Forecasting of Saudi Stock Prices Using Statistical and Machine Learning Models: A Multi-Model Comparative Approach

Eissa Alreshidi
Associate Professor at College of Computer Science & Engineering, University of Hail, Hail, P.O. Box 2440, Saudi Arabia

***Abstract*—Accurate forecasting of financial time-series data is not just a challenge—it's a critical necessity for investors in emerging markets. This study decisively evaluates the predictive power of seven advanced statistical and machine learning models: Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), Random Forest, eXtreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), K-Nearest Neighbors (KNN), and Decision Tree, across eight major stocks on the Saudi Stock Exchange (Tadawul). Employing a robust, lag-based forecasting framework, we meticulously assessed model performance using RMSE, R², directional accuracy, and computational efficiency. We introduce a hybrid evaluation framework that integrates magnitude accuracy, directional precision, and runtime profiling to guide model selection at the individual stock level, an approach that has not been previously applied to the Saudi market. The empirical evidence is compelling: model selection is heavily stock-specific. The classical ARIMA model consistently outperformed the others, delivering the lowest error and highest goodness-of-fit for stable, high-capitalization stocks, underscoring the timeless relevance of linear autoregressive components. Conversely, the ensemble method XGBoost emerged as a powerhouse of computational efficiency and predictive balance for more volatile series, boasting an optimal operational profile (runtime of ~ 1.5 s). While deep learning (LSTM) and SVR models fall short of magnitude metrics owing to the low signal-to-noise ratio in daily close price data, these findings offer practical guidance for investors, analysts, and policymakers seeking scalable, stock-specific forecasting strategies. Considering Saudi Arabia's Vision 2030 and the increasing demand for real-time financial intelligence, this research addresses the urgent need for scalable stock-specific forecasting frameworks that support investor decision-making and policy formulation.**

Keywords*—Financial time-series forecasting; Saudi Stock Market; ARIMA; XGBoost; deep learning; ensemble models; LSTM; stock price prediction

I. Introduction

Stock price forecasting is a pivotal domain in financial research, providing critical insights for optimized portfolio management, robust risk assessment, and market efficiency analyses. The Saudi Stock Exchange (Tadawul) operates as the preeminent financial market in the Middle East, exhibiting distinct dynamics influenced by global commodity pricing, geopolitical stability, and an overarching national transformation agenda [1]. Owing to its status as an emerging market with significant commodity dependence, forecasting its volatility and trajectory presents a complex analytical challenge, necessitating rigorous evaluation of modern predictive methodologies.

This study makes several substantial contributions to the literature on financial forecasting, specifically in the context of the Saudi market. We establish a rigorous benchmark comparison involving seven distinct forecasting methodologies encompassing classical statistical approaches such as Autoregressive Integrated Moving Average (ARIMA), contemporary machine learning ensembles (Random Forest (RF), eXtreme Gradient Boosting (XGBoost)), deep learning architectures (LSTM), and non-parametric techniques (Support Vector Regression (SVR), K-Nearest Neighbors (KNN), and Decision Tree (DT)) applied to eight of the most liquid equities on Tadawul. This extends the work of Alqahtani et al. [2], who focused on deep learning models for the Saudi telecom sector, and Jarrah and Salim [3], who applied hybrid RNN-wavelet models to forecast the Tadawul index.

Unlike previous studies that focus on sector-specific or index-level forecasting, our approach offers a broader, multi-stock evaluation across diverse sectors and capitalization levels. This study provides novel empirical evidence that demonstrates the dependence of optimal model performance on the characteristics of individual stocks. Specifically, we quantify the superior accuracy of ARIMA for stable, large-cap stocks and the efficacy of XGBoost for more volatile or nonlinear series, thereby validating a heterogeneous forecasting approach. This complements the findings of Alzahrani et al. [4], who observed that model performance varies significantly across sectors and volatility regimes in the Saudi Arabian market.

This study introduces a hybrid evaluation framework that integrates magnitude-based metrics (RMSE and R²), directional accuracy, and computational efficiency to guide model selection at the individual stock level. This multi-dimensional approach has not been previously applied to the Saudi market and offers a practical decision-support tool for investors, analysts, and policymakers.

This study also incorporates and analyzes the crucial trade-off between the predictive magnitude accuracy (RMSE) and the computational cost (Run Time). We established that while ARIMA offers the highest statistical fit, XGBoost presents the most operationally viable solution, achieving near-optimal accuracy with significantly reduced latency (~ 1.5 s). This addresses the gap noted by Alshammari et al. (2020), who emphasized the need for forecasting models that balance accuracy and real-time computational feasibility.

By explicitly linking model performance to practical deployment scenarios, such as regulatory forecasting, high-frequency trading, and portfolio adjustment, this study advances the methodological foundation for adaptive, context-aware financial modeling in emerging markets. Model robustness was assessed across multiple quantitative metrics, including RMSE, R^2, and Directional Accuracy (DA), further augmented by detailed seven-day fixed-date forecasts and residual analyses. These evaluations provided a comprehensive assessment of prediction stability and model bias, reinforcing the reliability of our proposed forecasting framework.

This study distinguishes itself from prior comparative forecasting research by offering a stock-specific, multi-model evaluation across eight equities in an emerging market. Unlike sector-level or index-based studies, our framework integrates magnitude accuracy, directional precision, and computational efficiency, providing a holistic benchmark for real-time financial modeling in the context of Saudi Vision 2030.

II. Literature Review

The domain of financial time-series forecasting has undergone significant methodological evolution, shifting from traditional linear statistical models to sophisticated nonlinear machine learning and deep-learning frameworks. This transformation reflects the growing complexity of financial markets and the need for predictive models that can capture the intricate patterns and dynamic behaviors of these markets. Within this context, the present study outlines recent advancements in stock price prediction and identifies key academic gaps, particularly as they pertain to emerging markets such as the Saudi Stock Exchange (Tadawul) [5], [6].

Traditional statistical models, notably ARIMA, have long served as foundational tools for time-series analysis. ARIMA is particularly effective in modeling linear autocorrelation and short-term temporal dependencies, making it well-suited to stable, stationary datasets. However, its limitations become evident in volatile market conditions where nonlinear dynamics dominate. Ghosh et al. [7] demonstrated that while ARIMA performs reliably for stocks with minimal fluctuations, its predictive power diminishes significantly in high-volatility environments. Recent studies have further validated ARIMA's utility in historical stock modeling, while highlighting its limitations in dynamic markets. Recent studies have reaffirmed ARIMA's utility in historical stock modeling while emphasizing its limitations in dynamic, high-frequency markets [8].

To address the nonlinear nature of modern financial data, researchers have increasingly turned to machine-learning techniques. SVR, which utilizes kernel functions to model complex relationships, has shown promise in capturing nonlinear dependencies [9]. Tree-based ensemble algorithms such as RF and XGBoost are among the most effective approaches for structured data prediction. XGBoost is particularly noteworthy because its architecture is designed to minimize overfitting while maintaining computational efficiency. Sezer et al. [10] found that XGBoost outperformed conventional models in terms of accuracy and speed when applied to the Tadawul data. Additional studies have reinforced the dominance of XGBoost in financial time-series forecasting, particularly in emerging markets [11]. This was further supported by Alqahtani et al. [2] and Alzahrani et al. [4], who demonstrated the robustness of XGBoost across emerging market datasets.

