






RESEARCH ARTICLE

Forecasting Minute-Level Stock Prices with Denoised Data: A Comparative Study of Developed vs. Developing Financial Markets.

[version 1; peer review: 1 not approved]

Hasan Mohammed Sami , Monjurul Hoque Bhuiyan , Dibbo kundu, Sohrab Ahmed 

Accounting & Finance, North South University, Dhaka, Dhaka Division, 1212, Bangladesh

V1 First published: 24 Apr 2026, 15:613
<https://doi.org/10.12688/f1000research.178859.1>

Latest published: 24 Apr 2026, 15:613
<https://doi.org/10.12688/f1000research.178859.1>

Abstract

Background

Developing economies offer compelling investment opportunities but are often overlooked due to perceived instability. This study challenges the bias toward developed markets by exploring whether deep learning, applied to denoised high-frequency data, can unlock superior intraday returns in developing economies.

Methods

We compared three developed markets (the USA, Japan, and Singapore) against three developing markets (India, Brazil, and Malaysia). A standardised pipeline retrieved minute-level data for ~20 stocks per market, selected via K-means clustering based on market capitalisation and beta. Final portfolio inclusion was validated based on model performance (adjusted R^2 scores and minimal overfitting). Data was denoised using wavelet transforms, autoencoders, and Kalman filters sequentially, with hyperparameters optimised via Optuna to filter microstructure volatility through noise. We trained multiple deep learning architectures, including LSTM, Bi-LSTM, GRU, TCN, Transformers, Conv1D-LSTM and CNN-LSTM, employing walk-forward validation to forecast the next 150-minute prices.

Results

The models showed high predictive power, with adjusted R^2 values

Open Peer Review

Approval Status 

1

version 1

24 Apr 2026


[view](#)

1. **Ewerton Alex Avelar**, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Any reports and responses or comments on the article can be found at the end of the article.

ranging from 78% to 95% (average 85%) and minimal overfitting (< 10%). In simulated investing portfolios weighted by volatility, developing markets beat developed markets by a significant margin. Developing markets had Sharpe ratios 1.2–1.6 times those of their developed brethren, with buy-side Sharpe ratios in the 1.9–5.3 range, compared to 1.0–1.4 for developed markets. Through rigorous walk-forward validation and simulated intraday trading on denoised minute-level data, this research empirically demonstrates that developing markets deliver Sharpe ratios 1.2–1.6 times higher than those of developed markets, validating their superiority for exploiting inefficiencies in high-frequency forecasting and alpha generation.

Conclusions

Deep learning models are a good way to exploit the inefficiencies and volatility in developing markets. These findings suggest that computational development can help overcome perceived risks associated with validating and developing markets as a reliable source of alpha for global investment strategies.

Keywords

deep learning; stock price forecasting; minute-level data; Optuna-based denoising; developing markets; Sharpe ratio; LSTM for forecasting; Adaptive Market Hypothesis.

Corresponding author: Hasan Mohammed Sami (hasan.sami@northsouth.edu)

Author roles: **Sami HM:** Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Supervision, Validation; **Bhuiyan MH:** Data Curation, Project Administration, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **kundu D:** Data Curation, Investigation, Project Administration, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Ahmed S:** Project Administration, Resources

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2026 Sami HM *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Sami HM, Bhuiyan MH, kundu D and Ahmed S. **Forecasting Minute-Level Stock Prices with Denoised Data: A Comparative Study of Developed vs. Developing Financial Markets. [version 1; peer review: 1 not approved]** F1000Research 2026, 15:613 <https://doi.org/10.12688/f1000research.178859.1>

First published: 24 Apr 2026, 15:613 <https://doi.org/10.12688/f1000research.178859.1>

Abbreviations

AMH: Adaptive Market Hypothesis
 Bi-LSTM: Bidirectional Long Short-Term Memory
 CNN: Convolutional Neural Network
 DL: Deep Learning
 ERP: Equity Risk Premium
 GRU: Gated Recurrent Unit
 LSTM: Long Short-Term Memory
 RMSE: Root Mean Square Error
 SNR: Signal-to-Noise Ratio
 TCN: Temporal Convolutional Network

1. Introduction

Global investors traditionally prefer mature, liquid, and information-efficient developed equity markets such as the US, Japan, and Singapore. In contrast, developing markets (India, Brazil, Malaysia, etc.) often do not find their way into global portfolios, as they are frequently overlooked due to volatility, currency risks, and perceived informational inefficiencies.^{71–73} This “flight to safety” persists despite the Adaptive Market Hypothesis, which suggests that some of the very inefficiencies and volatility clustering cited as risks in developing markets may create exploitable short-term patterns for sophisticated traders.^{1,2}

The challenge of exploiting these patterns is magnified at high frequencies. Minute-level stock data is inherently non-stationary and plagued by a low signal-to-noise ratio due to microstructure effects, such as bid-ask bounces. These issues are acute in developing economies, which tend to have thinner liquidity and external shocks play a bigger role; this can result in larger forecasting errors. Yet, sophisticated architectures of deep learning algorithms (from Long Short-Term Memory (LSTM) networks to Transformers and CNN-LSTM hybrid networks) do a great job of capturing nonlinear patterns in such noisy data. By exploiting mechanisms such as gating units (in LSTMs/GRUs) for long-term dependencies, dilated convolutions (in TCNs) for extended receptive regions and self-attention (in Transformers), these technologies have the potential to turn the perceived risks of insufficiently efficient environments into a trustworthy source of alpha.^{3,4,74}

This research addresses an important gap in the literature: no other study has systematically compared whether deep learning, when applied to denoised minute-level data, produces better risk-adjusted intraday returns in developing versus developed markets using an identical methodological pipeline. We provide the first controlled analysis across matched large-cap stocks, isolating pure price signals (excluding traditional technical indicators) and constructing distinct long (buy) and short (sell) portfolios to empirically challenge the persistent investor bias toward developed markets.

To ensure a rigorous comparison, our pipeline employs Optuna-optimised denoising (incorporating wavelet transforms, autoencoders, and Kalman filters) to filter microstructure noise before training models with walk-forward validation. We specifically chose India, Brazil, Malaysia as developing markets due to their balanced liquidity and stability evidenced by the strong inflows of FDI (e.g., US\$81.04 billion for India during FY2024–25; ~US\$84.1 billion for Brazil through Nov 2025; strong inflows for Malaysia) and stable projections of GDP growth (e.g., ~6.6% for India, ~2.5% for Brazil, ~4.4% for Malaysia).^{3–6,9} We deliberately omitted other developing regions (e.g., sub-Saharan Africa - e.g. South Africa; smaller Asian economies - e.g., the Philippines) because the literature shows that they have much higher risks (e.g., political instability and low FDI inflows; e.g., US\$8.9 billion for the Philippines in 2024) and this could be a source of bias in drawing comparisons (e.g.,^{7,8}). This selection enables fair, high-frequency comparisons with developed economies.

In sum, this study shows that Deep Learning based on denoised data at the min level can leverage inefficiencies in developing markets to deliver superior risk-adjusted returns at the intraday level vis-a-vis developed markets, challenging the conventional bias among investors that under-utilises opportunities to generate alpha. By exploiting the natural volatility of these markets with sophisticated computer methods, the results highlight the possibility of liquid emerging large-cap stocks as useful sources of diversification and enhanced returns within global portfolios. This work not only fills the gap between theoretical market inefficiencies and practical trading strategies but also offers a replicable framework for geo traders to exploit short-term opportunities in less efficient environments.

2. Research objectives

To address the prevailing investor bias and validate the efficacy of deep learning in these regions, this study poses the following research objectives:

- a) **To investigate the behavioral disparity in capital allocation:** We aim to understand why investors prioritize developed nations despite the theoretical high-growth potential and exploitable inefficiencies in developing ones. Specifically, we ask: *Why do investors considering investments in developed nations not also consider investments in developing ones?*
- b) **To evaluate the risk-return trade-off via computational methods:** We identify the development markets that have the same or better risk-return trade-offs to the intraday trading when using advanced denoising and deep learning. We ask the following question: *Are the developing countries equal or better off in terms of risk-return trade-offs due to current market conditions, available computational powers and favorable macroeconomic environment?*
- c) **To assess the viability of liquid developing markets:** We examine whether statistically significant markets, such as India, Brazil, and Malaysia, provide viable trading opportunities and diversification benefits comparable to major developed economies. This addresses the question: *Can investors invest in statistically significant markets like India, Brazil, and Malaysia, as well as similar, statistically stable, and opportune equity markets, and expect comparable returns or trading opportunities?*

These objectives highlight the importance of empirically analysing the interaction between market microstructure and deep learning, which is the main motivation for this work.

3. Research motivation and contribution

Motivated by the juxtaposition of the high growth potential of developing markets with their underrepresentation in global portfolios, this work aims to empirically challenge the “flight to safety” bias. Our main contribution is to show that it is possible to systematically extract alpha from these environments using deep learning, and to present solid evidence for the impact of computational advances in reducing the volatility of developing markets and converting it into actionable, high-quality returns.^{3,10} Furthermore, by providing a replicable framework for denoising high-frequency data, this research offers a practical methodology for institutional investors to cross the boundary of complex data challenges in developing economies and effectively bridge the gap between theoretical market inefficiencies and trading utility in the real world.

Our unique value is creating opportunities for intraday trading in developing economies. By applying market-standard statistical methods, we validate that investors can participate in the equity markets of developing economies to secure daily trading returns and effectively expand the global investment horizon beyond traditional developed markets.

4. Literature review

4.1 Theoretical framework: Risk premium and market efficiency

The disparate performance of developed and developing equity markets is one of the oldest and most basic theories of economics. At the centre of this distinction is the idea of the Equity Risk Premium (ERP) - the excess return over the risk-free rate that investors expect to be paid to compensate them for the higher level of systematic risk they are taking.^{5,11-13} In emerging markets (e.g., USA, Japan, Singapore) the risk will be of a higher nature due to several elements such as political instability, currency fluctuation, and liquidity.^{5,11-13} In a more technical manner, while developed markets are characterised by mature institutions and fast information diffusion, developing economies often have persistent inefficiencies and are acutely susceptible to external shocks.¹³⁻¹⁶ A prominent study published confirmed this delusion by showing that emerging markets exhibit much higher volatility clustering during global crises than their developed counterparts.

Recent publications further highlight these dynamics. For instance, a recent causality investigation of stock prices and macroeconomic indicators in the Indian stock market reveals a tight relationship between volatility in developing economies and macroeconomics, which, in turn, influences asset pricing.¹⁷ Similarly, research on behavioural biases in the Indian market shows that psychological factors, for instance, overconfidence and disposition effects, amplify under-allocation to developing markets.^{3,18} Furthermore, recent work on cointegration in the Indian stock market focuses on spillover effects, in line with the idea that the volatility of developing markets can spread but can also provide unique opportunities for diversification.¹⁹

However, the Adaptive Market Hypothesis (AMH) challenges the traditional Efficient Market Hypothesis (EMH) by positing that market efficiency is a dynamic process rather than an absolute state¹ further suggests that inefficiencies, e.g., time delays in price discovery and volatility clustering, are not only risks but also opportunities for exploitable patterns. This view has empirical support, suggesting that in developing markets, weak-form efficiency is very rare and, as a result,

arbitrage opportunities still exist and can be exploited with sophisticated computational models.^{20,21} This framework implies that behavioural biases may cause investors to under-allocate capital to developing markets, where informational inefficiencies could be exploited to generate alpha.^{2,18}

4.2 Deep learning architectures in financial forecasting

Traditional econometric models (e.g., ARIMA, GARCH) have been the standard tools for financial prediction.^{22,58} have extensively detailed the use of ARIMA models for stock trend forecasting. However, these linear models can struggle to account for the complex nonlinearities and regime shifts in high-frequency financial data.²³ Consequently, Deep Learning (DL) has emerged as a great alternative, with its adaptive feature learning and its superior generalization on out-of-sample data.

A recent bibliometric review on artificial intelligence and finance, published by,²⁴ charts this evolution, showing a proliferation of neural network applications for predictive modelling and volatility analysis after 2010. This corresponds to our move towards DL for processing noisy & high frequency data for time series forecasting.

