhancing decision-making processes in the stock market.

2. Materials and Methods

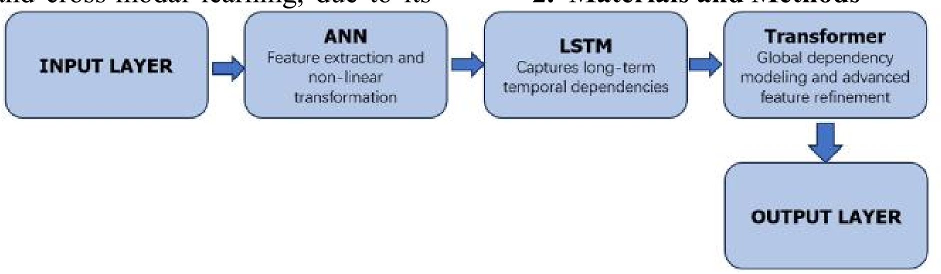

Figure 1. Flow Chart of the Hybrid Model

The flowchart figure 1. illustrates a multi-layered data processing pipeline designed to extract, transform, and model complex patterns in input data. The process begins with an “Input Layer,” which serves as the entry point for raw data.

Proceeding from there, the Artificial Neural Network (ANN) layer is responsible for feature extraction and non-linear transformation of the input data. This step enhances the richness and complexity of the features, allowing for more sophisticated analysis by subsequent layers.

Next, the Long Short-Term Memory (LSTM) network focuses on capturing long-term temporal dependencies within the data. By maintaining information across sequences, this layer can identify patterns that persist over extended periods, making it particularly useful for time-series data or sequences where context is essential.

Following LSTM, the Transformer layer implements global dependency modeling and advanced feature refinement. Leveraging attention mechanisms, this component captures relationships between all elements in the input at various scales, providing a comprehensive understanding of contextual dependencies.

Finally, the refined and enriched data is passed to the “Output Layer,” where results are generated based on the intricate analyses performed in preceding stages. This structured approach allows for robust and nuanced insights into the underlying data characteristics.

Table 1 Performance of Models

Model	APPLE	TSLA	JNJ
ANN+LSTM+Transformer:	0.0078	0.0112	0.0089
RF+SVM+XGBoost:	0.0004	0.0018	0.0001
LOSS/MAE			

In the evaluation of two competing models, ANN+LSTM+Transformer and RF+SVM+XGBoost, across the datasets of three leading companies—Apple, Tesla, and Johnson & Johnson (JNJ)—the ANN+LSTM+Transformer model is subjected to comparison with the RF+SVM+XGBoost model as a reference counterpart. For Apple's dataset, the ANN+LSTM+Transformer model yields a test loss of 0.0078, which is higher than the 0.0004 achieved by the RF+SVM+XGBoost model. Similarly, the ANN+LSTM+Transformer model's MAE for Apple is 0.0089, significantly greater than the 0.0001 reported by the RF+SVM+XGBoost model. In the context of Tesla's data, the ANN+LSTM+Transformer model's performance is again compared to the RF+SVM+XGBoost model, with the former recording a test loss of 0.0112 and an MAE of 0.0089, both of which are higher than the 0.0018 test loss and 0.0001 MAE of the RF+SVM+XGBoost model. For Johnson & Johnson (JNJ), the ANN+LSTM+Transformer model presents a test loss of 0.0089 and an MAE of 0.0089, which are notably higher than the 0.0001 test loss and MAE of the RF+SVM+XGBoost model. Across all three companies, the ANN+LSTM+Transformer model is consistently outperformed by the RF+SVM+XGBoost model in terms of both test loss and MAE. This comparative analysis indicates that the RF+SVM+XGBoost model has better performance than ANN+LSTM+Transformer model.