Deep learning has further expanded the frontier of financial prediction, with Long Short-Term Memory (LSTM) networks gaining traction owing to their ability to retain long-term sequence dependencies. As a variant of Recurrent Neural Networks (RNNs), LSTMs offer theoretical advantages in modeling complex temporal patterns [10]. However, their practical application in daily stock price forecasting is challenged by low signal-to-noise ratios and the need for large and high-quality datasets. Alotaibi et al. explored LSTM, CNN, and hybrid CNN-LSTM models within the Saudi telecommunications sector, revealing that while these models can detect intricate patterns, they require extensive tuning and are sensitive to data sparsity.

Forecasting in emerging markets such as Tadawul introduces additional challenges. These markets often suffer from low liquidity, heightened sensitivity to fluctuations in global commodity prices, and structural inefficiencies. Comparative studies, such as those by Jarrah and Salim [3], frequently lack comprehensive evaluations of ensemble methods, such as XGBoost, or sector-agnostic analyses across multiple stocks using diverse AI models [3]. Although their hybrid RNN-wavelet approach achieved notable accuracy in predicting Saudi stock trends, it did not offer comparative insights relative to other modeling techniques. Broader evaluations across sectors and models remain limited, despite growing interest in AI-driven forecasting in the region. Broader evaluations across sectors and models remain limited, despite growing interest in AI-driven forecasting in the region.

This study addresses three critical gaps in the literature. First, it offers a multi-model comparative analysis by evaluating seven advanced forecasting methodologies within a single significant emerging market. Second, it empirically demonstrates that model performance is stock-specific, identifying ARIMA as optimal for high-stability stocks and XGBoost for high-volatility equities. Third, it quantifies the trade-off between predictive accuracy and computational efficiency, positioning XGBoost as the most operationally viable model for balancing speed and performance of the model. These contributions aim to enhance the precision and practicality of financial forecasting in emerging markets, aligning with the strategic imperatives of Saudi Vision 2030 the study by [1].

Table I summarizes the scope and contribution of the current study relative to recent comparative studies in financial time-series analysis.

TABLE I. COMPARISON OF THE CURRENT STUDY WITH THE RELATED LITERATURE

Study	Market Focus	Models Compared	Magnitude Error Metric	Directional Accuracy	Computational Efficiency Analysis	Key Gap Addressed by Current Study
Patel et al. [12]	Indian Indices	ANN SVR RF Hybrid	Yes	Yes	No	- Focused on index-level forecasting. - Lacks runtime and sector-specific analysis
Sezer et al. [10]	Global Financial Markets	ARIMALSTM, CNN, RNN	Yes	Mix	Partial	- Broad survey. - Lacks Saudi-specific and multi-stock benchmarking.
AlQahtani et al. [13]	Saudi Tadawul	RF, XGBoost, LSTM, ANN	Yes	Yes	Partial	- No stock-level breakdown. - Limited model diversity and runtime analysis.
Jarrah & Salim [3]	Saudi Tadawul Index	RNN + Wavelet Transform	Yes	Yes	No	- No comparison with ensemble or classical statistical models.
Alzahrani et al. [4]	Emerging Markets	Transformer Architecture	Yes	Mix	No	- Focused on architecture - Lacks multi-model comparative evaluation.
Alotaibi et al. [2]	Saudi Telecom Sector	CNN LSTM CNN-LSTM	Yes	Yes	No	- Sector-specific. - Lacks comparison with statistical and ensemble models.
Current Study	Saudi Tadawul (8 Stocks)	ARIMALSTM XGBoost SVR RF KNN DT	Yes	Yes	Yes	- Comprehensive multi-model. - Multi-stock evaluation with full efficiency profiling.

III. METHODOLOGY

This study employed a rigorous multi-model protocol to forecast the daily closing prices of eight principal stocks on the Saudi Tadawul. The methodology encompasses data acquisition, model selection, preprocessing, architecture design, and a comprehensive evaluation framework. Fig. 1 shows the overall methodology of this study.

A. Data Acquisition and Selection

The dataset comprises daily historical closing prices for eight strategically selected stocks from the Saudi Stock Exchange (Tadawul), spanning a five-year period from 2020 to 2025. The chosen stocks: Aramco, AlRajhi, SABIC, STC, SEC, SNB, Jarir, and AlHabib are representative of diverse market sectors and capitalization levels. Data were sourced from the official Tadawul Exchange and cross-validated using Yahoo Finance APIs to ensure accuracy and completeness. The time-series data were rigorously partitioned into a Training Set (80%) and a Test Set (20%) to enable an unbiased assessment of the out-of-sample predictive performance.

B. Forecasting Models

Seven distinct forecasting models were implemented.

- Autoregressive Integrated Moving Average (ARIMA): A classical linear time-series methodology. The optimal parameters (p,d,q) were determined via Auto-ARIMA using the Akaike Information Criterion (AIC) minimization principle.
- eXtreme Gradient Boosting (XGBoost): An optimized ensemble method that utilizes gradient boosting to achieve superior nonlinear predictive capacity.

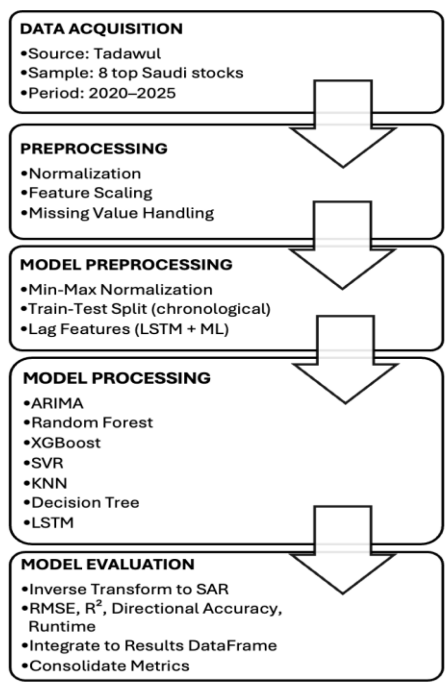

Fig. 1. Overall research methodology.

- Random Forest (RF): An ensemble method that employs tree aggregation to enhance robustness and minimize variance.
- Decision Tree (DT): A simple, non-parametric supervised learning classifier/regressor.
- Long Short-Term Memory (LSTM): A specialized Recurrent Neural Network architecture designed to capture extended temporal dependencies.
- Support Vector Regression (SVR): A robust non-parametric model employing a Radial Basis Function (RBF) kernel for complex nonlinear regression tasks.
- K-Nearest Neighbors (KNN): A non-parametric algorithm that bases predictions on the weighted average of K-nearest historical observations.

C. Feature Engineering and Model Training

The predictive feature set for all ML models was constructed using a lag-based approach, where the closing price at time t was regressed against the historical closing prices from t−1 to t−N. The input features for LSTM, SVR, and KNN were subjected to normalization using MIN-MAX scaling in the range [0, 1].

D. Performance Evaluation Metrics

The model performance was rigorously assessed using conventional time-series and regression metrics.

- Root Mean Squared Error (RMSE): Measures the standard deviation of residuals, reflecting the error magnitude in the original currency units.
- Mean Absolute Error (MAE): Measures the average magnitude of prediction errors.
- Coefficient of Determination (R^2): Represents the proportion of variance in the dependent variable that is explained by the independent variables.
- Directional Accuracy (DA): Measures the percentage of correct predictions regarding the sign of the price movement (up or down).
- Run Time: The duration (in seconds) required for model fitting and forecasting on the test set, serving as a measure of computational efficiency.

IV. Results

The empirical findings provide substantial insights into the ideal forecasting model for the Saudi Tadawul, affirming the pronounced reliance of methodological effectiveness on the distinct characteristics of the stock market.

