ADBA Information Technology in Economics and Business
Cutting-Edge Scientific Solutions

e-ISSN: 3023-8943
RESEARCH ARTICLE
Vol.3 / No.2 / 2026 / pp.85-90
https://doi.org/10.69882/adba.iteb.2026077

An Empirical Comparison of Deep Learning Models for Stock Direction Prediction: Evidence from Delta Air Lines

Alperen Aydın *,[1], **Murat Işık** *,[2] **and Mehmet Ali Yalçınkaya** *,[3]
*Department of Computer Engineering, Faculty of Engineering and Architecture, Kırşehir Ahi Evran University, 40100, Kırşehir, Türkiye.

ABSTRACT In this study, deep learning-based binary classification models were developed and compared in order to predict the closing direction (up/down) of Delta Air Lines (DAL) stock on the next business day. Three feature sets including technical indicators, competitor airline stocks, and market/sector representatives were derived from daily data for the period of April 30, 2015–April 17, 2026. Nine experiments were conducted with Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) architectures. The highest accuracy and F1 score were obtained by the MLP model using only DAL technical indicators (accuracy = 52.74%; F1 = 0.489), whereas the GRU model using only DAL technical indicators produced the marginally highest ROC-AUC (0.525). However, the performance values remained only slightly above the simple comparison benchmark; competitor and market variables did not provide a consistent improvement. The findings indicate that past price/volume-based technical indicators had limited predictive value for next-day direction prediction. Within the examined stock, period, feature space, and experimental design, the results provide no evidence against the weak-form Efficient Market Hypothesis. Thus, the study points to the limits of the examined technical analysis-based approach.

KEYWORDS
Deep learning
Stock direction prediction
LSTM
Gated recurrent unit
Multilayer perceptron
Efficient market hypothesis

INTRODUCTION

Predicting the direction of stock prices in financial markets is an important but difficult problem in terms of investment decisions, risk management, and portfolio construction processes. Daily price movements are affected not only by past price and volume values, but also by news flow, sector expectations, macroeconomic conditions, and investor behaviors. Therefore, stock direction prediction should be addressed as a problem of extracting a short-term and generalizable signal from past observations. The weak-form Efficient Market Hypothesis argues that past price and volume information is reflected in current prices, and therefore it is difficult to obtain consistently superior returns through technical analysis approaches based solely on historical price information (Fama 1970; Malkiel 2003). This study examines this issue on Delta Air Lines (DAL) stock using deep learning models and interprets the results in relation to the weak-form Efficient Market Hypothesis.

Manuscript received: 2 June 2026,
Revised: 8 July 2026,
Accepted: 21 July 2026.

[1]alperen.aydin@ahievran.edu.tr
[2]muratisik@ahievran.edu.tr (**Corresponding author**)
[3]mehmetyalcinkaya@ahievran.edu.tr

Technical analysis indicators are widely used in financial decision support systems by transforming trend, momentum, volatility, and volume information into numerical variables. Moving averages, MACD, RSI, ATR, and Bollinger Bands are among the most frequently used indicators in this context (Appel 2005; Bollinger 2002; Brock *et al.* 1992; Wilder 1978). Deep learning methods, on the other hand, stand out in financial forecasting studies due to their capacity to represent nonlinear relationships (Berat Sezer *et al.* 2019; Fischer and Krauss 2018; LeCun *et al.* 2015). In particular, gated recurrent networks such as LSTM and GRU were developed to model sequential dependencies in time series data (Bengio *et al.* 1994; Cho *et al.* 2014; Hochreiter 1997). However, when financial series have a low signal-to-noise ratio, it is not guaranteed that more complex architectures will provide higher test performance (Berat Sezer *et al.* 2019; LeCun *et al.* 2015; Srivastava *et al.* 2014).

The aim of this study is to show to what extent the closing direction of DAL stock on the next business day can be predicted using only DAL technical indicators, competitor airline stocks, and market/sector representatives. For this purpose, MLP, LSTM, and GRU models were compared on three feature sets under the same data split and the same evaluation metrics. Thus, rather than only reporting the best model, the study also makes visible the limits of technical indicator-based deep learning approaches in short-term direction prediction. The remainder of the paper is organized as follows: Section 2 reviews the related literature, Section 3 describes the dataset, Section 4 presents the methodology, Section 5 reports the findings, and Sections 6 and 7 provide the discussion and conclusion.

