Project Methods
Module: COMP8805
Student Name: Abdul Rashid Omeni
Kent Login: aoo60
Date: 6th May 2025

Critical Review of Deep Learning–Based Stock Price Prediction

Predicting stock prices is a notoriously tricky task due to financial markets' chaotic, time-dependent nature (Malkiel, 1989; Jin et al., 2017). The problem goes beyond mere fluctuations in daily or intraday prices: stock movements may reflect more extensive macroeconomic variables, industry-wide events, company fundamentals, or even investor psychology (Ding et al., 2015). Such interplay of shifting factors leads to nonlinear and rapidly evolving patterns that many conventional approaches (e.g., linear regression, ARIMA) often fail to capture (Box et al., 2015). In response, deep learning models have emerged as powerful contenders, promising better handling of complex, high-dimensional data and demonstrating an aptitude for unearthing subtle patterns in time series (Bengio et al., 2015).

Complexity of Stock Market Data

The nonlinearity and volatility in stock data present a formidable modelling challenge (Zhang et al., 2017). Price series often exhibit abrupt changes attributable to diverse triggers, such as earnings announcements, geopolitical tensions, technological innovations, or viral social media sentiments (Selvin et al., 2017). Traditional time-series models like ARIMA or GARCH assume stationarity or limited memory, which might hold temporarily in stable regimes but struggle when structural breaks or regime shifts occur (Box et al., 2015). As a result, short-term or local cycles are often mis-specified by these “one-size-fits-all” parametric approaches. Moreover, an individual stock is rarely self-contained; it may be influenced by correlated equities, sector-specific indices, commodity prices, and economic signals (Chen et al., 2017). Realistically, stock movements represent supply-demand dynamics, investor sentiment, reaction to news, and macro-trends. This interplay underscores a need for multi-factor approaches to unearth hidden relationships among these correlated variables (Li et al., 2018). Thus, modelling stock price fluctuations effectively requires algorithms flexible enough to handle multiple inputs and dynamic sufficient to track the system’s ever-changing equilibrium (Qin et al., 2017).

The Rise of Deep Neural Networks

Deep learning offers a framework where numerous layers of interconnected neurons can map highly nonlinear functions, capturing latent signals that might be invisible to shallow or purely linear models (Bengio et al., 2015). One impetus for using deep networks is their ability to learn relevant features from raw data automatically. In earlier machine learning methods, such as Support Vector Machines or Random Forests, the user often invests substantial effort in feature engineering or domain-specific transformations (Chang and Lee, 2017). By contrast, deep architectures theoretically "discover" these transformations through training, provided sufficient data and robust regularization exist.

Recent empirical evidence has shown that carefully tuned deep models can outperform older paradigms in stock trend classification, volatility forecasting, or price prediction (Heaton et al., 2017; Sunny et al., 2020). Success, nonetheless, demands large-scale and up-to-date data, specialized model designs (e.g., recurrent or convolutional layers), and extensive hyperparameter tuning. As financial markets are prone to unanticipated events—such as pandemic-era disruptions—models might quickly become obsolete, so continuous retraining is essential (Qin et al., 2017).

LSTM and Long-Term Dependencies

Among the deep models, **Long Short-Term Memory (LSTM)** networks have proven especially prevalent for sequential data like stock time series (Hochreiter and Schmidhuber, 1997). LSTMs were devised to mitigate the vanishing/exploding gradient problem, a key limitation of classic Recurrent Neural Networks (RNNs). By incorporating memory cells and gating functions (input, forget, and output gates), an LSTM can "decide" which older information to remember or discard (Greff et al., 2017).

In stock forecasting, the LSTM's gating structure can glean cyclical patterns (e.g., quarterly earnings cycles) or longer-term investor sentiment shifts (Fu et al., 2016). Studies such as Jia (2016) and Ding et al. (2015) report that LSTMs often surpass basic RNNs or feed-forward networks in predictive accuracy, particularly for multi-week or monthly horizons. However, LSTMs are not without flaws. They can overfit historical anomalies if the environment changes drastically. For example, pre-2019 air travel patterns might not reliably predict airline stocks post-2020 due to the pandemic shift in travel behaviour (Chen and Hao, 2017). Thus, while LSTMs offer advanced memory capabilities, they still need to incorporate mechanisms—like dropout, adaptive learning rates, or rolling retraining—to remain agile in fast-shifting market conditions (Qin et al., 2017).

Another extension is **Bi-Directional LSTM (BI-LSTM)**, which processes a sequence forward and backward, theoretically capturing context from both preceding and succeeding time points (Althelaya et al., 2018). Although "future" data is not normally available for real-time predictions, BI-LSTMs can be valuable for retrospective analyses or refining certain pattern-detection tasks. For actual day-to-day forecasting, only the forward pass is typically used (Sunny et al., 2020). Even so, the popularity of these memory-based networks underscores the ongoing interest in deeper, more flexible recurrent models that might glean causality or synergy across varied time intervals.