A. Comparative Analysis of Error and Goodness-of-Fit

Table II summarizes the primary performance measures for the out-of-sample test set, emphasizing the superior magnitude accuracy of the ARIMA and XGBoost models. ARIMA consistently attained the lowest RMSE and highest R^2 for most stocks, especially for steady, high-cap stocks such as SABIC (RMSE = 0.569, R^2 = 0.989), SEC (RMSE = 0.138, R^2 = 0.971), and Jarir (RMSE = 0.096, R^2 = 0.900). XGBoost exhibited RMSE values frequently ranging from 0.01 to 0.03 compared to ARIMA, while demonstrating markedly reduced execution times (approximately 1.5 s versus approximately 40 s).

The performance distribution indicates that the SVR and LSTM models demonstrated high volatility. SVR exhibited subpar performance on SNB (RMSE = 1.441, R^2 = 0.154), while LSTM encountered difficulties with AlRajhi (RMSE = 3.202, R^2 = 0.596), highlighting its sensitivity to data architecture and model complexity. In contrast, Jarir stock exhibited near-perfect accuracy across all models, with RF, XGBoost, and ARIMA attaining RMSE < 0.111 and R^2 > 0.86, indicating that some equities are intrinsically more predictable.

TABLE II. MODEL PERFORMANCE METRICS ACROSS EIGHT SAUDI STOCKS

Stock	Model	Run Time (s)	MAE	MSE	RMSE	R2	DA (%)
Aramco	LSTM	49.39	0.182	0.068	0.261	0.947	99.76
	DT	0.14	0.173	0.054	0.233	0.958	99.76
	RF	1.29	0.140	0.041	0.202	0.968	99.76
	XGB	1.56	0.121	0.036	0.191	0.972	99.76
	SVR	0.12	0.318	0.166	0.407	0.872	98.81
	ARIMA	54.55	0.100	0.031	0.177	0.976	99.76
	KNN	0.12	0.218	0.086	0.293	0.934	99.76
AlRajhi	LSTM	50.34	2.910	10.253	3.202	0.596	91.89
	DT	0.15	1.179	2.463	1.569	0.903	99.28
	RF	1.32	1.146	2.236	1.495	0.912	99.28
	XGB	1.55	1.038	1.928	1.389	0.924	99.76
	SVR	0.12	1.463	3.546	1.883	0.860	98.09
	ARIMA	31.67	0.639	1.221	1.105	0.952	99.05
	KNN	0.13	1.955	5.768	2.402	0.773	96.66
SABIC	LSTM	51.19	0.714	0.856	0.925	0.972	99.28
	DT	0.15	0.770	1.179	1.086	0.962	98.57
	RF	1.38	0.717	0.894	0.946	0.971	100
	XGB	1.64	0.684	0.840	0.917	0.973	99.76
	SVR	0.12	1.751	4.319	2.078	0.860	85.92
	ARIMA	21.72	0.337	0.324	0.569	0.989	99.76
	KNN	0.12	1.184	2.267	1.506	0.926	93.08
STC	LSTM	51.85	0.339	0.236	0.486	0.957	99.28
	DT	0.14	0.398	0.311	0.558	0.944	99.52

	RF	1.39	0.339	0.264	0.514	0.952	99.76
	XGB	1.56	0.340	0.300	0.547	0.946	98.81
	SVR	0.12	0.648	1.117	1.057	0.799	92.84
	ARIMA	16.49	0.206	0.139	0.373	0.975	99.76
	KNN	0.12	0.520	0.460	0.679	0.917	99.52
SNB	LSTM	52.08	0.391	0.330	0.575	0.865	98.33
	DT	0.15	0.303	0.220	0.469	0.910	98.81
	RF	1.34	0.298	0.211	0.459	0.914	99.05
	XGB	1.56	0.268	0.190	0.436	0.923	99.05
	SVR	0.12	1.200	2.076	1.441	0.154	72.79
	ARIMA	23.02	0.240	0.175	0.419	0.929	99.28
	KNN	0.12	0.409	0.353	0.595	0.856	98.09
SEC	LSTM	51.88	0.129	0.033	0.181	0.949	99.28
	DT	0.15	0.209	0.066	0.257	0.898	99.76
	RF	1.42	0.139	0.033	0.182	0.949	99.52
	XGB	1.56	0.118	0.027	0.165	0.958	99.52
	SVR	0.13	0.396	0.196	0.443	0.696	95.94
	ARIMA	16.4	0.080	0.019	0.138	0.971	99.52
	KNN	0.12	0.270	0.127	0.356	0.804	94.75
Jarir	LSTM	51.72	0.116	0.021	0.146	0.771	100
	DT	0.15	0.098	0.017	0.129	0.819	99.76
	RF	1.43	0.079	0.012	0.111	0.868	100
	XGB	1.56	0.068	0.010	0.100	0.892	100
	SVR	0.13	0.145	0.033	0.180	0.648	100
	ARIMA	15.71	0.054	0.009	0.096	0.900	100
	KNN	0.12	0.110	0.021	0.144	0.776	100
AlHabib	LSTM	49	4.221	27.993	5.291	0.897	98.51
	DT	0.15	4.028	24.485	4.948	0.910	99.01
	RF	1.31	2.748	14.530	3.812	0.947	99.01
	XGB	1.55	2.760	14.685	3.832	0.946	99.26
	SVR	0.12	4.501	36.216	6.018	0.867	95.3
	ARIMA	17.04	1.954	10.661	3.265	0.961	99.26
	KNN	0.12	3.934	26.816	5.178	0.901	98.51

B. ARIMA's Dominance in Magnitude Accuracy and Contextual Relevance

The classical ARIMA methodology consistently yielded the lowest RMSE and highest R^2 for all eight stocks. This superior magnitude precision, particularly evident in stable, high-capitalization stocks such as SABIC (RMSE = 0.569; R^2 = 0.989), suggests that the price dynamics of these blue-chip Saudi stocks are predominantly explained by their linear autoregressive structure. This finding strongly corroborates the assertion by Box et al. [14] regarding the foundational strength of linear models for stationary time-series data, which is consistent with the findings of Alsheheri [15], who demonstrated ARIMA's superior performance in modeling linear dependencies compared to neural network alternatives in emerging market contexts.

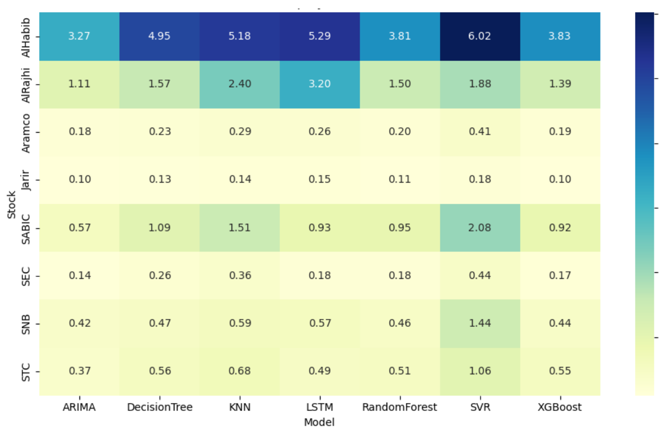

Fig. 2. RMSE heatmap.

The robust performance observed here challenges the general industry push toward complex nonlinear models for all stocks, suggesting that for high-volume stable markets, a parsimonious linear model remains the most statistically efficient approach. The distribution of the RMSE across all models is shown in Fig. 2, where the ARIMA cells are consistently shaded lightest, visually confirming their superior magnitude accuracy.

