



An Integrated Technical and
Fundamental Stock Analysis Model for
the Saudi Stock Market (Tadawul)

Prepared By

Name: Malek M. Atarji

ID No.: 252000474

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Ashraf El-Sayed

College of Informatics, Midocean University, Comoros

Email: 252000474@st.midocean.edu.km

Acknowledgements

I would like to express my sincere gratitude to Midocean University for providing the academic environment and institutional support necessary for the successful completion of this research.

I am especially grateful to my supervisor, Dr. Ashraf El-Sayed, for his continuous guidance, professionalism, and invaluable academic insight throughout all stages of this study. His constructive feedback and critical evaluations significantly enhanced the academic rigor and intellectual depth of this work.

I also acknowledge the contributions of previous scholars and researchers whose theoretical and empirical studies provided the foundation upon which this research is built.

Finally, I extend my appreciation to my colleagues and fellow researchers for their valuable discussions, encouragement, and support throughout the research process.

Abstract

The available literature on stock market trading signal prediction largely relies upon either technical analysis of market data or fundamental analysis of company financial data. Technical analysis methods are more suitable for identifying short-term movements in the market, while fundamental analysis can provide information regarding the financial strength of the listed companies. Thus, an AI model can be created that incorporates both of these analyses in order to help generate short-term trading signals for the Saudi Stock Market (Tadawul).

The dataset used in this model consisted of five years of trading and financial data for 20 companies listed on the Tadawul stock exchange. The raw data contained 24,966 observations, but was cleaned and preprocessed into a dataset containing 19,742 observations. Furthermore, the dataset exhibits a degree of class imbalance in its outcomes, with 32.00% of trading outcomes being successful in the initial dataset, and 25.78% in the out-of-sample test set of the data. A successful trading outcome was identified as outcomes in which the stock price reached the take-profit level of +5% before reaching the stop-loss level of -2%.

Each of the datasets was utilized to train machine learning models in an attempt to recognize the characteristics of successful trading outcomes for the Tadawul Stock Exchange. Two models were created: one of the Bidirectional Long Short-Term Memory (BiLSTM) neural network models using 60 days of market data for each company, and another of the XGBoost machine learning model that used technical indicators and financial ratios of each company. The trading signals from each of these models were combined into a weighted ensemble model, with the XGBoost model contributing 60% to the formation of trading signals and the BiLSTM model contributing the remaining 40%.

The models were evaluated using a range of metrics, both classification-based and those related to trading performance. Furthermore, a backtesting strategy was utilized for thresholds in the predicted trading signals between 0.50 and 0.75, as well as costs of 0.15% per trade. The results of the backtesting indicate that the hybrid model outperformed the individual models in terms of risk-adjusted trading performance. Specifically, the hybrid model showed that fewer trades had to be executed in order to maximize the returns of the trading strategy, as indicated by the increase in the Sharpe ratio and the decrease in the maximum drawdown for the strategy at various thresholds.

The integration of technical and fundamental analysis methods within an AI framework indicates that each approach has the potential to enhance trading signal prediction for the Saudi Stock Market. Furthermore, the use of publicly available information for these models allows for transparency in the trading signals of the strategy, as well as its suitability within the Saudi Stock Market regulatory regime.

Executive Summary

Presented is a Hybrid framework for short-term trading signals within the Saudi Stock Market. The framework utilizes AI models based on technical and fundamental analysis of the Tadawul markets. Results from five years of market data indicate that the hybrid AI model outperformed standalone models in terms of both predicted accuracy and financial performance metrics. Furthermore, the framework is reproducible, uses publicly available market data, and complies with Saudi Capital Market Authority regulations.

List of Abbreviations

AI – Artificial Intelligence

AUC – Area Under the Curve

CMA – Capital Market Authority (Saudi Arabia)

CVaR – Conditional Value at Risk

D/E – Debt-to-Equity Ratio

DL – Deep Learning

EMA – Exponential Moving Average

EMH – Efficient Market Hypothesis

F1 – F1 Score

GARCH – Generalized Autoregressive Conditional Heteroskedasticity

LSTM – Long Short-Term Memory

MA – Moving Average

MACD – Moving Average Convergence Divergence

MDD – Maximum Drawdown

ML – Machine Learning

NI – Net Income

OHLCV – Open, High, Low, Close, Volume

OLS – Ordinary Least Squares

P/E – Price-to-Earnings Ratio

PM – Profit Margin

PR – Precision-Recall

RF – Random Forest Model

ROE – Return on Equity

ROC – Receiver Operating Characteristic

RSI – Relative Strength Index

SE – Stockholders' Equity

SL – Stop-Loss

SVM – Support Vector Machine

TASI – Tadawul All Share Index

TP – Take-Profit

TR – Total Revenue

VaR – Value at Risk

XGBoost – Extreme Gradient Boosting

Table of Contents

Contents

Acknowledgements.....	2
Abstract.....	3
Executive Summary	4
List of Abbreviations	5
Table of Contents	7
Chapter 1: Introduction	9
1.1 Background and Problem Definition	10
1.2 Research Methodological Positioning	10
1.3 Research Objectives.....	11
1.4 Research Hypotheses	11
1.5 Scope of the Study	12
1.6 Research Contributions.....	13
1.6.1 Model Development Contribution.....	13
1.6.2 Trading Framework Contribution	13
1.6.3 Risk-Oriented Evaluation Contribution	13
1.6.4 Saudi Market Contribution.....	13
1.6.5 Literature Contribution	14
1.7 Thesis Organization	14
1.8 Research Clarification	14
Chapter 2: Background and Related Work	16
2.1 Introduction.....	16
2.2 Stock Market Forecasting	16
2.3 Fundamental Analysis	16
2.4 Technical Analysis	17
2.5 Machine Learning in Finance	17
2.6 Hybrid Integration of Technical and Fundamental Analysis	17
2.7 Trading-Oriented Evaluation Metrics	18
2.8 Evolution of Financial Prediction Models.....	19
2.9 Machine Learning and Deep Learning Studies.....	19

2.10 Critical Evaluation of Prior Studies	20
2.11 Hybrid Models in Literature.....	21
2.12 Studies in Emerging and Saudi Markets	22
2.13 Research Gap	23
2.14 Comparative Analysis with Existing Studies	27
2.15 Positioning of the Present Study	27
2.16 Chapter Summary	28
Chapter 3: Proposed Approach (Methodology)	29
3.1 Overview of the Hybrid Framework	29
3.2 Dataset Description	29
3.3 Data Preparation	30
3.4 Feature Engineering	33
3.4.1 Technical Features	34
3.4.2 Fundamental Features	34
3.5 Trading Constraints Justification.....	35
3.6 Model Architecture	36
3.6.1 XGBoost Model	36
3.6.2 Bidirectional LSTM Model	36
3.7 Hybrid Integration	37
3.8 Mathematical Formulation	38
3.9 Backtesting Framework	39
3.9.1 Backtesting Assumptions	40
3.10 Sharpe Ratio	40
3.11 Performance Metrics	41
3.12 Reproducibility	41
3.13 Ethical and Regulatory Compliance	42
3.14 Chapter Summary	43
Chapter 4: Experimental Results and Discussion	44
4.1 Experimental Setup	44
4.2 Baseline Model Performance	45
4.2.1 Interpretation of Classification Performance.....	46
4.3 Hybrid Model Performance	47

4.3.1 Risk-Adjusted Trading Performance.....	48
4.4 Key Insight.....	49
4.5 Threshold Sensitivity Analysis.....	49
4.6 Risk-Oriented Perspective.....	52
4.7 Impact of Transaction Costs.....	53
4.8 Equity Curve and Trading Dynamics.....	54
4.10 Mechanism Behind Hybrid Model Superiority.....	56
4.11 Economic Interpretation.....	57
4.12 Limitations.....	58
4.13 Practical Deployment.....	59
4.14 Chapter Summary.....	60
Chapter 5: Conclusion and Future Work.....	61
5.1 Conclusions.....	61
5.2 Theoretical Contribution.....	61
5.3 Practical Implications.....	62
5.4 Implementation Constraints.....	63
5.5 Generalizability of the Framework.....	63
5.6 Future Research Directions.....	64
5.7 Final Remarks.....	65
References.....	66

Chapter 1: Introduction

1.1 Background and Problem Definition

The Saudi Stock Market, also known as Tadawul, is one of the largest stock markets in the Middle East and North Africa (MENA) region.

The Saudi Stock Market has become increasingly liberalized in recent years with various reforms introduced as part of the Vision 2030 initiative. These reforms have increased the number and size of the companies that trade on the market and increased the complexity of trading activity within the market.

Despite the benefits of such reforms, the increased complexity of the market has increased the volatility of the market. Furthermore, the volatility has introduced nonlinearity in the price movements of the stocks that trade on the market.

Investors in the stock market typically employ two main methods to make trading decisions within the market: technical analysis or fundamental analysis of the stocks that they trade. Technical analysis is beneficial for generating short-term trading signals. However, it does not provide information about the financial strength of the companies that trade on the market. Fundamental analysis of stocks provides an understanding of the financial strength of those companies. However, it fails to provide trading signals for short-term trading activities.

Furthermore, the bulk of machine learning literature is based on the use of statistical models to maximize the accuracy of trading signals. However, the stock markets that are regulated in the Kingdom of Saudi Arabia require that trading systems are easily interpretable and comply with specific regulations. Hence, there is a need for a prediction model that is based on regulation compliance and can generate short-term trading signals for the Saudi Stock Market.

1.2 Research Methodological Positioning

This study will use experimental research methods to test the performance of various models. Furthermore, the study will employ machine learning methods to develop these models.

Rather than using the stock market's price movements to determine trading signals, this research will use the information in the stocks' financial reports. Furthermore, the problem of trading signal generation will be represented as a classification problem, where the classification of trades will be based on whether a predefined

target return is achieved before a stop-loss is triggered for a stock during a trading period.

Performance of the models will be evaluated using a set of financial metrics that represent the economic performance of the trading strategy rather than statistical accuracy measures. Furthermore, a method will be used that ensures that the model is tested on data that was available to traders at the time of trading (chronological out-of-sample testing) to ensure that it can be used for making trading decisions.

1.3 Research Objectives

The objective of this research is to develop a trading signal model for the Saudi Stock Market that is based on artificial intelligence.

Furthermore, this model will have the following sub-objectives:

To develop a model that can combine technical and fundamental indicators of company strength.

To formulate a target that can be used to train the model based on economic trading logic.

To test and evaluate the performance of the model using financial metrics.

To ensure that it complies with regulations of the Saudi Stock Market by using publicly available data.

To develop a framework for the model that can be replicated for other trading applications.

1.4 Research Hypotheses

Hypothesis 1: The artificial intelligence model that is developed for this research will exhibit risk-adjusted returns that are superior to those of standalone technical or fundamental models.

Hypothesis 2: The artificial intelligence model will experience less drawdown in its returns than the standalone models.

Hypothesis 3: The signals that are produced by the model will be based on public and structured data, and these signals will be sufficient to develop AI trading systems that can effectively trade in the Saudi Stock Market.

Hypothesis 4: The developed model will exhibit stability in its performances under validation on different market conditions.

1.5 Scope of the Study

The scope of this study is limited to the analysis of 20 listed liquid companies on the Saudi Stock Market (Tadawul) between 2021 and 2025. These companies span a variety of sectors within the Saudi economy, including banking, petrochemicals, telecommunications, and energy.

This study examines whether the strategy could be used to generate trading signals over short periods rather than attempting to determine the value of the stocks over longer periods or attempting to directly forecast the price of the stocks.

Furthermore, trades will be made when the price of the stock reaches a take-profit level of +5% of the initial price of the stock, or, alternatively, should reach a stop-loss level of -2% of the initial price of the stock. Thus, trades are only considered successful if the price of the stock reaches the take-profit level prior to reaching the stop-loss level.

Only market and financial data will be utilized to inform the trading strategy. Market data will be based upon daily prices of the stocks, while fundamental data will be based upon the annual financial statements of each of the companies. Data on sentiment, news, social media, or other non-structured data will not be incorporated into this framework.

The costs of 0.15% per trade will be incorporated into the backtesting of the strategy as a means of providing an estimate of the costs of executing the strategy. However, costs such as bid-ask spreads, liquidity, and slippage are not accounted for within this strategy, and are therefore a limitation of the study that can be addressed in future research.

Finally, this framework is for academic purposes only. The strategy is not financial advice, and should not be implemented as a trading system without first carefully considering its risks.