Multi-Factor and Attention-Based Approaches

Modern forecasting pipelines do more than feed a single price series into an LSTM. They might integrate multiple data streams—related equities, sector indices, futures, or textual sentiment (Ding et al., 2015). By combining these signals, the network can cross-check which exogenous variables consistently move in tandem with the target stock (Li et al., 2018). For instance, Sunny et al. (2020) found that incorporating macro indicators (like inflation or unemployment data) improved prediction accuracy for daily close prices in a regional market.

Yet, injecting a large variety of features can also hamper the model by introducing noise (Chen et al., 2017). An attention mechanism can mitigate this risk. **Attention-based LSTM** networks weigh each factor or time step differently, focusing the model's capacity on the most predictive inputs. Qin et al. (2017) introduced a "dual-stage" attention approach, learning which time windows are relevant and which input variables are pertinent at each step. This methodology often boosts performance, especially in multi-step forecasts or chaotic domains like the stock market. While these architectures show promise, their interpretability remains a work in progress—traders and regulators often request transparent rationales behind neural predictions, complicating the adoption of black-box attention networks (Heaton et al., 2017).

CNN with Sliding Window for Short-Term Patterns

Although RNN and LSTM variations typically dominate sequential modelling, some studies highlight **Convolutional Neural Network (CNN)** approaches for short-term, local patterns (Selvin et al., 2017). CNNs excel at capturing localized features within a fixed-size window and are less prone to forgetting older time steps or becoming confused by them. Instead, they treat each window of data as a "snapshot," extracting prominent signals or changes. This approach can be especially effective during intraday or minute-level data, where abrupt changes matter more than a slow-burn historical trend.

Selvin et al. (2017) and Jia (2016) reported that a CNN-sliding window could outperform RNN-based models when the market environment was highly erratic or featured recurring micro-patterns. However, CNN models might overlook relevant context beyond the set window, such as momentum that has been building up over several months (Zhang et al., 2017). Consequently, a practical approach might combine CNN modules (to detect local anomalies) with LSTM modules (for context-aware memory) or an attention layer that merges both perspectives (Qin et al., 2017).

Data Availability, Overfitting, and Other Challenges

1. **Data Quality Availability**

While deep networks thrive on large volumes of data, stock market data is extremely noisy. Corporate actions (stock splits, dividends), merges, or missing intraday trades can hamper model stability (Li et al., 2018). Additionally, smaller-cap stocks may have limited liquidity, producing patchy data. Balancing data volume against data reliability is crucial for robust model building (Chen and Hao, 2017).

2. **Overfitting Risk**

Overfitting arises when a network learns idiosyncrasies specific to the training set, failing to generalize to future states (Heaton et al., 2017). This risk is amplified in non-stationary environments: a model that performs well under one regime might become inaccurate if interest rates or global conditions shift. Techniques such as dropout layers, regularization, or early stopping are standard mitigations, but ongoing validation on the latest data remains essential (Qin et al., 2017).

3. **Interpretability and Regulatory Concerns**

The "black-box" nature of many deep learning methods raises concerns about accountability and transparency—particularly in high-stakes financial contexts. Regulators and institutional investors often demand interpretable or explainable systems. Some solutions involve post-hoc methods like LIME (Local Interpretable Model-Agnostic Explanations), although these remain approximate (Heaton et al., 2017).

4. **Computational Costs**

Training advanced architectures like multi-input attention-based LSTM can be computationally expensive and time-consuming, requiring GPU acceleration or distributed computing frameworks (Selvin et al., 2017). This overhead can hinder frequent model retraining, a necessity when market data changes continually.

Critical Appraisal and Future Directions

Deep learning undeniably offers powerful capabilities for modelling complex, nonlinear financial data. LSTM-based architectures have proven adept at capturing cyclical and persistent patterns, while attention layers or CNN-slicing capture shorter-term fluctuations or abrupt shifts (Ding et al., 2015; Qin et al., 2017; Selvin et al., 2017). Yet there is no single architecture that universally excels: a model that thrives in a bullish, stable market might fail under volatility or unexpected shocks (Chen and Hao, 2017). As markets shift, the deep learning system must adapt—ongoing retraining, feature updates, and hyperparameter tuning become integral parts of the pipeline.

Furthermore, the challenge extends beyond raw predictive power to the underlying strategy or usage scenario. High-frequency trading systems might prioritize microsecond-level data and extremely short horizon predictions, possibly benefiting from CNN windows. On the other hand, a portfolio manager with multi-day or monthly horizons might rely on LSTM memory or external macro signals (Akita et al., 2016). This indicates the necessity of problem-specific customization of deep neural models, rather than a one-size-fits-all approach (Qin et al., 2017).

Another pressing direction is the exploration of **transformer architectures** that rely wholly on attention, discarding recurrent or convolutional layers (Vaswani et al., 2017). Early experiments in language modelling show the superiority of such attention-based approaches for capturing long-range relationships. Whether these techniques will robustly manage real-time finance data is a subject of ongoing research (Qin et al., 2017).