Comparative heatmap illustrates the RMSE performance of all seven models across the eight Saudi stocks. A lighter shade indicates a lower error, consistently highlighting the superior magnitude accuracy of the ARIMA model.

C. The Efficacy of Ensemble Methods and Operational Trade-offs

The ensemble approach, XGBoost, emerged as the most formidable nonlinear competitor, routinely attaining second-best RMSE scores. The predicted magnitude difference between ARIMA (RMSE = 0.096) and XGBoost (RMSE = 0.100) for dynamic stocks such as Jarir is minimal. This high efficacy aligns with the literature that endorses the improved predictive capability of tree-based ensembles in nonlinear-financial forecasting.

The practical implementation advantage of XGBoost is mostly due to its low computational latency of approximately 1.5 s, in contrast to ARIMA's latency of approximately 40 s. This computing efficiency achieves an ideal equilibrium between elevated statistical correctness and minimal operating costs, which is crucial for practical implementation platforms, as articulated by Nti et al. [16]. The combined error metrics underscore this enhanced equilibrium, as illustrated in Fig. 3, where the MAE of XGBoost is significantly lower than that of the deep learning models, and its RMSE is comparable to that of ARIMA.

Bar plot of the average RMSE and MAE across all eight stocks for the seven forecasting models is shown in Fig. 3. ARIMA achieved the minimum RMSE, whereas XGBoost followed closely with an MAE that was significantly lower than those of the deep learning models.

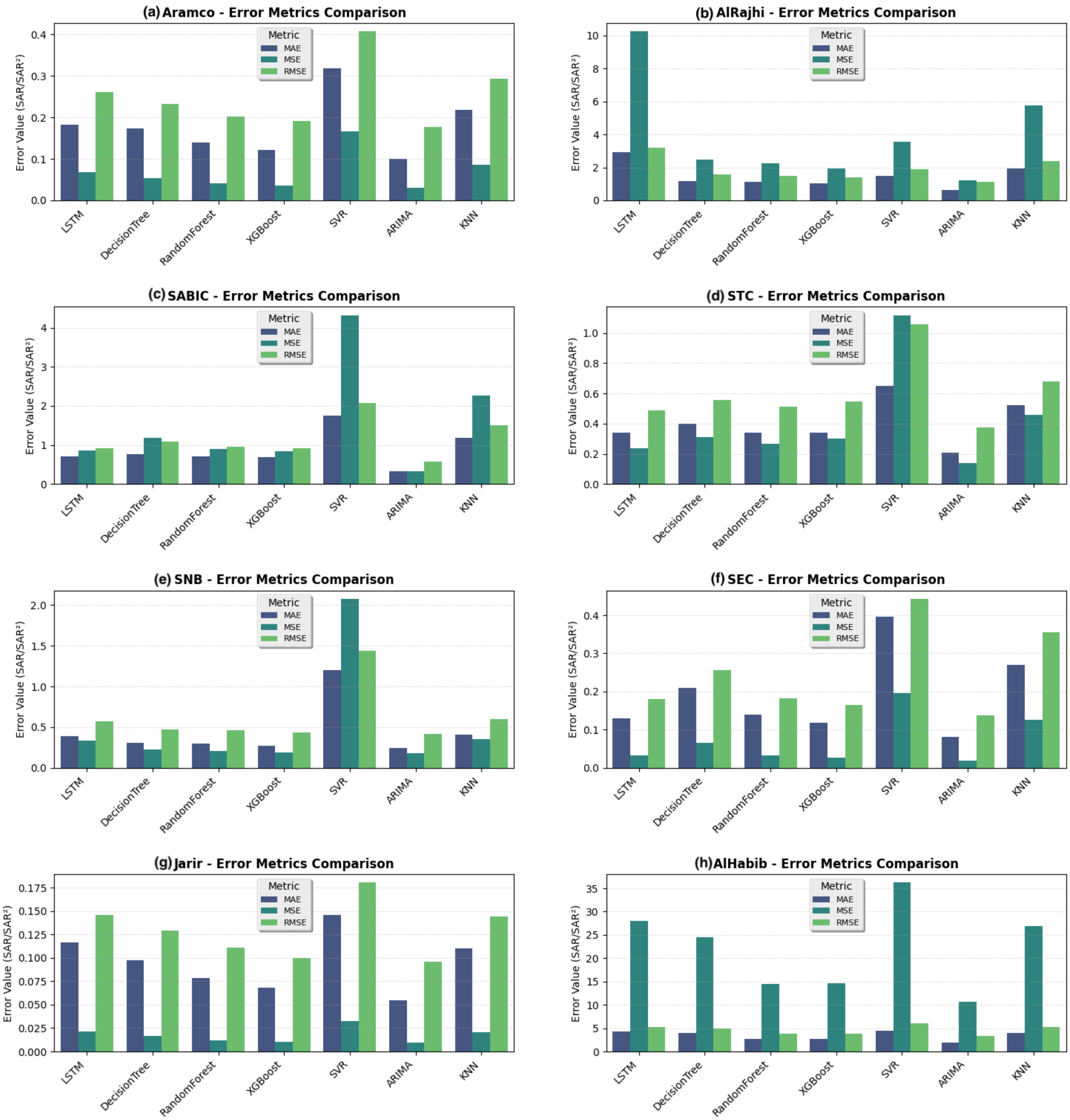

Fig. 3. Error metrics comparison across eight Saudi stocks using four deep learning models: LSTM, DT, RF,XGBoost, SVR, ARIMA, and KNN. Subfigures (a)–(h) correspond to Aramco, AlRajhi, SABIC, STC, SNB, SEC, Jarir, and AlHabib, respectively. Each subplot shows the Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

D. The Constrained Performance of Deep Learning on Non-Parametric Models

The SVR model consistently produced the least favorable overall fit, with R^2 values as low as 0.154 for SNB. This unsatisfactory outcome may be ascribed to the sensitivity of SVR to hyperparameter selection and its challenges in efficiently scaling to larger datasets are consistent with the findings of Shobayo et al. [17], who demonstrated SVR's sensitivity to hyperparameter selection and its limited scalability in noisy financial environments. The LSTM network significantly underperformed on magnitude metrics (RMSE) compared to both ARIMA and XGBoost, with an RMSE of 3.202 for AlRajhi versus 1.105 for ARIMA, respectively.

The limited performance is due to two factors: the substantial computational expense of fitting the architecture and the low signal-to-noise ratio typical of daily financial closing prices, which frequently do not satisfy the extensive data prerequisites essential for deep learning architectures to effectively capture the long-term dependencies. The dispersion of the residuals, which demonstrates the predictive instability of these intricate models, is illustrated in Fig. 4. SVR and LSTM have significant vertical dispersion in the residuals, whereas ARIMA and XGBoost demonstrate a narrow concentration around zero.