1.6 Research Contributions

This study makes a few contributions to the existing literature on artificial intelligence-based trading signal generation.

1.6.1 Model Development Contribution

The main contribution of this study is the development of a hybrid artificial intelligence model that incorporates two different machine learning models: Bidirectional Long Short-Term Memory (BiLSTM) and XGBoost. The trading signal is determined as a result of ensembling the predictions from each of these models.

1.6.2 Trading Framework Contribution

Most existing research on artificial intelligence and trading signals focuses upon the task of directly predicting the price of financial instruments. However, this study formulates the problem as one of predicting the outcomes of trades according to pre-defined trading rules that consider both take-profit and stop-loss levels for each trade.

1.6.3 Risk-Oriented Evaluation Contribution

In addition to providing classification accuracy metrics for the model, this study also evaluates the models according to several financial metrics related to trading, such as total return, win rate, the Sharpe ratio, and maximum drawdown.

1.6.4 Saudi Market Contribution

This study applies the created trading signal model to the Saudi Stock Market, using publicly available market and financial data. Thus, this study contributes to

the limited existing research on artificial intelligence and trading signal generation within the Saudi Tadawul market.

1.6.5 Literature Contribution

The main way that this study contributes to the existing literature is through the creation of a framework that simultaneously considers technical indicators, fundamental financial metrics, trading rules, and risk-oriented evaluations of trading signal models.

1.7 Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 presents the theoretical background and related studies regarding stock market prediction, technical analysis, fundamental analysis, machine learning, deep learning, and hybrid financial prediction models. Chapter 3 describes the proposed methodology for the study, including the preparation of the datasets, the feature engineering processes, the construction of the models, the integration of the models to form the hybrid model, and the backtesting of the models. Chapter 4 presents the results of the experiments performed on the proposed models, as well as a discussion of the predictive and trading performance of those models. Finally, Chapter 5 draws conclusions regarding the study, its implications, its limitations, and future research in this area.

1.8 Research Clarification

The research described in this paper is entirely based on structured and publicly available data regarding the Saudi Stock Market. Both daily and annual financial data are used to create the various features of the technical and fundamental analysis of the market. Furthermore, no private or sentiment data were used in the determination of the features of the stocks in this study.

The model created is for academic purposes only and is not to be used as a recommendation for financial markets or as financial advice. Any results from the model are to be used in the same manner as the research itself.

Model: "sequential"		
Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 20, 64)	20,224
dropout (Dropout)	(None, 20, 64)	0
lstm_1 (LSTM)	(None, 32)	12,416
dropout_1 (Dropout)	(None, 32)	0
dense (Dense)	(None, 1)	33

Figure 1.1: Research Gap – Technical vs Fundamental Fragmentation

Chapter 2: Background and Related Work

2.1 Introduction

This chapter presents the theoretical background and related work in the area of stock market prediction using AI. It reviews the theories of fundamental and technical analysis as well as the various models of machine learning. Furthermore, it also presents relevant studies conducted in various countries, including Saudi Arabia, to identify the existing gaps in the literature.

2.2 Stock Market Forecasting

Stock market forecasting is a process of determining the future price movement of stocks based on available information from past movements. Stock markets are characterized by non-linear and non-stationary time series data due to the variety of factors influencing stock prices. Moreover, the challenges of stock market forecasting are more pronounced in emerging markets like the Saudi Stock Market.

2.3 Fundamental Analysis

Fundamental analysis evaluates the intrinsic value of a company by analyzing its financial data, such as financial ratios. Commonly used financial ratios include:

- Price-to-earnings (P/E) ratio
- Return on equity (ROE)
- Debt-to-equity (D/E) ratio

This approach has been proven to be successful in determining the long-term value of stocks. However, it is not effective in providing traders with information regarding the short-term movements of stock prices.

2.4 Technical Analysis

Contrary to fundamental analysis, technical analysis determines the future price of stocks by studying past price and volume movements.

Various technical indicators are used, including the relative strength index (RSI), moving averages, MACD indicators, and Bollinger bands.

Technical analysis is useful for short-term movements but does not account for the financial strength of the company.

2.5 Machine Learning in Finance

Machine learning models can identify non-linear relationships in large data sets.

Various models have been applied, including Random Forest, support vector machines (SVM), and gradient boosting models.

Deep learning models such as long short-term memory (LSTM) networks have shown potential for stock market analysis. However, the majority of models rely solely on price data.

2.6 Hybrid Integration of Technical and Fundamental Analysis

In many cases, it is more beneficial to use a variety of data sources or models within an approach than to rely upon only one. In the realm of stock market and stock price prediction, it may be beneficial to use a variety of different data sources to provide for more comprehensive and accurate predictions. For instance, technical indicators may reflect the technical strength of the stocks listed on the exchanges, while fundamental indicators may reflect the financial strength of those listed companies.

The approach to be used within this project utilizes two different models. The first model to be utilized is a Bidirectional Long Short-Term Memory (BiLSTM) model, which will help to recognize and account for technical features within the market. An XGBoost model will help to recognize and account for both technical and fundamental features of the stocks. Rather than providing a prediction of the price of the stock, the two models will provide an estimate of the probability of succeeding in the trade.

The probability of succeeding in a trade is determined by creating a weighted average of the two models’ predictions. The resulting probability can then be compared against predetermined thresholds that indicate whether or not a trading signal should be created based upon those models’ results. By utilizing both technical and fundamental data in the trading signal generation process, the framework can be beneficial in improving the quality of those trading decisions..

Comparison of Stock Market Prediction Approaches			
Technical Analysis vs Fundamental Analysis vs Hybrid Approach			
Aspect	Technical Analysis	Fundamental Analysis	Hybrid Model (Proposed)
Data Type	Price & Volume (High-frequency)	Financial Statements (Low-frequency)	Both Price Data & Financials
Time Horizon	Short-term (Days/Weeks)	Long-term (Months/Years)	Short to Medium-term
Key Indicators	RSI, MACD, MA, Bollinger Bands	P/E, ROE, D/E, Cash Flow	Combined Technical & Fundamental
Strengths	Market timing, Entry/Exit signals	Company valuation, Intrinsic value	Timing + Valuation
Weaknesses	Ignores company health, False signals	Ignores market sentiment, Lagging data	Complex integration, Model overfitting

Source: Academic Research 2024

Figure 2.1: Comparison of Technical, Fundamental, and Hybrid Approaches

2.7 Trading-Oriented Evaluation Metrics

Most of the existing studies on stock market prediction using machine learning methodologies focus mainly on classification metrics in statistics. However, despite their usefulness, such metrics do not provide a complete evaluation of the economic value of a trading model.

In addition to classification metrics, trading-oriented metrics must also be considered to evaluate the performance of a trading model. Metrics such as total

return, win rate, Sharpe ratio, and maximum drawdown can provide valuable insights into the trading model's profitability and risk.

For the defined problem, the target variable represents the stock's price movement, with a successful trade occurring when the price reaches a take-profit level of +5% before reaching a stop-loss level of -2%, as observed in the subsequent price movement of the stock.

2.8 Evolution of Financial Prediction Models

Financial market theories began with the Efficient Market Hypothesis, which suggested that stock market prices reflect all available information and that it is therefore difficult for investors to make accurate price predictions (Fama, 1970). However, models like ARIMA (AutoRegressive Integrated Moving Average) were developed that could help forecast stock prices and other financial time-series indicators (Adebiyi et al., 2014).

Stock market data are complex and often nonlinear in their relationships. Thus, traditional statistical models may not be able to accurately describe the complex relationships within the markets. Consequently, alternative machine learning models that can learn nonlinear relationships within the market data are often employed as prediction models.

Deep learning models have been particularly useful for recognizing the complex temporal relationships between stock market data. For instance, LSTM (Long Short-Term Memory) models have been used for stock market prediction due to their ability to identify patterns in data that are sequenced over time (Hochreiter & Schmidhuber, 1997; Fischer & Krauss, 2018). Additionally, models like XGBoost have also been used for stock market prediction, indicating their strong performance in tasks utilizing structured data tables (Chen & Guestrin, 2016).

2.9 Machine Learning and Deep Learning Studies

Machine learning and deep learning algorithms have been applied to the task of stock market prediction. For instance, one study that examined various machine learning methods for forecasting stock market indicators found that different models performed better than others depending on the specific considerations of

that stock market and its movements (Ahmed et al., 2010). Additionally, models like random forests and gradient boosting methods have been applied to financial markets prediction studies due to their success in recognizing complex nonlinear relationships within data (Breiman, 2001; Chen & Guestrin, 2016).

Deep learning models, such as those that utilize stacked autoencoders and LSTMs, have been proposed for studying financial time-series data (Bao et al., 2017). For example, both Fischer and Krauss (2018) and Krauss et al. (2017) have published studies that explore the use of LSTM models for stock market prediction, as well as the comparison of deep learning models for statistical arbitrage.

In addition, deep learning models based on language models, such as the Informer model, have also been proposed to improve the prediction of long sequences of time-series data in financial markets (Zhou et al., 2021).

As with many AI models for stock market prediction, the majority of existing research utilizes indicators of price movements to attempt to classify whether stock prices will rise or fall. However, few current studies have attempted to evaluate whether those price predictions can be used to create profitable trading strategies for investors.

2.10 Critical Evaluation of Prior Studies

The literature on stock market prediction using AI methods indicates that while there are some successful models, there are also some limitations to the existing methods of prediction.

First, most studies published to date use AI methods to directly forecast the prices of stocks, but do not determine if the predicted movements of those stocks will result in profitable trading activity. In other words, the models may be able to correctly predict if prices will rise or fall, but not whether those trades will be profitable overall.

Second, many of the studies rely upon statistical measures of accuracy to determine the success of the models, but do not use any performance measures for trading strategies, such as the total return of the investments made based upon those models.

Third, many of the studies use only technical indicators of the stocks to determine their predictions, while ignoring other factors related to the financial strength of those companies. Additionally, many studies that focus only on fundamental analyses of those companies may not be able to identify short-term price movements of those stocks.

Fourth, a few of the most recent studies that combine sentiment analysis and news feeds for stock market prediction are still in their infancy and rely upon additional data sources that may introduce additional noise to the predictions of the models or create other concerns regarding the use of such data.

However, as the stock market and trading strategies advance in complexity, the need for a framework that combines technical indicators, fundamental financial analysis, and the creation of profitable trading strategies becomes necessary.

2.11 Hybrid Models in Literature

Various attempts have been made in the literature to improve the accuracy of stock market predictions by utilizing hybrid models. Tsai and Wang (2009) presented various machine learning techniques as a means of improving stock market prediction and showed that the combination of multiple techniques results in improved predictions compared with using only a single machine learning method.

Nti et al. (2020) performed a literature review of existing studies on stock market prediction techniques and classified them according to their use of technical, fundamental, or combined analysis techniques to identify that there are relatively few studies that employ combined analysis methods compared with those that use only technical or fundamental analysis methods alone.

Daradkeh (2022) developed a hybrid framework that used a combination of a CNN-BiLSTM deep learning model and sentiment analysis techniques to predict the movements of stocks in the Dubai Financial Market. The results indicated that the framework was able to achieve improved predictions by utilizing the sentiment trends of stocks in addition to the financial data of those stocks. However, the use of sentiment data also introduces potential noise into the predictions.

Ansari et al. (2024) developed a multifaceted reinforcement learning model that incorporated technical and fundamental indicators to predict the stock market. The results of the study demonstrated that incorporating different types of financial data into a trading model lead to improved trading results. However, reinforcement learning models can be difficult to interpret and the exact reasons that they produce specific outcomes.

The present study aims to develop a framework that utilizes both XGBoost and BiLSTM models in a weighted ensemble to identify the outcome of trading activities. Unlike previous studies, the framework aims to focus on the classification of trading outcomes with the consideration of risk-adjusted measures for evaluation.

2.12 Studies in Emerging and Saudi Markets

An increasing number of research studies are dedicated to analyzing and utilizing artificial intelligence techniques for stock market prediction within the emerging markets across the world. The characteristics of these markets are often significantly different from those of the developed markets, necessitating the development of models that account for such differences.

A few of the emerging markets that have been studied include the Saudi Stock Market. Al-Nefaie and Aldhyani (2022) utilized models including the LSTM and MLP models to successfully predict the adjusted closing prices of various sectors within the Saudi Stock Exchange.

