جــامـعـة الــشـرق الأوسـط
MIDDLE EAST UNIVERSITY
Amman - Jordan

The Prediction Power of Accounting Information: Evidence from Machine Learning Applications in Stock Price Forecasting for Listed Companies in Amman Stock Exchange.

Prepared by

Ahmad Saad Mohammad Aqra

Supervised by

Professor Dr. Asmaa Ibrahim Al-Amarna

A Thesis Submitted in Partial Fulfillment of the Requirements for the Master's Degree in Accounting

Department of Financial and Accounting Sciences

Faculty of Business

Middle East University

January, 2026

جامعة الشرق الأوسط
MIDDLE EAST UNIVERSITY
Amman - Jordan

القوة التنبؤية للمعلومات المحاسبية : أدلة من تطبيقات التعلم الآلي في توقع أسعار الاسهم للشركات المدرجة في بورصة عمان.

إعداد
احمد سعد محمد عقرة

إشراف
الاستاذة الدكتورة أسماء إبراهيم العمارنه

قدمت هذه الرسالة استكمالاً لمتطلبات الحصول على درجة الماجستير في المحاسبة

قسم العلوم المالية والمحاسبية
كلية الاعمال
جامعة الشرق الاوسط

كانون الثاني، 2026

Defense Committee Decision

This thesis, titled **"The Prediction Power of Accounting Information: Evidence from Machine Learning Applications in Stock Price Forecasting for Listed Companies in Amman Stock Exchange."** by researcher **Ahmad Saad Mohammad Aqra** and was successfully defended and approved on 18/01/2026.

Examination Committee Members

Name	Title	Place	Signature
Prof. Asma'a Ibrahim Al-Amarneh	Supervisor	Middle East University	....
Dr. Ahmad Zuhair Marei	Internal Member and Committee Head	Middle East University	....
Dr. Enas Mahmoud Alabood	Internal Member	Middle East University	....
Dr. Murad Yousef Abu Addous	External Member	Al-Balqa Applied University	....

Authorization

I, **Ahmad Saad Mohammad Aqra**, authorize Middle East University to provide copies of my thesis on paper and electronically, in whole or in part, to libraries, organisations, bodies, and institutions concerned with scientific research and studies upon request.

Name: Ahmad Saad Mohammad Aqra.

Date: 18/01/2026.

Signature:

Acknowledgment

In the name of Allah, the Most Gracious, the Most Merciful. Peace and blessings be upon the noblest of creation and messengers, our master and beloved Prophet Muhammad, and upon his family and companions.

I extend my deepest gratitude to Allah Almighty, whose countless blessings have enabled me to complete this work. To Him alone belongs all praise and thanks, first and last.

I would like to take this opportunity to recognize the scholarly contribution of my supervisor, Professor Dr Asmaa Ibrahim Saleem Al-Amarna, whose support was crucial to the completion of this thesis. The characteristics of this thesis are a reflection of her involvement and ought to be acknowledged. Her efforts and her professional standards, which are a rare to find, helped shape the value of the thesis and finish within the timeline. My wishes are may Allah reward her with the best among the good, and protect, and bless her with righteous success.

I also would like to extend my thanks to the members of the examination committee, and to faculty and staff of Middle East University, for the ease and support on my journey to completing this study.

Ahmad Saad Mohammad Aqra

Dedication

To everyone who believed in me, supported me, and stood beside me throughout this journey.

To every heart that offered kindness, every hand that extended help, and every voice that encouraged me forward.

To all those whose presence—near or far—added strength, patience, and meaning to this path.

To every teacher who guided me,
every colleague who shared the road with me,
and every person who contributed, even silently, to the completion of this work.

To all of you…
I dedicate this modest achievement.