- A. **Sequence Models (LSTM, Bi-LSTM, GRU):** Long Short-Term Memory (LSTM) networks established a benchmark for financial time series by effectively managing long-term dependencies²⁵ demonstrated that LSTMs-based predictions achieved, and the portfolio generated daily returns of 0.46% and a Sharpe ratio of 5.8 (before transaction costs) on S&P 500 data, significantly outperforming memory-free methods, such as random forests. Subsequent research has led to the optimization of such architectures³: showed that Bidirectional LSTMs (Bi-LSTM) can be used to increase context awareness in volatile series, with up to 70% directional accuracy²⁶; showed that Gated Recurrent Units (GRU) give similar performance with converging times of 20–30% faster than Bi-LSTMs.
- B. **Convolutional and Attention-Based Models:** To obviate the sequential processing limitations on recurrent architectures, Temporal Convolutional Networks or TCNs were introduced as a dilated convolution technology to solve the long runtime problems of long-range modelling in parallelizable computations, and could be superior to standard LSTMs in high-frequency tasks.^{27,28} Furthermore, Transformers can use a self-attention mechanism to capture global dependencies across long-term horizons, achieving the lowest Root Mean Square Error (RMSE) in intraday regimes.^{29,30}
- C. **Hybrid Architectures:** Recent surveys name hybrid models, e.g. CNN-LSTM or Conv1D-LSTM, as state-of-the-art. Zhang et al. (2019) showed that CNNs are particularly useful for extracting local features from high-frequency microstructure data (e.g., limit order books). These architectures mix convolutional layers for feature extraction and recurrent layers for sequence modelling, being top-ranked in empirical surveys for accuracy and being less prone to overfitting.^{23,31} However, issues such as overfitting in volatile regimes persist, which basically calls for a good validation technique.

4.3 Signal processing and optuna-optimised denoising

Minute-level financial data is almost always noisy, with microstructure effects, such as bid-ask bounces, that obscure fundamental price trends. To perform effective forecasting, rigorous denoising is necessary to improve the Signal-to-Noise Ratio (SNR).

- A. **Denoising Techniques:** Wavelet transforms are widely used to decompose signals and remove high-frequency irregularities.^{32,33} Furthermore,³⁴ provided influential evidence that integrating stacked auto-encoders (SAEs) with LSTMs significantly improves predictive performance by reconstructing clean price signals via neural compression.³⁵ While effective, these methods require careful parameter tuning to avoid signal loss.
- B. **Optimization with Optuna:** A major improvement to pipelines today is the automated tuning of denoising parameters. Optuna, which uses a Tree-structured Parzen Estimator (TPE), optimises hyperparameters (such as wavelet thresholds and latent dimensions) 30%–50% faster than grid search.³⁶ This guarantees reproducible and robust noise reduction, resulting in clearer signals for subsequent deep learning models. Critically, empirical review studies indicate that denoising optimisation strategies can reduce RMSE by 15–30% in noisy financial series, although the effectiveness of optimal denoising strategies can vary across market conditions.³⁷

4.4 Overfitting in time series forecasting

Overfitting is a major issue for the prediction of financial time series, as models can learn noise instead of generalizable patterns - a particular danger with volatile developing markets.^{10,23} This is especially true of deep learning models for regime shifts. To mitigate this, it is important to implement rigorous validation protocols.

This research follows the standards promoted by,³⁸ using Walk-Forward Validation rather than traditional k-fold cross-validation.⁷⁵ This way, temporal order is preserved without look-ahead bias, and the model is tested on “future” data that does not exist or that it had never seen during training.³⁹ Combined with regularisation techniques inherent in modern architectures (e.g., dropout layers in LSTMs), this validation strategy ensures that the alpha generated is robust and not merely an artefact of overfitting.

4.5 Stock selection and portfolio construction

To ensure a fair and representative comparison across markets, stock selection must account for the varying market structures⁴⁰ demonstrated that **K-means clustering** based on market capitalisation and beta effectively groups stocks by risk-return profiles, improving prediction accuracy by 10–20% compared to random selection.

For portfolio construction and validation, this study adheres to the rigorous standards advocated by,³⁸ employing Walk-Forward Validation rather than traditional k-fold cross-validation to preserve temporal order and prevent look-ahead bias. We employ Inverse Volatility Weighting for portfolio allocation. In high-frequency regimes, estimating stable covariance matrices is notoriously unstable; inverse volatility weighting prioritises risk control, dynamically allocating capital to lower-volatility assets to minimise drawdowns—a critical feature for short-term strategies.^{41,42} Compared to alternatives such as Mean-Variance Optimisation, this method offers greater robustness in volatile, developing markets, though it may underperform in low-volatility environments.

4.6 Gap analysis

Despite extensive literature on financial forecasting (e.g.,⁷⁶), significant gaps remain in the systematic comparison of developed and developing markets using high-frequency data.

Table 1 summarizes the current literature on financial forecasting, pointing out the methodological voids in the literature and indicating how the current paper addresses them by Optuna-tuned denoising pipelines.

5. Methodology

The proposed forecasting and portfolio construction pipeline comprises the following stages, as described by a lucid chart below:

Table 1. Gap analysis of relevant literature.

| Literature section | Summary | Gap identified | Contribution of this study |
|--------------------------------|---|---|--|
| Roy et al. (2025) | Reviews AI trends in finance, noting ML’s role in stock prediction. | Absence of specialized data denoising or Optuna-based hyperparameter tuning. | Integrates denoising and Optuna tuning to significantly enhance predictive accuracy. |
| Chauhan et al. (2025) | Studies macro indicators’ impact on Indian sectoral indices using ARDL. | Relies on traditional statistical methods; lacks focus on Beta-based evaluation for stability. | Emphasizes Beta-based statistical evaluation to enhance market stability and reduce risk. |
| Basireddy et al. (2024) | Develops hybrid ML models for predicting rankings. | Lacks a systematic methodology for Optuna-based hyperparameter optimization. | Implements Optuna-based tuning to ensure models achieve maximum predictive capability. |
| Ali et al. (2023) | Examines cointegration and causality between Indian and global markets. | Fails to address simulation risk mitigation through Beta-similarity in short-term trading. | Addresses Beta-based similarity to mitigate risks specifically within short-term trading simulations. |

The stepwise analysis is explained below:

- a) **Data Collection:** The minute price data from 6 countries (3 developed and 3 developing countries) for the last 30 days are collected.
 - Company Evaluation: K-Means clustering is used strictly to segregate companies with high market capitalisation and good market representation, with closeness of Beta to 1.
- b) **Price denoising & signal generation:** Selected companies are then fed into the denoising process to generate predictive pattern price signals, which could further be trained in DL (Deep Learning) models for price forecasting accuracy.
- c) **Prediction Model:** The denoised price signals are then fed into various DL methods with Optuna-defined hyperparameter optimisation that assures the best possible training accuracy for future signals.
- d) **Validation of Price Dataset:** We finally employ a selected set of future price datasets, which are tested in segments of 10 batches for 30 prices in each batch, for error-based walk forward prediction and finally correction of prediction based on corrected features.
- e) **Finally,** the predicted prices with the highest adjusted R^2 scores and low overfitting scores are selected for accuracy testing, which are used by price timestamp in portfolio creation for generating the best possible portfolio.

The overall research process involves support of the lowest possible volatility by two sequential steps:

1. The Denoising process, it reduces high-frequency signals through Optuna-based optimisation.
2. By DL based prediction process, the method ensures that Optuna-based hyperparameter optimisation only trains on low-frequency signals, which assures the best possible correction through walk-forward validation.

The research also assures that after successful price prediction confirmation of the predicted price signals, which are successful by directional accuracy, the predicted timestamp with the highest and lowest price is used for real prices for portfolio development, and both the increasing price options are used for buy and the decreasing price options are used for sell portfolio. Furthermore, an inverse price volatility-based allocation further assures risk reduction rather than profit maximisation, which also seems effective across developed and developing country investment environments.

Figure 1 shows the various steps of the extensive, step-by-step methodology used in this study, from initial data gathering and K-means clustering to signal generation and walk-forward validation.

A. Data collection and stock selection

We first assemble the dataset by targeting the six equity markets of interest. Using Yahoo Finance, minute-level OHLC and volume data are downloaded for each market's primary exchange. Within each country, we select the 20 largest companies by market cap and download their latest price series. To ensure representative sampling of market movements, the 30-day rolling beta of each stock relative to its relevant index (e.g., S&P 500, N225, STI) is computed, and stocks with high or low beta are noted. Only large-cap stocks with full data having consistent trading histories are retained for analysis.⁴⁰

Table 2 shows the balanced portfolio of high-liquidity stocks, selected across developing and developed markets (USA, Japan, Singapore, India, Brazil, Malaysia), and provides details of the tickers and their respective market cap ranges in each, which form the basis for the modelling presented later. Retrieve OHLC (Adjusted close) data at the minute level for a selection of stocks through Yahoo Finance and ensure that approx. 30 trading days are gathered for each stock in the market.

Table 3 gives a small glimpse of the raw, minute-level OHLC prices, showing the high-frequency temporal structure before the implementation of our denoising pipeline:

The method generates information for each of the six target markets (**developed: US, Japan, Singapore; developing: India, Brazil, Malaysia**), shortlists ~20 large-cap stocks by market capitalisation. Compute a 30-day rolling beta against

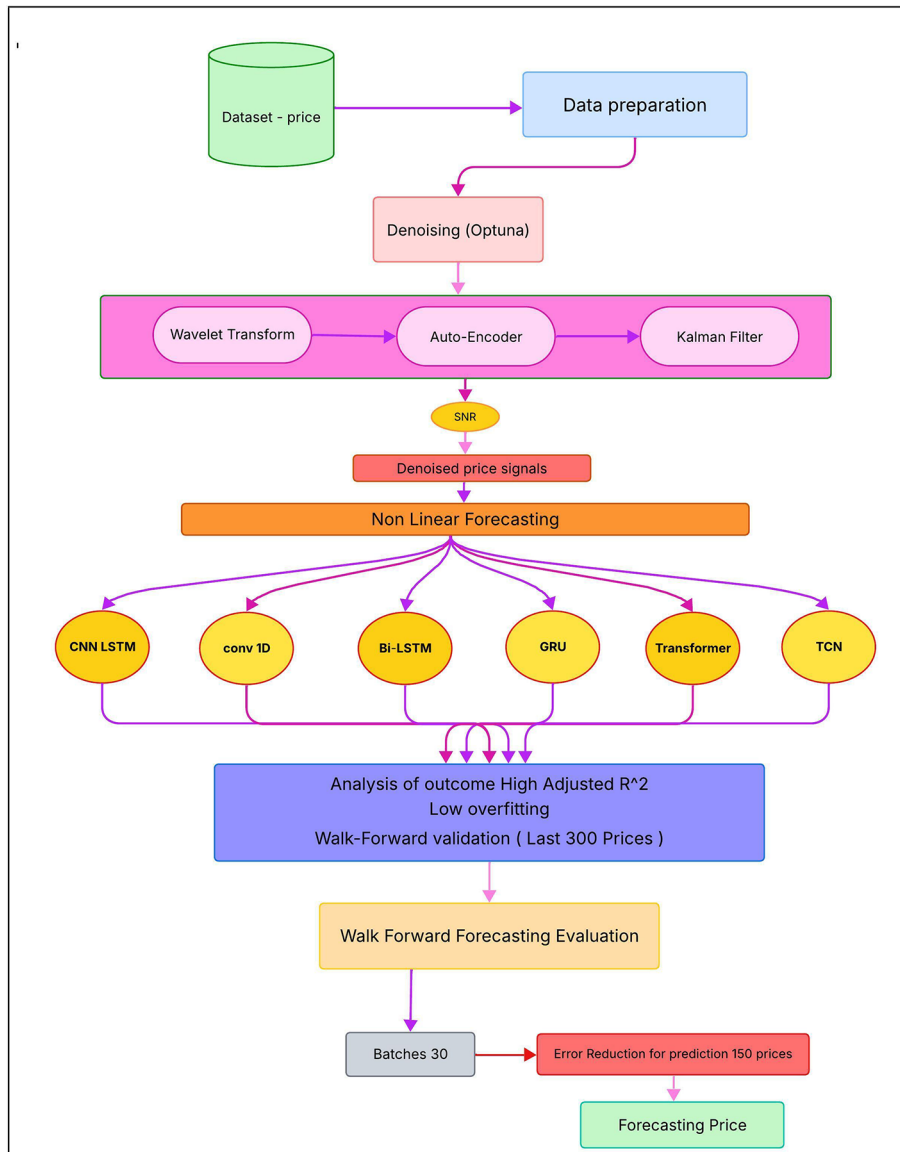


Figure 1. Methodology and forecasting pipeline. This flowchart illustrates the various steps of the extensive step-by-step methodology used in this study, from initial data gathering and K-means clustering to signal generation and walk-forward validation.

the local benchmark index to ensure broad market representation. Only stocks with complete, high-quality data were retained.⁴⁰ It's also observed that, for developing countries, selecting companies with strong financial ratios provides a sound basis for effective financial support and portfolio-based prediction.⁴³ An example of a sequential multi-stage denoising pipeline using Discrete Wavelet Transforms, a Variational Autoencoder, and Kalman Filters to produce a clean prediction is shown in [Figure 2](#).

B. Preprocessing and denoising

Prior to modelling, each raw price series is converted to log-returns and normalised. Minute-level data are notoriously noisy, so we implement a three-stage denoising strategy.