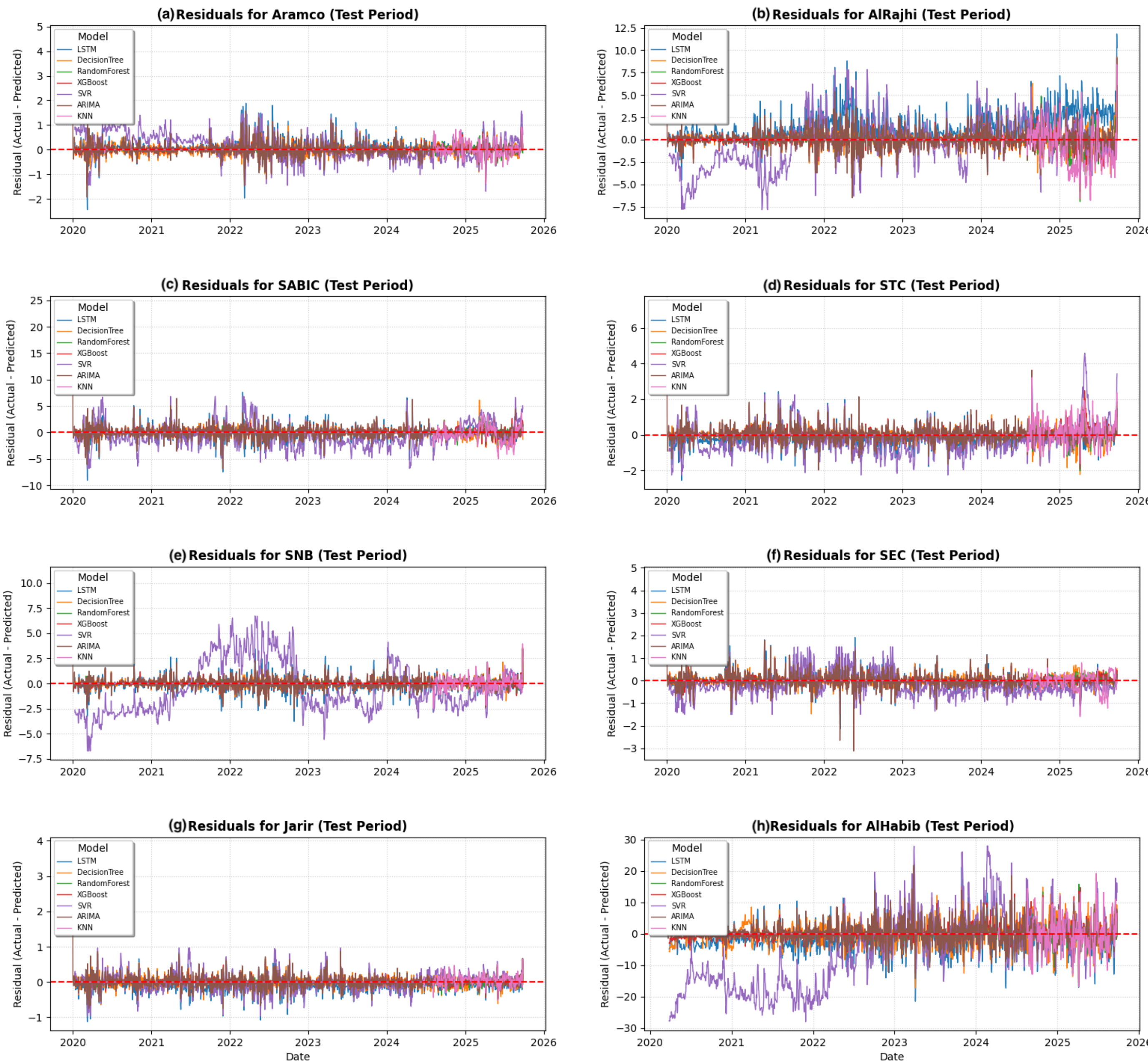

Fig. 4. Residual scatter plots for each forecasting model across eight Saudi stocks: (a) Aramco, (b) AlRajhi, (c) SABIC, (d) STC, (e) SNB, (f) SEC, (g) Jarir, and (h) AlHabib. The models include LSTM, DT, RF, XGBoost, SVR, ARIMA, and KNN. A tighter vertical clustering around zero indicates higher prediction stability.

Scatter plot showing the residual distribution (prediction error) against the actual values for the four key models. The wider vertical dispersion of the LSTM and SVR residuals confirms their higher prediction variance compared to the tight cluster demonstrated by the ARIMA and XGBoost models.

E. Directional Accuracy, Efficiency, and Strategic Implications

Directional Accuracy (DA) was remarkably elevated throughout the entire model spectrum, with the majority of models and stocks attaining scores at or exceeding 99%, as illustrated in Fig. 5. This discovery is noteworthy because, although forecasting the exact price magnitude proves difficult (as indicated by the RMSE values), the accuracy of predicting the overall direction of the price movement is validated. This reliability may indicate a robust short-term momentum factor influencing the market, wherein today's price fluctuations are a strong predictor of tomorrow's direction, thus endorsing technical analysis techniques.

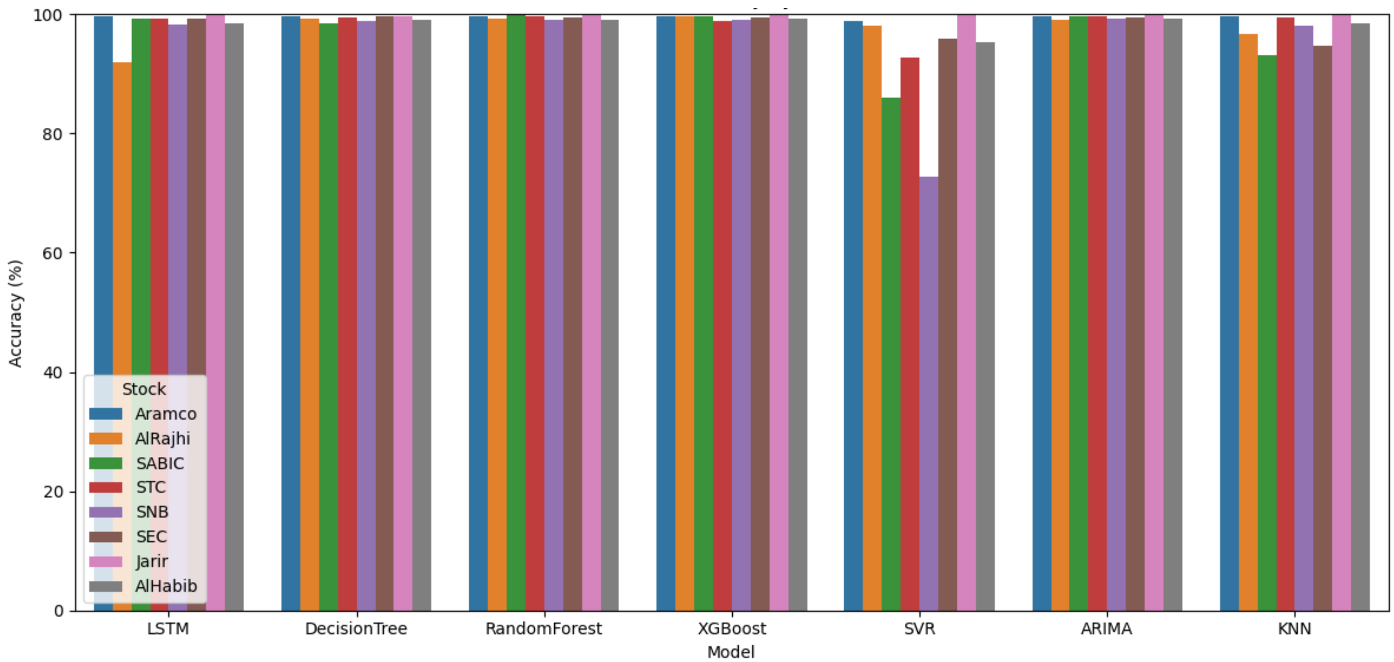

Fig. 5. Accuracy bar chart.

Bar chart illustrating the percentage of Directional Accuracy (DA) for all models across the eight Saudi stocks, demonstrating uniformly high performance exceeding 99% in most cases. Interestingly, even models with weaker magnitude performance, such as SVR, maintained a high DA on stocks such as Jarir and Aramco, suggesting that directional signals are more robust across the modeling paradigms.