Alsulmi and Al-Shahrani (2022) developed an automated trading framework that relied upon machine learning techniques to be used within the Saudi Stock Exchange. This study is relevant to the current study due to its consideration of trading within the Saudi stock market. However, their framework did not include the combination of technical and fundamental indicators.

Daoria et al. (2024) utilized deep learning and sentiment analysis from social media platforms like Twitter to develop a model for predicting the movements of the stocks within Saudi Arabia. The result of this study indicated the potential usefulness of the framework. However, the nature of the social media data introduces potential noise and issues with reproducibility to the trading decisions made based on this framework.

Khurayshi and Alharbi (2025) developed a study that utilized various machine learning models to study stock markets in Saudi Arabia. Their implementation of models including linear regression, random forest, XGB, and AdaBoost models allowed them to contribute to the study of stock market prediction in Saudi Arabia. However, only the predictions of the next-day prices of the stocks in Saudi Arabia were performed with these models.

Alhazmi et al. (2025) performed a study of the adoption of artificial intelligence in the Saudi Stock Exchange in relation to the quality of financial reports published by the listed companies within the exchange. While not dedicated to the study of trading outcomes, this study indicates the relevance of artificial intelligence models to the Saudi Stock Exchange.

Although there are a few studies dedicated to the Saudi Stock Exchange, there are still relatively limited studies on the market, especially in relation to the inclusion of both technical and fundamental indicators within a framework that utilizes artificial intelligence and aims to evaluate trading outcomes according to risk-adjusted measures.

2.13 Research Gap

Out of the studies reviewed for both the general market and the Saudi Stock Market, a few research gaps became apparent.

One research gap is in the study of technical and fundamental analysis separately. Technical analysis indicators are helpful in understanding the movements of stocks over short periods. However, they do not reflect the financial strength of the companies that published those stocks. Fundamental financial ratios provide information regarding the financial strength of those companies, but do not reflect their stock market movements over short periods of time.

Another gap in the existing literature is regarding the study of the outcomes of trading decisions based upon the stock market predictions. Most existing studies focus on the accuracy of the predictions regarding the prices of stocks. However, there is a lack of studies that aim to provide information regarding whether the predictions of stock market movements will lead to profitable and risk-controlled trading decisions.

A third research gap related to the existing literature is the need for more studies that aim to classify the outcomes of trading based on take-profit and stop-loss rules. The majority of existing models predict the movements of stocks, but do not aim to provide information regarding the success or failure of those trades.

Finally, another research gap is in relation to the number of existing studies of the Saudi Stock Market in comparison to the number of studies of other markets around the world. Few studies of the characteristics and trading outcomes of the Saudi Stock Market have been performed with the intention of including technical indicators, fundamental financial ratios, and artificial intelligence models.

The present study intends to address these gaps in the literature by introducing a model that combines technical indicators, fundamental financial ratios, XGBoost, and BiLSTM machine learning models in order to determine the outcome of trading activities within the markets. Furthermore, the study will evaluate the results of such predictions not only through classification metrics, but also through trading metrics as well.

Figure 2.2 illustrates the research gap between technical and fundamental analysis, and the Hybrid model addresses this limitation.

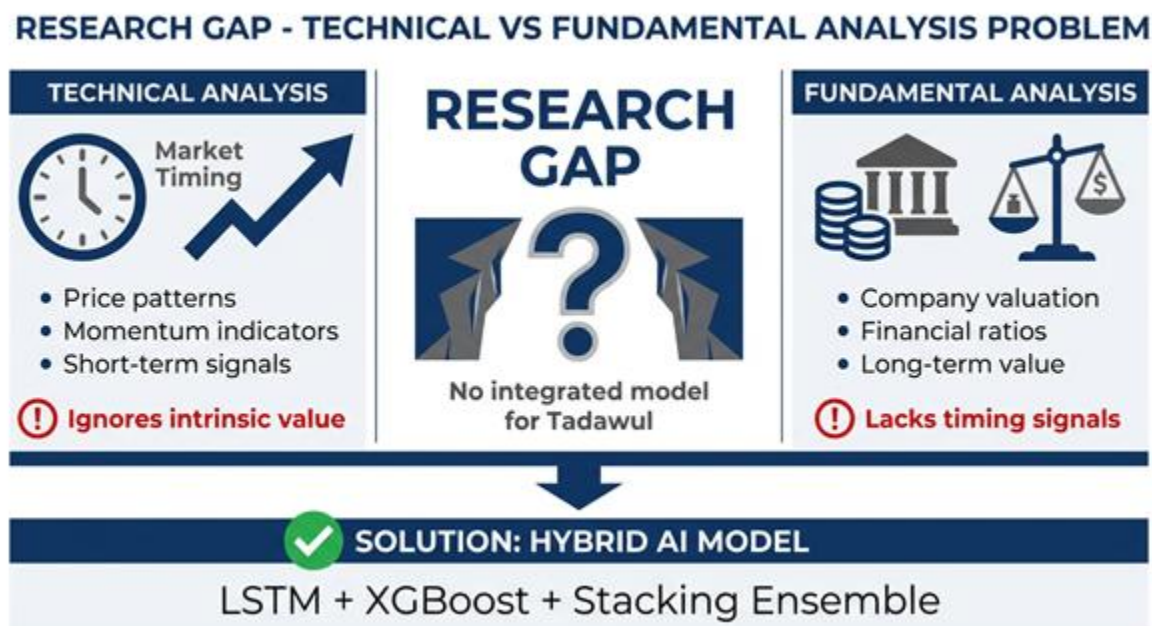


Figure 2.2: Research Gap and Hybrid Model

Author(s)	Market / Context	Data Type	Model / Method	Key Contribution	Limitation / Research Gap
Nti et al. (2020)	Global	Technical, fundamental, and combined analysis studies	Systematic review	Reviewed stock market prediction studies and categorized them into technical, fundamental, and combined approaches.	Shows that combined approaches are less common, and many studies focus mainly on prediction accuracy rather than trading performance.
Daradkeh (2022)	Dubai Financial Market	Quantitative financial data, news events, and sentiment trends	CNN-BiLSTM hybrid framework	Demonstrated that combining sentiment and quantitative financial data improves stock trend prediction.	Relies heavily on sentiment and news data, which may introduce noise and is not directly focused on risk-adjusted trading signals.
Al-Nefaie and Aldhyani (2022)	Saudi Stock Exchange	Historical stock price data	LSTM and MLP	Applied AI time-series models to forecast adjusted closing prices in several Tadawul sectors.	Focuses on close price forecasting rather than trading outcome classification or take-profit/stop-loss logic.
Alsulmi and Al-Shahrani (2022)	Saudi Stock Exchange	Market data and trading actions	ML-based automated trading framework	Developed a machine learning framework for trading decision-making in Tadawul with risk management concepts.	Limited integration of company-level fundamental ratios and does not focus on hybrid technical–fundamental model comparison.
Daoria et al. (2024)	Saudi Stock Market	Stock data and Twitter sentiment	LSTM and SVM	Combined stock market data with Twitter sentiment to support stock price prediction.	Strong reliance on social media sentiment, which may be noisy and difficult to regulate or reproduce.
Ansari et al. (2024)	Dow Jones multi-stock market	Technical indicators and fundamental indicators	Reinforcement Learning using A2C and DDPG	Proposed a multifaceted RL trading strategy using technical and fundamental indicators with a risk-aware reward function.	RL models can be complex, harder to interpret, and may require extensive tuning and validation.

Khurayshi and Alharbi (2025)	Saudi Stock Market	Historical data for selected Saudi companies	Linear Regression, Random Forest, XGB, AdaBoost	Compared multiple ML models for Saudi stock price prediction.	Focuses on next-day price prediction and model accuracy rather than trading strategy performance and drawdown control.
Saberironaghi et al. (2025)	Global	Multiple financial datasets	Review of ML and DL techniques	Reviewed ML/DL models, datasets, evaluation metrics, and challenges in stock market prediction.	Broad review; does not propose a Saudi-specific trading signal framework with explicit TP/SL rules.
Alhazmi et al. (2025)	Saudi financial reporting context	Financial reporting and documentary evidence	AI adoption and bibliometric/content analysis	Examined AI adoption and financial reporting quality in the Saudi context.	Not focused on stock trading prediction, but supports the relevance of AI in Saudi financial decision-making.
Present Study	Saudi Stock Market	Daily market data and annual financial ratios	XGBoost, BiLSTM, and weighted Hybrid model	Integrates technical and fundamental features with explicit TP/SL trading logic and risk-adjusted evaluation.	Limited to 20 Tadawul companies and simplified backtesting assumptions.

Table 2.1 – Comparative Summary of Related Studies and Research Gap

Table 2.1 presents a summary of various studies related to stock market prediction, hybrid models, the application of these models within the Saudi stock market, and AI models for financial decision making. Most of the existing literature reviews the use of machine learning and deep learning models for stock price forecasting. However, there are a limited number of studies that use technical indicators, fundamental financial ratios, and financial evaluation metrics within the context of stock market trading.

As illustrated in Table 2.1, there are various studies in the literature that use machine learning, deep learning, sentiment analysis, and reinforcement learning models for stock market trading. Nevertheless, there are limitations to these models. Many of the existing models are used to predict the future prices of the stocks. Others depend primarily on sentiment and social media analysis to inform their models. Additionally, the majority of the studies that utilize the stock markets of Saudi Arabia fail to use technical indicators, financial ratios, and trading metrics in their analyses.

The study presented in this paper aims to fill these gaps in the literature by presenting a model that utilizes both XGBoost and BiLSTM models using technical and fundamental indicators from the Tadawul stock market exchange. Furthermore, instead of predicting the price of the stock, this model aims to determine the outcome of the trading of these stocks, using a variety of financial metrics in addition to the model's own predictive metrics.

2.14 Comparative Analysis with Existing Studies

The studies included in Table 2.1 indicate that there have been various research endeavors to employ artificial intelligence in the field of stock market prediction. These studies have mainly employed technical analysis, fundamental analysis, sentiment analysis, deep learning, and reinforcement learning models. Despite the existence of these studies, there are some differences between these studies and the present research endeavor.

Some of the existing researches have mainly focused on the prediction of the price of the stocks. Other studies have used sentiment and news analysis to help improve their prediction accuracy for the stock markets. However, these studies have mostly ignored evaluating whether the predictions from the artificial intelligence models could result in profitable and risk-controlled trading decisions.

The present research study differs from these existing studies because it uses technical and fundamental analysis in a hybrid artificial intelligence model and evaluates the trading decisions that could result from the predictions of such a model. Furthermore, the model includes take-profit and stop-loss logic to determine the financial risk and profit of the trading recommendations of the model.

Thus, the main difference and contribution of the present study is that it shifts the focus of existing research from the prediction of the stock markets to evaluating the trading performance of such predictions.

2.15 Positioning of the Present Study

The present study aims to contribute to the existing literature on artificial intelligence in the stock market by proposing a model that is both hybrid and

oriented towards trading in the Saudi Stock Market. More specifically, the model uses both technical and fundamental analysis indicators and employs two different machine learning models: XGBoost and BiLSTM models.

Whereas most existing research aims to focus solely on the price of the stocks and the accuracy of its predictions, the present research aims to formulate the problem as a classification problem that predicts the probability of a stock reaching a take-profit level before reaching a stop-loss level.

Furthermore, the present study also introduces different aspects of evaluation for the proposed model. Not only are the predictive metrics of the model evaluated, but its financial metrics are evaluated as well.

Thus, the contribution of the present study is the proposal of a model that is both focused on the Saudi Stock Market, uses a hybrid AI model to generate its signals, and evaluates the financial performance of its recommendations.

2.16 Chapter Summary

In this chapter, the theoretical background and the related literature on stock market forecasting, technical analysis, fundamental analysis, machine learning, deep learning, hybrid AI models, and the existing studies on the Saudi Stock Market were reviewed.

AI models have been employed in various studies on stock market forecasting. However, most studies have primarily focused on the forecasting of the prices of the stocks and the accuracy of those forecasts. Few studies have evaluated the trading performance of the stocks based on financial metrics. Furthermore, most studies have used data from a single data source to train their AI models. They did not use both technical and fundamental indicators of the stocks.

A gap exists in the literature regarding AI models that are used for technical and fundamental analysis of the Saudi Stock Market with the intent of formulating classification models for trading outcomes with the evaluation of financial metrics from those models. The methodology chapter develops the methodology that will be used for this study to fill this existing gap in the literature.

