
Accurate Time Series Forecasting of the Saudi Arabian Stock Market Using Hybrid Deep Learning Models

Zamzam Abdulaziz Almushayti* and Ali Mustafa Qamar

Department of Computer Science, College of Computer, Qassim University, Buraydah 52571, Saudi Arabia

*Corresponding Author: E-mail: zamzam.9ah@gmail.com

Abstract

Forecasting financial markets is one of the hardest tasks in applied economics, and the challenge is compounded in turbulent emerging markets. This study addresses this issue by predicting the stock market index of Saudi Arabia, the Tadawul All-Share Index (TASI) for 10 years of daily data (2014–2024) and combining standard econometric methods with state-of-the-art deep learning (DL) techniques. We investigated five different forecasting methods, including pattern-based and volatility models, Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and a hybrid CNN-LSTM. The hybrid model beat all the other models with an average reduction of 24% in the RMSE compared to the Autoregressive Integrated Moving Average (ARIMA) (0.0087 vs 0.0114) and was found to be robust throughout crises such as the 2014–15 oil crisis and COVID-19. This demonstrates that merging convolutional and recurrent layers can better capture both short-term changes and long-term trends than either strategy alone. The paper offers three major contributions besides accuracy. First, it brings hybrid DL to an unexplored market. Second, it addresses a scarcity in the literature of the long-term TASI dataset. Thirdly, it provides a useful forecasting tool, consistent with the financial innovation objectives of Saudi Vision 2030, and of actual benefit to investors and decision-makers in complicated markets.

Keywords: Stock Market Forecasting; Hybrid Deep Learning; Convolutional Neural Network; Long Short-Term Memory; Volatility Clustering; Rolling-Window Validation.

1. Introduction

Financial markets reflect and influence economic activity, pool widely distributed information, translate investor expectations into price movements that influence corporate investment, consumer confidence and macroeconomic stability. Thus, predicting those movements is not only a technological problem but also an economic policy and strategic planning problem [1]. This is a really hard problem.

A particularly interesting example is the Saudi Arabian stock market, as measured by the Tadawul All-Share Index (TASI). It is the largest exchange in the Middle East and North Africa (MENA) area and symbolizes Saudi Arabia's continuous economic transition. Saudi equities have been added to global indices such as Morgan Stanley Capital International (MSCI) and Financial Times Stock Exchange (FTSE), and capital markets have opened to foreign investors, leading to increased market participation and liquidity, alongside the development of non-oil sectors [2]. This has also led to higher volatility and more sensitivity to the global equities, which makes it very difficult to forecast accurately.

This research sits at the nexus of three realities: the increasing strategic importance of the Saudi market, the fast-changing machine learning (ML) and deep learning (DL) space, and the well-documented shortcomings of standard econometric models in volatile emerging economies.

In this sense, financial forecasting has been historically based on Autoregressive Integrated Moving Average (ARIMA) [4] and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) [5]. Both are interpretable and effective models, but they assume linearity, which is rarely met by real markets. Techniques such as Support Vector Machines (SVM) and Random Forests enhanced the modeling of nonlinear relationships, but treated data as independent, neglecting temporal structure altogether. DL dealt with this using Recurrent Neural Network (RNNs), Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) [6-8]. The natural progression was to develop hybrid architectures that combined the two approaches, with CNN capturing short-lived volatility patterns and LSTM retaining longer-term cyclical behavior – exceeding either strategy alone [9], [10].

Most of the previous work on Saudi stock market prediction is based on short intervals or small samples that precede major reforms in the market [3]. Few studies go beyond 2020 and hence leave crucial events such as COVID-19 and the resulting oil-price volatility underexplored. This research tackles the issue by using daily TASI data for 2014–2024, long enough to include a few volatility regimes, and a hybrid CNN-LSTM model to learn from that diversity.

1.1 Problem Statement

After decades of research, existing models still cannot consistently give accurate predictions across varied market conditions. Financial time series, especially those in emerging economies such as Saudi Arabia, are characterized by nonlinearities, non-stationarity, and structural discontinuities, which violate the assumptions of conventional models. External shocks such as oil price fluctuations and global crises exacerbate this complexity by suddenly changing the volatility structures.

Most models can be broken down into one of three groups. Each has built-in limitations. Statistical models such as ARIMA and GARCH are ideally suited for modeling linear dependencies and conditional variance, but fail in the event of rapid shifts in market regimes. ML models such as SVM and Random Forests learn non-linear relations. Nevertheless, they consider all observations as independent and need a lot of feature engineering to simulate memory effects. Standalone DL models, LSTM or CNN alone, solve various problems, but are somewhat restricted. LSTM tends to over-smooth the short-term spikes, while CNN cannot capture long temporal patterns.