Table 2. Stock selection across markets.

| Market category | Country | Stock ticker | Market cap |
|-----------------|-----------|---|------------------------------------|
| Developed | USA | WBD, VTRS, PCG, KIM, KEY, IVZ, INTC, HPE, HBAN, HAL, F, DOC, AMCR, WBA, NCLH, HST, APA, AES | 8.54B to 177.44B (USD) |
| | Japan | 4043.T, 4324.T, 5332.T, 5411.T, 5631.T, 6724.T, 6762.T, 6841.T, 7011.T, 7267.T, 8031.T, 8058.T, 8253.T, 9021.T, 7203.T, 7272.T, 4543.T, 4704.T, 6702.T | 1.047 T to 15.039 T (JPY) |
| | Singapore | C09.SI, Z74.SI, C2PU.SI, BS6.SI, Z77.SI, V03.SI, U96.SI, U14.SI, S68.SI, AIY.SI, O39.SI, H78.SI, H02.SI, A17U.SI | 2.714B to 75.649B (SGD) |
| Developing | India | COALINDIA, CENTRALBK, IDEA, ITC, IOB, IRB, JPPOWER, LLOYDSENT, MAHABANK, MSUMI, NSLNISP, OLAELEC, PCJEWELLER, PSB, SAGILITY, TATAMOTORS, UCOBANK, UJJIVANSFB, YESBANK, NTPC | 85.503B to 5.27 T (INR) |
| | Malaysia | 7090.KL, 5311.KL, 5302.KL, 0348.KL, 0309.KL, 0276.KL, 8176.KL, 7579.KL, 6399.KL, 5326.KL, 5238.KL, 5185.KL, 5139.KL, 5099.KL, 2488.KL, 0277.KL, 0181.KL | 93.288 M to 26.124B (MYR) |
| | Brazil | TOTS3, SBSP3, PFRM3, PSSA3, EQTL3, MDNE3, VLID3, CURY3, TGMA3, IRBR3, AMBP3 | 1.705B to 874.213B (BRL) |

Table 3. Sample minute-level dataset.

| Timestamp | Ticker - (F) |
|-----------------------------|--------------|
| 2025-05-01 13:30:00 + 00:00 | 10.15009975 |
| 2025-05-01 13:31:00 + 00:00 | 10.17500019 |
| 2025-05-01 13:32:00 + 00:00 | 10.20499992 |
| 2025-05-01 13:33:00 + 00:00 | 10.13370037 |
| 2025-05-01 13:34:00 + 00:00 | 10.11999989 |
| 2025-05-01 13:35:00 + 00:00 | 10.11999989 |
| 2025-05-01 13:36:00 + 00:00 | 10.09500027 |
| 2025-05-01 13:37:00 + 00:00 | 10.10499954 |
| 2025-05-01 13:38:00 + 00:00 | 10.08500004 |
| 2025-05-01 13:39:00 + 00:00 | 10.07499981 |
| 2025-05-01 13:40:00 + 00:00 | 10.02999973 |
| 2025-05-01 13:41:00 + 00:00 | 10.03989983 |

- I. **Denoising:** The study converts raw minute-bar price series to log-returns and normalises. Apply a sequential multi-stage denoising pipeline combining
- (i) **Discrete Wavelet Transform** (Daubechies family with soft-thresholding),
 - (ii) **Variational Autoencoder** (deep learning reconstruction), and
 - (iii) **Kalman Filter** (state-space smoothing).

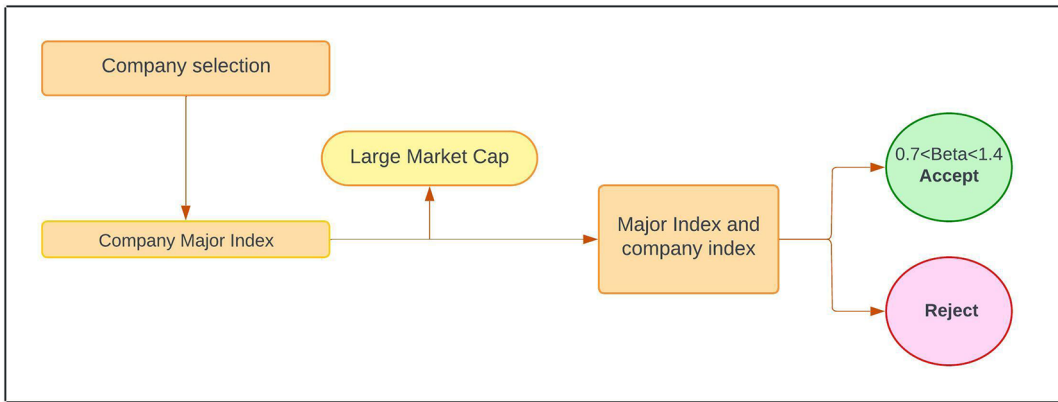


Figure 2. Data preprocessing and denoising flowchart. A conceptual flowchart of the sequential multi-stage denoising pipeline combining Discrete Wavelet Transforms, Variational Autoencoders, and Kalman Filters.

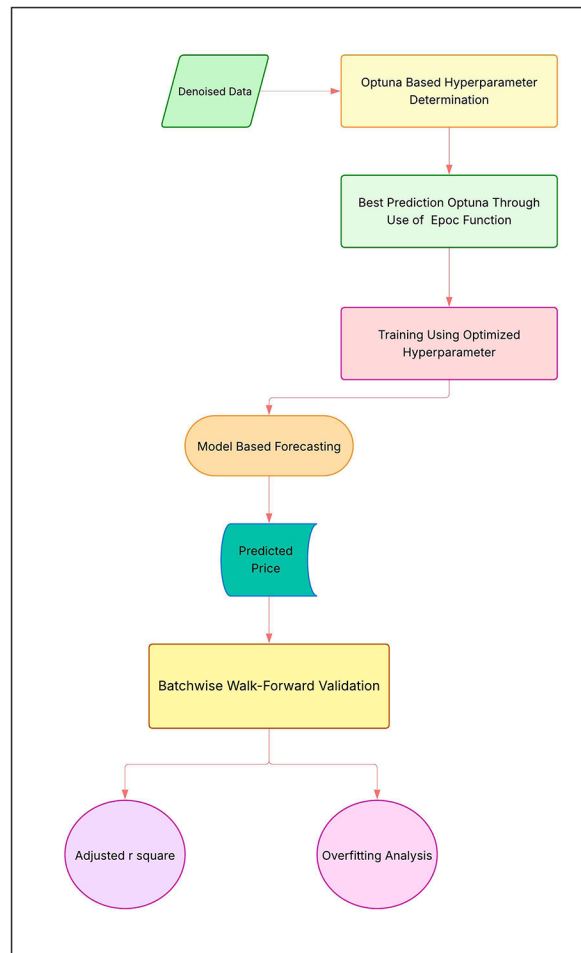


Figure 3. Denoising process of stock price data.

The hyperparameters for each denoising stage (wavelet type, decomposition level, threshold λ ; VAE architecture; Kalman noise covariance) are automatically optimised using the TPE sampler in the Optuna library to maximise the signal-to-noise ratio.³⁶

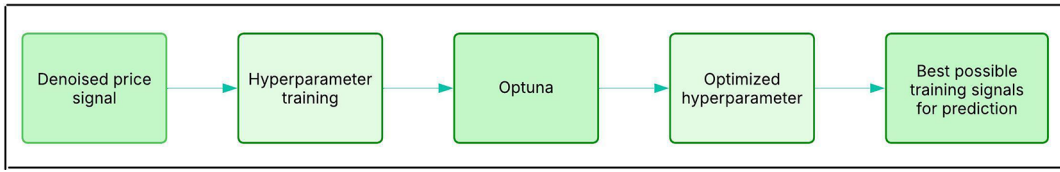


Figure 4. Model training architecture.

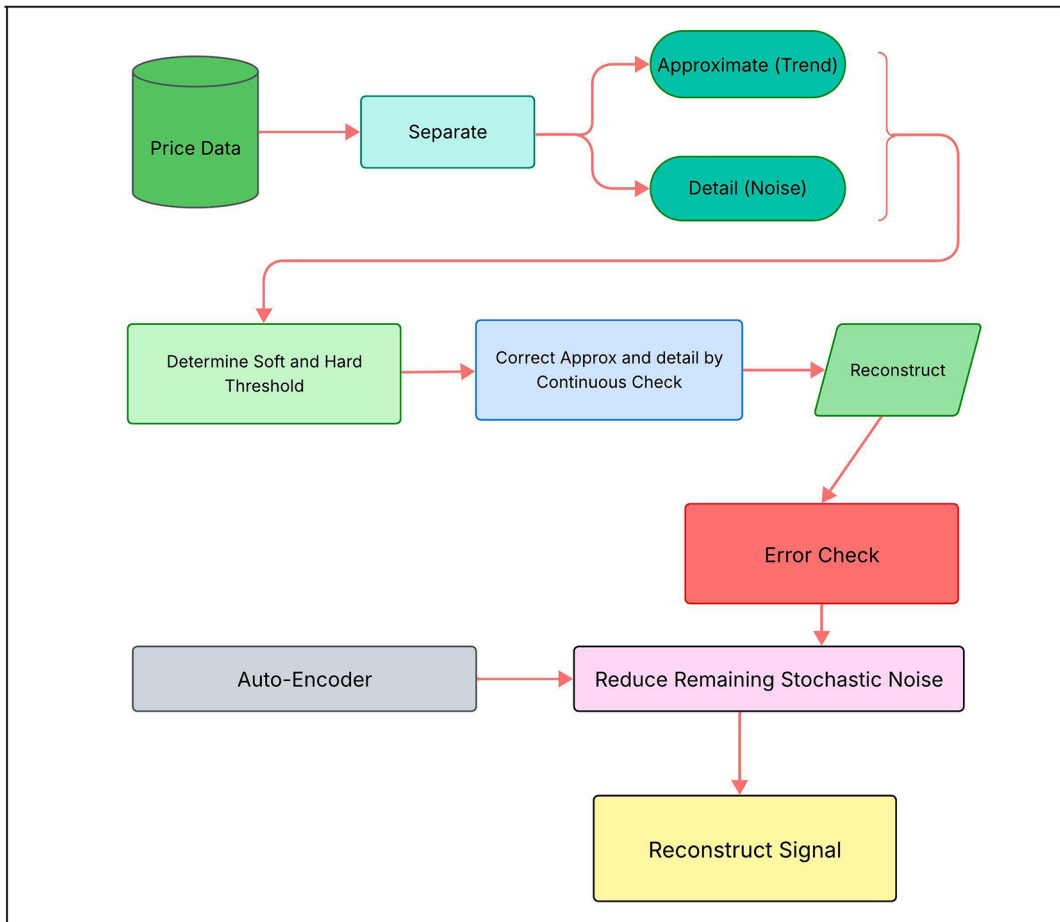


Figure 5. Real-trend analysis through denoising.

Figure 3 shows how, during the optimisation phase, the specific data transformations are applied, and how Optuna and Wavelet transforms help to filter out microscope noise.

To effectively capture the underlying market dynamics without the distortion caused by microstructure noise, this study extracts the true trend using the Smoothed Sequence approach, following the methodology established by.³² By continuously adjusting to the market, this method filters out transient volatility and isolates the actual price trajectory. Specifically, for a given time series of raw stock prices $\{x_1, x_2, \dots, \dots, x_n\}$, the smoothed sequence $\{y_1, y_2, \dots, \dots, y_n\}$ utilising a window size of k is computed as follows³²:

$$y_i = \frac{1}{K} \sum_{\{j=i-\frac{k}{2}\}_j}^{\{i+\frac{k}{2}\}} X_j \tag{Equation 1}$$

Where, y_i represents the smoothed value; it moves in the same direction as the market to reflect the true trend. By extracting this smoothed sequence, we ensure that the deep learning models that follow are fed data that reflects true market movements, not erratic movements driven by high-frequency noise.

II. Model Training: The research then tests six advanced deep learning models on the denoised univariate price series - 1D Conv-LSTM, CNN-LSTM, Gated Recurrent Unit (GRU), Bidirectional LSTM (Bi-LSTM), Temporal Convolutional Network (TCN) and Transformer (self-attention). Each model has a very simple architecture: 30 past log-returns with a 30-minute time window to forecast the next-minute intensity, with 2 stacked hidden layers, 20% dropout, Adam optimiser, and Mean Squared Error loss. No traditional technical indicators are used; only the pure price signal is focused on.^{2,22}

Figure 4 outlines the structural framework of the six deep learning models utilised in this study, detailing the 30-minute input windows, hidden layers, and evaluation checkpoints.

Figure 5 visually illustrates the discrete wavelet transform (DWT) process, showing how the raw signal is split into coarse and detail coefficients.

III. Walk-Forward Validation: We then employ an expanding-window (rolling origin) walk-forward validation: at each step (~300-minute trading day), retrain all models on all past denoised data and test on the next out-of-sample block. The process records performance metrics at each step: adjusted R^2 (typically **78–95%**) and overfitting ratio (**<10%**) on the validation block. Finally, generate the held-out 150-point forecast and evaluate its risk-adjusted return.