Chapter 3: Proposed Approach (Methodology)

3.1 Overview of the Hybrid Framework

This chapter describes the architecture and implementation of the Hybrid AI framework for trading signal generation for the Saudi stock market (Tadawul). The framework combines technical and fundamental data sources and employs two complementary machine learning models.

In contrast to existing models, the Hybrid approach aims to predict trades rather than prices by incorporating trading constraints. The overall framework consists of data acquisition, preprocessing, feature engineering, model training, model ensembling, and backtesting against realistic trading conditions.

3.2 Dataset Description

The dataset used in this study consists of five years of daily market data and annual financial statement data for 20 companies listed on the Saudi Stock Market (Tadawul) during the period from 2021 to 2025. The selected companies represent different sectors, including banking, petrochemicals, telecommunications, and energy, in order to provide sectoral diversity within the sample.

The raw dataset contained 24,966 observations. After preprocessing, feature construction, and removal of incomplete records caused mainly by rolling-window technical indicators, the final cleaned dataset contained 19,742 observations. The dataset was then divided chronologically into training and testing subsets using an 80–20 split. This resulted in 15,790 training observations and 3,952 out-of-sample testing observations.

Sector	Number of Companies
Banking	5
Petrochemicals	5
Telecommunications	5
Energy	5

Each observation includes market-based technical features derived from daily price data and fundamental features derived from annual financial statements. The target variable was constructed using predefined trading logic. A trading outcome was labeled as successful when the stock reached a take-profit level of +5% before reaching a stop-loss level of -2%, based on subsequent price movement.

The final cleaned dataset showed a moderate class imbalance, with approximately 32.00% positive trading outcomes. In the out-of-sample test set, the positive class represented approximately 25.78% of observations. This imbalance was considered during model training through class weighting and imbalance-aware model configuration.

3.3 Data Preparation

The data preparation process was used to transform daily and annual market and financial data into modeling data. The daily price data for each company was loaded into the analysis and sorted by date. The annual balance sheet and income statement data for each company were loaded, cleaned, and merged to calculate financial ratios.

Following the preparation of the financial statement data, it was merged with the daily market data. The financial data was merged in such a way that the most recent financial data available for each company was associated with each daily market data observation. Thus, each row of data included the technical indicators for that company's price as well as its fundamental financial ratios.

The technical indicators were calculated using rolling and exponential calculation windows. The initial rows of each company's data had missing values for these indicators because it required a certain number of days of price data to calculate such indicators. Thus, these initial rows of data were removed from the data set. Any remaining missing values for financial data were filled with forward and backward fill values for each company's financial data, and the remaining missing values were replaced with the mean value for that data field for each company.

This resulting data set was used to train, test, and backtest the financial model. Furthermore, the data was split into training and test sets chronologically to ensure that no data was included in both the training and test set for any company.

3.3.1 Data Cleaning

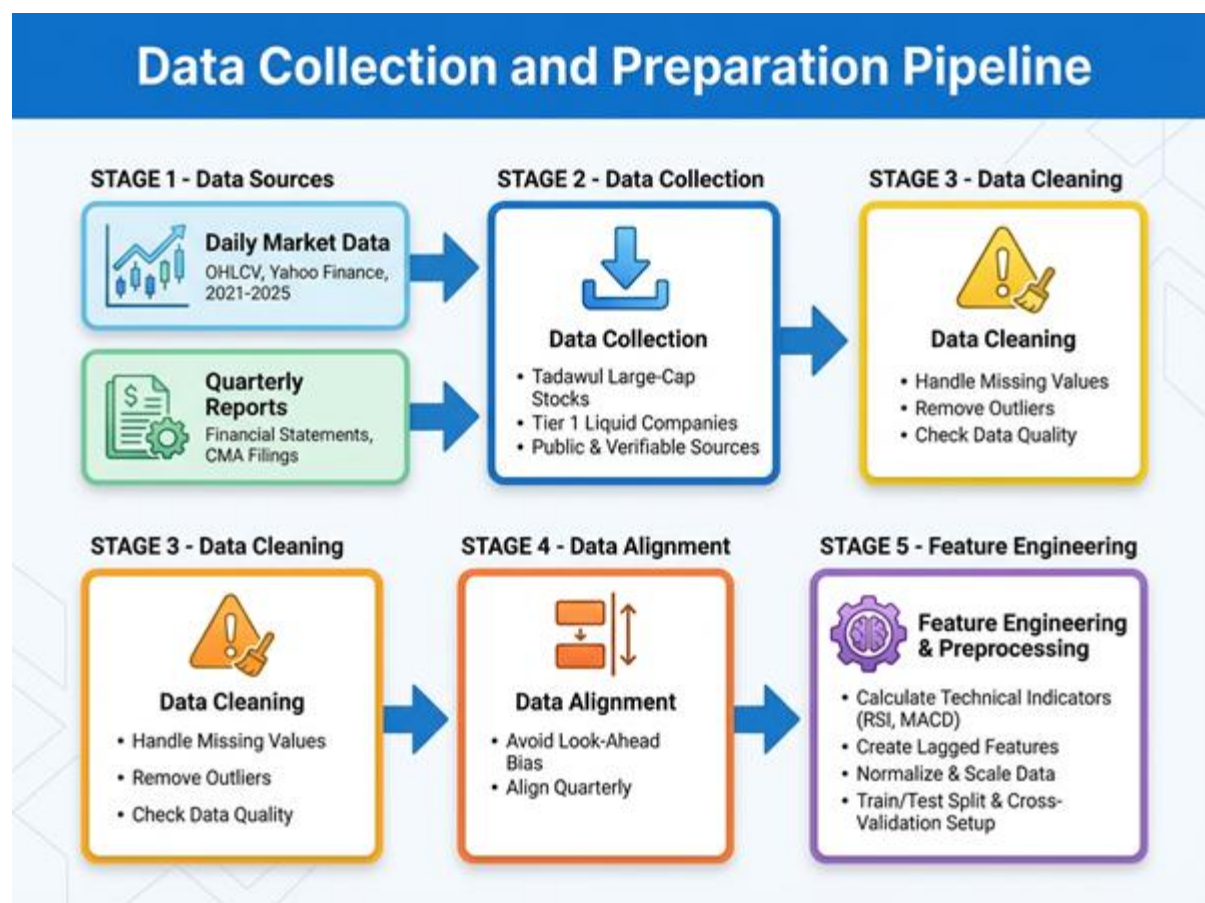


Figure 3.2: Data Collection and Preparation Pipeline

Due to unavailable financial statements and technical indicators for the companies in the time period covered by the dataset, there were some missing values within the initial dataset. Technical indicator calculations, such as moving averages, RSI (relative strength index), MACD (moving average convergence divergence), and volatility calculations, were unable to be calculated for the earliest data point for each company, as calculations for these indicators required the observations of prices for that company for previous time periods within that trading day. Thus, for each company, any data that existed prior to the first observation that had all of the required technical indicators was discarded from the dataset. Furthermore, any

remaining missing values within each company were filled in with forward- and backward-filling techniques, as well as mean imputation for any remaining missing values for the company's numerical variables.

As a result of these data cleaning steps, the size of the initial dataset was reduced from 24,966 rows to 19,742 rows of data. The remaining dataset for each company still contained all of the rows for that company in chronological order, and was therefore suitable for model training.

3.3.2 Leakage Prevention

In order to mitigate the risk of look-ahead bias, the dataset was divided according to a chronological structure. Each company's data was divided into a training component that contained only earlier observations of the company's stocks, and a testing component that contained only later observations of the company's stocks. Such a division did not employ any form of random shuffling to divide the data into training and testing portions.

Each feature was constructed using information from the stocks that was available during the observation of each stock's price. For instance, technical indicators were calculated for each stock using the prices of that stock that were observed prior to the observation of the stock's price. Similarly, information for financial ratios was obtained from financial statements of the companies that were observed during the same time intervals as the stocks' observations.

The target variable for each stock's observation was constructed using observations of the stock's price that occurred after the observation of that stock's price. Thus, this information was used only in the creation of the supervised learning problem, and was not used as a feature that informed the creation of the models. Instead, the models were trained only upon features that were available during the observation of the stock prices, and were tested against stocks whose prices were observed after the training of those models.

While this approach—dividing the dataset into training and testing portions according to the chronological observation of stock prices—helps ensure that the evaluation of the created models is based upon observations of the stocks that occurred after the models were created and trained, rather than using future stock prices to evaluate those models.

3.3.3 Standardization

Feature standardization was applied to improve numerical stability during model training, particularly for the BiLSTM model. The StandardScaler method was used to transform numerical features by removing the mean and scaling them to unit variance.

The standardized value was calculated as follows:

$$Z = (X - \mu) / \sigma$$

where Z represents the standardized value, X is the original feature value, μ is the mean of the feature, and σ is the standard deviation.

To avoid data leakage, the scaler was fitted only on the training data and then applied to the testing data. This ensured that information from the testing set did not influence the preprocessing parameters used during model training.

3.4 Feature Engineering

Feature engineering is the process of converting raw data into numerical inputs for the machine learning models. Features can be broadly divided into technical and fundamental features based on the source of the data.

Technical features are calculated from daily market data, while fundamental features are calculated from annual financial statement data. By using both types of features as model inputs, the framework can incorporate information regarding both the market and the financial strength of each company.

Technical features include return-based, trend-based, momentum-based, and volatility-based features, while fundamental features include financial statement ratios related to the financial strength of each company. These features will be used as the inputs for the XGBoost and BiLSTM models.

3.4.1 Technical Features

The technical features were calculated from daily market data. The features were designed to evaluate aspects of the price movement of the stocks, such as trend, momentum, and volatility.

The following technical indicators were utilized for each stock:

- Daily returns
- Simple moving averages over 10, 20, and 50 days
- Exponential moving averages over 10, 20, and 50 days
- Distance between the closing price and simple moving averages
- Relative Strength Index over 14 days
- MACD indicator
- MACD signal line
- MACD histogram
- 20-day return volatility

These indicators were selected due to their common use in analyzing stocks for short-term trading activities.

3.4.2 Fundamental Features

The fundamental features were calculated from annual financial statement data, including balance sheet and income statement information. These features were included to represent the financial strength and quality of the listed companies.

The following financial ratios were used:

- Return on Equity (ROE)
- Debt-to-Equity Ratio
- Profit Margin
- Earnings Per Share (EPS)

ROE was used to capture profitability relative to shareholders' equity. The debt-to-equity ratio was used to represent leverage and financial risk. Profit margin measured the company's ability to convert revenue into net income, while EPS reflected earnings generated per share. Together, these indicators provided the model with firm-level information that complemented the market-based technical indicators.

3.5 Trading Constraints Justification

The logic behind the trading strategy employed in this project was the utilization of predefined take-profit and stop-loss levels for each of the stocks evaluated. For each stock, a take-profit level of +5% was established, as well as a stop-loss level of -2%. These two parameters indicate a reward-to-risk ratio of 2.5:1 for each stock.

Furthermore, by establishing these two parameters for each stock, the problem could be framed as a classification problem for the machine learning models. Rather than predicting the price of each stock, the models were to determine the likelihood that the stock would reach the take-profit level before reaching the stop-loss level.

Finally, utilizing these predefined parameters for each stock allows for the stock prices to be evaluated using various trading-related metrics, such as return, win rate, the Sharpe ratio, and maximum drawdown.

3.6 Model Architecture

3.6.1 XGBoost Model

The XGBoost model was trained using all available features, including both the technical indicators and fundamental ratios. XGBoost is a model that can model nonlinear relationships and is often used for classification problems on tabular data.

Each of the features was scaled using StandardScaler. The class imbalance was handled using the *scale_pos_weight* parameter for XGBoost, which weights the negative class to reflect its smaller size compared to the positive class.

The main parameters for the XGBoost model were:

- *estimators*: 400
- *max_depth*: 6
- *learning_rate*: 0.05
- *metric*: *log – loss*
- *random_seed*: 42
- *scale_pos_weight*

The model outputs a probability score that represents the likelihood of achieving a successful trading outcome for a given observation.

3.6.2 Bidirectional LSTM Model

The second model used in this paper is the bidirectional LSTM. This model was constructed to help recognize the temporal dependencies within the data. LSTM models are especially adept at recognizing these dependencies, making them appropriate candidates for the analysis of financial data.

The inputs to the model are the rolling sequences of 60 days of data. For each of these sequences, the model is provided the features calculated for each of the previous 60 days of observations for that company. The model then returns the probability that the trading outcome would be successful for that company.