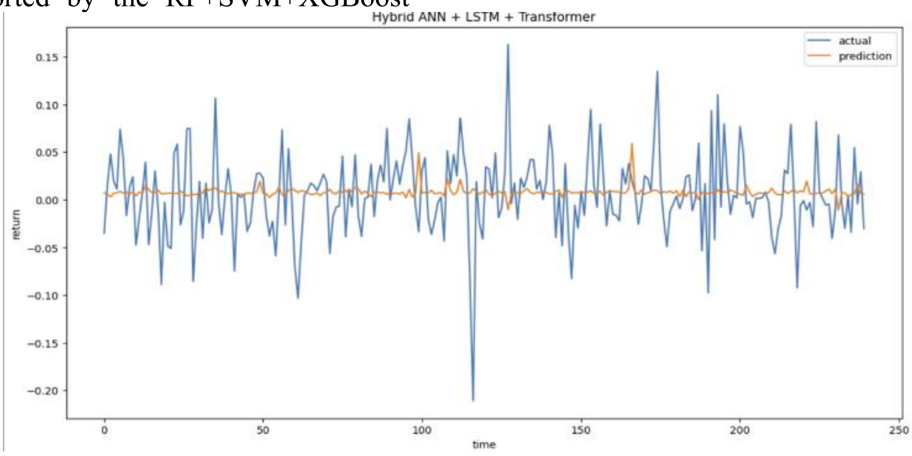

Figure 2. Comparism of Two Conditions

The Fig.2 provided showcases the efficacy of a Hybrid Artificial Neural Network, which integrates ANN, LSTM, and Transformer architectures, in practical settings. On the graph, the horizontal axis is a timeline, and the vertical axis measures the return rate. The solid blue line tracks the actual return rate, while the orange dashed line plots the forecasted return rate by the model. A notable convergence is observed between the predicted and actual return rates, with the two lines frequently converging or nearly coinciding across the majority of the period. Despite this alignment, it is crucial to examine the discrepancies between the actual and predicted values. The model demonstrates a high level of precision in identifying and following market trends. However, the analysis also highlights the model's performance in responding to abrupt market changes. In some instances, the predicted return rate lags slightly behind the actual rate, indicating that while the model is generally reliable, there is room for improvement in its ability to react instantaneously to market volatility.

3. Conclusion

Stock price prediction is among the most complex challenges in the financial domain due to the high volatility, nonlinearity, and susceptibility of financial markets to external influences such as breaking news and unexpected events. While our research explored the potential of a hybrid model integrating ANN, LSTM, and Transformer architectures, the practical performance did not meet expectations. Despite initial optimism, the model exhibited notable shortcomings in prediction accuracy and robustness when applied to real-world datasets.

Several factors contributed to the underperformance of the hybrid model. Key issues included insufficient hyperparameter tuning, inadequate data preprocessing, and challenges in balancing the complexity of integrated architectures with the available training data. These factors limited the model's ability to fully exploit the strengths of its components and led to suboptimal results in comparison to baseline models.

Nevertheless, this study highlights the importance of combining diverse architectures to tackle the inherent complexity of stock prediction. Future efforts should focus on optimizing hyperparameters through systematic techniques, improving preprocessing workflows to enhance feature representation, and incorporating additional data sources such as market sentiment and macroeconomic indicators. Regularization strategies should also be employed to prevent overfitting and improve the generalizability of hybrid models.

Despite its limitations, the concept of hybrid models for stock prediction holds significant promise. With continued refinement and innovation, integrating architectures like ANN, LSTM, and Transformer can pave the way for more effective and robust predictive frameworks. These advancements have the potential to transform financial analytics, offering valuable tools for investors and contributing to a deeper understanding of market dynamics.

References

[1] Meyler, A., Kenny, G., & Quinn, T. (1998). Forecasting Irish inflation using ARIMA models.

[2] Ma, Q. (2020). Comparison of ARIMA, ANN and LSTM for stock price prediction. In *E3S Web of Conferences* (Vol. 218, p. 01026). EDP Sciences.

[3] Tan, L., Liu, S., Gao, J., Liu, X., Chu, L., & Jiang, H. (2024). Enhanced self-checkout system for retail based on improved YOLOv10. *Journal of Imaging*, *10*(10), 248.

[4] Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, *5*(2), 241-259.

[5] .Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*, *37*(1), 479-489.

[6] Yetis, Y., Kaplan, H., & Jamshidi, M. (2014, August). Stock market prediction by using artificial neural network. In *2014 world automation congress (WAC)* (pp. 718-722). IEEE.

[7] Gurjar, M., Naik, P., Mujumdar, G., & Vaidya, T. (2018). Stock market prediction using ANN. *International Research Journal of Engineering and Technology*, *5*(3), 2758-2761.

[8] Ghosh, A., Bose, S., Maji, G., Debnath, N., & Sen, S. (2019, September). Stock price prediction using LSTM on Indian share market. In *Proceedings of 32nd international conference on* (Vol. 63, pp. 101-110).

[9] Pawar, K., Jalem, R. S., & Tiwari, V. (2019). Stock market price prediction using LSTM RNN. In *Emerging Trends in Expert Applications and Security: Proceedings of ICETEAS 2018* (pp. 493-503). Springer Singapore.

[10] Zhu, W., & Hu, T. (2021, July). Twitter Sentiment analysis of covid vaccines. In *2021 5th International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 118-122).

[11] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, *270*(2), 654-669.

[12] `Xu, Y., Sushmit, A., Lyu, Q., Li, Y., Cao, X., Maltz, J. S., ... & Yu, H. (2022). Cardiac CT motion artifact grading via semi-automatic labeling and vessel tracking using synthetic image-augmented training data. *Journal of X-Ray Science and Technology*, *30*(3), 433-445.

[13] Kim, T., & Kim, H. Y. (2019). Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. *PloS one*, *14*(2), e0212320.

[14] Choi, H. K. (2018). Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model. *arXiv preprint arXiv:1808.01560*.

[15] Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.