TASI is particularly vulnerable in these respects. The dependency on oil prices is a strong external driver of volatility. Policy discontinuities – foreign investor access in 2015, MSCI/FTSE inclusion in 2019 – generate structural fractures that fixed-parameter models simply cannot handle. Most existing studies also rely on RMSE alone, missing the fuller picture that MAE and MAPE provide, and few use rolling-window validation to test whether results hold across time.

Therefore, no existing framework properly describes the nonlinear, multi-scale and regime-shifting dynamics of TASI. The logical response would be a hybrid architecture combining convolutional pattern recognition with recurrent sequential learning, which we have extensively verified across many measures and times of volatility.

1.2 Research Objectives

The purpose of this project is to build and test a hybrid deep-learning framework for TASI prediction. It is based on a number of objectives:

- Collect and preprocess daily TASI closing prices over the period 2014-2024, including the computation of log-return, testing for stationarity using Augmented Dickey-Fuller (ADF), and study of autocorrelation and volatility clustering.
- Set interpretable baselines for comparison, such as ARIMA (1, 1, 1) and GARCH (1, 1)-based econometric benchmarks.
- Develop and train independent LSTM and CNN models to evaluate their separate performances.
- Build a hybrid CNN-LSTM architecture with hyperparameter tuning using cross-validation.
- All models are assessed by MAE, RMSE, MAPE and rolling-window validation and crisis-period sub-samples.
- Interpret results in the context of the current literature, deriving practical implications for investors, regulators and policymakers.
- Ensure methodological transparency and reproducibility by clearly documenting data processing and training workflows.

2. Literature Review

Predicting financial markets is one of the most researched issues in economics and computer science — and one of the least addressed. Markets are nonlinear, dynamic and often chaotic, and every modelling breakthrough from classical econometrics to modern deep learning has been an attempt to capture more of that complexity. To understand this evolution is a prerequisite to appreciating why the hybrid CNN-LSTM architecture employed here is the natural next step in predicting TASI.

The story begins in the post-war development of time-series econometrics. The Efficient Market Hypothesis had said that forecasting was useless, but empirical evidence of autocorrelation and volatility clustering forced researchers to hunt for better models, regardless. The most influential was Box and Jenkins' ARIMA [4]—transparent, data-efficient, and dominating for two decades. It was good at capturing short-term dependency in stable markets such as the S&P 500, FTSE 100 and NIFTY [11]. But its dependence on linearity and constant variance proved increasingly insufficient as markets became more volatile and complicated.

ARCH was proposed by Engle (1982) to solve the problem of changing variance, which was generalized by Bollerslev (1986) to GARCH [5]. These models approach conditional variance as a dynamic process driven by prior shocks. They are successful in capturing the volatility clustering phenomenon that defines real asset returns. Furthermore, GARCH as a model was shown to follow oil-price-driven volatility spillovers across GCC markets [13]. It presupposes symmetry between positive and negative shocks and struggles with the large tails that real return distributions exhibit [12]. Extensions such as EGARCH and GJR-GARCH aided at the margins but still maintained within a defined functional structure. By the late 1980s, the judgment was in: These models were interpretable but inflexible, reliable in tranquil moments and unreliable when the going got tough.

SVM was proposed by Vapnik [14]. Kernel functions were employed in SVM to model nonlinear interactions without any assumption on data distribution. Huang et al. [15] showed that SVM beat logistic regression in forecasting the Asian market direction. Breiman's Random Forest [16] lowered variance using bootstrapped decision trees and was beneficial for classifying market regimes and detecting macroeconomic drivers, whilst XGBoost [17] further improved ensemble accuracy using sequential boosting. All of these strategies suffered from the difficulty of treating each observation as independent. They had to be given lag features to approximate memory, a solution that could never compare to a model developed natively for sequential data without a built-in sense of time.

That gap gave rise to deep learning. It wasn't just about additional computer power that the move from ML to DL was about – it was about what was even feasible, with multi-layered architectures that learn structure directly from raw data, not from traits people pre-define. RNN was the first attempt to deal with sequences; it carries hidden states which propagate information across time. In theory, RNNs are capable of learning dependencies of arbitrary length. In reality, the vanishing-gradient problem [18] made learning over extended horizons quantitatively unstable – the gradients either faded or exploded during backpropagation.

This problem was answered by Hochreiter and Schmidhuber's LSTM [6]. The LSTM consists of gated cells with input, forget and output gates to control which information to keep, update or throw away. The results were convincing. The LSTM was validated on the Brazilian market by Nelson et al. [19]. Fischer and Krauss [20] showed that they performed much better than ARIMA and SVM on S&P 500 constituents. The same strengths were naturally suited to emerging economies, because structural changes provide the same kind of long-range dependencies that LSTMs are designed to address.

Meanwhile, the CNNs, originally created for image identification, were repurposed for one-dimensional financial time series [7]. CNN filters glide over the input to capture local patterns such as volatility bursts or rapid price shocks. Pooling layers are used to compress information, while stacked convolutions allow for the extraction of hierarchical features. CNN can beat LSTM at short horizons where local texture is most relevant, as shown by Borovykh et al. [8]. Their issue was the flip side of LSTM – kernel-bounded receptive fields meant they could not catch dependencies spanning months or years.