Lastly, interpretability remains an essential factor. While advanced attention-based models can highlight important time steps or input factors, bridging the gap between neural signals and actual trading rationale is an unresolved concern. As institutional usage grows, so will demands for validating and explaining model decisions in compliance with financial regulations (Heaton et al., 2017).

Conclusion

Deep learning provides an appealing solution for the nonlinear, chaotic world of stock price prediction, surpassing many older machine learning and statistical methods in accuracy. Architectures such as LSTMs, Bi-LSTMs, and CNNs with sliding windows each contribute unique strengths: LSTMs excel at retaining relevant historical information, while CNNs capture abrupt short-term shifts. Incorporating multi-factor data and attention further refines predictive power. Yet, the environment in which these networks operate is uncommonly volatile. Continuous retraining and robust risk management are essential because a small mismatch between learned historical patterns and new regimes can lead to significant forecast errors. Future research stands to benefit from more interpretative attention mechanisms, hybrid deep learning pipelines, and domain-driven feature engineering, ensuring these models address the persistent unpredictability that defines modern financial markets.

References

- Akita, R., Yoshihara, A., Matsubara, T. and Uehara, K. (2016) 'Deep learning for stock prediction using numerical and textual information', *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1–6.
- Althelaya, K.A., El-Alfy, E.-S.M. and Mohammed, S. (2018) 'Evaluation of bidirectional LSTM for short-and long-term stock market prediction', *2018 9th International Conference on Information and Communication Systems (ICICS)*, pp. 151–156.
- Bengio, Y., Courville, A. and Vincent, P. (2015) 'Representation learning: A review and new perspectives', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), pp. 1798–1828.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015) *Time Series Analysis: Forecasting and Control*. 5th edn. New Jersey: Wiley.
- Chang, Y.-H. and Lee, M.-S. (2017) 'Incorporating markov decision process on genetic algorithms to formulate trading strategies for stock markets', *Applied Soft Computing*, 52, pp. 1143–1153.
- Chen, K., Zhou, Y. and Dai, F. (2015) 'A LSTM-based method for stock returns prediction: A case study of China stock market', *2015 IEEE International Conference on Big Data (Big Data)*, pp. 2823–2824.
- Chen, Q. et al. (2017) 'Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention', *Proceedings of the 40th International ACM SIGIR Conference*, pp. 335–344.
- Chen, Q. and Hao, Y. (2017) 'A feature weighted support vector machine and k-nearest neighbor algorithm for stock market indices prediction', *Expert Systems with Applications*, 80, pp. 340–355.
- Ding, X., Zhang, Y., Liu, T. and Duan, J. (2015) 'Deep learning for event-driven stock prediction', *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2327–2333.
- Fu, R., Zhang, Z. and Li, L. (2016) 'Using LSTM and GRU neural network methods for traffic flow prediction', *31st Youth Academic Annual Conference of Chinese Association of Automation*, pp. 324–328.
- Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R. and Schmidhuber, J. (2017) 'LSTM: A search space odyssey', *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), pp. 2222–2232.
- Heaton, J., Polson, N. and Witte, J. (2017) 'Deep learning in finance', *arXiv preprint arXiv:1602.06561*.
- Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural Computation*, 9(8), pp. 1735–1780.
- Jia, H. (2016) 'Investigation into the effectiveness of long short-term memory networks for stock price prediction', *arXiv preprint arXiv:1603.07893*.
- Jin, F., Wang, W., Chakraborty, P., Self, N., Chen, F. and Ramakrishnan, N. (2017) 'Stock market volatility and learning', *Journal of Finance*, 71(1), pp. 33–82.
- Li, H., Shen, Y. and Zhu, Y. (2018) 'Stock price prediction using attention-based multi-input LSTM', *Proceedings of Machine Learning Research*, 95, pp. 454–469.

- Malkiel, B.G. (1989) ‘Efficient market hypothesis’, *The New Palgrave: Finance*. New York: Norton.
- Qin, Y., Song, D., Cheng, H., Jiang, G. and Cottrell, G.W. (2017) ‘A dual-stage attention-based recurrent neural network for time series prediction’, *arXiv preprint arXiv:1704.02971*.
- Selvin, S., Vinayakumar, R., Gopalakrishnan, E.A., Menon, V.K. and Soman, K.P. (2017) ‘Stock price prediction using LSTM, RNN and CNN-sliding window model’, *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1643–1647.
- Sunny, M.A.I., Maswood, M.M.S. and Alharbi, A.G. (2020) ‘Deep learning-based stock price prediction using LSTM and bi-directional LSTM model’, *Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, IEEE.
- Vaswani, A. et al. (2017) ‘Attention is all you need’, *Advances in Neural Information Processing Systems (NeurIPS)*, 30, pp. 5998–6008.
- Zhang, L., Aggarwal, C. and Qi, G.-J. (2017) ‘Stock price prediction via discovering multi-frequency trading patterns’, *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2141–2149.