First, a **Discrete Wavelet Transform (DWT)** using Daubechies wavelets decomposes the return series $x(t)$ into coarse (approximation) and detail coefficients. Specifically, at the level j , the transform is given by convolutions with a low-pass $h[n]$ and high-pass $g[n]$ filters^{32,44}:

$$cA_j[n] = \sum_k x[k] h[n-k], cD_j[n] = \sum_k x[k] g[n-k] \quad \text{Equation (2)}$$

The process essentially classifies the true noise sequence referenced by high-pass $g[n]$ as high-frequency price signals from real low-pass $h[n]$ signals. The methodology trains on real price movements that are repetitive and train-worthy and is intended to be retained in the neural network for real sequence-based price movement applications. Soft-thresholding is then applied to the detail coefficients to suppress high-frequency noise⁴⁵:

$$\widetilde{cD}_j[n] = \text{sign}(cD_j[n]) \max(|cD_j[n]| - \lambda, 0), \quad \text{Equation (3)}$$

where the threshold λ is tuned via Optuna optimisation. The denoised series is reconstructed by inverting the wavelet transform using $\widetilde{cD}_j[n]$ and $cA_j[n]$. In this method λ supports the positional appropriation, where continuous training helps determine the optimal price signals for denoising. The price signal thus ensures a repetitive frequency rather than high levels of unconventional noise.

1) Variational Autoencoder (VAE)

Next, the partially denoised series is passed through a Variational Autoencoder for nonlinear noise removal. The VAE consists of an encoder mapping the input return x to a latent distribution $q(z|x)$, and a decoder reconstructing x from the latent sample z . Training maximises the evidence lower bound (ELBO) on the log-likelihood^{46,47}:

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{q(z|x)} [\log p(x|z)] - D_{\text{KL}}(q(z|x) \| p(z)), \quad \text{Equation (4)}$$

which balances the reconstruction loss (first term) and the KL divergence between the learned latent distribution and the prior $p(z)$. After training, the decoder output yields a reconstructed time series with much of the residual noise removed. This ensures the price signal contains only filtered trainable signals for KL.

2) Kalman Filter

Finally, we apply a Kalman filter to the VAE output to smooth any remaining microstructure noise. We model the log-return series as a linear Gaussian state-space system: the state (true return) x_k evolves as $x_k = Ax_{k-1} + w_k$ and the observation y_k is $y_k = Cx_k + v_k$, where w_k, v_k are Gaussian noises. At each time step, the filter performs a.^{48,49}

Prediction:

$$\hat{x}_k^- = A\hat{x}_{k-1}, P_k^- = AP_{k-1}A^T + Q, \quad \text{Equation (5)}$$

Update:

$$K_k = P_k^- C^T (CP_k^- C^T + R)^{-1}, \hat{x}_k = \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-), P_k = (I - K_k C)P_k^- \quad \text{Equation (6)}$$

Here, P_k is the error covariance, Q, R are the process/measurement covariances, and K_k is the Kalman gain. This recursive filtering yields a final smooth state estimate \hat{x}_k that forms the clean price signal used for forecasting. Overall, the combination of DWT, VAE and Kalman filtering is an effective approach in amplifying the true signal by absorbing noise sequentially, both high-frequency and model-based noise.

C. Predictive models

In the forecasting stage, we deploy six state-of-the-art deep learning architectures, each designed to capture temporal patterns in the denoised price series. Specifically, we implement:

1) Convolutional Neural Network (CNN-1D)

CNN-1D s extract short-term local features via convolutional filters. For an input sequence X and a filter K of size s , the convolution output (feature map) at time t is^{50,51}:

$$C_t = f\left(\sum_{i=0}^{s-1} K_i X_{t+i} + b\right), \quad \text{Equation (7)}$$

Where, f is a nonlinear activation (e.g., ReLU) and b is a bias term. Stacking such filters allows the model to detect salient intraday patterns (e.g., recent spikes or dips).

2) Gated Recurrent Unit (GRU)

The GRU is a recurrent network that controls information flow via update z_t and reset r_t gates. The equations are⁵²:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r), z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \quad \text{Equation (8)}$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h), h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \quad \text{Equation (9)}$$

Where, x_t is the input at time t , h_{t-1} the previous hidden state, and σ the sigmoid function. The update gate z_t blends the old and candidate states, while the reset gate r_t controls how much past information to forget.

3) Long Short-Term Memory (LSTM)

The LSTM is similar but uses separate input, forget, and output gates to mitigate vanishing gradients. Its core equations are⁵³:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad \text{Equation (10)}$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c), c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, h_t = o_t \tanh(c_t) \quad \text{Equation (11)}$$

Here, c_t is the cell state storing long-term memory, updated by the forget (f_t) and input (i_t) gates, and h_t is the hidden output gated by o_t .

4) Temporal Convolutional Network (TCN)

The TCN uses causal dilated convolutions to capture very long-range dependencies. For a dilation factor d and filter size k , the dilated convolution at position s is²⁷:

$$F^d(s) = \sum_{i=0}^{k-1} K_i \cdot x_{s-d \times i} \quad \text{Equation (12)}$$

By stacking layers with exponentially increasing d , the receptive field grows exponentially, allowing the TCN to learn long-term patterns without recurrences.

5) Transformer (Self-Attention)

The Transformer model relies on self-attention to weight all past observations when forecasting. The core scaled dot-product attention is⁵⁴:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad \text{Equation (13)}$$

where Q , K , V are the query, key, and value matrices derived from the input embeddings, and d_k is the key dimension. This mechanism enables the model to adaptively focus on relevant past time steps for each prediction.

All denoising hyperparameters (wavelet family, autoencoder bottleneck size, and Kalman covariances) and model hyperparameters (learning rates, layer sizes, and dropout rates) are tuned using Optuna to ensure optimal performance on a held-out validation split. In total, these six architectures encompass both convolutional, recurrent, and attention-based approaches, each expected to capture different aspects of intraday dynamics. We exclude other methods (e.g., technical indicators or simpler ML models) to focus on these well-established deep learning models, which are known to capture nonlinear patterns in financial time series.

D. Model architecture and training

Each forecasting model is trained end-to-end on the Optuna-denoised price series. We use a consistent configuration: two hidden layers (e.g., two LSTM layers or two convolutional stacks), 20% dropout to mitigate overfitting, and the Adam optimiser with mean-squared-error loss (as in²⁵). The input to each model is a fixed-length window of 30 past log-returns, and the output is the predicted next-minute return. Training proceeds on the expanding set of historical data, as in walk-forward validation; no cross-market data leakage is allowed. This uniform architecture facilitates a fair comparison of model performance across markets and ensures that any differences arise from market characteristics rather than model design.

E. Walk-forward validation and performance evaluation

To rigorously assess forecasting accuracy in a simulated real-time setting, we use expanding-window (walk-forward) evaluation. At each validation step (approximately 1 trading day, ≈ 300 minutes), all models are retrained on the cumulative denoised data up to that point and then tested on the immediately following out-of-sample block. This process is repeated until the end of the sample, yielding an out-of-sample forecast for each step. We record the adjusted R^2 of each one-day forecast (typically 78–95%) and monitor overfitting by comparing training vs validation error (overfitting ratio $< 10\%$ indicates well-generalised models).

Finally, the last 150-minute forecast (simulating the final day) is used to compute trading performance. In each market, we generate long/short signals based on the predicted returns and form a capital-weighted portfolio. We then compute the risk-adjusted return using the Sharpe ratio, defined as

$$\text{Sharpe} = \frac{\text{Portfolio Return} - \text{Risk} - \text{Free Rate}}{\sigma_{\text{returns}}} \quad \text{Equation (14)}$$

Market-specific risk-free rates are based on the prevailing 10-year government bond yields (e.g., $\sim 6.65\%$ in India, 13.88% in Brazil, and 3.58% in Malaysia at the end of 2025). All yield estimates are drawn from^{5,55} for consistency.

| Country | Buy/Sell | Stocks | Buy price | Sell price | Buy time | Sell time | Revenue | Overall Portfolio Risk | Buy Portfolio return | Sell Portfolio return |
|---------|--------------------------|--|-----------|------------|----------|-----------|---------|------------------------|----------------------|-----------------------|
| Brazil | Buy | IOIV3 S.A. | 41.03 | 41.93 | 80 | 94 | 0.48 | 0.0440 | 1.24% | 1.37% |
| | Buy | Povo Seguro S.A. | 33.01 | 33.97 | 28 | 14 | 0.39 | | | |
| | Buy | Vale Soluções S.A. | 24.95 | 25.38 | 7 | 19 | 1.92 | | | |
| | Buy | Duv Computação e Incorporadora S.A. | 23.42 | 23.97 | 5 | 12 | 0.54 | | | |
| | Sell | Prodama Distribuidora de Produtos Farmacêuticos S.A. | 0.39 | 0.64 | 140 | 12 | 3.39 | | | |
| | Sell | Equatorial S.A. | 35.15 | 35.34 | 128 | 5 | 0.78 | | | |
| | Sell | Moura Dubois Engenharia S.A. | 21.68 | 22.13 | 118 | 11 | 0.91 | | | |
| | Sell | Tegma Cargas Logística S.A. | 34.23 | 34.46 | 133 | 74 | 1.11 | | | |
| Sell | FIE Brasil Recursos S.A. | 44.21 | 44.87 | 190 | 34 | 0.85 | | | | |

Figure 6. Portfolio optimization for Brazil.

A Sharpe ratio above 1 is considered a favourable risk-return tradeoff. The above procedures ensure that our evaluation reflects true out-of-sample performance, free from look-ahead bias.

6. Technical background

6.1 Why predicting minute-level price data is difficult

Forecasting on minute-level stock prices is inherently challenging because the raw price series are dominated by short-term noise - for example, in the case of short-term noise due to bid-ask spreads and imbalances in the order book - more so than by stable predictive structures.⁵⁶ Our raw data from Yahoo Finance exhibit erratic behaviour and a high degree of non-stationarity in mean and variance, which limits the ability of standard linear time-series models to a great extent.⁵⁷⁻⁵⁹ As a result, we utilise nonlinear deep learning architectures and a growing walk-forward retraining strategy to extract weak yet economically relevant signals and maintain forecast validity during rapid regime changes.^{2,25,60}

6.2 Optuna for denoising and hyperparameter tuning

To explore the complex parameter spaces in our pipeline, we use Optuna, an automated hyperparameter optimisation tool based on Tree-structured Parzen Estimators (TPE).³⁶ Optuna has been used in two separate stages: **first**, to optimise the signal-to-noise ratio for denoising (optimise the parameters of wavelets, autoencoders and Kalman Filters) and **second** optimize the prediction error on a dataset not included in the training (optimise the learning rate, the number of layers and the dropout procedure in the Deep Learning models). This automated, define-by-run approach is crucial for managing mutually dependent factors and ensuring resource efficiency when manual grid search would be impractical computationally.

6.3 Denoising processes and their importance

Three complementary denoising techniques are used in order to suppress microstructure noise while retaining economic signals. **First**, Wavelet transformations (e.g. Daubechies) to deconstruct the series to suppress high frequency coefficients by soft-thresholding (for the purpose of effectively filtering the irregular noise but retaining underlying dynamics).^{32,61} **Second**, denoising autoencoders generate a compressed latent representation because the reconstruction process imposes compression, separating main price signals from random variations.^{62,63} **Third**, Kalman filters are recursive state-space estimators that smooth the series by adapting to changes in volatility without using fixed window lengths.^{64,65} Collectively, these methods, tuned with Optuna, improve the signal-to-noise ratio by amplifying meaningful short-term movements that are important for the model's predictive accuracy.

6.4 Deep learning models for minute-level forecasting

We use deep sequence models (including such models as TCN, GRU, LSTM (both standard and bidirectional), and hybrid architectures of CNN-LSTM/Transformer). Through such modeling we can forecast denoised minute-level movements. By contrast with the traditional linear methods of financial time series analysis (e.g., ARIMA), these nonlinear models account for nonstationary dynamics and complex dependencies, which have long been neglected by old-fashioned statistical tools or shallow neural networks.^{25,66}

RNN-based models (LSTM, GRU) use gates to learn long-term temporal dependencies while filtering out noise. Hybrid CNN-LSTM architectures are more effective than these: with convolutional layers, they can transform inputs into deeply learned local views before processing them as sequence data.³⁴ Temporal Convolutional Networks (TCN) leverage both dilated causal convolutions for efficient long-range memory (Bai et al., 2018) and Transformers, which utilise self-attention to capture volatility clustering and regime transitions.⁶⁶ Studies have now confirmed that fitting these methods to transient patterns characteristic of intraday markets has significantly improved risk-adjusted returns and reduced RMSE.⁶⁷

7. Results

The deep learning models, i.e., LSTM, Bi-LSTM, GRU, TCN, Transformers, Conv1D-LSTM, and CNN-LSTM, were developed using denoised minute-level stock price data in a walk-forward validation. Hyperparameters were optimised using the Optuna optimisation library, and for microstructure noise denoising, wavelet transform, autoencoders, and Kalman filters were used sequentially.