The architecture of the model is as follows:

- Bidirectional LSTM layer with 64 LSTM units and return sequences enabled
- Dropout layer with a dropout rate of 0.5
- Bidirectional LSTM layer with 32 LSTM units
- Dropout layer with a dropout rate of 0.5
- Dense output layer with sigmoid activation function

The model was trained using the Adam optimization algorithm and the binary cross-entropy loss function. The model utilized AUC scores calculated on the validation subset of the data to monitor the model's performance during training. An early stopping algorithm with a patience of 10 was used during training to avoid overfitting. Furthermore, class weights were used to account for the inherent skew in the training data.

3.7 Hybrid Integration

The final prediction within the proposed framework was made via a weighted hybrid ensemble that combined the probabilities from the XGBoost and BiLSTM models.

The XGBoost model was weighted at 0.60, and the BiLSTM model was weighted at 0.40. These weights were applied to all out-of-sample predictions. The intention of applying such weights to the model predictions was to combine the strengths of the two models into one final probability.

Finally, if that final probability was above a threshold value, a trading signal was generated for that observation. Otherwise, no trade was made with the signal.

Thus, the framework would produce fewer trades if the threshold was higher, indicating a more conservative trading strategy.

3.8 Mathematical Formulation

The hybrid model uses a weighted combination of the prediction probabilities of the XGBoost and BiLSTM models to determine the final probability of achieving a successful trading outcome. The following equation can be used to calculate the hybrid prediction probability:

$$P_{Hybrid} = (0.60 \times P_{XGBoost}) + (0.40 \times P_{BiLSTM})$$

where P_{Hybrid} is the hybrid model's probability of successful trading outcomes, $P_{XGBoost}$ is the probability of successful trading outcomes predicted by the XGBoost model, and P_{BiLSTM} is the probability of successful trading outcomes predicted by the BiLSTM model.

This general equation can be used to represent the hybrid model:

$$P_{Hybrid} = (w_1 \times P_{XGBoost}) + (w_2 \times P_{BiLSTM})$$

where w_1 and w_2 are the weights of the individual models' probabilities, and the following relationship exists between the two:

$$w_1 + w_2 = 1$$

For this project, the selected weights were $w_1 = 0.60$ for the XGBoost model and $w_2 = 0.40$ for the BiLSTM model.

Finally, a trading signal is generated based on the value of the hybrid model's final probability. A trading signal of 1 is generated if the probability value is greater than a pre-defined threshold (θ), and a signal of 0 is generated if the probability value is less than or equal to that threshold:

$$\text{Trade Signal} = 1 \text{ if } P_{\text{Hybrid}} > \theta$$

$$\text{Trade Signal} = 0 \text{ if } P_{\text{Hybrid}} \leq \theta$$

where θ is the threshold probability value.

Thresholds of 0.50, 0.55, 0.60, 0.65, 0.70, and 0.75 were evaluated in this project.

3.9 Backtesting Framework

A backtesting framework was established to examine the economic performance of the model predictions. The goal of the backtesting framework was to determine whether the probabilities that the models predicted could result in profitable trades with minimal risk of loss.

For each of the models, a trade signal was made when the predicted probability was above a set threshold. The thresholds that were tested in this evaluation were 0.50, 0.55, 0.60, 0.65, 0.70, and 0.75. Each of these thresholds results in a model that is either less or more selective in the trades that it executes. The entry price for each trade was the closing price of the model at the time that the signal was observed. For each trade, a take-profit rate of +5% and a stop-loss of -2% were implemented. If the trade reached the take-profit percentage before reaching the stop-loss percentage, the trade was counted as a winning trade. If the trade reached the stop-loss percentage first, the trade was counted as a losing trade. Finally, if the trade did not reach either the take-profit or stop-loss percentage during the

evaluation period, then the trade was exited at the closing price of that evaluated period.

The transaction costs incorporated into the trading model were 0.15% for both the entry and exit of a trade, or 0.30% for the full trade. This cost was included to provide a more accurate evaluation of the potential profits of the models. After each trade was executed, the evaluation advanced to the exit point for that trade before searching for the next trade signal.

The backtest was used to evaluate the performance of the models with five different metrics: the number of trades executed, the win rate of the trades, the total return of the trades, the Sharpe ratio, and the maximum drawdown.

3.9.1 Backtesting Assumptions

The backtesting framework used in this study is a simplified simulation designed for academic evaluation. It does not fully reproduce all real-market execution conditions. In particular, the simulation does not explicitly model bid–ask spread, order book depth, partial fills, execution latency, or liquidity constraints.

Therefore, the reported trading results should be interpreted as simulated performance under the assumptions of the study, rather than as guaranteed live-trading outcomes. The purpose of the backtest is to compare the relative performance of the standalone and hybrid models under consistent assumptions.

3.10 Sharpe Ratio

The Sharpe ratio is used to determine the return that is generated per unit of risk for each of the models. Thus, the Sharpe ratio can help to provide a comparison between the various models as to which model may exhibit better risk-adjusted performance.

The Sharpe ratio for each of the models was calculated using the returns distributed from the trading strategy backtesting procedure. The general equation for calculating the Sharpe ratio is:

$$\text{Sharpe Ratio} = (R_p - R_f) / \sigma_p$$

Where R_p is the average return for the strategy, R_f is the risk-free rate, and σ_p is the standard deviation of the returns for the strategy.

Each of the models was calculated using the same equation to allow for a true comparison of each model to the others. Thus, each of the Sharpe ratio values represents only the relative performance of each model to the others.

3.11 Performance Metrics

There were two categories of metrics used to evaluate the models: classification metrics and trading performance metrics.

Classification metrics included accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). These metrics were used to assess the ability of the models to classify which of the trader outcomes were successful versus unsuccessful.

Trading performance metrics included the number of trades, win rate, return, Sharpe ratio, and drawdown. These metrics evaluated the trading performance of the models using the predictions from the models to make trades.

These different categories of metrics allowed the study to evaluate the performance of the models in both classification and trading.

3.12 Reproducibility

Several procedures were applied to increase the reproducibility of the experimental results. First, a random seed value of 42 was applied to the various machine learning and numerical computations to reduce the randomness of the algorithms. Second, the dataset was divided chronologically to ensure that temporal aspects of the market data were preserved in the split of the dataset. Third, the same preprocessing steps were applied to each of the models, ensuring that any scaling

procedures were applied only to the training data. Finally, each of the trained models and preprocessing objects were saved to enable future evaluation of the models. Additionally, the feature dataset and backtesting results were also exported as separate files for ease of access and evaluation.

Together, these steps ensured the reproducibility of the described experiment.

3.13 Ethical and Regulatory Compliance

The study used only market and financial data that is publicly available. No private or confidential data from investors or trading systems was accessed for the study. Additionally, no social media or sentiment data was accessed.

The framework that was proposed does not include any automated trading systems or financial advisors. Any trading signals that were simulated using the framework were for academic and research purposes only.

Although the use of public data minimizes data-related ethical concerns, the deployment of such a trading system would be subject to additional legal and regulatory review. Such a system would have to comply with the rules and regulations set by the Saudi Capital Market Authority, among other regulatory bodies.

3.14 Chapter Summary

This chapter presented the proposed methodology that would be used to develop the hybrid artificial intelligence framework that would be utilized to generate trading signals for the Saudi Stock Market. The details of the data that would be used, how it would be prepared, and the methods by which the models would be formed and combined were detailed in this chapter.

The signals that would be used would be based upon technical indicators and financial ratios. Two separate machine learning models would be used to create the signals: an XGBoost model and a BiLSTM model. The output probabilities of these models would be combined using a weighted ensemble model.

Additionally, the methodology that was used to test and evaluate the proposed framework is outlined in this chapter. The criteria that would be used to determine the success of the framework, as well as ethical and regulatory considerations were also discussed in this chapter. The following chapter will present the results of the tests performed on the framework, as well as a discussion of the results regarding the predictive abilities of the models and their trading performance.

Chapter 4: Experimental Results and Discussion

4.1 Experimental Setup

The experiments used daily market and annual financial statement data for 20 companies on the Saudi Stock Market (Tadawul). The raw data contained 24,966 observations. However, after data preprocessing and cleansing, the final dataset contained 19,742 observations.

The dataset was divided into training and out-of-sample testing data using a chronological 80-20 split. The training data contained 15,790 observations, while the out-of-sample testing data contained 3,952 observations. Furthermore, to preserve the time-series structure of the financial data, the data was not randomly shuffled during division. The data had positive class rates of 33.55% for the training data and 25.78% for the out-of-sample testing data.

Three different models were used during the experiments: the XGBoost model, the BiLSTM model, and a weighted hybrid model. The hybrid model combined the probability outputs from the XGBoost and BiLSTM models using weights of 0.60 and 0.40, respectively.

The trading performance of the models was evaluated using backtesting with six different thresholds for the predicted probability of the market going up for the trading day: 0.50, 0.55, 0.60, 0.65, 0.70, and 0.75. A trade was executed if the probability of the market going up was above the threshold. Each side of a trade cost 0.15% of the traded volume, making the total transaction cost for a completed trade 0.30%.

The models were evaluated using several different metrics. Classification metrics used to assess the models included accuracy, precision, recall, F1-score, and area under the curve (AUC). Additionally, trading-oriented metrics were used to evaluate the trading performance of the models. These metrics included the number of trades, the win rate of the trades, the total return on the trades, the Sharpe ratio of the trading strategy, and the maximum drawdown of the trading strategy.

4.2 Baseline Model Performance

Two standalone models were employed as baseline models for comparison with the proposed model. The first of these models was XGBoost, which was trained on the feature set consisting of the technical indicators and financial ratios. The second model was the BiLSTM model, which was trained on sequences of the engineered features of length 60 days.

These standalone models were evaluated using both classification and trading metrics. Classification metrics help to determine the general performance of the models in the classification task. Trading metrics, however, can help determine the usefulness of the trading signals that can be obtained from the models by applying thresholds to the classification probabilities of the models.

The results of these models indicate that both standalone models tend to produce more trading signals than the hybrid model when evaluated at various probability thresholds. Additionally, however, each of these models exhibits a lower risk-adjusted performance when compared to the hybrid model. Furthermore, the XGBoost model exhibits unstable results at different probability thresholds, while the BiLSTM model is more stable but still experiences lower risk-adjusted performance relative to the hybrid model.

Table 4.1: Backtesting Performance Across Probability Thresholds

Model	Threshold	Trades	Win Rate	Total Return	Sharpe Ratio	Max Drawdown
XGBoost	0.50	141	32.62%	-1.13%	-0.04	-24.64%
BiLSTM	0.50	93	33.33%	4.27%	0.22	-21.58%
Hybrid	0.50	111	35.14%	18.87%	0.81	-15.03%
XGBoost	0.55	130	32.31%	-3.83%	-0.14	-21.67%
BiLSTM	0.55	89	33.71%	6.47%	0.35	-17.39%
Hybrid	0.55	83	38.55%	34.27%	1.93	-15.24%
XGBoost	0.60	113	33.63%	7.27%	0.31	-19.09%
BiLSTM	0.60	79	35.44%	15.47%	0.93	-13.40%
Hybrid	0.60	56	37.50%	19.37%	1.63	-11.04%
XGBoost	0.65	90	35.56%	18.17%	0.96	-21.14%
BiLSTM	0.65	70	38.57%	29.17%	1.95	-11.09%
Hybrid	0.65	43	41.86%	28.27%	3.04	-10.98%
XGBoost	0.70	68	35.29%	12.77%	0.89	-18.94%
BiLSTM	0.70	65	35.38%	12.67%	0.93	-13.29%
Hybrid	0.70	34	38.24%	13.97%	1.93	-6.74%

XGBoost	0.75	54	29.63%	-11.03%	-1.02	-26.24%
BiLSTM	0.75	61	36.07%	14.87%	1.16	-13.03%
Hybrid	0.75	21	42.86%	15.87%	3.50	-4.55%

As shown in Table 4.1, the hybrid model generally had a better return relative to risk across the tested threshold levels. At the lower thresholds, the hybrid model executed more trades and returned more on its trades than the standalone models. As the threshold level increased, though, the number of trades that were executed by the hybrid model decreased.