Ahmad Saad Mohammad Aqra

Table of Contents

Table of Tables

Table of Figures

List of Abbreviations

Abbreviation	Full Term
AI	Artificial Intelligence
AMH	Adaptive Markets Hypothesis
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
ASE	Amman Stock Exchange
CNN	Convolutional Neural Network
EFB	Exclusive Feature Bundling
EMH	Efficient Market Hypothesis
EPS	Earnings Per Share
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GBDT	Gradient Boosting Decision Tree
GDP	Gross Domestic Product
GOSS	Gradient-based One-Side Sampling
IFRS	International Financial Reporting Standards
JOD	Jordanian Dinar
LightGBM	Light Gradient Boosting Machine
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
R^2	Coefficient of Determination
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
ROA	Return on Assets
ROE	Return on Equity
SD	Standard Deviation
SHAP	SHapley Additive exPlanations
VIF	Variance Inflation Factor
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

The Prediction Power of Accounting Information: Evidence from Machine Learning Applications in Stock Price Forecasting for Listed Companies in the Amman Stock Exchange

Prepared by
Ahmad Saad Mohammad Aqra
Supervisor by
Professor Dr. Asmaa Ibrahim Al-Amarna

Abstract

This study evaluates whether accounting-based financial ratios retain explanatory and predictive relevance for stock outcomes in the Amman Stock Exchange (ASE) and whether an AI-enhanced framework improves performance under out-of-sample validation. Using a balanced monthly panel of 15 ASE-listed firms observed over 2022–2023 (N = 360) across banking, industrial, and services sectors, we model stock price level, monthly returns, and price direction as functions of profitability (ROA, ROE), liquidity (current and quick ratios), leverage (debt-to-equity, interest coverage), and growth (EPS growth), while controlling for firm size (market capitalization), trading activity (volume), sector dummies, and macroeconomic conditions (inflation, GDP growth, interest rate). The empirical design tests three hypotheses: (H1) direct linear relevance of ratios for stock outcomes; (H2) incremental explanatory value of a non-linear signal extracted by LightGBM in a stacked framework; and (H3) superior out-of-sample performance of the AI-enhanced approach.

Results show a clear outcome-dependent pattern. The direct model provides strong and stable performance for price levels (training R^2 = .927; 5-fold CV R^2 = .916; RMSE = 1.870), indicating substantial accounting-based value relevance for valuation levels in the ASE. In contrast, performance is weak for returns (training R^2 = .134; CV R^2 = .015; RMSE = 0.054) and does not generalize for price direction (training R^2 = .100; CV R^2 = −.021; RMSE = 0.457), consistent with the high noise of short-horizon market movements. LightGBM learns a strong in-sample non-linear signal (training R^2: price level .937; returns .530; direction .449), but cross-validation reveals limited generalization for returns and direction (CV R^2 = −.056 and −.157, respectively), indicating sensitivity to overfitting in a data-constrained setting. Sectoral patterns suggest comparatively better model coherence in industrial firms than in banking and services, while macro controls exhibit limited incremental contribution over the short analytical horizon.

Overall, the evidence supports a pragmatic conclusion: accounting ratios remain highly informative for valuation levels, whereas predicting short-term returns and direction is substantially harder even with advanced ML. AI methods are most useful as signal/feature extractors within robust validation and governance practices, rather than as standalone forecasting engines, supporting hybrid modelling strategies for analysts and policymakers in semi-efficient emerging markets.

القوة التنبؤية للمعلومات المحاسبية: أدلة من تطبيقات التعلم الآلي في توقع أسعار الأسهم للشركات المدرجة في بورصة عمّان