Table 4. SNR value table (Developing markets).

| Malaysia (Ticker) | SNR | India (Ticker) | SNR | Brazil (Ticker) | SNR |
|-------------------|------|----------------|-------|-----------------|-------|
| 6399.KL | 2.66 | COALINDIA.NS | 3.59 | PSSA3.SA | 6.49 |
| 5238.KL | 6.00 | CENTRALBK.BO | 7.50 | EQTL3.SA | 42.55 |
| 5099.KL | 5.46 | IDEA.BO | 4.54 | MDNE3.SA | 5.83 |
| 2488.KL | 3.64 | ITC.NS | 3.92 | IRBR3.SA | 46.27 |
| 5139.KL | 5.74 | IOB.BO | 7.14 | TOTS3.SA | 1.80 |
| 5326.KL | 7.44 | JPOWER.BO | 6.29 | SBSP3.SA | 42.25 |
| 7579.KL | 2.65 | MAHABANK.BO | 8.14 | PFRM3.SA | 4.62 |
| 8176.KL | 2.18 | MSUMI.BO | 5.69 | CURY3.SA | 6.41 |
| 0276.KL | 3.97 | OLAELEC.BO | 5.51 | AMBP3.SA | 41.08 |
| 0348.KL | 4.80 | PSB.BO | 5.45 | — | — |
| 5311.KL | 6.78 | TATAMOTORS.NS | 45.16 | — | — |
| — | — | UCOBANK.BO | 7.63 | — | — |

Table 5. SNR value table (Developed markets).

| USA (Ticker) | SNR | Japan (Ticker) | SNR | Singapore (Ticker) | SNR |
|--------------|-------|----------------|-------|--------------------|------|
| WBD | 4.50 | 4543.T | 39.89 | BS6.SI | 4.41 |
| KEY | 5.23 | 5411.T | 44.85 | O39.SI | 4.24 |
| INTC | 4.09 | 6724.T | 43.54 | Z77.SI | 5.49 |
| HPE | 45.64 | 7011.T | 47.88 | V03.SI | 3.89 |
| HBAN | 48.32 | 8031.T | 44.21 | AIY.SI | 5.93 |
| AMCR | 6.19 | 8058.T | 49.12 | H78.SI | 6.28 |
| PCG | 42.93 | 8253.T | 48.75 | Z74.SI | 4.72 |
| KIM | 43.27 | — | — | C2PU.SI | 2.34 |
| IVZ | 46.30 | — | — | — | — |
| HAL | 43.38 | — | — | — | — |
| F | 41.53 | — | — | — | — |
| AES | 48.82 | — | — | — | — |

Table 6. Simulated portfolio performance and sharpe ratios.

| Market group | Market | Return (%) | Final amount (\$\$) | Profit (\$\$) | Sharpe ratio |
|--------------|-----------|------------|---------------------|---------------|--------------|
| Developing | Malaysia | 5.60 | 1,056.00 | 56.00 | 5.89 |
| | India | 1.95 | 1,019.50 | 19.50 | 1.90 |
| | Brazil | 1.29 | 1,012.90 | 12.90 | 5.32 |
| Developed | USA | 0.20–0.99 | 1,002.00–1,009.90 | 2.00–9.90 | 1.2–1.6 |
| | Japan | 0.31–0.83 | 1,003.10–1,008.30 | 3.10–8.30 | ~1.0 |
| | Singapore | 0.62 | 1,006.20 | 6.20 | ~1.4 |

Figure 6 shows the detailed breakdown of deep learning forecasting results in both developed and developing markets. It shows the top performing stocks based on the country, the ticker, the company name, the date and the adjusted R², the best model, predicted price for the stock, the real price, the high predictive accuracy (Adjusted R²) and low overfitting rates of the various models in terms of predicting and comparing forecasted prices with actual intraday market data. Although a

high adjusted R^2 indicates good predictive performance, financial processes with high variance can be justified by good predictive hyperparameters.⁶⁸

Stock selection for inclusion in the portfolio was based on the highest adjusted R^2 values (indicating the strength of the model's predictions against real prices) and the lowest overfitting value (the difference between the training and validation R^2 , kept below 10%).

7.1 Signal-to-Noise Ratio (SNR) Analysis

To further ensure that the quality of stocks and the strength of exploitable patterns in denoised data were sufficient, we calculated the Signal-to-Noise Ratio (SNR) for selected stocks. In the context of finance, SNR measures the ratio between structured, predictable price movements (signal) against the random microstructure movement (noise), and is borrowed from signal processing. Higher SNR implies cleaner, more reliable forecasting signals - especially important in developing markets where there is typically greater inherent volatility, but also exploitable inefficiencies.

Table 4 compares the initial Signal-to-Noise Ratios (SNR) across the selected assets in developing markets' imperialistic environments, illustrating the micro-structure noise present in volatile environments.

Table 5 summarises the Signal-to-Noise Ratios (SNR) of large-cap stocks from developed markets as a baseline for clarity, compared with those from developing economies.

We used inverse-volatility weighting to construct portfolios from the best models' buy/sell signals. In simulated trading, there was a set amount of money to start with (for example, \$1,000) over 30 days. The forecasts were for 150 minutes of the day, and the rules were realistic.

Table 6 details the outcomes of the 30-day simulated intraday trading strategy. The results demonstrate that developing markets yielded substantially higher Sharpe ratios and raw profits compared to their developed counterparts.

8. Portfolio optimization

In this part, we present the results of the portfolio optimisation procedure applied to the predicted minute-level stock prices in both developed (USA, Japan, Singapore) and emerging (India, Brazil, Malaysia) markets. We used a conventional pipeline based on denoised high-frequency data and deep learning predictions, using models including LSTM, Bi-LSTM, GRU, TCN, Transformers, Conv1D-LSTM, and CNN-LSTM to optimise. The final portfolios included stocks with the best adjusted R-squared values (78%–95%, with an average of about 85%) and the least overfitting (less than 10%). This was confirmed using walk-forward testing.

The hyperparameter optimization within our deep learning architectures was specifically designed to test and lock onto the true market trajectory provided by the preprocessed data. Because the models were trained on the Smoothed Sequence,³² they captured the underlying trends rather than overfitting to random market fluctuations. This ability to map real price movements directly accounts for the robust predictive performance, yielding Adjusted R^2 scores ranging from 75% to 95% across the tested assets. Without this smoothed sequence isolating the true trend, predicting the real price movements with such high accuracy would not have been possible.

The portfolios were built separately for buy (long) and sell (short) positions. To lower risk and raise returns, they were weighted by inverse volatility. We did conduct some behaviour, such as simulating intraday trading with a projection horizon of 150 minutes, starting from an initial amount included in the model and considering realistic limitations, such as transaction costs (implicit in returns) and liquidity. The standard deviation of returns is used to estimate the portfolio's total risk. Buy-and-sell portfolio returns are shown as percentages. Developing markets consistently showed higher Sharpe ratios (1.2–1.6 times those of developed markets), with buy-side Sharpe ratios of 1.9–5.3 versus 1.0–1.4 in developed markets, highlighting superior risk-adjusted performance driven by exploitable inefficiencies. We include the best portfolios for each country below. These include buy selections, company names, buy/sell prices, timing (in minutes), revenues, overall portfolio risk, and buy/sell portfolio return.

Figures 7 through **12** are visual maps that plot the optimised buy-and-sell portfolio for each market. These charts show the exact timestamps of each trade, how assets were allocated to them, and the expected intraday returns over the simulated 150-minute horizon.

| Country | Buy/Sell | Stocks | Buy price | Sell price | Buy time | Sell time | Revenue | Overall Portfolio Risk | Buy Portfolio Return | Sell Portfolio Return |
|---------|----------|---------------------------------------|-----------|------------|----------|-----------|---------|------------------------|----------------------|-----------------------|
| India | Buy | Vodafone Idea Limited | 6.88 | 7.02 | 22 | 112 | 6.03 | 0.0679 | 2.17% | 1.28% |
| | Buy | IRB Infrastructure Developers Limited | 50.67 | 51.34 | 6 | 132 | 0.66 | | | |
| | Buy | Jai Prakash Power Ventures Limited | 15.46 | 16.08 | 1 | 45 | 1.84 | | | |
| | Buy | Yes Bank Limited | 21.04 | 21.54 | 4 | 24 | 1.98 | | | |
| | Sell | NMDC Steel Limited | 38.02 | 38.38 | 147 | 34 | 0.73 | | | |
| | Sell | Sagility India Limited | 39.32 | 40.15 | 144 | 33 | 0.83 | | | |
| | Sell | Liivian Small Finance Bank I limitad | 43.02 | 43.62 | 142 | 1 | 0.60 | | | |

Figure 7. Portfolio optimization for India.

| Country | Buy/Sell | Stocks | Buy price | Sell price | Buy time | Sell time | Revenue | Overall Portfolio Risk | Buy Portfolio Return | Sell Portfolio Return |
|----------|----------|--------------------------------------|-----------|------------|----------|-----------|---------|------------------------|----------------------|-----------------------|
| Malaysia | Buy | Apex Healthcare Berhad | 2.10 | 2.16 | 88 | 129 | 0.79 | 0.1553 | 6.82% | 4.95% |
| | Buy | Cape EMS Berhad | 0.21 | 0.22 | 63 | 150 | 6.98 | | | |
| | Buy | Aurelius Technologies Berhad | 3.24 | 3.31 | 59 | 103 | 0.67 | | | |
| | Buy | Chemite Innovation Berhad | 0.20 | 0.21 | 35 | 79 | 7.84 | | | |
| | Buy | Agicore GS Holdings Berhad | 0.47 | 0.52 | 11 | 37 | 5.07 | | | |
| | Buy | Aircount Datcom Berhad | 1.11 | 1.17 | 86 | 135 | 1.51 | | | |
| | Sell | Wavefront Berhad | 0.28 | 0.30 | 150 | 51 | 5.12 | | | |
| | Sell | AWC Berhad | 0.59 | 0.62 | 91 | 49 | 2.49 | | | |
| | Sell | Astro Malaysia Holdings Berhad | 0.18 | 0.21 | 149 | 76 | 9.29 | | | |
| | Sell | 99 Speed Mart Retail Holdings Berhad | 2.05 | 2.09 | 85 | 25 | 0.79 | | | |
| | Sell | ArAsia X Berhad | 1.68 | 1.73 | 146 | 60 | 1.25 | | | |
| | Sell | AFPI Bank Berhad | 2.67 | 2.72 | 132 | 30 | 0.92 | | | |
| | Sell | AEON Credit Service (M) Berhad | 5.71 | 5.82 | 146 | 1 | 0.22 | | | |
| | Sell | Capital A Berhad | 0.85 | 0.88 | 150 | 62 | 2.29 | | | |

Figure 8. Portfolio optimization for Malaysia.

| Country | Buy/Sell | Stocks | Buy price | Sell price | Buy time | Sell time | Revenue | Overall Portfolio Risk | Buy Portfolio Return | Sell Portfolio Return |
|-----------|----------|---|-----------|------------|----------|-----------|---------|------------------------|----------------------|-----------------------|
| Singapore | Buy | City Developments Limited | 4.82 | 4.85 | 105 | 128 | 0.524 | 0.0186 | 1.18% | 0.68% |
| | Buy | Singapore Telecommunications Limited | 3.93 | 3.98 | 17 | 137 | 0.721 | | | |
| | Buy | Parkway Life Real Estate Investment Trust | 4.01 | 4.09 | 2 | 81 | 0.944 | | | |
| | Buy | Yangzijiang Shpbuilding | 2.09 | 2.12 | 7 | 36 | 1.426 | | | |
| | Sell | Singapore Telecommunications Limited | 3.82 | 3.88 | 150 | 20 | 0.697 | | | |
| | Sell | Venture Corporation Limited | 11.08 | 11.15 | 147 | 55 | 0.256 | | | |
| | Sell | Sembcorp Industries Ltd | 6.84 | 6.87 | 35 | 4 | 0.409 | | | |
| | Sell | UOL Group Limited | 5.70 | 5.73 | 35 | 7 | 0.565 | | | |
| | Sell | Singapore Exchange Limited | 14.26 | 14.32 | 38 | 13 | 0.115 | | | |
| | Sell | iFAST Corporation Ltd | 6.50 | 6.59 | 150 | 42 | 0.270 | | | |
| | Sell | Oversea-Chinese Banking Corporation Limited | 16.29 | 16.37 | 150 | 5 | 0.37 | | | |
| | Sell | Hongkong Land Holdings Limited | 5.19 | 5.22 | 150 | 81 | 0.508 | | | |
| | Sell | Haw Par Corporation Limited | 11.75 | 11.83 | 85 | 43 | 0.212 | | | |
| | Sell | CapitaLand Ascendas REIT | 2.83 | 2.85 | 138 | 1 | 0.864 | | | |

Figure 9. Portfolio optimization for Singapore.