LITERATURE REVIEW

The literature relevant to this study can be grouped into three strands: the debate on market efficiency and the informational content of past prices, machine learning approaches that frame stock movement as a classification problem, and the more recent wave of deep learning applications in financial forecasting. The question of whether future price movements can be inferred from past market data has been debated since the formalization of the Efficient Market Hypothesis (Fama 1970; Malkiel 2003). Timmermann and Granger (2004) argue that predictability in financial markets tends to be self-limiting: once a profitable pattern becomes widely known, its exploitation erodes the signal that produced it, so that forecasting success is expected to be local and transient rather than persistent.

The empirical evidence on technical analysis is correspondingly mixed. Brock *et al.* (1992) found that simple moving-average and trading-range rules carried predictive content for the Dow Jones Industrial Average, and Lo *et al.* (2000) showed, using nonparametric pattern recognition, that several widely used chart patterns provide incremental information about the conditional distribution of returns, although this does not necessarily translate into profitable trading. Reviewing approximately one hundred modern studies, Irwin (2007) report that positive findings on the profitability of technical trading rules are frequently weakened by data-snooping bias, ex-post rule selection, and transaction costs. The present study is situated within this debate, as its results are interpreted in relation to the weak form of the hypothesis.

A second strand of the literature frames next-period stock movement as a binary classification problem rather than a level-estimation problem. Leung *et al.* (2000) compared classification and level-estimation models on major stock indices and found that direction-based classification models generally led to better trading performance than models forecasting the level of returns, which motivates the direction-prediction design adopted in this study. Kara *et al.* (2011) predicted the daily direction of the Istanbul Stock Exchange National 100 Index using artificial neural networks and support vector machines fed with ten technical indicators and reported average accuracies of around 75% for the neural network model.

Patel *et al.* (2015) showed on Indian stocks and indices that converting continuous technical indicators into trend-deterministic (discretized) signals improved the performance of four machine learning classifiers. Ballings *et al.* (2015) benchmarked ensemble methods against single classifiers for one-year-ahead price direction over a large sample of European stocks and found that random forests ranked among the top performing algorithms. Zhong and Enke (2017) combined dimensionality reduction techniques with neural networks to forecast the daily direction of the S&P 500 ETF and obtained accuracies only modestly above benchmark levels. It should be noted that the comparatively high accuracies in this strand are typically obtained at the index level, over longer horizons, or with discretized inputs, whereas daily direction prediction for an individual stock tends to yield accuracies much closer to 50% (Fischer and Krauss 2018; Krauss *et al.* 2016).

The third strand concerns deep learning. Early surveys of soft computing methods for market forecasting (Atsalakis and Valavanis 2009) already documented the dominance of neural approaches, and more recent systematic reviews (Berat Sezer *et al.* 2019; Bustos and Pomares-Quimbaya 2020; Jiang 2021; Kumbure *et al.* 2022) show that LSTM has become the most frequently used architecture in financial time series forecasting, with technical indicators remaining among the most common input categories (Kumbure *et al.* 2022). On the empirical side, Fischer and Krauss (2018) found that LSTM networks outperformed memoryless models such as random forests and logistic regression for daily directional prediction over the S&P 500 universe, although the advantage weakened in later subperiods; notably, their reported daily directional accuracy of approximately 54% illustrates the narrow margins attainable in this problem. In an earlier study on the same universe, Krauss *et al.* (2016) found that individual deep neural networks were outperformed by gradient-boosted trees and random forests, with profitable signals largely disappearing in recent years.

Nelson *et al.* (2017) and Bao *et al.* (2017) reported favorable results for LSTM-based direction prediction and hybrid wavelet-autoencoder-LSTM frameworks, respectively, while Shen and Shafiq (2020) emphasized the role of extensive feature engineering rather than architecture alone. Chong *et al.* (2017) examined deep feature representations on intraday Korean stock returns and reported only modest gains over autoregressive baselines, concluding that deep learning offers no automatic improvement. More recently, attention-based architectures have also entered this literature: Wang *et al.* (2022) applied a deep Transformer model to the prediction of major stock market indices and reported that it can outperform recurrent baselines in back-testing experiments, while the systematic survey of Olorunnimbe and Viktor (2023) cautions that many reported successes in this field rest on backtesting protocols whose realism and reproducibility vary considerably.