إعداد

أحمد سعد محمد عقرة

إشراف

الأستاذة الدكتورة أسماء إبراهيم العمارنة

الملخص

سعت هذه الدراسة إلى تقييم ما إذا كانت النسب المالية المحاسبية تحتفظ بقيمة تفسيرية وتنبؤية لنتائج الأسهم في سوق عمّان المالي، وإلى فحص ما إذا كان توظيف إطار مدعوم بالذكاء الاصطناعي يُحسّن الأداء عند التحقق خارج العينة. اعتمدت الدراسة على لوحة بيانات شهرية متوازنة شملت 15 شركة مدرجة خلال الفترة 2022–2023(N = 360) ، موزعة على قطاعات البنوك والصناعة والخدمات. وتم نمذجة مستوى السعر والعائد الشهري واتجاه السعر باستخدام نسب الربحية (العائد على الأصول، العائد على حقوق الملكية)، والسيولة (النسبة الجارية والسريعة)، والرافعة المالية (الدين إلى حقوق الملكية، تغطية الفوائد)، والنمو (نمو ربحية السهم)، مع التحكم في حجم الشركة (القيمة السوقية)، ونشاط التداول (حجم التداول)، ومتغيرات القطاع، والمؤشرات الكلية (التضخم، نمو الناتج المحلي، سعر الفائدة). واختبرت الدراسة ثلاث فرضيات (H1) :الأهمية الخطية المباشرة للنسب المحاسبية لنتائج الأسهم؛ (H2) القيمة التفسيرية الإضافية لإشارة غير خطية مستخلصة بواسطة نموذج LightGBMضمن إطار تراكبي؛ (H3) تفوق الأداء التنبؤي خارج العينة للنموذج المعزز بالذكاء الاصطناعي.

أظهرت النتائج نمطًا يعتمد على نوع المتغير التابع. فقد حقق النموذج الخطي المباشر أداءً قويًا ومستقرًا لمستويات الأسعار R^2)داخل العينة = 0.927؛ R^2 بالتحقق المتقاطع خماسي الطيات = 0.916؛ RMSE = 1.870)، بما يعكس أهمية محاسبية عالية لتفسير مستويات التقييم في السوق. في المقابل، كان الأداء ضعيفًا للعوائد R^2)داخل العينة = 0.134؛ R^2 بالتحقق المتقاطع = 0.015؛ (RMSE = 0.054 ولم يُظهر اتجاه السعر قابلية تعميم R^2) داخل العينة = 0.100؛ R^2 بالتحقق المتقاطع = −0.021؛(RMSE = 0.457 ، وهو ما يتسق مع ارتفاع الضوضاء في الآفاق الزمنية القصيرة. وتمكن نموذج LightGBM من تعلّم إشارة غير خطية قوية داخل العينة :R^2) مستوى السعر 0.937؛ العوائد 0.530؛ الاتجاه 0.449)، إلا أن التحقق خارج العينة كشف تراجعًا في القابلية للتعميم بالنسبة للعوائد والاتجاه R^2 = −0.056) و−0.157 على التوالي)، بما يشير إلى حساسية لفرط المواءمة في بيئة محدودة البيانات. كما أظهرت الأنماط القطاعية تماسكًا أفضل نسبيًا في القطاع الصناعي مقارنة بقطاعي البنوك والخدمات، في حين كان الأثر الإضافي للمتغيرات الكلية محدودًا خلال الأفق الزمني القصير.

تخلص الدراسة إلى نتيجة عملية مفادها أن النسب المحاسبية تظل شديدة الأهمية لتفسير مستويات التقييم، بينما يبقى التنبؤ بالعوائد القصيرة الأجل واتجاهات الأسعار تحديًا كبيرًا حتى مع استخدام تقنيات تعلم الآلة المتقدمة. وتُعد أدوات الذكاء الاصطناعي أكثر فاعلية بوصفها مستخرِجات للإشارة والميزات ضمن أطر تحقق صارمة وحوكمة منهجية، لا كأدوات تنبؤ مستقلة، بما يدعم اعتماد استراتيجيات هجينة تجمع بين الاستخلاص الذكي للإشارة والنماذج الإحصائية الأبسط والأكثر استقرارًا في الأسواق شبه الكفؤة الناشئة.