| Country | Buy/Sell | Stocks | Buy price | Sell price | Buy time | Sell time | Revenue | Overall Portfolio Risk | Buy Portfolio Return | Sell Portfolio Return |
|---------|----------|----------------------------------|-----------|------------|----------|-----------|---------|------------------------|----------------------|-----------------------|
| Japan | Buy | Tokuyama Corporation | 2799.51 | 2815.62 | 23 | 127 | 0.3117 | 0.0019 | 0.0033% | 0.002% |
| | Buy | Dentsu Group Inc | 3111.12 | 3129.05 | 14 | 143 | 0.4002 | | | |
| | Buy | TOYO LTD | 3683.83 | 3701.05 | 101 | 117 | 0.1137 | | | |
| | Buy | JFE Holdings, Inc | 1695.36 | 1708.46 | 1 | 150 | 0.5337 | | | |
| | Buy | The Japan Steel Works, Ltd | 6838.99 | 6891.87 | 67 | 123 | 0.1627 | | | |
| | Buy | Saiko Eppon Corporation | 1897.99 | 1919.66 | 101 | 150 | 0.5627 | | | |
| | Buy | TDK Corporation | 1584.69 | 1595.99 | 13 | 108 | 0.6725 | | | |
| | Buy | Yokogawa Electric Corporation | 3519.77 | 3555.96 | 35 | 144 | 0.2421 | | | |
| | Buy | Mitsubishi Heavy Industries, Ltd | 3283.93 | 3345.56 | 5 | 148 | 0.2671 | | | |
| | Buy | Honda Motor Co., Ltd | 1464.60 | 1475.31 | 2 | 111 | 0.3307 | | | |
| | Buy | Mitsui & Co., Ltd. | 3005.66 | 3031.20 | 63 | 140 | 0.2830 | | | |
| | Buy | Mitsubishi Corporation | 2930.20 | 2946.29 | 61 | 138 | 0.3674 | | | |
| | Buy | Credit Saison Co., Ltd | 3985.46 | 3936.31 | 29 | 134 | 0.2038 | | | |
| | Buy | West Japan Railway Company | 3128.48 | 3143.92 | 13 | 145 | 0.3242 | | | |
| | Buy | Toyota Motor Corporation | 2455.77 | 2486.96 | 48 | 150 | 0.4519 | | | |
| | Buy | Yamaha Motor Co., Ltd | 1106.48 | 1116.74 | 7 | 106 | 0.6320 | | | |
| | Sell | Tenaro Corporation | 2992.66 | 2921.51 | 150 | 36 | 0.4426 | | | |
| | Sell | Trend Micro Incorporated | 10772.91 | 10870.44 | 33 | 9 | 0.1006 | | | |
| | Sell | Fujitsu Limited | 3318.32 | 3333.68 | 146 | 128 | 0.3643 | | | |

Figure 10. Portfolio optimization for Japan.

| Country | Buy/Sell | Stocks | Buy price | Sell price | Buy time | Sell time | Revenue | Overall Portfolio Risk | Buy Portfolio Return | Sell Portfolio Return |
|---------|----------|------------------------------------|-----------|------------|----------|-----------|---------|------------------------|----------------------|-----------------------|
| USA | Buy | Warner Bros. Discovery Inc | 9.78 | 9.99 | 20 | 150 | 0.61208 | 0.0125 | 0.97% | 0.89% |
| | Buy | Viatrix Inc | 8.72 | 8.80 | 91 | 129 | 0.71063 | | | |
| | Buy | PG&E Corp. | 16.79 | 16.94 | 31 | 127 | 0.28401 | | | |
| | Buy | Kimco Realty Corporation | 21.15 | 21.29 | 60 | 128 | 0.28681 | | | |
| | Buy | Keycorp | 15.76 | 15.89 | 1 | 142 | 0.24205 | | | |
| | Buy | Invesco Ltd | 14.42 | 14.53 | 2 | 118 | 0.33738 | | | |
| | Buy | Intel Corp | 19.41 | 19.61 | 1 | 133 | 0.20193 | | | |
| | Buy | Hewlett Packard Enterprise Co | 17.11 | 17.27 | 21 | 130 | 0.15795 | | | |
| | Buy | Huntington Bancshares, Inc. | 15.56 | 15.66 | 1 | 122 | 0.29954 | | | |
| | Buy | Halliburton Co. | 19.49 | 19.70 | 30 | 126 | 0.20693 | | | |
| | Buy | Ford Motor Co | 10.31 | 10.44 | 18 | 128 | 0.50711 | | | |
| | Buy | Healthpeak Properties Inc. | 17.33 | 17.47 | 20 | 126 | 0.43463 | | | |
| | Buy | Ammco P/c | 9.05 | 9.16 | 1 | 129 | 0.73065 | | | |
| | Sell | Walgreens Boots Alliance Inc | 11.21 | 11.32 | 146 | 118 | 0.98670 | | | |
| | Sell | Paramount Global | 12.06 | 12.18 | 138 | 72 | 0.25437 | | | |
| | Sell | Norwegian Cruise Line Holdings Ltd | 17.62 | 17.76 | 150 | 118 | 0.40730 | | | |
| | Sell | Host Hotels & Resorts Inc | 15.45 | 15.59 | 150 | 27 | 0.39769 | | | |
| | Sell | APA Corporation | 16.98 | 17.17 | 150 | 8 | 0.38195 | | | |
| | Sell | AES Corporation | 10.07 | 10.14 | 150 | 120 | 0.60798 | | | |

Figure 11. Portfolio optimization for the USA.

| Country | Ticker | Company Name | Date | Adjusted R ² | Model | Predicted Price | Real Price | Low Overfitting |
|-----------|---|---|-----------|-------------------------|-----------|-----------------|------------|-----------------|
| USA | WBD | Warner Bros. Discovery Inc | 5/1/2025 | 0.9011 | GRU | 9.82 | 8.56 | -0.025755 |
| | KEY | KeyCorp | 5/1/2025 | 0.8441 | TCN | 15.764 | 14.81 | 0.082306 |
| | INTC | Intel Corporation | 5/1/2025 | 0.819661 | Bi-LSTM | 19.408 | 20.18 | 0.02312 |
| | HPE | Hewlett Packard Enterprise Company | 5/1/2025 | 0.8876 | TCN | 17.145 | 16.44 | 0.015127 |
| | HBAN | Huntington Bancshares Incorporated | 5/1/2025 | 0.8487 | Bi-LSTM | 15.556 | 14.48 | 0.036611 |
| | AMCR | America's Choice | 5/1/2025 | 0.3528 | GRU | 3.052 | 3.17 | 0.009338 |
| | PCG | PG&E Corporation | 5/1/2025 | 0.8207 | TCN | 16.846 | 16.62 | 0.013624 |
| | KIM | Kimco Realty Corporation | 5/1/2025 | 0.8105 | GRU | 21.223 | 20.22 | 0.010211 |
| | IVZ | Invesco Ltd. | 5/1/2025 | 0.8029 | TCN | 14.422 | 14.01 | 0.01335 |
| | HAL | Halliburton Company | 5/1/2025 | 0.8378 | TCN | 19.526 | 19.81 | 0.008956 |
| | F | Ford Motor Company | 5/1/2025 | 0.8542 | TCN | 10.325 | 10.15 | 0.00313 |
| | AES | The AES Corporation | 5/1/2025 | 0.7927 | TCN | 10.084 | 9.32 | 0.083356 |
| Japan | 4543.T | Terumo Corporation | 6/4/2025 | 0.8875 | GRU | 2608.55 | 2628 | 0.000875 |
| | 5411.T | JFE Holdings, Inc | 5/7/2025 | 0.3537 | TCN | 1635.36 | 1670.5 | 0.001063 |
| | 6124.T | Seiko Epson Corporation | 5/7/2025 | 0.7822 | GRU | 1910.38 | 1954.5 | 0.000842 |
| | 7011.T | Mitsubishi Heavy Industries, Ltd | 5/7/2025 | 0.8227 | LSTM | 3287.03 | 2783 | 0.000194 |
| | 8031.T | Mitsui & Co., Ltd. | 5/7/2025 | 0.3425 | GRU | 3006.03 | 2822 | 0.000161 |
| | 8058.T | Mitsubishi Corporation | 5/7/2025 | 0.816 | GRU | 2932.06 | 2675.5 | 0.002338 |
| Singapore | 8253.T | Credit Saison Co., Ltd | 5/7/2025 | 0.8373 | GRU | 3914.34 | 3360 | 0.025331 |
| | ES6.SI | Yangzijiang Shipbuilding (Holdings) Ltd | 5/5/2025 | 0.8154 | GRU | 2.0963 | 2.15 | 0.07172 |
| | Q38.SI | Oversea-Chinese Banking Corporation Limited | 5/5/2025 | 0.3208 | TCN | 16.3689 | 16.21 | 0.003251 |
| | Z71.SI | Singapore Telecommunications Limited | 5/5/2025 | 0.8658 | GRU | 3.8482 | 3.8 | 0.000958 |
| | Y93.SI | Venture Corporation Limited | 5/5/2025 | 0.8004 | GRU | 11.021 | 10.99 | 0.011472 |
| | AY1.SI | IFAST Corporation Ltd | 5/5/2025 | 0.9433 | GRU | 6.5928 | 6.33 | 0.010418 |
| India | HPS.SI | Hongkong Land Holdings Limited | 5/5/2025 | 0.7192 | Bi-LSTM | 5.2111 | 4.92 | 0.014453 |
| | Z14.SI | Singapore Telecommunications Limited | 10/6/2025 | 0.8597 | GRU | 3.3935 | 3.88 | 0.034024 |
| | C2PU.SI | Parkway Life Real Estate Investment Trust | 5/5/2025 | 0.7867 | GRU | 4.018 | 4.18 | 0.003444 |
| | COALINDIA.NS | COALINDIA.NS | 6/5/2025 | 0.7284 | Bi-LSTM | 387.8037 | 336.8 | 0.0087 |
| | CENTRALEK.BO | Central Bank of India | 5/5/2025 | 0.8838 | Bi-LSTM | 37.3275 | 36.85 | 0.13546 |
| | IDEA.BO | Vodafone Idea Limited | 5/5/2025 | 0.8157 | Bi-LSTM | 6.8923 | 7.08 | 0.00305 |
| | ITC.NS | ITC Limited | 6/6/2025 | 0.8912 | GRU | 414.011 | 420.2 | 0.00533 |
| | IOB.BO | Indian Overseas Bank | 5/5/2025 | 0.8705 | Bi-LSTM | 37.9749 | 37.82 | 0.13436 |
| | JPOWER.BO | Jaiprakash Power Ventures Limited | 5/5/2025 | 0.8714 | GRU | 15.4631 | 14.35 | 0.09112 |
| | MAHABANK.BO | Bank of Maharashtra | 5/5/2025 | 0.3538 | GRU | 51.7723 | 50.36 | 0.02637 |
| | MSUMILBO | Motherson Sumi Wiring India Limited | 5/5/2025 | 0.3045 | GRU | 58.2978 | 56.34 | 0.00672 |
| | OLAELEC.BO | Ola Electric Mobility Limited | 5/5/2025 | 0.3210 | GRU | 50.6176 | 48.7 | 0.01738 |
| Malaysia | PSB.BO | Punjab & Sind Bank | 5/5/2025 | 0.7863 | Bi-LSTM | 30.6136 | 29.46 | 0.00723 |
| | TATAMOTORS.NS | Tata Motors Limited | 6/6/2025 | 0.3688 | GRU | 694.4481 | 709.2 | 0.02916 |
| | UCOBANK.BO | UCO Bank | 5/5/2025 | 0.3636 | GRU | 31.3274 | 31.36 | 0.02525 |
| | 5311.KL | Cape EMS Berhad | 5/5/2025 | 0.7372 | TCN | 0.215219 | 0.26 | 0.09316 |
| | 0348.KL | Chionite Innovation Berhad | 5/5/2025 | 0.7396 | GRU | 0.197696 | 0.24 | 0.1092 |
| | 0276.KL | Autocount Dotcom Berhad | 5/5/2025 | 0.8316 | GRU | 1.122414 | 1.04 | 0.0434 |
| | 8176.KL | Wavfront Berhad | 5/5/2025 | 0.3068 | GRU | 0.295191 | 0.28 | 0.0274 |
| | 1797.KL | A/WC Berhad | 5/5/2025 | 0.8937 | GRU | 0.620636 | 0.75 | 0.138 |
| | 5326.KL | 33 Speed Mart Retail Holdings Berhad | 5/5/2025 | 0.7954 | Bi-LSTM | 2.092055 | 2.28 | 0.0523 |
| | 5139.KL | AEON Credit Service (M) Berhad | 5/5/2025 | 0.3046 | GRU | 5.82206 | 6.03 | 0.0321 |
| | 2488.KL | Alliance Bank Malaysia Berhad | 5/5/2025 | 0.8787 | GRU | 4.37127 | 4.47 | 0.00139 |
| | 5099.KL | Capital A Berhad | 5/5/2025 | 0.8088 | GRU | 0.869896 | 0.85 | 0.02156 |
| Brazil | 5238.KL | AirAsia X Berhad | 5/5/2025 | 0.8012 | GRU | 1.707445 | 1.74 | 0.01543 |
| | 6399.KL | Astro Malaysia Holdings Berhad | 5/5/2025 | 0.8958 | LSTM | 0.2049393 | 0.18 | 0.05775 |
| | PSSA3.SA | Porto Seguro S.A | 5/27/2025 | 0.8414 | GRU | 53.64106 | 50.68 | 0.0092 |
| | EQTL3.SA | Equatorial S.A. | 5/27/2025 | 0.7378 | GRU | 35.32008 | 37.33 | -0.0008 |
| | MDNE3.SA | Moura Dubaux Engenharia S.A | 5/27/2025 | 0.3659 | TCN | 22.09676 | 18.92 | 0.006 |
| | IRBR3.SA | IRE-Brazil Resseguros S.A | 5/27/2025 | 0.7485 | GRU | 44.11720 | 48.27 | -0.005 |
| Brazil | TOTS3.SA | TOTVS S.A. | 5/27/2025 | 0.9098 | GRU | 41.12463 | 43.46 | 0.03307 |
| | SBSP3.SA | Companhia de Saneamento Básico do Estado de São Paulo | 5/26/2025 | 0.754582 | GRU | 114.97316 | 118.38 | 0.0077 |
| | PFPM3.SA | Pharmas Distribuidoras de Produtos Farmacêuticos S.A. | 5/27/2025 | 0.96 | Bi-LSTM | 8.61987 | 8.42 | -0.0026 |
| | CURY3.SA | Curry Concretora e Incorporadora S.A | 5/27/2025 | 0.3655 | GRU | 23.42388 | 23.55 | 0.0061 |
| AMBR3.SA | Ambipar Participações e Empreendimentos S.A | 5/26/2025 | 0.8825 | GRU | 166.34934 | 131.77 | 0.026 | |

Figure 12. Comprehensive predictive performance and accuracy metrics across markets.