Finally, a growing body of work cautions against expecting complex models to dominate in low signal-to-noise settings. Makridakis *et al.* (2018) showed in a large-scale forecasting competition setting that classical statistical methods outperformed popular machine learning methods across horizons. In the asset pricing domain, Gu *et al.* (2020) found that although machine learning improves out-of-sample return prediction mainly through regularization and the capture of nonlinear interactions, the monthly out-of-sample predictive R-squared for individual stocks remains below one percent, underscoring how weak the exploitable signal is at the level of single securities. These findings are consistent with the argument that more expressive architectures do not guarantee better generalization in financial data (Berat Sezer *et al.* 2019; LeCun *et al.* 2015; Srivastava *et al.* 2014). On the other hand, Kelly *et al.* (2024) have recently argued, both theoretically and empirically, that heavily parameterized models can improve out-of-sample return prediction despite apparent overfitting, so the relationship between model complexity and predictive success in finance remains an open question.

Taken together, the literature shows strong but heterogeneous evidence: reported accuracies vary widely with the target market, aggregation level, forecast horizon, input representation, and validation protocol, and several influential results rely on random data splits or tuning choices that are vulnerable to information leakage. Comparatively few studies offer a controlled comparison of feedforward and recurrent architectures on a single liquid stock under a strictly chronological split, leakage-free scaling, and an identical training protocol. The present study addresses this setting for Delta Air Lines, additionally testing whether competitor and market/sector context variables add predictive value, and interprets the resulting near-benchmark performance in relation to the weak-form Efficient Market Hypothesis (Fama 1970; Timmermann and Granger 2004).

DATASET

Raw data were downloaded from Yahoo Finance in daily OHLCV format. The initial download was performed for the 2010–2026 range; however, due to the starting date of the JETS ETF data, all symbols were aligned on the common trading days of April 30, 2015–April 17, 2026. After removing missing rows resulting from technical indicators and the last-day label, the final modeling data set consisted of 2,736 observations. The target variable is a binary label indicating whether the closing price of DAL stock on day t+1 is higher than on day t; 1 represents an increase, while 0 represents a decrease or a flat closing. The data sets used and the split structure are summarized in Table 1.

Table 1 Summary of Feature Sets and Data Split

Dataset	Content	Features	Total	Train/Val./Test	Seq. Test
Dataset 1	Only DAL technical indicators	22	2,736	1,915 / 273 / 548	528
Dataset 2	DAL + UAL, AAL, LUV + competitor summaries	38	2,736	1,915 / 273 / 548	528
Dataset 3	DAL + competitors + JETS/SPY + difference variables	42	2,736	1,915 / 273 / 548	528

The class distribution is balanced: 1,349 observations were calculated as class-0 (49.31%) and 1,387 observations as class-1 (50.69%). Therefore, no additional class balancing procedure was applied. The training, validation, and test split was not performed randomly; in order to preserve the time series structure, it was divided chronologically as 70%/10%/20%. The closing prices of DAL, competitor stocks, and market/sector representatives on common trading days are given in Figure 1.

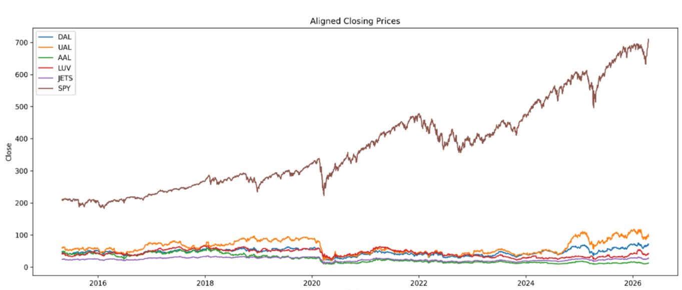

Figure 1 Closing prices aligned according to common trading days

Feature engineering is three-layered. In the first layer, daily return, gap return, intraday high-low range, moving averages, EMA12/EMA26, MACD, RSI-14, ROC-10, Bollinger bandwidth, ATR-14, and volume change ratio were calculated for DAL. In the second layer, daily return, volume change, RSI-14, and MA5/MA20 ratio were derived for UAL, AAL, and LUV stocks; in the third layer, JETS and SPY returns and the difference between DAL return and the competitor and sector averages were added. Normalization was performed using StandardScaler fitted only on the training data, and the same scaler was applied to the validation and test data (Pedregosa *et al.* 2011).