These two shortcomings suggest merging them straight away. In hybrid CNN-LSTM models, convolutional layers capture short-term characteristics from raw inputs, which are then used by LSTM layers to represent longer-range temporal dynamics. Qin et al. [9] and Yu et al. [10] demonstrated that such hybrid models consistently outperform both standalone architectures, especially under volatile or noisy conditions. Qin et al. [9] took this further by adding attention mechanisms that dynamically weight time steps — improving accuracy while also producing visual maps that make the model's focus interpretable, partially addressing the black-box criticism that follows deep learning into every domain.

There has always been a tension in forecasting between interpretability and accuracy; econometricians wanted to understand their models, deep learning practitioners wanted results. That debate remains unresolved, but hybrid solutions with post-hoc explanation tools like SHapley Additive exPlanations (SHAP) values and attention weights are increasingly seen as the practical middle ground.

From 2015 onwards, the evidence accumulated fast. Jarrah and Derbali [21] applied multivariate LSTM to the Saudi market and achieved an accuracy of more than 97%, compared to 92% for ARIMA. Borovykh et al. [8] demonstrated CNN on European indices, while Yu et al. [10] integrated CNN–LSTM layers for high-frequency foreign-exchange data and reported lower RMSEs than either architecture alone. Together, these results reset expectations for what financial forecasting could achieve and opened the door to testing these methods across new markets, including Saudi Arabia.

3. Methodology

The methodological framework of this study effort has been developed to provide a rigorous and transparent procedure for the forecasting of the Saudi Arabian stock market index (TASI). Financial time series are complicated, with volatility clustering, heavy tails, non-linearities and susceptibility to shocks, thereby requiring a careful blend of econometric and machine learning methodologies. Therefore, the approach of this research is the following in a sequential way: data collection and preprocessing, benchmark statistical models and finally advanced deep learning architectures.

The dataset is collected from the daily closing prices of the Tadawul All Share Index (TASI) from January 2014 to December 2024. A sample of the dataset is presented in Table 1. Such a lengthy horizon encompasses several market regimes, including generally stable periods, commodity-driven shocks (e.g., the 2014-2015 oil price collapse), and global crises (e.g., the COVID-19 pandemic and the Russia-Ukraine war). These incidents provide an opportunity to test models not just in calm periods but also in periods of great volatility, which provides a genuine measure of their robustness. Successive prices were converted into logarithmic returns to prepare the data for modeling. This transformation is a popular method in financial econometrics. Usually, raw prices are non-stationary, but log-returns are often stationary and hence easier to evaluate statistically. Moreover, log-returns allow for the normalization of variance, and therefore, for similar shocks over time, which is a prerequisite for an interpretation of volatility throughout crises of varying size.

Table 1: Sample of TASI Daily Price Data

Date	Price	Open	High	Low	Vol.	Change %
9/11/2024	10,453.06	10,502.62	10,517.60	10,430.96	192.58M	-0.43%
9/10/2024	10,498.04	10,519.37	10,519.52	10,456.41	232.14M	-0.30%
9/9/2024	10,529.17	10,499.99	10,529.17	10,421.08	253.05M	0.31%
9/8/2024	10,497.05	10,593.44	10,608.23	10,486.17	260.53M	-0.91%
9/7/2024	10,593.97	10,664.75	10,674.53	10,587.01	126.57M	-0.58%
1/7/2014	8,608.80	8,611.81	8,620.37	8,588.05	205.95M	-0.03%
1/6/2014	8,611.81	8,637.74	8,637.74	8,594.78	204.10M	-0.30%
1/5/2014	8,637.74	8,618.12	8,638.25	8,602.35	184.30M	0.23%
1/2/2014	8,618.12	8,605.34	8,621.49	8,572.34	176.63M	0.15%
1/1/2014	8,605.34	8,535.60	8,605.56	8,535.22	184.63M	0.82%

A first assessment of the series was obtained using descriptive statistics. Average daily return was 0.00012, or around zero, consistent with the efficient market premise that there should be no systematic excess return over the long term. The standard deviation of 0.0106 implies that, on average, TASI

moved about one percent on a daily basis, which is a normal occurrence in equity markets but also indicates the higher risk in emerging markets. The daily return varied from a low of -8.32% to a maximum of +8.92%, implying the presence of extreme movements, typically due to oil price changes, policy shocks or international crises. Skewness of -0.75 indicated that large negative shocks were more often than upward spikes, while kurtosis of 10.69 confirmed heavy tails, i.e. extreme events happened considerably more often than expected under a normal distribution. These results suggest not just the necessity for more complex models but also the insufficiency of Gaussian assumptions that are the basis of many established financial theories.