9. Conclusions/Discussion

The results confirm that deep learning on denoised minute-level data translates inefficiencies in developing markets into superior risk-adjusted intraday returns, thereby challenging investor bias toward the stability of developed markets.⁶⁹ Equivalent forecasting accuracy, but significantly greater tradability in India, Brazil and Malaysia, demonstrating the Adaptive Market Hypothesis, which posits higher practicability of computational improvements in exploiting anomalies in less efficient environments.^{1,2}

This is an advantage over weaker returns that come with higher equity risk premiums in emerging markets,⁵ for risk that is well compensated for (volatility, etc) but that provides alpha where noise is managed. The absence of technical indicators isolates pure price dynamics and makes the study more generalizable, which is rare for intraday.²⁵

The empirical results of this study validate the efficacy of data denoising combined with advanced deep learning architectures for high-frequency price forecasting. For all six of the markets analyzed, the predictive models provided a good level of accuracy with adjusted R² scores ranging from 78–95% (average ~85%). Crucially, the walk-forward validation implementation ensured that overfitting remained well below 10%, ensuring that the models reflected generalizable market patterns rather than ephemeral market noise.

When these predictive signals were converted into inverse volatility-weighted trading portfolios over a 150-minute intraday horizon, there was a clear difference in performance between the market types. The portfolio optimisation analysis showed that developing markets (India, Brazil and Malaysia) achieved a significant risk-adjusted gain relative to developed markets (USA, Japan and Singapore). Developing markets produced buy-side Sharpe ratios that ranged from 1.9 to 5.3, which were 1.2 to 1.6 times higher than the 1.0 to 1.4 Sharpe ratios recorded in mature markets. Ultimately, these results from the portfolio bear out that the microstructure volatility inherent in the developing world can be managed

and exploited, providing an alternative, highly lucrative allocation of short-term alpha for investors to counter traditional 'flight to safety' biases.

In practice, the framework offers retail and institutional investors a replicable tool to reallocate to liquid emerging large-cap stocks, which could potentially improve diversification and returns given saturation in developed markets. Limitations include the short 30-day test period and reliance on historical data; future work could consider longer horizons or account for transaction costs.

Overall, deep learning breaks down creative-market jargon because in these developing economies, it shows underserved qualities and there is a lot for more exposure to be fluent for better, more efficient portfolio.^{3,4}

9.1 Limitations

Although this study provides empirical evidence of the efficacy of deep learning on denoised minute-level data for intraday forecasting in a developing market, there are certain limitations that must be recognised.

First, the analysis is based on a rather limited 30-day trading period. Although the look-ahead bias is reduced by the walk-forward validation method, this limited horizon does not necessarily eliminate regime shifts, structural changes in markets, or exceptional events (e.g., global crises) that may affect the patterns of inefficiencies.^{1,69} The data's robustness would improve by extending it over a multi-year span.

Second, simulated trading portfolios do not account for transaction costs, slippage, or market impact, which are particularly pronounced in high-frequency, minute-level strategies. In less developed financial markets, such as India, Brazil, and Malaysia, lower liquidity and wider bid-ask spreads can severely erode realised returns.⁴¹ In the real world, some portions of the portfolio might not actually generate the Sharpe ratios quoted here.

Third, the denoising pipeline (combination of wavelet transforms, variational autoencoders and Kalman filters), although optimized using Optuna, the risk of over-denoising or removing economically meaningful high frequency signals (e.g. microstructures dynamics critical for intraday alpha)^{32,56} are adjusted with proper trend pick up and drop off which is essentially achieved as well but heavy market shocks cannot preserve the require information; Parameter sensitivity and possible signal distortion are still challenges in sequential denoising methods.

Fourth, stock selection targeted the large-cap stocks in the liquid markets and excluded thinner or more volatile segments. This can be an overestimate of exploitable inefficiencies, as the smaller the stock or what is excluded (e.g., from developing regions), the greater the political risk and the lower the predictability.⁵ This process essentially ignores Fama French model, suggest against small market gains are against large-cap stocks in the long run.⁷⁰

Finally, adjusted R^2 and Sharpe ratios show strong performance; the results are based on historical data from Yahoo Finance, which may include survivorship bias or adjustments that are not 100% indicative of live trading conditions. Overfitting (even with walk-forward validation and reported small gaps (<10%)) is a potential issue for deep learning models trained on noisy financial series.^{10,38}

These limitations point to directions for future research, including the inclusion of transaction costs, longer horizons, testing in real time for execution, and extending to additional markets or hybrid denoising techniques.

Ethics approval and consent to participate

Not applicable. This study utilises publicly available, secondary market data and does not involve human participants or animal subjects.

Software availability

- **Source Code:** The full code pipeline for data denoising (Wavelet/Autoencoder/Kalman), model training, and portfolio simulation is available at: <https://github.com/Monjur1841/minute-level-stock-forecasting>
- Archived source code at time of publication: <https://doi.org/10.5281/zenodo.19014187>
- License: OSI-approved MIT License.

Availability of data and materials underlying data

The historical stock price data used in this study are publicly available from Yahoo Finance.

- **Source:** <https://finance.yahoo.com/>.
- Source code is available from: <https://github.com/Monjur1841/minute-level-stock-forecasting>.
- The source code is provided under an OSI-approved MIT License.
- **Retrieval Method:** Data can be retrieved programmatically using the open-source yfinance Python library.
- **Specifics:** The specific ticker symbols for the developed (USA, Japan, Singapore) and developing (India, Brazil, Malaysia) markets analyzed in this study are listed in Table 1 of the manuscript.

References

- Lo AW: **The adaptive markets hypothesis.** *J. Portf. Manag.* 2004; **30**(5): 15–29.
[Publisher Full Text](#)
- Gu S, Kelly B, Xiu D: **Empirical asset pricing via machine learning.** *Rev. Financ. Stud.* 2020; **33**(5): 2223–2273.
[Publisher Full Text](#)
- Kumar S, Sharma P, Garg R: **Stock market forecasting using deep learning in emerging markets.** *Expert Syst. Appl.* 2024; **238**: Article 121987.
[Publisher Full Text](#)
- Zhang X, Li Y, Wang Q: **Deep learning for financial time series forecasting: Evidence from Chinese and Indian markets.** *Financ. Res. Lett.* 2023; **52**: 103543.
[Publisher Full Text](#)
- Damodaran A: **Equity risk premiums (ERP): Determinants, estimation, and implications – The 2025 edition.** *SSRN Electron. J.* 2025.
[Publisher Full Text](#)
- MSCI: **MSCI Market Classification Framework and Emerging Markets Index.** Morgan Stanley Capital International; 2025.
[Reference Source](#)
- International Monetary Fund: **World Economic Outlook Update: Global Economy: Steady amid Divergent Forces.** IMF; 2025.
[Reference Source](#)
- Panigrahi S, Panda A: **Factors Influencing FDI Inflow to India, China and Malaysia: An Empirical Analysis.** *Asia Pac J Manag Res Innov.* 2012; **8**(2): 121–131.
[Publisher Full Text](#)
- World Bank: **Global Economic Prospects, June 2025.** The World Bank Group; 2025.
[Reference Source](#)
- Sezer OB, Gudelek MU, Ozbayoglu AM: **Financial time series forecasting with deep learning: A systematic literature review: 2005–2019.** *Appl. Soft Comput.* 2020; **90**: 106181.
[Publisher Full Text](#)
- Sharpe WF: **Capital asset prices: A theory of market equilibrium under conditions of risk.** *J. Financ.* 1964; **19**(3): 425–442.
[Publisher Full Text](#)
- Lintner J: **The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets.** *Rev. Econ. Stat.* 1965; **47**(1): 13–37.
[Publisher Full Text](#)
- Bekaert G, Harvey CR: **Emerging equity markets in a globalizing world.** *SSRN Electron. J.* 2017.
[Publisher Full Text](#)
- Bekaert G, Harvey CR: **Emerging markets finance.** *J. Empir. Financ.* 2003; **10**(1–2): 3–55.
[Publisher Full Text](#)
- Lee M-J, Choi S-Y: **Comparing market efficiency in developed, emerging, and frontier equity markets: A multifractal detrended fluctuation analysis.** *Fractal Fract.* 2023; **7**(6): 478.
[Publisher Full Text](#)
- Bakry W, Nghiem X-H, Bhatti MI, et al.: **Digital finance and sustainable development: Evidence from developing nations.** *Sci. Prog.* 2024; **107**(3).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chauhan SS, Suri P, Twala B, et al.: **Exploring the relationship between macroeconomic indicators and sectoral indices of Indian stock market [version 2; peer review: 2 approved, 1 approved with reservations].** *F1000Res.* 2025; **14**: 180.
[Publisher Full Text](#)
- Kumar P: **Large language models (LLMs): survey, technical frameworks, and future challenges.** *Artif. Intell. Rev.* 2024; **57**: 260.
[Publisher Full Text](#)
- Ali NP, Saha A: **Market Efficiency of Indian Capital Market: An Event Study Around the Announcement of Results of Lok Sabha Election 2019.** *Int J Financ Res.* 2021; **12**(1): 60–70.
[Reference Source](#)
- Ghazani MM, Ebrahimi SB: **Testing the adaptive market hypothesis in the Iranian stock market.** *Phys A Stat Mech Appl.* 2023; **611**: 128415.
[Publisher Full Text](#)
- Urquhart A, Hudson R: **Efficient or adaptive markets? Evidence from major stock markets using very long run historic data.** *Int. Rev. Financ. Anal.* 2013; **28**: 130–142.
[Publisher Full Text](#)
- Zhang GP: **Time series forecasting using a hybrid ARIMA and neural network model.** *Neurocomputing.* 2003; **50**: 159–175.
[Publisher Full Text](#)
- Hosseini SM, et al.: **Deep learning architectures for financial forecasting: A comprehensive survey.** *Expert Syst. Appl.* 2023; **212**: Article 118769.
[Publisher Full Text](#)
- Yunita A, Pratama MI, Almuzakki MZ, et al.: **Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models.** *MethodsX.* 2025; **15**: 103462.
[Publisher Full Text](#)
- Fischer T, Krauss C: **Deep learning with long short-term memory networks for financial market predictions.** *Eur. J. Oper. Res.* 2018; **270**(2): 654–669.
[Publisher Full Text](#)
- Lu Y, Zhang Y, Chen X: **Time series forecasting using LSTM and GRU models for stock prediction.** *Appl. Intell.* 2021; **51**(12): 8975–8991.
[Publisher Full Text](#)
- Bai S, Kolter JZ, Koltun V: **An empirical evaluation of generic convolutional and recurrent networks for sequence modeling.** *arXiv.* 2018.
[Publisher Full Text](#)
- Kong X, Chen Z, Liu W, et al.: **Deep learning for time series forecasting: A survey.** *Int. J. Mach. Learn. Cybern.* 2025.
[Publisher Full Text](#)
- Lim B, Arik SO, Loeff N, et al.: **Temporal fusion transformers for interpretable multi-horizon time series forecasting.** *Int. J. Forecast.* 2021; **37**(4): 1748–1764.
[Publisher Full Text](#)
- Wu N, Green B, Ben X, et al.: **Deep transformer models for time series forecasting: The influenza prevalence case.** *arXiv.* 2024.
[Publisher Full Text](#)
- Alzubaidi L, Zhang J, Humaidi AJ, et al.: **Time series forecasting in financial markets using deep learning models.** *World Journal of*