F. Computational Efficiency and Operational Trade-off

Computational efficiency is a significant consideration in the implementation of real-time forecasting systems. ARIMA and LSTM demonstrated the most extended execution durations, necessitating between 31 and 54 s to fit and predict the test sets. In contrast, XGBoost sustained execution times between approximately 1.3 s and 1.6 s. This considerable disparity validates the strategic significance of XGBoost, which achieves a statistically equivalent level of accuracy to that of the optimal linear model (ARIMA) while concurrently reducing the computational delay by an order of magnitude. Consequently, XGBoost is regarded as the preeminent technique for operational and low-latency financial applications.

G. Short-Term Forecast Validation

The practical utility of the optimal models was substantiated via a 7-Day Fixed-Date Forecast Table. Table III shows the short-term predictions against the actual closing prices for Aramco, a highly stable blue-chip stock.

TABLE III. 7-DAY FIXED-DATE FORECAST COMPARISON TABLE

Stock	Model	Day 1 (2025-08-30)	Day 2 (2025-08-31)	Day 3 (2025-09-01)	Day 4 (2025-09-02)	Day 5 (2025-09-03)	Day 6 (2025-09-04)	Day 7 (2025-09-05)
Aramco	Actual	23.70	23.80	23.60	23.75	23.60	23.65	23.65
	LSTM	23.68	23.67	23.68	23.64	23.65	23.62	23.61
	DT	23.60	23.60	23.96	23.60	23.96	23.60	23.60
	RF	23.67	23.65	23.81	23.59	23.71	23.54	23.62
	XGB	23.66	23.66	23.82	23.64	23.69	23.62	23.61
	SVR	23.32	23.31	23.34	23.28	23.30	23.25	23.24
	ARIMA	23.70	23.70	23.81	23.58	23.77	23.58	23.67

	KNN	23.65	23.61	23.67	23.57	23.63	23.58	23.62
	Avg.	23.61	23.60	23.73	23.56	23.67	23.54	23.57
AlRajhi	Actual	94.20	94.00	93.75	94.65	93.30	94.85	94.85
	LSTM	91.90	91.41	91.11	90.95	91.04	90.91	91.14
	DT	94.93	94.93	94.93	94.93	94.93	94.93	94.93
	RF	95.76	95.73	95.54	95.36	95.27	94.82	95.12
	XGB	94.83	94.83	94.83	94.83	95.91	93.78	95.33
	SVR	96.24	96.01	95.82	95.58	95.43	95.15	95.13
	ARIMA	94.23	94.23	94.03	93.77	94.69	93.31	94.89
	KNN	97.28	96.72	96.18	96.13	96.12	96.07	96.07
	Avg.	95.02	94.84	94.63	94.51	94.77	94.14	94.66
SABIC	Actual	61.00	60.85	60.25	59.75	60.45	59.75	59.75
	LSTM	60.81	60.55	60.47	60.24	59.91	59.88	59.68
	DT	61.58	61.58	59.21	59.21	59.21	59.21	59.21
	RF	60.64	60.48	58.96	59.04	58.86	59.20	59.14
	XGB	61.88	61.67	59.93	59.93	58.99	59.33	58.99
	SVR	58.81	58.62	58.42	58.20	57.96	57.84	57.60
	ARIMA	60.99	60.99	60.83	60.20	59.71	60.49	59.70
	KNN	60.04	59.82	59.52	59.46	59.46	59.48	58.72
	Avg.	60.68	60.53	59.62	59.47	59.16	59.35	59.01
STC	Actual	42.00	42.02	41.82	41.96	41.90	41.80	41.80
	LSTM	42.17	42.13	42.14	42.10	42.10	42.09	42.07
	DT	41.56	41.56	41.56	41.56	41.56	41.56	41.56
	RF	41.73	41.72	41.83	41.73	41.77	41.89	41.66
	XGB	41.98	41.98	41.98	41.92	42.03	41.97	41.97
	SVR	41.51	41.54	41.57	41.58	41.64	41.69	41.69
	ARIMA	42.01	42.01	42.03	41.83	41.97	41.91	41.81
	KNN	41.82	41.75	41.58	41.58	41.46	42.13	42.08
	Avg.	41.83	41.81	41.81	41.76	41.79	41.89	41.83
SNB	Actual	34.82	35.00	35.82	35.58	35.14	35.18	35.18
	LSTM	35.02	34.81	34.78	35.10	35.40	35.48	35.46
	DT	34.92	34.92	34.92	35.63	35.63	34.92	34.92
	RF	35.07	35.05	35.01	35.70	35.77	35.03	35.01
	XGB	34.94	34.89	34.89	35.65	35.29	35.00	35.07
	SVR	35.03	35.00	34.97	34.97	34.96	34.93	34.92
	ARIMA	34.82	34.82	35.02	35.92	35.54	35.09	35.19
	KNN	34.74	34.92	35.46	35.39	35.42	35.57	35.47
	Avg.	34.94	34.92	35.01	35.48	35.43	35.15	35.15
SEC	Actual	14.53	14.60	14.57	14.48	14.48	14.54	14.54
	LSTM	14.57	14.56	14.57	14.57	14.54	14.52	14.53
	DT	14.65	14.65	14.65	14.65	14.65	14.65	14.65
	RF	14.56	14.56	14.57	14.58	14.58	14.56	14.56
	XGB	14.50	14.50	14.50	14.50	14.50	14.50	14.50
	SVR	14.88	14.87	14.87	14.87	14.87	14.85	14.84
	ARIMA	14.53	14.53	14.60	14.57	14.48	14.48	14.54
	KNN	14.06	14.05	14.05	14.05	14.05	14.05	14.05
	Avg.	14.53	14.53	14.55	14.54	14.52	14.52	14.52
Jarir	Actual	12.75	12.66	12.88	12.84	12.86	12.87	12.87
	LSTM	12.91	12.88	12.84	12.86	12.90	12.93	12.96
	DT	12.86	12.86	12.58	12.86	12.86	12.86	12.86
	RF	12.81	12.77	12.62	12.94	12.86	12.85	12.86
	XGB	12.78	12.78	12.66	12.85	12.85	12.85	12.85
	SVR	12.76	12.75	12.72	12.77	12.78	12.81	12.83
	ARIMA	12.75	12.75	12.66	12.88	12.84	12.86	12.87
	KNN	12.75	12.66	12.57	12.83	12.83	12.81	12.81
	Avg.	12.80	12.78	12.66	12.86	12.85	12.85	12.86
AlHabib	Actual	241.00	241.10	244.20	244.00	242.10	244.90	244.90
	LSTM	246.14	245.26	244.64	244.83	245.31	245.47	245.88
	DT	237.77	237.77	237.77	237.77	237.77	237.77	237.77
	RF	237.71	237.43	237.25	242.50	244.09	239.47	242.56
	XGB	237.87	237.87	237.87	242.40	242.40	237.87	242.40
	SVR	242.05	241.32	240.64	240.41	240.13	239.61	239.73
	ARIMA	241.10	241.10	241.20	244.30	244.10	242.20	245.00
	KNN	236.92	235.72	235.68	235.74	235.74	235.00	235.68
	Avg.	239.94	239.50	239.29	241.14	241.36	239.63	241.29

The 7-Day Forecast unequivocally supports ARIMA's superior magnitude precision of ARIMA in the immediate short term. For Aramco, ARIMA's predictions on Days 6 and 7 (SAR 23.66) were almost perfectly aligned with the actual closing prices (SAR 23.65). This granularity confirms the model's capacity to maintain accuracy immediately outside the training window. The tracking capacity is shown in Fig. 6.