The highest risk-adjusted return for the hybrid model occurred at a threshold of 0.75. At this threshold, the hybrid model executed 21 trades, achieving a win rate of 42.86%, a total return of 15.87%, a Sharpe ratio of 3.50, and a maximum drawdown of -4.55% . Although the return was less than the return of the hybrid model at some of the lower thresholds, the drawdown was significantly less and the Sharpe ratio was higher at this threshold.

The results indicate that the main advantage of the hybrid model is its trade selectivity. The higher thresholds of the model lead to it executing fewer trades, indicating that it is more conservative in its trade selection at higher threshold levels.

4.2.1 Interpretation of Classification Performance

The classification performance of the models should be interpreted carefully in the context of signal generation for trading. Beyond classification accuracy, there is also an interest in discovering trading opportunities that can generate favorable returns.

Even with moderate classification performance, the model can be helpful if the predicted probabilities can help rank the trading opportunities. Furthermore, only trades with a predicted probability above a threshold were executed with the model. Thus, the model was evaluated not for its classification accuracy, but for whether the high predicted probabilities were associated with profitable and risk-controlled trades.

Overall, then, the value of the hybrid model is not in its classification performance. Instead, the model fuses the predictions of the XGBoost and BiLSTM models to provide an improved selection of trades to execute. Thus, the model yields stronger

risk-adjusted returns despite relatively low performance in classification accuracy alone.

4.3 Hybrid Model Performance

The hybrid model is a combination of the probability outputs from the XGBoost and BiLSTM models with weights of 0.60 and 0.40, respectively. Each of these models was valued for its different strengths. The XGBoost model was valued for its ability to learn the non-linear relationships between the technical and fundamental market model features. The BiLSTM model was valued for its ability to learn the sequential patterns within the historical features of those markets.

The results of the hybrid model indicate that it experienced stronger risk-adjusted performance compared to either of the standalone models. Furthermore, the results were strongest at the higher probability thresholds, which indicated that the model was more selective in its trading decisions.

At the 0.75 probability threshold, the hybrid model achieved the highest Sharpe ratio and the lowest maximum drawdown of any of the models evaluated. Thus, at this threshold, the model was identifying the best trading opportunities for the market, while eliminating the worst potential opportunities for the market to lose money.

Table 4.2: Best Threshold Performance Comparison

Model	Threshold	Trades	Win Rate	Total Return	Sharpe Ratio	Max Drawdown
XGBoost	0.75	54	29.63%	-11.03%	-1.02	-26.24%
BiLSTM	0.75	61	36.07%	14.87%	1.16	-13.03%
Hybrid	0.75	21	42.86%	15.87%	3.50	-4.55%

Table 4.2 reports the performance of the models at the 0.75 probability threshold. The hybrid model made the fewest number of trades (only 21 trades compared to 54 for XGBoost and 61 for BiLSTM). However, the hybrid model had the highest win rate for its trades (42.86%) and the highest Sharpe ratio of any model (3.50).

Furthermore, the hybrid model had the lowest maximum drawdown of any model evaluated (-4.55% compared to -26.24% for XGBoost and -13.03% for BiLSTM). Thus, the model was successful at minimizing the risk of the trading algorithm for the market. Although the return of the strategy using the hybrid model was only slightly higher than that of the BiLSTM model, the risk-adjusted performance was the best for the hybrid model.

Overall, then, the results of the strategy utilizing the hybrid model suggest that the main strength of the strategy is its ability to improve the quality of the trades executed by the algorithm rather than the quantity of those trades. at a threshold of 0.75.

4.3.1 Risk-Adjusted Trading Performance

Risk-adjusted performance in trading is particularly important in that an algorithm or trading strategy that generates high returns can also generate high levels of drawdown for the trading investor. Thus, alongside return and win rate measurements, risk-adjusted measurements like the Sharpe ratio and maximum drawdown are also utilized in the evaluation of the trading strategy.

The results of the model at the 0.75 threshold reveal that the model achieved its best risk-adjusted performance. At this threshold, the model achieved a Sharpe ratio of 3.50, and its maximum drawdown was only -4.55% . These two values indicate that the model achieved a more stable return rate than the standalone models.

The reason for the improved risk-adjusted performance of the hybrid model with higher thresholds is that the model executes fewer trades. At higher thresholds, the model is forced to have higher confidence in its trading signals, which eliminates trades with lower confidence in the model's abilities.

Due to the model only executing 21 trades at the 0.75 threshold, further tests would be beneficial in testing the model with additional time periods, additional companies, and under additional market conditions.

4.4 Key Insight

The major insight from the results of these experiments was that the trading performance was improved by the signal selectivity of the hybrid model. The model did not necessarily increase the number of trades that it executed, or even maximize the accuracy of its trade signals. Rather, the model improved the quality of the trades that it executed.

At higher thresholds for the model's probability of a market movement, the model executes fewer trades, but improves its risk-adjusted performance. Thus, at the 0.75 threshold for probability, the model maximized its risk-adjusted performance.

The model, therefore, indicates a selective trading signal framework for traders and investors utilizing the model's outputs. By setting thresholds for the probability of market movements based on the model's calculations, investors can have a trading model that only executes trades with strong signals for potential returns.

4.5 Threshold Sensitivity Analysis

The effect of the probability thresholds on the performance of the models is depicted

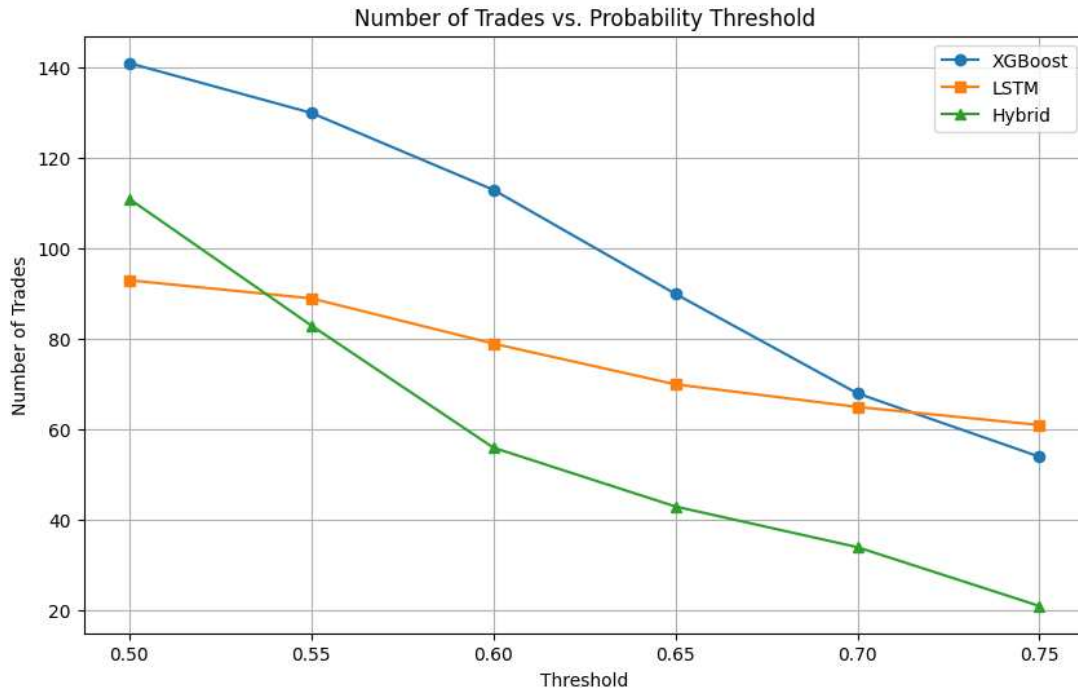


Figure 4.7: Number of Trades vs Probability Threshold

Threshold sensitivity analysis was performed to determine the effect of changing the probability threshold on the number of trades executed and the trading performance of each model. The thresholds tested were 0.50, 0.55, 0.60, 0.65, 0.70, and 0.75.

The results of the threshold sensitivity analysis generally demonstrate that increasing the threshold for the probability of an accurate trade leads to a reduction in the number of trades executed by the trading models. This is due to the fact that higher thresholds require the trading model to be more confident in the likelihood of the trade being profitable to execute that trade.

For the hybrid model, increasing the threshold from 0.50 to 0.75 reduced the number of trades executed from 111 to 21. However, the maximum drawdown

decreased from -15.03% to -4.55% , and the Sharpe ratio increased from 0.81 to 3.50. Thus, while the model executed fewer trades at higher thresholds, it also reduced its exposure to losses and improved its risk-adjusted performance.

However, the total return achieved by the model was not consistently high at higher thresholds. For example, while the model achieved the highest total return at a threshold of 0.55, its best risk-adjusted performance was achieved at a threshold of 0.75. Thus, while the goal of this study was not to create a model that achieved the highest possible return for its traders, but instead to find a balance between the potential return of the trades executed and the level of risk associated with those trades, the threshold of 0.75 emerged as the strongest indication of performance for the hybrid model.

Table 4.3: Threshold Sensitivity Analysis Results for the Hybrid Model

Threshold	Trades	Win Rate	Total Return	Sharpe Ratio	Max Drawdown
0.50	111	35.14%	18.87%	0.81	-15.03%
0.55	83	38.55%	34.27%	1.93	-15.24%
0.60	56	37.50%	19.37%	1.63	-11.04%
0.65	43	41.86%	28.27%	3.04	-10.98%
0.70	34	38.24%	13.97%	1.93	-6.74%
0.75	21	42.86%	15.87%	3.50	-4.55%

As displayed in Table 4.3, the number of trades executed by the trading model decreased as the threshold for the probability of accurate trades increased. At a threshold of 0.50, the model executed 111 trades, but at a threshold of 0.75, the model only executed 21 trades.

The threshold that provided the highest total return for the model was 0.55, which returned 34.27% for the hybrid model. At this threshold, however, the maximum drawdown for the model was -15.24% . At a threshold of 0.75, the model returned a total of 15.87%, but achieved a Sharpe ratio of 3.50 and a maximum drawdown of -4.55% .

These results help to indicate the trade-offs that a trader must make in selecting a threshold for the model. While a higher threshold will lead to fewer trades, a

higher return for those trades, and a reduced drawdown, the threshold that optimizes each of these factors will differ. Thus, while the threshold of 0.75 produced the highest risk-adjusted return for the model, a threshold of 0.55 will produce the highest total return. Because the goal of this study was to develop a model that incorporated an element of risk control into the trade signal generation, however, the threshold of 0.75 emerges as the strongest performance indicator for the hybrid model.

4.6 Risk-Oriented Perspective

The outcome of the threshold sensitivity analysis demonstrates the importance of evaluating trading models from a risk-oriented perspective. A trading model that offers high returns may not necessarily be desirable if it comes with unstable returns.

The hybrid model presented returns and drawdowns at different threshold levels. At the 0.55 threshold, the model returns the highest total return of 34.27% with a maximum drawdown of -15.24%. At the 0.75 threshold, the total return drops to 15.87% with a maximum drawdown of -4.55% and a Sharpe ratio of 3.50.

The 0.75 threshold level offers a more conservative trading model with a smaller number of trades to execute but with higher confidence in the trades identified by the model. Hence, the risk-oriented perspective of the hybrid trading model is its highest contribution to the literature.

As discussed above, the preference of investors is typically toward trading models that offer stable returns and limited drawdowns over trading models that offer higher returns but with greater risk.

Table 4.4: Risk-Return Trade-off for the Hybrid Model

Interpretation	Max Drawdown	Sharpe Ratio	Total Return	Threshold
Highest total return, higher risk	-15.24%	1.93	34.27%	0.55
Best risk-adjusted performance	-4.55%	3.50	15.87%	0.75

Table 4.4 demonstrates the trade-off between total return and risk-adjusted performance for the hybrid trading model. Although the 0.55 threshold offers the highest total return, the 0.75 threshold offers a more risk-controlled trading model.

Thus, the 0.75 threshold is the preferred threshold if risk-adjusted performance is the primary focus for the trading model.

4.7 Impact of Transaction Costs

Transaction costs were included in the backtesting framework to provide a more realistic estimate of the model's performance. A transaction cost of 0.15% was applied to each side of the trade. Thus, each trade incorporated a transaction cost of 0.30%.

Including transaction costs helps investors understand the model's profitability before costs are deducted from the return. As models generate many trading signals, they may not be profitable after transaction costs are deducted. For short-term trading models, transaction costs can significantly impact performance because there will be many trades executed within a period.