The correlation structure between selected technical and contextual variables is given in Figure 2. The matrix shows that some indicators naturally move together, whereas the added competitor and market variables may not produce a strong discriminative signal on their own for target direction prediction.

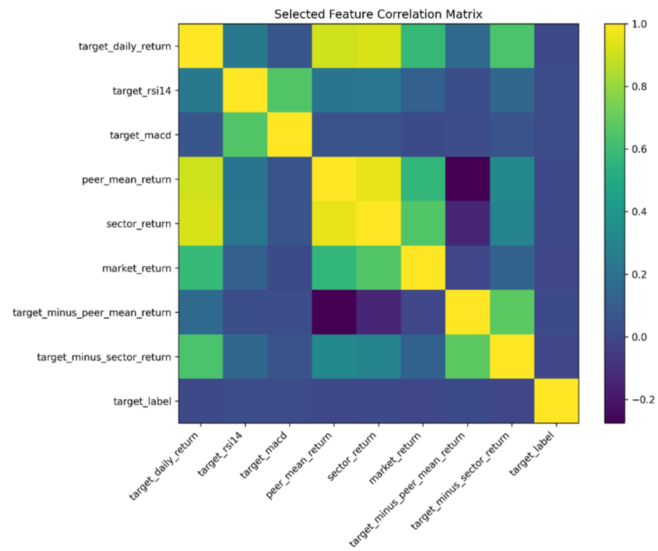

Figure 2 Correlation matrix for selected features

METHOD

Three architectures were compared in the study. MLP is a two-hidden-layer model that classifies the normalized tabular features belonging to the same day without constructing a sequence structure. LSTM is a gated recurrent network that takes a 20-day feature window as input and models temporal dependencies through its memory cell and input/forget/output gates (Bengio *et al.* 1994; Hochreiter 1997). GRU, on the other hand, has a simpler structure than LSTM with its update and reset gates (Cho *et al.* 2014). Thus, a memoryless tabular structure, a more complex memory mechanism, and a simpler gated structure were compared on the same problem.

All experiments were implemented with PyTorch, and the Adam optimizer was used (Kingma and Ba 2014; Paszke *et al.* 2019). The loss function was set as BCEWithLogitsLoss, the decision threshold as 0.50, the learning rate as 0.001, the batch size as 32, the maximum number of epochs as 100, the early stopping patience value as 15, and the random seed as 42. In the MLP, two fully connected layers with 128 and 64 neurons, ReLU activation, and dropout were used; in the LSTM/GRU models, a single recurrent layer with 64 hidden units, dropout, and a fully connected layer with 32 neurons were used. Performance was measured with accuracy, precision, recall, F1, and ROC-AUC metrics. The weights at the epoch with the lowest validation loss were saved, and the test results were reported only with this model. The parameters used and their values are given in Table 2.

FINDINGS

The test results of the nine experiments consisting of three models and three feature sets are given in Table 3. The highest accuracy

Table 2 Training Protocol and Hyperparameters

Parameter	Value
Maximum epochs	100
Lookback window	20 days
Learning rate	0.001
Loss function	BCEWithLogitsLoss
Data split	70% / 10% / 20% (chronological)
Early stopping	15 epochs
Batch size	32
Optimizer	Adam
Feature scaling	StandardScaler
Random seed	42

and F1 score were obtained in Exp-01, in which the MLP was trained only with DAL technical indicators (accuracy = 52.74%; F1 = 0.489). In contrast, Exp-07, which used a GRU with the same DAL-only feature set, produced the marginally highest ROC-AUC (0.525), compared with 0.524 for Exp-01. Nevertheless, all of these values indicate limited discriminative performance in a problem with an approximately balanced class distribution.