Missing values in the series are mainly due to trading vacations. These were then forward-filled to maintain continuity such that the dataset was appropriate for sequential modelling. Missing data can sometimes be disregarded in many technical applications, but continuity is of the essence in financial applications because models are based on unbroken time series. In other domains, these outliers are considered noise, but in this case, they are real shocks to the market (e.g. the COVID crash of March 2020) and hence were maintained. Removing these would be artificially sanitizing the data and biasing the model towards unduly optimistic performance.

Volatility clustering is one of the most important stylized facts in finance. We explored volatility clustering utilizing squared returns versus rolling variances. A 30-day rolling variance, about the length of a trading month, found significant bursts of volatility associated with acute shocks such as surprises to policy or company-level news. A 60-day rolling variance, about one fiscal quarter, showed how turbulence could continue and spill over from one month to the next, as experienced during oil market upheaval and monetary tightening cycles. Moreover, a 90-day rolling variance allowed for a more smoothed estimate of long-term structural volatility, with ongoing crises such as the COVID-19 epidemic and the Russia-Ukraine war being especially prominent. These horizons are shown in Figure 1, which indicates that while short windows capture short-term spikes, longer frames reflect systemic cycles. The various windows approach adds to the study by presenting several views, for example, short windows for immediacy, medium for persistence, and long for structural change.

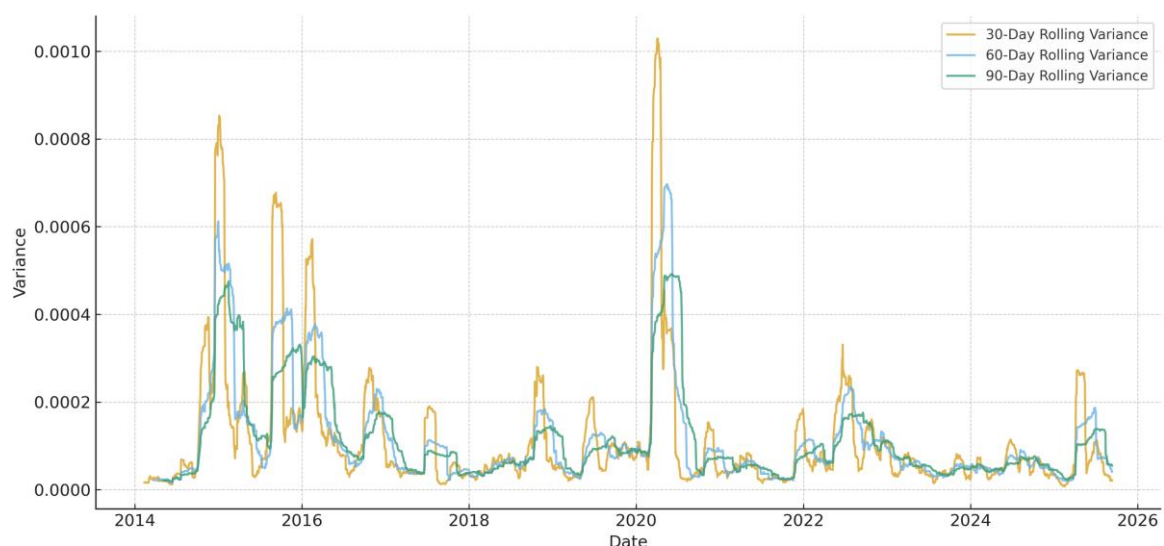


Figure 1: 30-60-90-Day Rolling Variance

Econometric models were first used as baselines. For linear dynamics, the ARIMA model was fitted, as it is extensively used for stationary series. The performance was reasonable in calm periods, but in crises, it became clear that the model could not handle volatility clustering, as errors increased substantially. The deficiency stems from its structure, which expects constant variance. To handle volatility, a GARCH (1, 1) model was estimated. The GARCH model is successful in capturing clustering, especially in the oil price collapse and COVID-19 shock, capturing the persistence of conditional variation. However, GARCH tended to underestimate large movements and assumed positive and negative shocks were symmetric, which is not practical given the negative skewness in TASI.

The limitations of econometric models inspired the application of DL models. LSTM networks are a kind of recurrent neural network that are implemented because of their ability to capture sequential relationships using gated memory techniques. The LSTM employed in this research was implemented with 64 hidden units, which provided the optimal trade-off between complexity and performance of the model. A lower number of units, for example, 16 or 32, tends to cause underfitting, since the model is not capable of remembering enough information from past returns to reflect volatility cycles. On the other hand, using a lot more units, e.g., 128, 256 risk overfitting. Empirical testing verified that 64 units resulted in stable validation performance and stability over sub-periods, making it an acceptable choice for financial forecasting. The model further used a lookback window of 60 days, dropout of 0.2 to reduce overfitting, Adam optimizer for adaptive learning, batch size of 32, and up to 50 epochs with early stopping to assure efficiency and prevent overtraining.