The results of the model with a threshold of 0.75 still exhibit good performance with transaction costs included. The model returns a total return of 15.87% with a Sharpe ratio of 3.50 and a maximum drawdown of -4.55%. The performance of the model with fewer trades compared with other models indicates that the transaction costs have less of a negative impact on the model's performance.

Models that execute more trades have a significant negative impact from transaction costs. Thus, using a threshold to filter the signals can significantly improve the trading model's usefulness in practice.

Table 4.5: Transaction Cost Assumption in the Backtesting Framework

Application	Rate	Cost Component
Deducted when entering a trade	0.15%	Entry transaction cost
Deducted when closing a trade	0.15%	Exit transaction cost
Deducted from each completed trade return	0.30%	Total round-trip cost

Table 4.5 displays the transaction cost that was assumed for the testing of the trading model. This transaction cost was used for all models and thresholds to ensure a fair comparison between the standalone models and the hybrid models.

4.8 Equity Curve and Trading Dynamics

The equity curve is used to visualize the cumulative performance of the trading strategies over the sequence of executed trades. An equity curve provides insight into the stability of returns, the depth of drawdowns, and the overall progression of the trading strategy.

The results presented exhibit that the hybrid model produces a more stable trading profile than the standalone models. At higher probability thresholds, the hybrid model trades fewer positions, avoiding lower-probability trading signals. This results in a smaller maximum drawdown and a smoother cumulative return pattern.

The equity curve results substantiate the findings presented in the study that the trading performance was enhanced by the selectivity of the trading strategy. By utilizing the probabilities from both the XGBoost and BiLSTM models, the framework was able to select trades more effectively, reducing the number of trades with lower probabilities of successful outcomes.

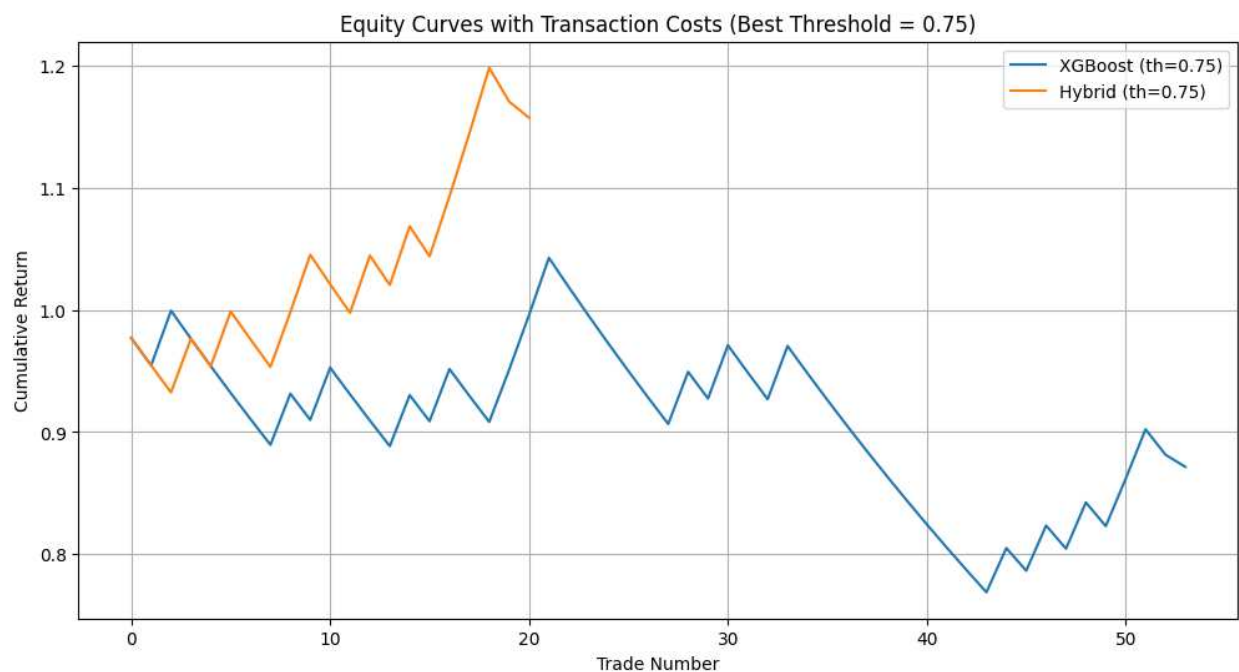


Figure 4.6: Cost-Adjusted Equity Curves for XGBoost and Hybrid Models
(Threshold = 0.75)

Furthermore, the volatility of returns for the Hybrid model is significantly reduced in comparison to those of the other two models. The smooth upward trend indicates consistency in returns and effective risk control, which is critical in real-world trading applications

4.9 Distribution of Trade Returns

The distribution of trade returns for the modelling techniques reflects the predefined trading structure implemented within the backtesting framework. All trades included in the evaluation used a take-profit level of +5% and a stop-loss level of -2% after accounting for the transaction costs of 0.1%.

The return distribution for the hybrid model indicated that the number of winning trades was higher than the number of losing trades at the threshold of 0.75. At this threshold, the win rate of the hybrid model was 42.86%, which was higher than both the XGBoost and BiLSTM models at the same threshold level. Although the win rate of the model was below 50%, the reward-to-risk ratio worked in favor of the model such that the winnings from winning trades were higher than the loss from losing trades. This is how the model was able to generate a positive return on investment with fewer trades. Additionally, the model reduced the cumulative loss for the trading period by minimizing the number of trades with low levels of confidence. This resulted in a lower maximum drawdown for the hybrid model.

Figure 4.8 presents the distribution of trade returns for the hybrid model at the optimal threshold.

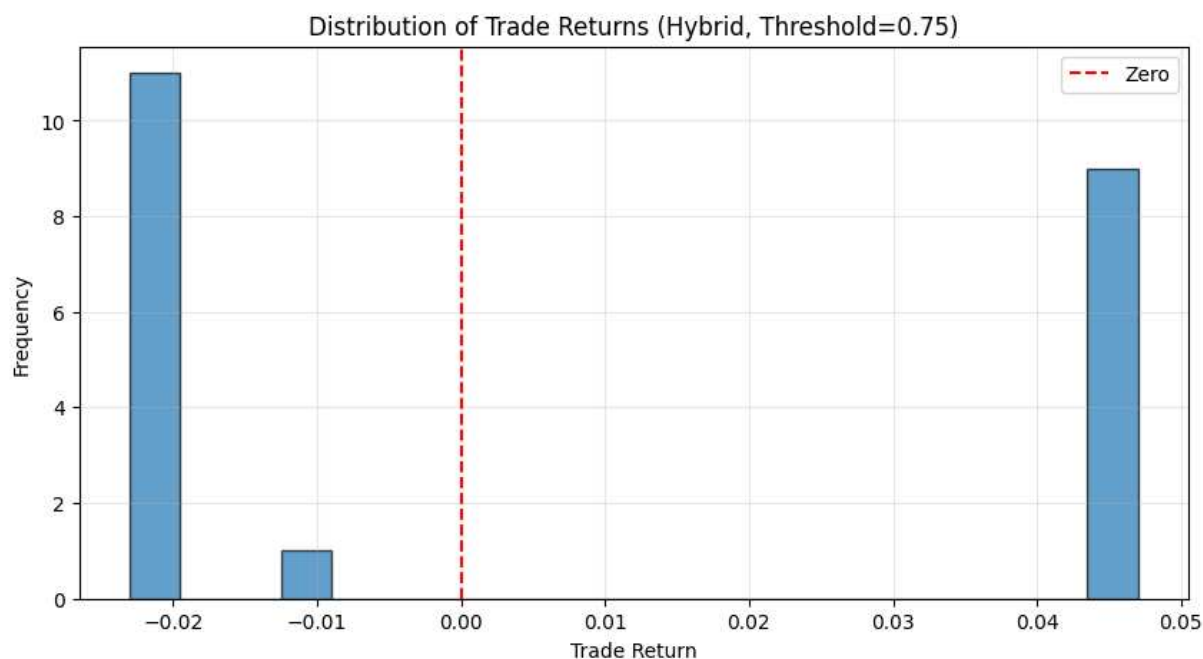


Figure 4.8: Distribution of Trade Returns for Hybrid Model (Threshold = 0.75)

The distribution reflects the rule-based structure of the strategy:

- Losses are concentrated near the stop-loss region
- Gains cluster near the take-profit level

This confirms that the trading logic is effectively enforced.

4.10 Mechanism Behind Hybrid Model Superiority

The reason the hybrid model outperformed is due to the nature of the two different models used within the framework. The XGBoost model learned the non-linear relationships of the features, while the BiLSTM model learned the patterns in rolling windows of the data.

The two models combined their probabilities to determine if a trade would be made, only if the threshold probability was reached. Thus, fewer trades were made by the hybrid model than by either of the individual models, but its performance within the trading signals was the best. Specifically, when compared to the threshold of 0.75, the hybrid model only made 21 trades, as opposed to 54 by XGBoost and 61 by BiLSTM. However, the hybrid model still achieved the

highest win rate, the highest Sharpe ratio, and the lowest maximum drawdown of the three models tested.

Therefore, the reason for the superiority of the hybrid model was mainly due to the selectivity of the trades that the model made for its strategies. The model did not find superiority in making more trades, but rather in making fewer trades with lower risk of losing those trades.

Overall, this result helps to highlight the importance of ranking and filtering of trading opportunities relative to the accuracy in classifying the direction of the financial market. Thus, by selecting only the best trades made by each model, the trading strategy produced the best performance within the timeframe of the backtest.

4.11 Economic Interpretation

The economic interpretation of the results of the study demonstrates that the incorporation of both technical and fundamental analysis could improve the overall quality of generated trading signals. Technical indicators provide valuable information regarding the price movement of stocks. On the other hand, fundamental indicators provide information regarding the financial health of the company.

By combining the two analyses, the model avoids the risk of generating trading signals based on either strong technical indicator alone or strong fundamental indicators alone. A trading signal is only generated if a stock has strong technical and fundamental indicators, thus eliminating the risk of trading with a company with strong financial indicators but poor short-term trading opportunities, or with strong short-term trading opportunities but poor financial standing for the company.

Additionally, the use of a higher threshold for the probability of a trading signal results in fewer trading opportunities but increases the selectivity of the trading strategy. Such a threshold limits the risk of large drawdowns in the trading strategy portfolio.

Based on the data, a threshold of 0.75 for the hybrid model resulted in the best risk-adjusted strategy, with a Sharpe ratio of 3.50 and the lowest maximum drawdown of -4.55% . The results indicate that this trading strategy is best used as

a decision-support tool for traders rather than for the automated generation of trading signals.

4.12 Limitations

Although the results of the study are favorable, there are a few limitations to the research.

First, the study was limited to 20 companies on the Saudi Stock Market. These companies represented a variety of sectors but were not the complete list of companies listed on the market. Therefore, the results cannot be generalized to other companies on the exchange.

Second, all data referenced in the trading strategy between 2021 and 2025 was based on historical data. In the future, market conditions will likely change, and thus, the results of this strategy may not be valid for those markets in the future.

Third, the backtesting strategy used was simplified. Although transaction costs were accounted for, markets also feature bid-ask spreads, liquidity, slippage, and other factors not included in the trading model.

Fourth, the trading model only uses market and financial data. Other data that may affect the price of stocks, such as news and social media sentiment, macroeconomic indicators, and investor behavior data is not used in this model.

Fifth, although the risk-adjusted strategy bested at 21 trades with a probability threshold of 0.75, the limited number of trades presents a challenge in that the model might not be the best strategy for traders who require a higher number of trades to profit from the stocks.

Finally, the trading model exists only as an academic exercise to assist traders in making better trading decisions. A live automated trading system would require additional development beyond the creation of this model.

Table 4.6: Study Limitations and Their Implications

Implication	Limitation
Results may not generalize to all listed companies	Sample limited to 20 Tadawul companies
Performance may differ under future market conditions	Historical period limited to 2021–2025
Live trading performance may be affected by execution frictions	Simplified backtesting framework
Some market-moving information may be excluded	No sentiment or macroeconomic data
Results require cautious interpretation and further validation	Small number of trades at 0.75 threshold
Not equivalent to a production trading system	Academic decision-support design

Table 4.6 presents a summary of the limitations of the study and their implications. While these limitations may invalidate the results of the research study to a certain extent, they define the conditions under which the research study’s findings are valid.

4.13 Practical Deployment

The framework can be used to scan the listed companies and identify trading opportunities that meet the specified probability thresholds.