Chapter One
Background and Problem Statement

1.1 Introduction

Forecasting stock prices represented a fundamental problem in financial markets. This problem is particularly the case in developing markets. In developing markets predictive uncertainties are amplified due to market inefficiencies, lack of liquidity, and volatility in the macro-economy (Al-Qaisi & Al-Rdaydeh, 2020). The Amman stock exchange (ASE) is Jordan's leading financial market. It demonstrates the aforementioned problem due to the diverse sectors, the CPU, and the structures that rely on the economy in the region (Hendawi, 2022). Classical financial theories, i.e. Hypothesis EMH, argued that the prices of the stock uniformly reflected all available information. However, the Amman stock exchange markets records revealed a clear inefficiency of a semi-strong level where the anomalies in the accounting ratios indicators: profitability, liquidity, and leverage, in macroeconomic variables, and ratios maintained a predictive level (Al-Zubi et al. 2010, Hendawi 2022). For example, Al-Shiab (2006) showed that ARIMA models predicted the stock prices of ASE with a level of moderate accuracy. In addition, Al-Zubi et al. (2010) indicated that there are ANN that outperformed the linear regression in determining the volume of the trade and the interest rate. The evidences revealed the limitation of the classical methods and the level of the machine that can address the problem of the ASE.

Linear regression and the autoregressive integrated moving average approach still published what could be classified as empirical research in the domain of finance; however, they were beginning to receive quantitatively and qualitatively better critiqued documents published in the domain of finance regarding the linearity and types of data inputs. Guo et al. (2023) and Bozkurtturanyigit (2021) pointed out that these studies were chronically, negatively and positively impacted by black swans (especially ones that the region experienced) and type and class of data disruption, ignoring the abundance of high-dimensional data in the analysis (technicals). As they stated, these types of analyses did not have the capacity to predict on the rapid fluctuations of the value of entire assets across the class and type of data inputs for the model that was used to integrate the technical, and it did not have the capacity to predict the overall value of a cluster of assets in the month or two of the banking sector liquidity crises that rapidly shifted the entirety of the Integrated data. These types of analyses justified the increased portfolios with value predictive potential.

The rise of advanced machine learning algorithms transformed the financial forecasting landscape and provided researchers with frameworks such as Light Gradient Boosting Machine (LightGBM).

LightGBM was a type of gradient-boosting algorithm; every machine learning algorithm of a decision tree. The difference with LightGBM was the way it built and grew trees. In particular, it used a much faster computationally histogram method that led to trees of much greater depth (which is a good thing, as will be discussed). With that approach, it was able to go beyond the trivial leaf-to-leaf construction in a depth-first fashion by using a breadth-first methodology to grow and optimize the trees. This explains LightGBM's sizeable contribution to financial forecasting where sector-specific volatility significantly increased the predictive power of leverage ratios.

The particular circumstances surrounding the ASE demanded innovative solutions. ASE exhibited very low liquidity and was highly dependent on the performance of the economy of the surrounding region, which created circumstances under which traditional methodologies failed (Al-Zubi et al., 2010). For example, in Al-Qaisi and Al-Rdaydeh (2020) ASE banking stocks were attempted to be modeled with a hybrid ARIMA-wavelet model, albeit unsuccessfully. They reported a MAPE of 4.22% which is not a very impressive result and attempted to incorporate volatility jumps in the model. They recognized some of the limitations of the model in forecasting price movements to the extent that a MAPE of the model would be of such a magnitude that it would not possess the exegesis to which it would be entitled. In the case of behavioral finance, the situation was aggravated on the grounds that the analyses of sentiment (which were introduced for unrelated and justifiable reasons) were excluded in the bulk of the AI model building endeavours.

Though significant progress had been achieved, there remained fundamental shortcomings in the literature. First, most of the applications of LightGBM were in developed markets or in Forex datasets, while there was a gap in the literature regarding applications in emerging markets like ASE (Jiatangzhi, 2025; Bozkurtturanyigit, 2021).