When the results in Table 3 are evaluated together in terms of model family and feature set, it is seen that the addition of competitor and market variables did not provide a consistent improvement. The F1 value achieved by the MLP using only DAL data decreased in the broader feature sets. The detailed outputs of Exp-01, which is the best model, also confirm the limited discriminative power. The confusion matrix shows 165 true negatives and 113 false positives for class-0; and 124 true positives and 146 false negatives for class-1. In other words, the model missed a significant portion of the upward days. Based on the same confusion matrix, the class-0 F1 score was 0.560 and the class-1 F1 score was 0.489, resulting in a macro-F1 score of 0.525. The balanced accuracy, calculated as the mean of the class-1 recall (0.459) and the class-0 recall, or specificity (0.594), was 0.526. These class-balanced metrics also indicate that the model had limited discriminative performance across the two classes. This finding is presented in Figure 3 together with the ROC curve. In Figure 4, the loss and accuracy curves of the MLP model are given. While the training loss decreases, the fluctuating course of the validation loss shows that the model could not provide a stable improvement on the validation data. Although the training accuracy increases, the validation accuracy remaining limited supports that the model's generalization power is weak.

Table 3 Test Set Performance Results

Exp.	Model	Feature Set	Acc.	Prec.	Recall	F1	AUC
Exp-01	MLP	Only DAL	52.74	52.32	45.93	0.489	0.524
Exp-02	MLP	DAL + Competitors	50.00	48.94	34.07	0.402	0.500
Exp-03	MLP	DAL + Competitors + Market	47.81	46.85	44.07	0.454	0.467
Exp-04	LSTM	Only DAL	49.81	46.34	22.27	0.301	0.520
Exp-05	LSTM	DAL + Competitors	47.73	43.42	25.78	0.324	0.508
Exp-06	LSTM	DAL + Competitors + Market	51.14	49.40	32.03	0.389	0.516
Exp-07	GRU	Only DAL	48.30	44.14	25.00	0.319	0.525
Exp-08	GRU	DAL + Competitors	51.33	49.68	30.08	0.375	0.511
Exp-09	GRU	DAL + Competitors + Market	49.62	47.47	36.72	0.414	0.488

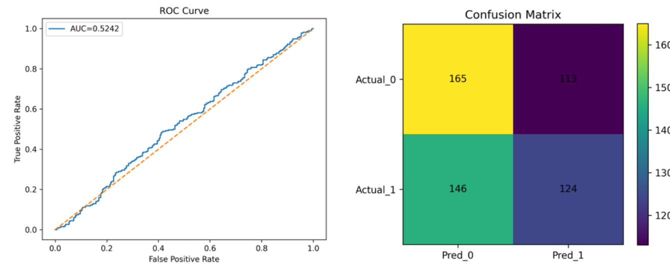

Figure 3 ROC curve and confusion matrix for Exp-01

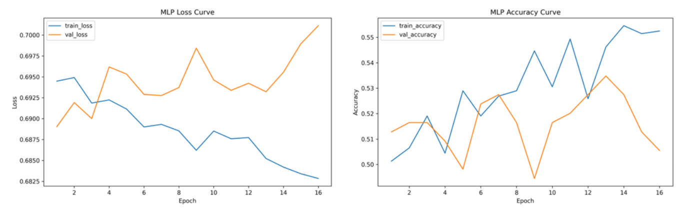

Figure 4 Training loss and accuracy curves for Exp-01

DISCUSSION

The findings provide no evidence against the weak-form Efficient Market Hypothesis within the examined stock, period, feature space, and experimental design. The technical indicators derived from past price and volume information did not strongly distinguish the next-day direction of DAL stock (Fama 1970; Malkiel 2003). This result does not mean that technical indicators are entirely dysfunctional; however, it shows that, within the examined period, target definition, and modeling framework, they did not produce a sufficient prediction signal on their own. In addition, the inability of LSTM and GRU to provide a clear superiority over MLP indicates that complex sequential architectures may not automatically be more successful in financial direction prediction (Berat Sezer *et al.* 2019; LeCun *et al.* 2015; Srivastava *et al.* 2014). When compared with literature, the results indicate differences in conditions rather than contradiction. Fischer and Krauss (Fischer and Krauss 2018) reported that LSTM could provide superiority in some periods within a broad S&P 500 universe; Nelson et al. (Nelson *et al.* 2017) reported the use of LSTM for stock movement direction; and Bao et al. (Bao *et al.* 2017) reported a hybrid framework that supports LSTM with preprocessing and representation learning components. In contrast, studies using news, events, or multi-source data show that additional information domains beyond price history may strengthen prediction (Ding *et al.* 2015; Zhang *et al.* 2018). This comparison is summarized in Table 4.

