King Fahd University of Petroleum & Minerals
Mathematics Department
Master of Science in Data Science & Analytics

جامعة الملك فهد للبترول والمعادن
King Fahd University of Petroleum & Minerals

MATH-619

A CNN-LSTM Hybrid Model for Multivariate Stock Price Forecasting Using Technical and Sentiment Indicators

Thursday, 18th September 2025

Student Name & ID	**Advisor**
ABDULLAH ALHARBI g202415480	Dr. Mousa Ahmad Al-Bashrawi **Co-advisor** Mr. Mohammed Agbawi

1. Introduction

In today's highly dynamic and sentiment-driven financial markets, accurate stock price prediction remains one of the most challenging yet impactful problems in both academic research and applied finance. With the rapid expansion of financial data sources and the rise of advanced artificial intelligence (AI) methods, deep learning models have become powerful tools for uncovering complex nonlinear relationships in financial time-series data.

Among modern deep learning architectures, **Convolutional Neural Networks (CNNs)** and **Long Short-Term Memory (LSTM)** networks stand out for their ability to capture spatial and temporal patterns, respectively. CNNs efficiently learn relationships across multiple technical indicators, while LSTMs model sequential dependencies in market behavior. A **hybrid CNN-LSTM architecture** combines these strengths, enabling the model to detect both short-term fluctuations and long-term trends in stock price movement.

Alongside technical indicators, **sentiment analysis** has emerged as an important dimension of market modeling. Public sentiment, social media discussions, and news events often reflect investor psychology and expectations—factors that can influence price movements before they appear in technical patterns. Incorporating sentiment data alongside technical signals allows the model to capture both quantitative and behavioral aspects of market dynamics.

Despite significant progress in machine learning, stock price forecasting remains difficult due to its noisy, chaotic, and highly nonlinear nature. Traditional models such as ARIMA or SVMs struggle to integrate unstructured signals like public sentiment. Deep learning models, while more powerful, require well-designed feature integration to capture both structural (technical) and psychological (sentiment) aspects of markets.

This project addresses this gap by integrating **technical indicators**, **market-level volatility (VIX)**, and **pre-scored Twitter sentiment** into a unified deep learning forecasting framework.

The sentiment data used in this study is sourced from a Kaggle dataset covering **30 September 2021 to 29 September 2022**, where each tweet is already assigned a sentiment score in the range **−1 to +1**. This allows direct incorporation of sentiment without the need for additional NLP preprocessing.

Finally, the model includes **Explainable AI (XAI)** techniques to provide transparency into which technical and sentiment features contribute most to the model's predictions.

2. Specific Topic of Study and Context

This study proposes a **hybrid CNN-LSTM model** to predict the **direction of stock price movement** (up or down) for a selected group of major technology companies. The model incorporates three key categories of features:

1. **Technical Indicators**

   Extracted from TradingView for the period **2015–2025**, including RSI, MACD, EMA(20 and 50),Bollinger Bands, and Volume.

2. **Market Sentiment Proxy (VIX)**

   The CBOE Volatility Index (VIX) is included to capture market-wide fear, uncertainty, and volatility.

3. **Twitter Sentiment (Pre-Scored Kaggle Data)**

   Twitter sentiment scores and tweet volumes for the period **30 September 2021 – 29 September 2022**, obtained from the Kaggle dataset *"Stock Tweets for Sentiment Analysis and Prediction"* (Equinxx, 2023). These scores were **already computed** by dataset providers.

The study focuses specifically on **technology-sector companies**—Apple (AAPL), Google (GOOG), Amazon (AMZN), Microsoft (MSFT), and Tesla (TSLA). These firms were chosen because they are:

- **Fundamentally strong and highly liquid**,
- **Characterized by high volatility**, which makes them suitable for deep learning forecasting,
- **Heavily discussed on social media and news platforms**, resulting in richer sentiment information.

To evaluate the effect of sentiment on predictive performance, the study adopts a **dual-dataset experimental design**:

- **Model A (Technical-only model):**

  Technical indicators + VIX (2015–2025)
- **Model B (Sentiment-enhanced model):**

  Technical indicators + VIX + Twitter sentiment (2021–2022)

This design allows for a direct comparison between traditional technical models and sentiment-augmented models.

Additionally, this research benefits from the guidance of Mr. Mohammed Agbawi, the CEO of Sindbad.Tech, who serves as a co-advisor. His expertise provides valuable practical insight into financial forecasting and strengthens the applied relevance of the project

References

Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv*. https://doi.org/10.48550/arXiv.1908.10063

Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, *183*(2), 181–192. https://doi.org/10.1016/j.jeconom.2014.05.008

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Chicago Board Options Exchange (CBOE). (2025). *CBOE Volatility Index: VIX*. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/VIXCLS

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

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. https://doi.org/10.1016/j.ejor.2017.11.054

Greyling, T., & Rossouw, S. (2022). Twitter sentiment and stock market movements: The predictive power of social media. *VoxEU CEPR*. https://cepr.org/voxeu/columns/twitter-sentiment-and-stock-market-movements-predictive-power-social-media

Kumar, A., Das, S. R., & Chattopadhyay, M. (2025). Significance of predictors: Revisiting stock return predictions using explainable AI. *Annals of Operations Research*. https://doi.org/10.1007/s10479-025-06717-2

Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. *Complexity*, *2020*, Article 6622927. https://doi.org/10.1155/2020/6622927

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates, Inc.

MarketPsych. (2023). *Refinitiv MarketPsych indices: Whitepaper*. Refinitiv, LSEG Group. https://www.refinitiv.com

Mittal, A., & Goel, A. (2011). Stock prediction using Twitter sentiment analysis. *Stanford CS229 Course Project*. http://cs229.stanford.edu

Pagolu, V. S., Challa, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System* (pp. 1345–1350). IEEE. https://doi.org/10.1109/SCOPES.2016.7955659

RavenPack. (2022). *Media sentiment index documentation*. RavenPack Analytics. https://www.ravenpack.com

Raza, A., Bukhari, S., Asghar, M. Z., Bangash, J. I., & Hussain, A. (2024). Enhancing profitability and investor confidence through interpretable AI models for investment decisions. *arXiv*. https://doi.org/10.48550/arXiv.2312.16223

Shi, Z., Hu, Y., Mo, G., & Wu, J. (2023). Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction. *arXiv*. https://doi.org/10.48550/arXiv.2204.02623

Stock tweets for sentiment analysis and prediction [Dataset]. (2023). Kaggle. https://www.kaggle.com/datasets/equinxx/stock-tweets-for-sentiment-analysis-and-prediction

TradingView. (2025). *TradingView financial data and indicators documentation*. https://www.tradingview.com

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2022). A comprehensive review on multiple hybrid deep learning approaches for stock prediction. *Intelligent Systems with Applications*, *16*, Article 200111. https://doi.org/10.1016/j.iswa.2022.200111

Zhang, J., Chan, W., & Lin, Y. (2024). Stock price prediction research based on CNN-LSTM. *Highlights in Business, Economics and Management*, *41*, 667–676. https://doi.org/10.54097/cyqrxb84