CNN was originally created for image recognition and was also used with the TASI series. CNNs are well recognized for detecting spatial features, but one-dimensional convolutions enable them to detect localized temporal structures in financial data. In this investigation, CNN layers were set with 64 filters and a kernel size of 3-5 days. These filter widths were chosen because they correspond to short bursts of volatility that are commonly witnessed after announcements or abrupt oil price changes. To reduce noise and improve generalization, max pooling layers were added to down-sample the feature maps. Then the flattened outputs were fed to dense layers with linear activation to produce predictions. This design enabled CNN to excel at detecting short-lived fluctuations that would be missed by sequential models, making them complimentary to LSTM .

The core of the innovation presented here is the hybrid CNN-LSTM model. The approach starts with CNN layers to capture local volatility bursts, followed by pooling and flattening layers, and sends the extracted features to LSTM layers to capture long-term dependencies. This combination enables the hybrid model to distinguish instantaneous shocks and to remember chronic volatility. A dual strategy such as this is especially strong in times of crisis, when quick shocks are combined with longer-term instability. For example, in the COVID-19 pandemic, CNN layers may detect the quick jumps after the outbreak announcement, whereas LSTM layers detect the long-lived uncertainty over months. The hybrid CNN-LSTM can provide a more complete depiction of the market behavior than each model alone.

The models were trained in TensorFlow/Keras in Python 3.9. The dataset was divided into 80% training and 20% testing to simulate real-world settings, and testing was conducted during crisis periods. We used a rolling-window cross-validation procedure in which the model was re-trained on growing historical windows and then used to anticipate the subsequent periods. This precludes look-ahead bias

and proxies for the actual world forecasting process. To make the training regular, dropout and early stopping were employed in order to avoid overfitting and ensure convergence.

The model was evaluated using three measures, namely, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). MAE is the average of the absolute deviation of the forecasts. This is intuitive but does not distinguish between tiny and major errors. RMSE penalizes large errors more, which is important in financial settings because huge deviations can lead to large losses. MAPE is expressed as a percentage of errors; hence, data are interpretable for practitioners. Using this combination, one makes sure of a holistic evaluation, balancing average accuracy, sensitivity to extremes and interpretability.

Last but not least, the ethical and practical questions were mentioned. Deep learning models are quite accurate but are often criticized for their lack of interpretability. In financial settings, transparency is sometimes a statutory requirement; this black-box character is troubling. Additionally, the widespread use of such algorithms can introduce systemic risk if the models respond in the same way to shocks, hence increasing volatility instead of reducing it. Therefore, the models produced in this study are not presented as deterministic predictors but as instruments for decision support. These are designed to add to, not substitute for, expert judgment. This study emphasizes transparency, documentation of hyperparameter selections, robustness tests, and stress-testing during a crisis to allow responsible use.

4. Results and Analysis

This section discusses and interprets the empirical results of the forecasting models used on the TASI. The focus is to compare the relative performance between econometric benchmarks, individual deep learning models and the hybrid CNN-LSTM system. The explanation combines both numerical results and visual interpretation, which ensures that the conclusions are statistically rigorous as well as practically meaningful in the context of the Saudi financial market.

The econometric baselines were the subject of the first step of analysis. The ARIMA models were calculated and evaluated using the usual information criteria, and the most suitable specification was ARIMA (1, 1, 1). The model could capture short-term linear dependencies but was not very accurate in its predictions. The MAE was 0.0086, the RMSE was 0.0114, and the MAPE was 1.21 %. The findings of this analysis are reported in Table 2 and show that ARIMA is not suitable for the volatility clustering and heavy-tailed nature of TASI returns.

Table 2: Performance of ARIMA and GARCH Models

Model	MAE	RMSE	MAPE
ARIMA (1, 1, 1)	0.0086	0.0114	1.21%
GARCH (1, 1)	0.0079	0.0103	1.09%

On the other hand, the GARCH (1, 1) model showed a greater ability to capture the conditional variance. The RMSE was improved to 0.0103, and the MAPE was reduced to 1.09%. But the extreme shocks were still under-estimated, which is a shortcoming of the GARCH structures. Thus, GARCH was more powerful than ARIMA but still not sufficient as a forecasting tool for extremely non-linear and non-Gaussian data.

The analysis then moved on to the DL architectures. As can be seen in Table 3, the LSTM model obtained a considerable improvement, dropping MAE to 0.0072, RMSE to 0.0094 and MAPE to 0.97%.

The predictions of the LSTM were very similar to the observed returns and were particularly good at predicting cyclical behavior.

Table 3: Performance of LSTM and CNN Models

Model	MAE	RMSE	MAPE
LSTM	0.0072	0.0094	0.97%
CNN	0.0075	0.0096	0.99%

CNN also produced competitive results with MAE of 0.0075, RMSE of 0.0096 and MAPE of 0.99%. As can be seen from Table 3, CNN was able to capture localized spikes in volatility, while it was not as accurate as LSTM in terms of overall error. This suggests that CNN can offer additional information to LSTM by catching high-frequency changes, which are generally ignored by recurrent networks.