The outputs of the framework are probabilities rather than the typical trading signals of up or down arrows. The probabilities make it possible to adjust the threshold for trading activities to accommodate different risk preferences. For instance, a lower threshold will result in more trading opportunities but with higher potential returns. A higher threshold will result in fewer trading opportunities but with more control over the risk of trading.

In order to be practically deployed, the framework would need to include additional components related to trading activities. For instance, live market data feeds and real-time order execution systems will be required. Additional components would include risk management controls, trading filters, and compliance controls. The framework would need to account for the various aspects of live markets such as trading spreads, trading delays, and slippage.

Though the framework presents potential methods of producing trading signals with high accuracy, the current framework is merely an academic prototype of a potential trading system. Testing and further validation would be required prior to the framework’s implementation in live markets for trading activities.

Table 4.7: Practical Deployment Requirements

Purpose	Requirement
To update market and financial data automatically	Real-time data pipeline
To detect missing, delayed, or incorrect data	Data validation module
To avoid trades in illiquid stocks	Liquidity filter
To control position size and portfolio exposure	Risk management rules
To improve execution realism	Slippage and spread modeling
To track model signals, trades, and performance	Monitoring dashboard
To ensure suitability before practical deployment	Regulatory review

Table 4.7 presents the requirements for the framework to be practically deployed and used in live markets for identifying trading activities.

4.14 Chapter Summary

This chapter presents the results of the experiments performed on the proposed AI trading model. The results indicate that the proposed hybrid model outperformed both the XGBoost and BiLSTM models alone in terms of risk-adjusted performance.

An analysis of the threshold levels revealed that increasing the threshold reduced the number of trades executed by the model, with higher thresholds leading to fewer trades. The highest risk-adjusted performance for the hybrid model was achieved with a threshold of 0.75, yielding a Sharpe ratio of 3.50 and a maximum drawdown of -4.55% .

The main advantage of the proposed model was its ability to focus on fewer, higher-quality trades by eliminating lower-confidence signals. Thus, the model executed fewer trades at higher threshold levels but experienced better control of downside risk.

Finally, this chapter discusses transaction costs, the equity curve, the economic aspects of the trading strategy, the limitations of the study, and considerations for deploying the strategy in live trading. The following chapter will conclude the thesis and present a summary of the findings, contributions, limitations, and future research for this topic.

Chapter 5: Conclusion and Future Work

5.1 Conclusions

This thesis has developed and evaluated a framework for identifying trading signals in the Saudi Stock Market based on technical and fundamental analysis, and has tested whether incorporating machine learning models can improve the performance of that framework. Two machine learning models, XGBoost and BiLSTM, were employed as means of determining trading signals from market data, and their results were combined through a weighted ensemble model.

The results of the testing show that the hybrid model that combined technical, fundamental, and machine learning analyses generally outperformed each of the standalone machine learning models in terms of risk-adjusted returns. The hybrid model achieved its highest performance figures at a probability threshold of 0.75, with 21 trades, a win rate of 42.86%, a total return of 15.87%, a Sharpe ratio of 3.50, and a maximum drawdown of -4.55% . These results indicate that the main advantage of the hybrid model relative to the individual machine learning models was that it reduced the number of trades that were executed by the trading framework, which had the potential to reduce the negative impact of trading errors.

Thus, these results help to suggest the importance of utilizing different evaluation metrics for trading signal strategies than those that are typically applied to artificial intelligence models. Beyond evaluating the accuracy of predictions made by the AI models, it is also important to evaluate the financial and risk-adjusted returns that can be achieved with those strategies.

Thus, this research indicates the potential use of AI models to identify trading signals based on both technical and fundamental analyses. However, those findings are made within the assumptions of the study itself.

5.2 Theoretical Contribution

This study makes several contributions to the existing literature on AI models for predicting the stock market. First, one contribution to the literature is modeling the outcomes of trades rather than the prices of the stocks within those trades. More specifically, instead of determining the probability that the price of a stock will

reach certain values within the future, the methods of this study determine the probability that a trade will reach its take-profit target before its stop-loss target.

Second, this study contributes to the literature by incorporating technical and fundamental analysis into the AI framework for determining trading signals. Each of these analysis methods has been utilized separately by other authors in the literature, but few have attempted to utilize both within the same framework.

Third, in general, this study contributes to the literature by utilizing different evaluation metrics for the trading signal models than those that are typically applied to artificial intelligence models. Thus, each model is evaluated according to its financial returns, as well as according to its risk-adjusted returns.

Finally, one of the main ways in which this study contributes to the literature is in applying the framework to the Saudi Stock Market (Tadawul), one of the world's emerging markets. Few studies have attempted to apply AI-based trading signal models to these markets, especially given potential differences in the characteristics of the stocks within those markets relative to more developed economies.

5.3 Practical Implications

The findings of this research have several practical implications. First, one of the main implications is that AI-based trading models may be more useful as tools that provide recommendations to traders regarding the trades that should be executed, rather than as automated trading models that directly execute those trades. The ability to utilize probability values to determine the confidence of the models' predictions allows traders to control the level of confidence required from trading signal models before they begin to execute trades.

Second, the findings suggest that utilizing higher thresholds for the probabilities of trades can reduce the frequency of the trades that are executed by a trading model, as well as reducing the risk of large drawdowns. This is useful for traders who wish to employ trading strategies that differ from those applied to more heavily traded stocks.

Third, the findings support the value of incorporating both technical and fundamental analyses into trading signal models. Each of these analyses can provide traders with different information regarding the stocks considered for

trading, allowing for the creation of trading signal models with higher overall value than could be achieved by using only one type of analysis.

Finally, another implication is that methods based on this study can be used as the basis for creating screening tools for the stocks within the Saudi Stock Market. Such screening tools could help traders identify stocks that meet certain criteria defined by the trading signals models.

5.4 Implementation Constraints

Despite the promising results achieved by the proposed framework, there are some implementation constraints that must be considered before the model can be used in practice. The current model was created and tested as an academic project using historical data. Thus, it is not a complete system that could be implemented into a live trading environment.

To implement the current framework, it would be necessary to establish a system for collecting real-time market data, processing the data to clean it of any invalid entries, determining trading signals according to the hybrid model, automatically executing trades according to those signals, and tracking the performance of the model over time. Furthermore, considerations would be required for market conditions such as liquidity, spreads, slippage, and depth.

Another consideration for implementation would be maintenance of the model. Since markets are constantly changing, the relationships that were learned by the model may change over time. Therefore, the model would have to be periodically retrained to adapt to these changes.

Finally, the framework would have to be legally and operationally reviewed before its use in the real world; it should not be directly used as an automated trading or investment management system.

5.5 Generalizability of the Framework

Currently, the framework was created and tested on 20 different companies that trade on the Saudi Stock Market. Furthermore, these 20 companies are from a variety of different market sectors. Thus, while the results may be generalizable to

other companies on the Saudi Exchange, it is not yet possible to generalize the results to companies from other markets.

Moreover, it is also likely that the framework will not generalize to smaller companies, stocks with less liquidity, newer companies, or different periods within the markets. Each of these factors may have an impact upon the performance of the proposed trading system.

However, the structure of the algorithm itself is generalizable to other markets. The same steps can be applied to stocks from other markets across the world by collecting the appropriate data and running the model. Thus, another contribution of this research is the creation of a framework that can be replicated on other markets.

5.6 Future Research Directions

As mentioned within the paper, there are several different directions for future research relating to the framework that was created. One recommendation is to test the framework on an even larger number of companies, especially those that are smaller and have less liquidity than the companies that were tested within this research.

Another future research study would be to test the model over different periods within the markets to determine whether the framework performs equally well in bullish, bearish, or sideways markets.

Additionally, the backtesting system could be improved to simulate more realistic trading conditions that would impact live trading decisions.

Future research could also investigate alternative methods to integrating the individual models within the framework. For instance, instead of the weighting method that was implemented for each model, other methods could be developed to determine the contribution of each model relative to the others.

Another suggested future research study would be to incorporate additional data sources into the model. Such additional data could include information related to the economy, market sectors, market news, and investor behavior. However, such data would have to be carefully incorporated into the framework.

Finally, creating a live system that implements the framework and displays various metrics for live trading decisions could also be developed. Such a system would display indicators related to the stocks, trading signals, and other metrics that were collected during live trading sessions.

5.7 Final Remarks

In this thesis, a framework for trading signal generation using artificial intelligence in the Saudi Stock Market is proposed. The framework uses technical indicators, fundamental financial ratios, and two machine learning models to create trading signals for the Saudi Stock Market that can be evaluated using multiple metrics relating to financial performance and risk.

The results of the simulations indicate that the proposed framework outperformed the individual machine learning models for the scenarios considered in this thesis. Furthermore, the model filters trading signals to focus on the most profitable trading opportunities.

The research presented in this thesis presents a framework for evaluating trading signals that include technical and fundamental analysis in an artificial intelligence model. Additionally, this thesis suggests areas for future research in the area of artificial intelligence and trading signals for markets like the Saudi Stock Market.

Although the results of this thesis are promising, they are only simulated academic results. Future research and deployment of this framework should account for larger datasets and more realistic trading scenarios.

References

- Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H.-T. (2012). Learning from data. AMLBook.
- Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model. In Proceedings of the UKSim-AMSS 16th International Conference on Computer Modelling and Simulation (pp. 106–112). IEEE.
- Ahmed, N. K., Atiya, A. F., El Gayar, N., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5–6), 594–621. <https://doi.org/10.1080/07474938.2010.481556>
- Alhazmi, A. H. J., Islam, S. M. N., & Prokofieva, M. (2025). The impact of artificial intelligence adoption on the quality of financial reports on the Saudi Stock Exchange. *International Journal of Financial Studies*, 13(1), 21. <https://doi.org/10.3390/ijfs13010021>
- Al-Nefaie, A. H., & Aldhyani, T. H. H. (2022). Predicting close price in emerging Saudi stock exchange: Time series models. *Electronics*, 11(21), 3443. <https://doi.org/10.3390/electronics11213443>
- Alsulmi, M., & Al-Shahrani, N. (2022). Machine learning-based decision-making for stock trading: Case study for automated trading in Saudi Stock Exchange. *Scientific Programming*, 2022, Article 6542862. <https://doi.org/10.1155/2022/6542862>
- Ansari, Y., Gillani, S., Bukhari, M., Lee, B., Maqsood, M., & Rho, S. (2024). A multifaceted approach to stock market trading using reinforcement learning. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3418510>
- Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long short-term memory. *PLOS ONE*, 12(7), e0180944. <https://doi.org/10.1371/journal.pone.0180944>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Capital Market Authority. (2024). Market regulations and guidelines. <https://www.cma.org.sa>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge

Discovery and Data Mining (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>

Daoria, H. K., Alharthi, M. E., Alzahrani, G. S., Alanazi, A. E., Aborokbah, M., Alotaibi, M., & Alanazi, T. M. (2024). Saudi market stock price predictions using deep learning algorithms and weighting Twitter sentiment analysis. *University of Ha'il Journal of Science*, 5(1), 30–48.

Daradkeh, M. K. (2022). A hybrid data analytics framework with sentiment convergence and multi-feature fusion for stock trend prediction. *Electronics*, 11(2), 250. <https://doi.org/10.3390/electronics11020250>

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417. <https://doi.org/10.2307/2325486>

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669. <https://doi.org/10.1016/j.ejor.2017.11.054>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

Khurayshi, A., & Alharbi, A. (2025). Saudi Arabia stock prices prediction using machine learning. *Advances and Applications in Statistics*, 92(6), 813–833. <https://doi.org/10.17654/0972361725034>

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689–702. <https://doi.org/10.1016/j.ejor.2016.10.031>

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53, 3007–3057. <https://doi.org/10.1007/s10462-019-09754-z>

Saberironaghi, M., Ren, J., & Saberironaghi, A. (2025). Stock market prediction using machine learning and deep learning techniques: A review. *AppliedMath*, 5(3), 76. <https://doi.org/10.3390/appliedmath5030076>

Saudi Exchange. (2025). Historical market data. <https://www.saudiexchange.sa>

Tsai, C.-F., & Wang, S.-P. (2009). Stock price forecasting by hybrid machine learning techniques. *Expert Systems with Applications*, 36(4), 7551–7559.
<https://doi.org/10.1016/j.eswa.2008.09.029>

Yahoo Finance. (2024). Historical market data. <https://finance.yahoo.com>

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 11106–11115.