Second, there was a gap in the literature on how much LightGBM could improve the relationship between accounting ratios and stock prices, while there was plenty of literature on ensemble models and technical indicators (Hendawi, 2022). For example, very few studies, if any, empirically tested or validated the theory proposed by Cao (2022, as cited in Hendawi, 2022) that there are unobserved relationships among financial variables and models of AI could reveal them. Third, and in the same context, the very few studies that documented this relationship (Hendawi, 2022) did not consider another important problem: the moderating role that industry sectors or macroeconomic variables on the performance of the model.

1.2 Problem Statement

Financial literature suggests that the efficiency of financial markets varies according to factors such as institutional development, market liquidity, and the transparency of financial disclosure. Consequently, in many emerging markets, stock prices do not fully or promptly incorporate available accounting information. Within this context, the Amman Stock Exchange (ASE) provides a relevant and important setting for examining the relationship between accounting information and stock prices, as prior evidence indicates that the market operates at a relatively low level of efficiency. The ASE has been characterized by high volatility, prolonged price stagnation, irregular investor participation, and weak alignment between stock prices and fundamental financial indicators, reflecting a structurally inefficient market environment (Al Ikhabariyah, 2010).

These conditions raise concerns about the adequacy of traditional accounting-based forecasting models in explaining stock price behavior in the ASE. In theory, financial ratios such as return on assets, liquidity ratios, and leverage ratios should provide meaningful signals for stock valuation; however, empirical observations in the ASE contradict this theoretical expectation. This inconsistency represents a clear anomaly when compared to developed markets, where accounting information typically exhibits stronger predictive and explanatory power. As a result, a significant research gap exists regarding whether conventional analytical approaches are sufficient to capture stock price dynamics in the Jordanian capital market (Jansen et al., 2018).

At the same time, advances in artificial intelligence and machine learning have demonstrated substantial success in improving financial forecasting accuracy in developed economies (Kumar et al., 2024; Tian et al., 2022). Despite this progress, such advanced methods remain largely underutilized in developing markets, including the ASE. The limited application of machine learning techniques has created a “black box” for both practitioners and academics, leaving unanswered questions about whether these methods can enhance the predictive power of accounting information in structurally inefficient markets. Moreover, existing studies have rarely examined how artificial intelligence techniques interact with accounting ratios to influence stock prices in the ASE over time, further deepening the research gap (Cao, 2022; Ajiga et al., 2024).

In response to these unresolved issues, this study seeks to investigate whether the application of advanced analytical techniques—specifically the Light Gradient Boosting Machine (LightGBM)—can strengthen and mediate the relationship between accounting information and stock price movements in the Amman Stock Exchange. By addressing both the limitations of traditional models and the underexplored role of machine learning in an emerging market context, the study aims to provide empirical evidence on the extent to which advanced methods can improve the interpretation and predictive power of accounting information in the ASE.

1.3 Study Objectives

1.3.1 Main Aim

This research aimed to determine the degree to which Light Gradient Boosting Machine (LightGBM) mediated the impact of financial statements analysis on the prediction of stock prices in the Amman Stock Exchange (ASE) and to build a stronger prediction model that was designed to tackle the problems specific to this developing market.

1.3.2 Specific Objectives, Questions and Hypotheses

This study sought the integration of Artificial Intelligence (AI) in the field of financial forecasting of the Amman Stock Exchange (ASE) through the handling of a theoretical, analytical, and practical study. The objectives have been designed to be baseline relationships and targeted for predictive value, and for prediction to be understandable and useful in the context of investment decisions.

Research Objective (RO1): To study the predictive power of accounting information with regard to stock price of the listed companies at Amman Stock Exchange (ASE) after controlling for firm-specific characteristics, industry affiliation, and macroeconomic conditions, Al-Haddad; Al-Zoubi, (2023).