Table 4 shows that the result of this study should be interpreted not as "deep learning does not work in financial forecasting," but as "short-term direction prediction for DAL remained limited when only technical indicators and nearby market context were used." The addition of competitor stock and ETF variables did not produce information gain; on the contrary, in some experiments, it may have increased noise and model complexity. Therefore, in future studies, richer sources such as rolling-window validation, feature selection, backtests including transaction costs, and news/sentiment data should be tested (Berat Sezer *et al.* 2019; Ding *et al.* 2015; Zhang *et al.* 2018).

Table 4 Comparison of the findings with the literature

Study	Data / Approach	Main Emphasis in the Literature	Relationship with the Present Study
Brock et al. (Brock *et al.* 1992)	Technical trading rules on the DJIA	Shows that some technical rules may carry signals against the random walk.	In the DAL case, technical features did not produce a strong signal in the short term.
Fischer and Krauss (Fischer and Krauss 2018)	S&P 500 constituents, LSTM	LSTM may outperform memoryless models in some periods.	No superiority of LSTM/GRU was observed for a single stock in the 2015–2026 period.
Nelson et al. (Nelson *et al.* 2017)	LSTM for stock direction	Direction prediction based on price history is possible, but the problem is difficult.	In this study, LSTM recall remained low.
Bao et al. (Bao *et al.* 2017)	WT + SAE + LSTM hybrid framework	Hybrid preprocessing/representation learning may improve performance.	Standard technical indicators and a simple LSTM were not sufficient.
Ding et al. (Ding *et al.* 2015); Zhang et al. (Zhang *et al.* 2018)	News/event/social media and multi-source data	Additional information sources may contribute to the prediction of price movement.	News, sentiment, and macro data should be added in future studies.

CONCLUSION

In this study, MLP, LSTM, and GRU models were systematically compared on three different feature sets in order to predict the closing direction of DAL stock on the next business day. The highest accuracy and F1 score were obtained by the MLP model trained only with DAL technical indicators (accuracy = 52.74%; F1 = 0.489), whereas the GRU model using the same DAL-only feature set produced the marginally highest ROC-AUC (0.525). However, these results indicated limited predictive power across all examined models. The addition of competitor stock and market/sector ETF variables did not provide a consistent improvement, and the LSTM and GRU models did not show a clear superiority over MLP. As a result, past price/volume-based technical indicators did not provide strong predictive performance on their own in the examined short-term direction prediction setting. Within the examined stock, period, feature space, and experimental design, the findings provide no evidence against the weak-form Efficient Market Hypothesis.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Appel, G., 2005 *Technical analysis: power tools for active investors*. FT Press.

Atsalakis, G. S. and K. P. Valavanis, 2009 Surveying stock market forecasting techniques–part ii: Soft computing methods. Expert Systems with Applications **36**: 5932–5941.

Ballings, M. *et al.*, 2015 Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications **42**: 7046–7056.

Bao, W., J. Yue, and Y. Rao, 2017 A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS ONE **12**: e0180944.

Bengio, Y., P. Simard, and P. Frasconi, 1994 Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks **5**: 157–166.

Berat Sezer, O., M. Ugur Gudelek, and A. Murat Ozbayoglu, 2019 Financial time series forecasting with deep learning: A systematic literature review: 2005-2019. arXiv e-prints p. arXiv:1911.13288.

Bollinger, J., 2002 *Bollinger on Bollinger bands*. McGraw-Hill, New York.

Brock, W., J. Lakonishok, and B. LeBaron, 1992 Simple technical trading rules and the stochastic properties of stock returns. The Journal of Finance **47**: 1731–1764.