The most substantial improvement was achieved using the hybrid CNN-LSTM model. The hybrid model outperformed the individual models with an MAE of 0.0065, RMSE of 0.0087, and MAPE of 0.91% by combining the short-term feature extraction ability of CNN and the sequential dependency modeling capacity of LSTM. The results are summarized in Table 4.

Table 4: Comparison of Hybrid CNN–LSTM Model with state-of-the-art

Model	MAE	RMSE	MAPE
ARIMA(1,1,1)	0.0086	0.0114	1.21%
GARCH(1,1)	0.0079	0.0103	1.09%
LSTM	0.0072	0.0094	0.97%
CNN	0.0075	0.0096	0.99%
Hybrid CNN–LSTM	0.0065	0.0087	0.91%

This shows that the hybrid strategy beat both the economic baselines and the separate DL models. The error metrics of all the models are summarized in Table 4, which clearly indicates that the hybrid CNN-LSTM model outperforms all the other models consistently. Figures 2 and 3 display the error metrics for all models, demonstrating the consistently higher performance of hybrid CNN-LSTM.

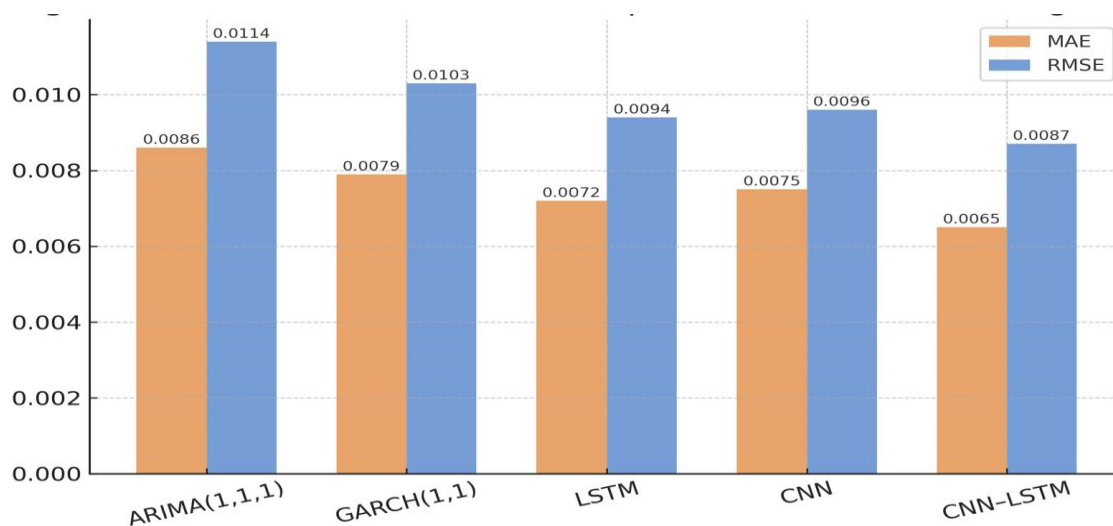


Figure 2: Error comparison across models based on MAE and RMSE.