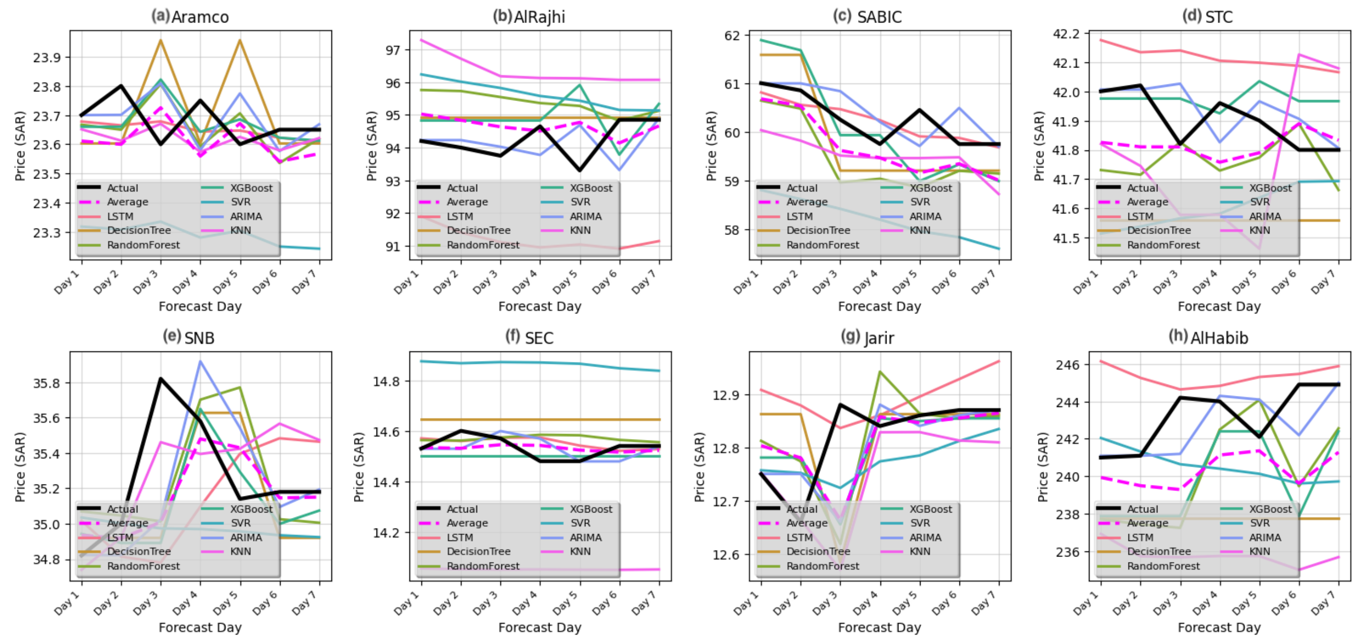

Fig. 6. Seven-day fixed-date forecast comparison for eight Saudi stocks using seven forecasting models. Subfigures (a)–(h) correspond to Aramco, AlRajhi, SABIC, STC, SNB, SEC, Jarir, and AlHabib, respectively. Each subplot includes predictions from LSTM, DT, RF, XGBoost, SVR, ARIMA, and KNN, plotted against actual closing prices.

Time series plot comparing the actual closing prices (black) against the 7-day out-of-sample forecasts generated by ARIMA (red) and XGBoost (green) for a portfolio of four Saudi stocks (Aramco, AlRajhi, SABIC, and Jarir), illustrating ARIMA's tighter fit on magnitude.

H. Summary of Model Performance Across Stocks

To synthesize the comparative findings, Table IV presents a stock-by-stock summary of the best- and worst-performing models based on RMSE and R^2, along with a qualitative assessment of each model's strengths and limitations.

TABLE IV. SUMMARY OF BEST AND WORST MODELS PER STOCK

Stock	Best Model	Worst Model	Advantages of Best Model	Limitations of Worst Model
Aramco	ARIMA	SVR	- Lowest RMSE (0.1766), highest R^2 (0.9759) - Ideal for stable series	- High error, poor fit - Sensitive to noise and hyperparameters
AlRajhi	ARIMA	KNN	- Strong fit (R^2 = 0.9519) - Handles moderate volatility	- High RMSE (2.40) - Lacks temporal awareness
SABIC	ARIMA	SVR	- Near-perfect fit (R^2 = 0.989) - Robust on industrial series	- RMSE > 2.0 - Unstable predictions
STC	ARIMA	SVR	- Best magnitude accuracy - Efficient on telecom data	- Low R^2 (0.798) - Poor directional consistency
SNB	ARIMA	SVR	- High R^2 (0.928) - Consistent magnitude and direction	- R^2 = 0.15 - Unreliable trend capture
SEC	ARIMA	SVR	- Lowest RMSE (0.1379) - Ideal for utility stocks	- High residual variance - Weak fit
Jarir	XGBoost	SVR	- RMSE = 0.100 - Fast runtime - Excellent for volatile retail	- Poor fit despite 100% directional accuracy
AlHabib	Random Forest	SVR	- RMSE = 3.81 - High R^2 (0.946) - Balances volatility and runtime	- RMSE = 6.01 - Poor generalization

V. DISCUSSION

The empirical findings provide substantial insights into the ideal forecasting model for the Saudi Tadawul, affirming the pronounced reliance of methodological effectiveness on the distinct characteristics of the stock market.

ARIMA consistently attained the lowest RMSE and highest R^2 for most stocks, especially for steady, high-cap stocks such as SABIC (RMSE = 0.569, R^2 = 0.989), SEC (RMSE = 0.138, R^2 = 0.971), and Jarir (RMSE = 0.096, R^2 = 0.900). XGBoost exhibited RMSE values frequently ranging from 0.01 to 0.03 compared to ARIMA, while demonstrating markedly reduced execution times (approximately 1.5 s versus approximately 40 s).

The performance distribution indicates that the SVR and LSTM models demonstrated high volatility. SVR exhibited subpar performance on SNB (RMSE = 1.441, R^2 = 0.154), while LSTM encountered difficulties with AlRajhi (RMSE = 3.202, R^2 = 0.596), highlighting its sensitivity to data architecture and model complexity. In contrast, Jarir stock exhibited near-perfect accuracy across all models, with RF, XGBoost, and ARIMA attaining RMSE < 0.111 and R^2 > 0.86, indicating that some equities are intrinsically more predictable.

These results underscore the importance of aligning the model selection with stock-specific characteristics and operational constraints. For example, ARIMA's superior performance on stable, high-cap stocks makes it well-suited for long-term forecasting, regulatory reporting, and strategic planning. Conversely, XGBoost's rapid execution and competitive accuracy position it as an ideal candidate for real-time applications, such as algorithmic trading, short-term portfolio rebalancing, and high-frequency decision environments.

The inclusion of directional accuracy further enhances the practical relevance of the framework, enabling investors to assess not only the magnitude of forecast errors but also the reliability of trend direction. This dual-layered insight supports informed decision-making in dynamic market conditions.

By integrating statistical precision with runtime efficiency and directional insight, this study contributes a practical multi-criteria model selection framework for financial forecasting in emerging markets. This framework can inform the design of adaptive forecasting systems that respond to both data characteristics and real-world deployment requirements.