- Advanced Engineering Technology and Sciences*. 2025.
[Publisher Full Text](#)
32. Tang Q, Fan T, Shi R, *et al.*: **Prediction of financial time series using LSTM and data denoising methods.** *arXiv*. 2021.
[Publisher Full Text](#)
 33. Vogl M, Rötzel PG, Homes S: **Forecasting performance of wavelet neural networks and other neural network topologies: A comparative study based on financial market data sets.** *Mach. Learn. Appl.* 2022; **8**: 100302.
[Publisher Full Text](#)
 34. Bao W, Yue J, Rao Y: **A deep learning framework for financial time series using stacked autoencoders and long-short term memory.** *PLoS One*. 2017; **12**(7): e0180944.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 35. Singh A, Ogunfunmi T: **An overview of variational autoencoders for source separation, finance, and bio-signal applications.** *Entropy*. 2022; **24**(1): 55.
[Publisher Full Text](#)
 36. Akiba T, Sano S, Yanase T, *et al.*: **Optuna: A next-generation hyperparameter optimization framework.** *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019; pp. 2623–2631.
[Publisher Full Text](#)
 37. Oprea S-V, *et al.*: **Deep learning-based time series forecasting.** *Artif. Intell. Rev.* 2024; **57**(12): Article 272.
[Publisher Full Text](#)
 38. Huang W, Gao X: **Evaluating Hierarchical Equal Risk Contribution Portfolios in the Chinese Stock Market.** *J Math Finance*. 2022; **12**(1): 179–195.
[Publisher Full Text](#)
 39. Bergmeir C, Benítez JM: **On the use of walk-forward validation for time series prediction.** *Inf. Sci.* 2012; **187**: 184–197.
[Publisher Full Text](#)
 40. Saenz JV, Quiroga FM, Bariviera AF: **Data vs. information: Using clustering techniques to enhance stock returns forecasting.** *Int. Rev. Financ. Anal.* 2023; **88**: 102657.
[Publisher Full Text](#)
 41. Ang A: **Asset management: A systematic approach to factor investing.** Oxford University Press; 2014.
[Publisher Full Text](#)
 42. Chaves D, Hsu J, Li F, *et al.*: **Risk parity portfolio vs. other asset allocation heuristic portfolios.** *J. Invest.* 2012; **21**(1): 108–118.
[Publisher Full Text](#)
 43. Bhuiyan MH, Moni AK, Halim MB, *et al.*: **Evaluating Portfolio Performance Using Technical Indicators and Financial Ratio for Stocks in NSE by PyPortfolioOpt.**
[Publisher Full Text](#)
 44. Mallat SG: **A theory for multiresolution signal decomposition: The wavelet representation.** *IEEE Trans. Pattern Anal. Mach. Intell.* 1989; **11**(7): 674–693.
[Publisher Full Text](#)
 45. Donoho DL, Johnstone IM: **Adapting to unknown smoothness via wavelet shrinkage.** *J. Am. Stat. Assoc.* 1995; **90**(432): 1200–1224.
[Publisher Full Text](#)
 46. Kingma DP, Welling M: **Auto-encoding variational Bayes.** *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*. 2014. arXiv:1312.6114.
[Reference Source](#)
 47. Wang Z, Wang C, Li Y: **Variational autoencoder based on knowledge sharing and correlation weighting for process-quality concurrent fault detection.** *Eng. Appl. Artif. Intell.* 2024; **133**: 108051.
[Publisher Full Text](#)
 48. Kalman RE: **A new approach to linear filtering and prediction problems.** *J. Basic Eng.* 1960; **82**(1): 35–45.
[Publisher Full Text](#)
 49. Tusell F: **Kalman filtering in R.** *J. Stat. Softw.* 2011; **39**(2): 1–27.
[Publisher Full Text](#)
 50. LeCun Y, Boser B, Denker JS, *et al.*: **Backpropagation applied to handwritten zip code recognition.** *Neural Comput.* 1989; **1**(4): 541–551.
[Publisher Full Text](#)
 51. Kiranyaz S, Avci O, Abdeljaber O, *et al.*: **1D convolutional neural networks and applications: A survey.** *Mech. Syst. Signal Process.* 2021; **151**: 107398.
[Publisher Full Text](#)
 52. Cho K, van Merriënboer B, Gulcehre C, *et al.*: **Learning phrase representations using RNN encoder-decoder for statistical machine translation.** *arXiv preprint arXiv:1406.1078*. 2014.
[Publisher Full Text](#)
 53. Hochreiter S, Schmidhuber J: **Long short-term memory.** *Neural Comput.* 1997; **9**(8): 1735–1780.
[Publisher Full Text](#)
 54. Vaswani A, Shazeer N, Parmar N, *et al.*: **Attention is all you need.** *Advances in Neural Information Processing Systems (NeurIPS)*. 2017; pp. 5998–6008.
[Reference Source](#)
 55. Trading Economics: **10-year government bond yields – December 2025.** 2025.
[Reference Source](#)
 56. Hasbrouck J: *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading.* Oxford University Press; 2007.
[Publisher Full Text](#)
 57. Cont R: **Empirical properties of asset returns: stylized facts and statistical issues.** *Quant. Financ.* 2001; **1**(2): 223–236.
[Publisher Full Text](#)
 58. Tsay RS: *Analysis of Financial Time Series.* Hoboken: John Wiley & Sons; 3rd ed. 2010.
[Publisher Full Text](#)
 59. Zhang X: **Financial Viability Analysis and Capital Structure Optimization in Privatized Public Infrastructure Projects.** *J. Constr. Eng. Manag.* 2005; **131**: 656–668.
[Publisher Full Text](#)
 60. Lo AW, MacKinlay AC: **When Are Contrarian Profits Due to Stock Market Overreaction?** *Rev. Financ. Stud.* 1990; **3**: 175–205.
[Publisher Full Text](#)
 61. Ramsey JB: **Wavelets in Economics and Finance: Past and Future.** *Studies in Nonlinear Dynamics & Econometrics*. 2002; **6**(3): 1–29.
[Publisher Full Text](#)
 62. Vincent P, Larochelle H, Lajoie I, *et al.*: **Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion.** *J. Mach. Learn. Res.* 2010; **11**: 3371–3408.
[Reference Source](#)
 63. Cao H, Song Y, Guan X: **Quality Investment With Information Acquisition Transparency.** *Manag. Decis. Econ.* 2024; **46**(7): 3998–4010.
[Publisher Full Text](#)
 64. Chen T, Guestrin C: **XGBoost: A Scalable Tree Boosting System.** *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM; 2016; pp. 785–794.
[Publisher Full Text](#)
 65. Wang W, He N, Yao K, *et al.*: **Improved Kalman filter and its application in initial alignment.** *Optik*. 2021; **226**: 165747.
[Publisher Full Text](#)
 66. Lim B, Zohren S: **Time-series forecasting with deep learning: a survey.** *Philos Trans A Math Phys Eng Sci.* 5 April 2021; **379**(2194): 20200209.
[Publisher Full Text](#)
 67. Giantsidi S, Tarantola C: **Deep learning for financial forecasting: A review of recent trends.** *Int. Rev. Econ. Financ.* 2025; **104**: 104719.
[Publisher Full Text](#)
 68. Sami HM, Jalal MS, Kabir MF, *et al.*: **IntPort: An Intelligent Portfolio Construction Technique Based on Financial Forecasting by Statistical Average Method.** *IEEE Access.* 2025; **13**: 35355–35375.
[Publisher Full Text](#)
 69. Bekaert G, Harvey C: **Emerging Equity Markets in a Globalizing World.** *SSRN Electron. J.* 2014.
[Publisher Full Text](#)
 70. Fama EF, French KR: **A five-factor asset pricing model.** *J. Financ. Econ.* 2015; **116**(1): 1–22.
[Publisher Full Text](#)
 71. Advisor Perspectives: **Global market volatility indices 2025 update.** 2025.
[Reference Source](#)
 72. Ernst & Young: **EY global alternative fund survey 2025. EY Reports.** 2025.
 73. J.P. Morgan Asset Management: **Long-term capital market assumptions 2025. J.P. Morgan Reports.** 2025.
 74. International Monetary Fund: *World economic outlook, October 2025.* IMF Publications; 2025.
 75. Arlot S, Celisse A: **A survey of cross-validation procedures for model selection.** *Stat Surv.* 2010; **4**: 40–79.
[Publisher Full Text](#)
 76. Muhammad D, Ahmed I, Naveed K, *et al.*: **An explainable deep learning approach for stock market trend prediction.** *Heliyon.* 2024; **10**(21): e40095.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status: 

Version 1

Reviewer Report 03 June 2026

<https://doi.org/10.5256/f1000research.197298.r482181>

© 2026 Alex Avelar E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ewerton Alex Avelar

Universidade Federal de Minas Gerais, Belo Horizonte, State of Minas Gerais, Brazil

Dear Authors,

The paper entitled "Forecasting Minute-Level Stock Prices with Denoised Data: A Comparative Study of Developed vs. Developing Financial Markets" is fascinating and has great potential for development.

However, several significant theoretical and methodological issues, as well as the limited exploration of the results, should be addressed.

I hope this report is helpful to you.

REPORT

The following aspects should be improved:

GENERAL

"In emerging markets (e.g., USA, Japan, Singapore) the risk will be of a higher nature due to several elements such as political instability, currency fluctuation, and liquidity" (p. 4). Please correct this statement to "... (e.g., Brazil, India, and Malaysia) ...".

INTRODUCTION

The justification for selecting developed and developing countries should be strengthened. It is recommended that the selection criteria be presented in the Methodology section and supported with additional evidence. For example, why was China not included?

The manuscript does not appear to provide sufficient support for the various objectives presented.

"Our main contribution is to show that it is possible to systematically extract alpha from these environments using deep learning, and to present solid evidence for the impact of computational advances in reducing the volatility of developing markets and converting it into actionable, high-

quality returns" (p. 4). Could this not be considered the primary objective of the study?

I suggest presenting the main objective as follows: "to analyze the possibility of systematically extracting alpha from developing-country markets using deep learning."

The current objective "a" does not seem to be an actual objective of the study, but rather an underlying assumption. Likewise, objectives "b" and "c" appear not to be genuine research objectives. Therefore, I suggest specific objectives (which could also be stated as research questions), such as:

- _ To develop deep learning models tailored to both developing and developed countries.
- _ To evaluate and compare the predictive performance of the models developed for both groups of countries.
- _ To discuss the implications of the findings from the perspective of investors and other market participants.

Also, broaden the discussion of the study's potential contributions to different stakeholders, such as analysts, market regulators, and the economies of developing countries as a whole.

LITERATURE REVIEW

Overall, this section is highly fragmented. Consider improving paragraph articulation and reducing the number of subsections.

"For instance, a recent causality investigation of stock prices and macroeconomic indicators in the Indian stock market reveals a tight relationship between volatility in developing economies and macroeconomics, which, in turn, influences asset pricing" (p. 4). It is important to support this argument with broader evidence (i.e., general conclusions from multiple studies) rather than relying on a single study.

In subsection 4.2, provide additional evidence that deep neural networks outperform memory-free algorithms (such as Random Forest, for example). Furthermore, explain why these specific models were selected and employed.

"Combined with regularisation techniques inherent in modern architectures (e.g., dropout layers in LSTMs), this validation strategy ensures that the alpha generated is robust and not merely an artefact of overfitting" (p. 6). Please provide a more detailed explanation of why this occurs. Strengthen the rationale for the use of K-means clustering by providing additional supporting arguments and studies.

Clarify the criteria used for selecting the papers presented in Table 1.

METHODOLOGY

Provide support from the literature for the specific methodological choices related to "Data Collection" (e.g., why exactly 30 days and 20 firms were selected), "Price Denoising and Signal Generation", "Prediction Model specification", and other key methodological decisions.

Please explain the decision to avoid using traditional technical indicators more clearly. What are the potential implications of excluding such indicators from the analysis?

Improve the quality of Figure 6. Additionally, why is this figure presented at this stage of the manuscript?

RESULTS

Why was RMSE selected instead of other widely used machine learning performance metrics, such as MAPE, for model evaluation?

Please improve the quality of the figures, as they are currently almost illegible.

I suggest replacing the specific term "accuracy" with a more general term, such as "performance,"

since accuracy is technically a classification metric. In contrast, the authors employed RMSE, which is a regression metric.

CONCLUSIONS / DISCUSSION

The practical implications of the findings should be discussed more thoroughly, extending beyond their relevance to investors.

There is no need to create subsection 9.1 if subsection 9.2 does not exist.

Regarding the limitations, it is important to discuss more carefully why not all developing countries may benefit equally from the findings and to address potential sampling biases in greater depth.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Machine learning applied to the financial market, algorithmic management, corporate finance, management control systems.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research