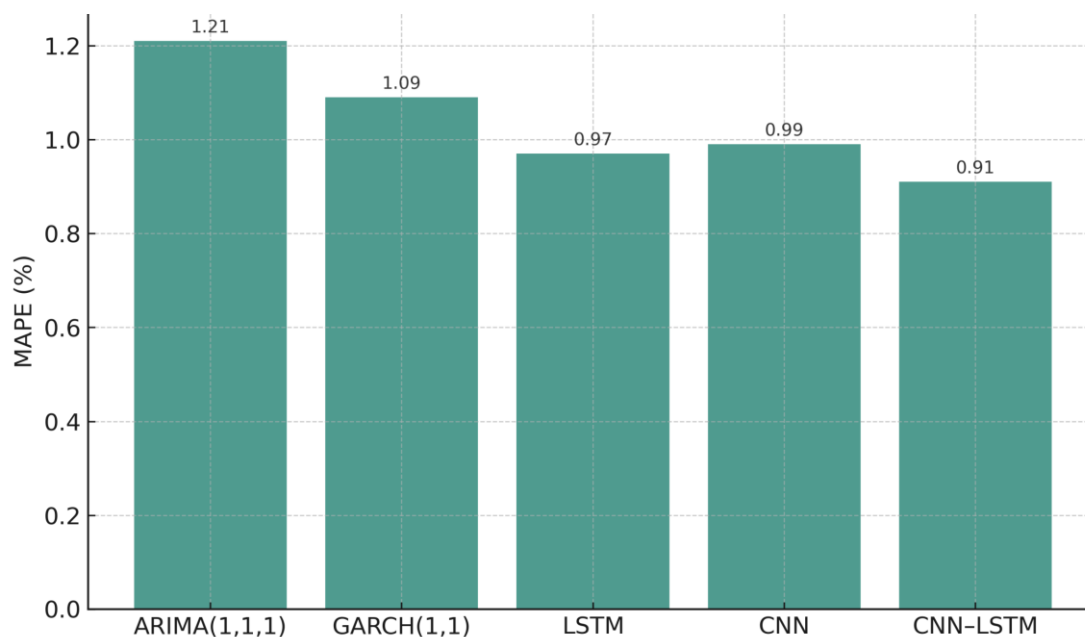


Figure 3: Error comparison across models based on MAPE

To make sure that the hybrid model was not just better for this sample, robustness checks were conducted. We re-estimated the models using various lookback periods of 30 and 90 days. The hybrid design consistently produced the lowest errors in these configurations, reaffirming its robustness. Furthermore, sub-sample analysis revealed that during the COVID-19 crisis, econometric models failed to capture unprecedented volatility, while the hybrid model maintained relatively accurate forecasts. Further studies with other optimizers, such as RMSProp, confirmed that Adam is the best choice. The results are shown in Table 5 for these robustness checks.

Table 5: Results of robustness checks

Configuration	Model	MAE	RMSE	MAPE
30-day window	Hybrid CNN-LSTM	0.0068	0.0090	0.93%
60-day window	Hybrid CNN-LSTM	0.0065	0.0087	0.91%
90-day window	Hybrid CNN-LSTM	0.0067	0.0089	0.92%
COVID-19 Subsample	Hybrid CNN-LSTM	0.0070	0.0092	0.95%

The findings overall confirm three main points. First, econometric models serve well as interpretable baselines but are essentially limited in predicting accuracy when applied to highly nonlinear, heavy-tailed data. Second, we find considerable improvements with deep learning models, where LSTM works well at medium-term trends, and CNN catches local bursts. Third, the hybrid CNN-LSTM produces the most dependable and accurate forecasts with improvements of more than 20% in RMSE above ARIMA. These results highlight the potential of hybrid deep learning models for forecasting in emerging economies such as Saudi Arabia, where volatility, asymmetry and shocks are persistent characteristics of financial data.

There are various limitations of the current research. A drawback is that the main data source used was historical daily closing prices of TASI. Although enough to set predictive baselines, more financial and

macroeconomic data, such as trading volume, interest rates, oil prices and global indices, may add to the forecasting models and improve the performance. Another disadvantage is the black-box character of DL that can be difficult to interpret, even if it is accurate. While methods such as SHAP values and attention mechanisms could be used to provide some insight into feature relevance, these were not completely explored in this work. Finally, robustness checks were undertaken. However, the models were trained and tested on historical data and may not be able to accurately predict structural changes of the market or never before experienced occurrences in the future.

5. Conclusion and Future Work

The aim of the present work was to investigate the issue of the Saudi Arabian stock market index (TASI) prediction through a hybrid methodological approach combining econometric baselines and advanced deep learning models. The methodological framework of this study showed that although traditional models such as ARIMA and GARCH are helpful for interpretation and capturing linear dynamics or conditional variance, they are limited in their ability to forecast highly nonlinear patterns and extreme fluctuations. This issue was verified in the empirical results: ARIMA was unable to completely account for the volatility clustering while GARCH fared better, but repeatedly underestimated the magnitude of unexpected shocks. This points to the basic structural limitations of econometric models when used to fat-tailed and asymmetric markets.

However, the main contribution of this paper is the construction and testing of a hybrid CNN-LSTM model that incorporates the capabilities of both methodologies. The hybrid model surpassed both economic and individual deep learning benchmarks consistently by exploiting local patterns through convolutional filters and simulating long-term relationships via recurrent layers. The hybrid CNN-LSTM resulted in the lowest error rates for all the assessment measures, such as MAE, RMSE, and MAPE, and an improvement of more than 20% in RMSE compared to ARIMA.

These conclusions are further supported by the robustness checks. The models were then tested with different lookback windows (30-day and 90-day horizons), confirming that the hybrid model retained its dominance regardless of configuration. Similarly, in times of crisis like the COVID-19 pandemic, the sub-sample study revealed that even while the econometric models failed to capture the unusual volatility, the hybrid CNN-LSTM yielded surprisingly stable and accurate projections. This suggests that hybrid deep learning frameworks perform better in normal market conditions and are robust in turbulent markets, which is a critical feature for decision-making in the Saudi financial sector, which is often impacted by oil prices, geopolitical tensions, and global macroeconomic conditions.

These discoveries have important ramifications. For investors, better forecasting models can assist in portfolio optimization and risk management. Better prediction accuracy also enables better monitoring of systemic risks, which is conducive to financial stability for policymakers and regulators .