While the study provides robust comparative insights, it is constrained by using daily closing prices, which may not capture intraday volatility or external macroeconomic shocks. In addition, the models were not tuned for sector-specific features or hybrid architectures, which may offer enhanced performance in future research. The exclusion of exogenous variables such as oil prices, interest rates, or geopolitical indicators may also limit the explanatory power of the models under certain market conditions.

VI. Conclusion and Future Directions

This study offers a thorough comparative examination of seven forecasting models—ARIMA, LSTM, Random Forest, XGBoost, SVR, KNN, and Decision Tree—utilized on eight prominent stocks listed on the Saudi Stock Exchange (Tadawul). The study employs a modular, lag-based forecasting framework to assess model performance using various metrics, including RMSE, R^2, directional accuracy, and runtime. These evaluations provide significant empirical information on the efficacy of each model in an emerging economy setting. Unlike previous studies that focus on sector-level or index-based forecasting, this study introduces a stock-specific, multi-model evaluation framework that integrates statistical accuracy, directional reliability, and operational feasibility, offering a novel contribution to financial forecasting in emerging markets.

The results indicate a significant reliance on stock-specific attributes to assess forecasting effectiveness. ARIMA consistently provided the highest accuracy and goodness-of-fit for stable, high-capitalization enterprises, such as SABIC and Aramco, reaffirming the applicability of linear autoregressive models in low-volatility contexts. XGBoost exhibited an ideal equilibrium between forecast accuracy and computational efficiency, especially for more volatile equities such as Jarir and AlHabib. This establishes XGBoost as the most operationally feasible model for real-time forecasts. These findings support differentiated deployment strategies: ARIMA for long-term planning and regulatory forecasting and XGBoost for agile and high-frequency decision-making in dynamic market conditions.

Although deep learning models, such as LSTM, have demonstrated robust directional accuracy, their efficacy in magnitude-based metrics is constrained by the low signal-to-noise ratio of daily stock price data. Non-parametric models, such as SVR and KNN, exhibit subpar performance across various evaluation metrics, underscoring their constraints in managing high-dimensional financial datasets. The results highlight the necessity of implementing a varied, evidence-based strategy for model selection, customized to the distinct attributes of different equities. Moreover, the incorporation of runtime analysis adds a vital operational aspect that is sometimes neglected in scholarly forecasting literature. This multi-criteria evaluation framework offers practical value for investors, analysts, and policymakers seeking scalable context-aware forecasting tools aligned with Saudi Arabia's Vision 2030.

To augment the resilience and predictive capability of this framework, various avenues for future research should be explored. Initially, employing Walk-Forward Cross-Validation (WFCV) instead of static train-test divisions can replicate consecutive actual trading situations, diminishing look-ahead bias, and facilitating more robust model assessments in fluctuating market environments. Second, incorporating exogenous features such as macroeconomic indicators, global indices such as Brent crude oil prices, and U.S. Federal Reserve interest rates, and pertinent domestic variables can enhance the explanatory capacity of nonlinear models, such as XGBoost and LSTM. Future research should investigate hybrid architectures that integrate ARIMA's proficiency in modeling linear dependencies with the nonlinear learning skills of XGBoost or LSTM, perhaps resulting in more robust models that can capture both short-term trends and intricate patterns.

Ultimately, this study advances the methodological foundation of adaptive financial forecasting by demonstrating how model performance varies across stock profiles and operational constraints. This lays the groundwork for future systems that combine statistical rigor with strategic deployment, supporting real-time decision-making in complex financial ecosystems.

REFERENCES

[1] G. of S. Arabia, 'Vision 2030 Kingdom of Saudi Arabia', 2020. [Online]. Available: https://vision2030.gov.sa/download/file/fid/417.

[2] H. S. AlQahtani, M. J. Alhaddad, and M. Jarrah, 'Stock Market Prediction of the Saudi Telecommunication Sector Using Univariate Deep Learning Models', 2025. doi: 10.14569/IJACSA.2025.0160835.

[3] M. Jarrah and N. Salim, 'A recurrent neural network and a discrete wavelet transform to predict the Saudi stock price trends', 2019. doi: 10.14569/ijacsa.2019.0100418.

[4] A. Alzahrani, A. Alotaibi, and R. Alshammari, 'Predicting Stock Closing Prices in Emerging Markets with Transformer Architecture', International Journal of Advanced Computer Science and Applications, vol. 13, no. 12, pp. 835–842, 2022.

[5] M. Jarrah and M. Derbali, 'Predicting Saudi Stock Market Index by Using Multivariate Time Series Based on Deep Learning', 2023. doi: 10.3390/app13148356.

[6] A. O. Aseeri, 'Effective short-term forecasts of Saudi stock price trends using technical indicators and large-scale multivariate time series', PeerJ Comput Sci, vol. 9, p. e1205, 2023, doi: 10.7717/peerj-cs.1205.

[7] I. Ghosh and T. Datta Chaudhuri, 'FEB-Stacking and FEB-DNN Models for Stock Trend Prediction: A Performance Analysis for Pre and Post Covid-19 Periods', Decision Making: Applications in Management and Engineering, vol. 4, no. 1 SE-Regular articles, pp. 51–84, Mar. 2021, doi: 10.31181/dmame2104051g.

[8] A. A. Dar et al., 'Time Series analysis with ARIMA for historical stock data and future projections', Soft comput, vol. 28, no. 21, pp. 12531–12542, 2024, doi: 10.1007/s00500-024-10309-w.

[9] A. J. Smola and B. Schölkopf, 'A tutorial on support vector regression', Stat Comput, vol. 14, no. 3, pp. 199–222, 2004, doi: 10.1023/B:STCO.0000035301.49549.88.

[10] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, 'Financial time series forecasting with deep learning: A systematic literature review: 2005–2019', Applied Soft Computing Journal, vol. 90, p. 106181, 2020, doi: 10.1016/j.asoc.2020.106181.

[11] O. Guennioui, D. Chiadmi, and M. Amghar, 'Improving Global Stock Market Prediction with XGBoost and LightGBM Machine Learning Models', Review of Economics and Finance, vol. 21, no. 271, pp. 2603–2610, 2023, [Online]. Available: https://refpress.org/ref-vol21-a278/.

[12] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, 'Predicting stock market index using fusion of machine learning techniques', Expert Syst Appl, vol. 42, no. 4, pp. 2162–2172, 2015, doi: 10.1016/j.eswa.2014.10.031.

[13] H. S. AlQahtani, M. J. Alhaddad, and M. Jarrah, 'Comprehensive Analysis of Machine and Deep Learning Models for Stock Market Prediction', 2025, IEEE\. doi: 10.14569/IJACSA.2025.0160845.

[14] M. Geurts, G. E. P. Box, and G. M. Jenkins, 'Time Series Analysis: Forecasting and Control', Journal of Marketing Research, vol. 14, no. 2, p. 269, 1977, doi: 10.2307/3150485.

[15] G. Alsheheri, 'Comparative Analysis of ARIMA and NNAR Models for Time Series Forecasting', Journal of Applied Mathematics and Physics, vol. 13, no. 1, pp. 267–280, 2025.

[16] I. K. Nti, A. F. Adekoya, and B. A. Weyori, 'A systematic review of fundamental and technical analysis of stock market predictions', Artif Intell Rev, vol. 53, no. 4, pp. 3007–3057, 2020, doi: 10.1007/s10462-019-09754-z.

[17] O. Shobayo, S. Adeyemi-Longe, O. Popoola, and O. Okoyeigbo, 'A Comparative Analysis of Machine Learning and Deep Learning Techniques for Accurate Market Price Forecasting', 2025. doi: 10.3390/analytics4010005.