Bustos, O. and A. Pomares-Quimbaya, 2020 Stock market movement forecast: A systematic review. Expert Systems with Applications **156**: 113464.

Cho, K. *et al.*, 2014 Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Chong, E., C. Han, and F. C. Park, 2017 Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications **83**: 187–205.

Ding, X. *et al.*, 2015 Deep learning for event-driven stock prediction. In *IJCAI*.

Fama, E. F., 1970 Efficient capital markets: A review of theory and empirical work. The Journal of Finance **25**: 383–417.

Fischer, T. and C. Krauss, 2018 Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research **270**: 654–669.

Gu, S., B. Kelly, and D. Xiu, 2020 Empirical asset pricing via machine learning. The Review of Financial Studies **33**: 2223–2273.

Hochreiter, S., 1997 Long short-term memory. Neural Computation MIT-Press .

Irwin, S. H., 2007 What do we know about the profitability of technical analysis? Journal of Economic Surveys .

Jiang, W., 2021 Applications of deep learning in stock market prediction: recent progress. Expert Systems with Applications **184**: 115537.

Kara, Y., M. A. Boyacioglu, and Ö. K. Baykan, 2011 Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. Expert Systems with Applications **38**: 5311–5319.

Kelly, B., S. Malamud, and K. Zhou, 2024 The virtue of complexity in return prediction. The Journal of Finance **79**: 459–503.

Kingma, D. P. and J. Ba, 2014 Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Krauss, C., X. A. Do, and N. Huck, 2016 Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. Technical report, FAU Discussion Papers in Economics.

Kumbure, M. M. *et al.*, 2022 Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications **197**: 116659.

LeCun, Y., Y. Bengio, and G. Hinton, 2015 Deep learning. Nature **521**: 436–444.

Leung, M. T., H. Daouk, and A.-S. Chen, 2000 Forecasting stock indices: a comparison of classification and level estimation models. International Journal of Forecasting **16**: 173–190.

Lo, A. W., H. Mamaysky, and J. Wang, 2000 Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. The Journal of Finance **55**: 1705–1765.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos, 2018 Statistical and machine learning forecasting methods: Concerns and ways forward. PLoS ONE **13**: e0194889.

Malkiel, B. G., 2003 The efficient market hypothesis and its critics. Journal of Economic Perspectives **17**: 59–82.

Nelson, D. M., A. C. Pereira, and R. A. De Oliveira, 2017 Stock market's price movement prediction with LSTM neural networks. In *2017 International Joint Conference on Neural Networks (IJCNN)*, IEEE.

Olorunnimbe, K. and H. Viktor, 2023 Deep learning in the stock market—a systematic survey of practice, backtesting, and applications. Artificial Intelligence Review **56**: 2057–2109.

Paszke, A. *et al.*, 2019 Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems **32**.

Patel, J. *et al.*, 2015 Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications **42**: 259–268.

Pedregosa, F. *et al.*, 2011 Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research **12**: 2825–2830.

Shen, J. and M. O. Shafiq, 2020 Short-term stock market price trend prediction using a comprehensive deep learning system. Journal of Big Data **7**: 66.

Srivastava, N. *et al.*, 2014 Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research **15**: 1929–1958.

Timmermann, A. and C. W. Granger, 2004 Efficient market hypothesis and forecasting. International Journal of Forecasting **20**: 15–27.

Wang, C. *et al.*, 2022 Stock market index prediction using deep transformer model. Expert Systems with Applications **208**: 118128.

Wilder, J. W., 1978 *New concepts in technical trading systems*. Greensboro, NC.

Zhang, X. *et al.*, 2018 Stock market prediction via multi-source multiple instance learning. IEEE Access **6**: 50720–50728.

Zhong, X. and D. Enke, 2017 Forecasting daily stock market return using dimensionality reduction. Expert Systems with Applications **67**: 126–139.

How to cite this article: Aydin A., Isik, M., and Yalcinkaya, M. A. An Empirical Comparison of Deep Learning Models for Stock Direction Prediction: Evidence from Delta Air Lines. *Information Technology in Economics and Business*, 3(2), 85-90, 2026.