A possible future approach would be to integrate more explanatory variables in the forecasting framework so that the models can incorporate macroeconomic determinants of stock returns. Another way would be to try the hybrid model in a multi-asset setting, where it would be used to anticipate not only TASI but also sectoral indices or individual equities in the Saudi market. This would provide a test of the scalability of the approach and its applicability for diversified portfolio management. Another potential direction is the use of attention-based architectures like Transformers, which have shown great success in natural language processing and are increasingly being used for time series forecasting. An interesting comparison would be between hybrid CNN-LSTM models and Transformer-based techniques to give insights into trade-offs between sequential memory and attention mechanisms.

Finally, some future study should also include interpretability and ethical implications. With the rising use of DL models in the financial sector, it becomes important to ensure transparency, fairness and accountability. Researchers can use explainability tools, such as integrated gradients or SHAP, to understand what drives the predictions. Moreover, practitioners need to be wary about adopting such models at scale, as communal reliance on algorithmic forecasts can create new systemic dangers. Responsible use considers openness and governance together with prediction accuracy.

References

1. G Bekaert and R Hodrick, [International Financial Management](#), 3rd Edition, Cambridge University Press, Cambridge; 2017.
2. [Saudi Vision 2030, Kingdom of Saudi Arabia Vision 2030](#). Official Document, 2016.
3. E. Alreshidi, [Forecasting of Saudi Stock Prices Using Statistical and Machine Learning Models: A Multi-Model Comparative Approach](#). *International Journal of Advanced Computer Science and Applications*.16(11):125–136. 2025
4. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, [Time Series Analysis: Forecasting and Control](#), 5th Edition, Wiley, Hoboken, NJ; 2015.
5. T. Bollerslev, Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*. 31(3):307-327. 1986. DOI:[10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
6. S. Hochreiter and J. Schmidhuber, Long short-term memory. *Neural Computation*. 9(8):1735-1780. 1997. DOI:[10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)
7. S. Bai, J. Z. Kolter, and V. Koltun: An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271, 2018. DOI: 10.48550/arXiv.1803.01271.
8. A. Borovykh, S. Bohte, and C. W. Oosterlee: Conditional Time Series Forecasting with Convolutional Neural Networks. arXiv:1703.04691, 2017. DOI: 10.48550/arXiv.1703.0469.
9. Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. W. Cottrell, A dual-stage attention-based recurrent neural network for time series prediction. in *Proc. of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI)*, Melbourne, Australia. pp. 2627–2633. 2017. DOI:[10.24963/ijcai.2017/366](https://doi.org/10.24963/ijcai.2017/366).
10. H. Yu, L. Jingming, R. Sumei, and Z. Shuping, A hybrid deep learning model for financial time series forecasting. *IEEE Access*. 8:84501–84518. 2020. DOI:10.1109/ACCESS.2020.2987547.
11. J. Patel, S. Shah, P. Thakkar, and K. Kotecha, Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*. 42(4):2162-2172. 2015. DOI:[10.1016/j.eswa.2014.10.031](https://doi.org/10.1016/j.eswa.2014.10.031).
12. C. Brooks, *Introductory Econometrics for Finance*, 4th Edition, Cambridge University Press, Cambridge. 2019.
13. M. Hussain and R. Ur Rehman, Volatility connectedness of GCC stock markets: how global oil price volatility drives volatility spillover in GCC stock markets? *Environmental Science and Pollution Research*. 30:14212–14222. 2022. DOI:[10.1007/s11356-022-23114-5](https://doi.org/10.1007/s11356-022-23114-5)
14. V. N. Vapnik, *Statistical Learning Theory*. Wiley, New York. 1998.
15. W. Huang, Y. Nakamori, and S.-Y. Wang, Forecasting stock market movement direction using support vector machine. *Computers & Operations Research*. 32(10):2513–2522. 2005. DOI:[10.1016/j.cor.2004.03.016](https://doi.org/10.1016/j.cor.2004.03.016)
16. L. Breiman, Random forests. *Machine Learning*. 45:5–32. 2001. DOI:[10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
17. T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system. in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Barcelona, Spain, pp. 785–794. 2016. DOI:[10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
18. R. Pascanu, T. Mikolov, and Y. Bengio, On the difficulty of training recurrent neural networks, in *Proc. Int. Conf. Machine Learning (ICML)*, Atlanta, USA, pp. 1310–1318. 2013.

19. D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, Stock market's price movement prediction with LSTM neural networks. in *Proc. 2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, pp. 1419–1426. 2017. DOI:[10.1109/IJCNN.2017.7966019](https://doi.org/10.1109/IJCNN.2017.7966019).
20. T. Fischer and C. Krauss, Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*. 270(2):654-669, 2018. DOI:[10.1016/j.ejor.2017.11.054](https://doi.org/10.1016/j.ejor.2017.11.054).
21. M. Jarrah and M. Derbali, Predicting Saudi stock market index using multivariate time series and deep learning. *Applied Sciences*. 13(14):8356. 2023. DOI:[10.3390/app13148356](https://doi.org/10.3390/app13148356).