This study is looking to evaluate the direct relationship among the financial statement parameters profitability, liquidity, and leverage (control) ratios, and stock price of ASE listed companies after controlling for firm-specific characteristics, industry affiliation, and macroeconomic conditions. This is the most important in terms of setting a value of forecasting with AI (Al-Haddad et al., 2022; Al-Zoubi, 2023).

Research Objective (RO2): To investigate how machine learning application provide incremental explanatory power over traditional accounting-based models in explaining stock prices of ASE listed companies after controlling for firm-specific characteristics, industry affiliation, and macroeconomic conditions. Specifically, to investigate the impact of LightGBM's ability to convert traditional financial ratios to stock price predictive ratios, and the extent to which that enhancing of prediction power (Kumar et al., 2024).

Research Objective (RO3): To compare the explanatory power of advanced analytical models and traditional accounting-based valuation models in explaining stock prices. Cao (2022) as well as Ajiga et al. (2022) focused on where the goal was to create mechanisms of transparency through model interpretability (SHAP and feature importance) in order to explain the pathways of the model's decision and to improve the model's transparency to the analyst as well as the investors.

Achievement of the above objectives go on to provide a dual contribution of enriching the understanding of merger and acquisition in advanced machine learning as well as enhancing the investment forecasting in the emerging market for synthesized financial theory.

1.4. Research Questions and Hypotheses

1.4.1 Research Questions

This study aims to answer the following research questions (RQs) aiming to understand how LightGBM adds value to the prediction of financial statements of firms listed on the Amman Stock Exchange (ASE) and the prediction of stock prices:

RQ1: To what extent does accounting information predict the stock price of companies listed on the Amman Stock Exchange (ASE) after controlling for firm-specific characteristics, industry affiliation, and macroeconomic conditions (Al-Haddad & Al-Zoubi, 2023; Al-Haddad et al., 2022; Al-Zoubi, 2023)?

RQ2: To what extent does applying machine learning—specifically LightGBM—provide incremental explanatory power over traditional accounting-based models in explaining stock prices of ASE-listed companies after controlling for firm-specific characteristics, industry affiliation, and macroeconomic conditions, and how effectively does LightGBM transform traditional financial ratios into stronger predictive signals (Kumar et al., 2024)?

RQ3: How do advanced analytical models compare with traditional accounting-based valuation models in explaining stock prices on the ASE, and to what extent can interpretability mechanisms (e.g., SHAP values and feature importance) enhance transparency by clarifying the pathways underlying model decisions for analysts and investors (Cao, 2022; Ajiga et al., 2022)?

1.4.2 Research Hypotheses

Leveraging existing theory in financial analysis, machine learning, and AI-enhanced forecasting, we advance the following hypotheses to be empirically tested:

H-0-1: Accounting information of companies listed on the Amman Stock Exchange does not have significant explanatory power in predicting stock prices after controlling for firm-specific characteristics, industry affiliation, and macroeconomic conditions.

H0-2: Accounting information of companies listed on the Amman Stock Exchange does not influence stock prices primarily through a non-linear signal extracted by LightGBM, after controlling for firm characteristics, industry sector, and macroeconomic conditions.

H0-3: LightGBM doesn't provide superior predictive performance in forecasting stock prices for listed companies in ASE compared to traditional models.

These hypotheses are grounded in theories of Fundamental Analysis (Graham & Dodd, 1934), Efficient Market Hypothesis (Fama, 1970), and Ensemble Learning Theory from the AI literature (Cao, 2022; Kumar et al., 2024). They reflect an interdisciplinary approach that bridges classical finance with computational techniques to improve stock price forecasting in emerging markets such as Jordan.

1.5.3 Study Design

In Figure 1, the conceptual framework demonstrates the correlation between the financial ratios and the stock price as mediated by the LightGBM model, as well as the industry sector and macroeconomic variables moderating the framework. This figure depicts the direct and indirect relationships.

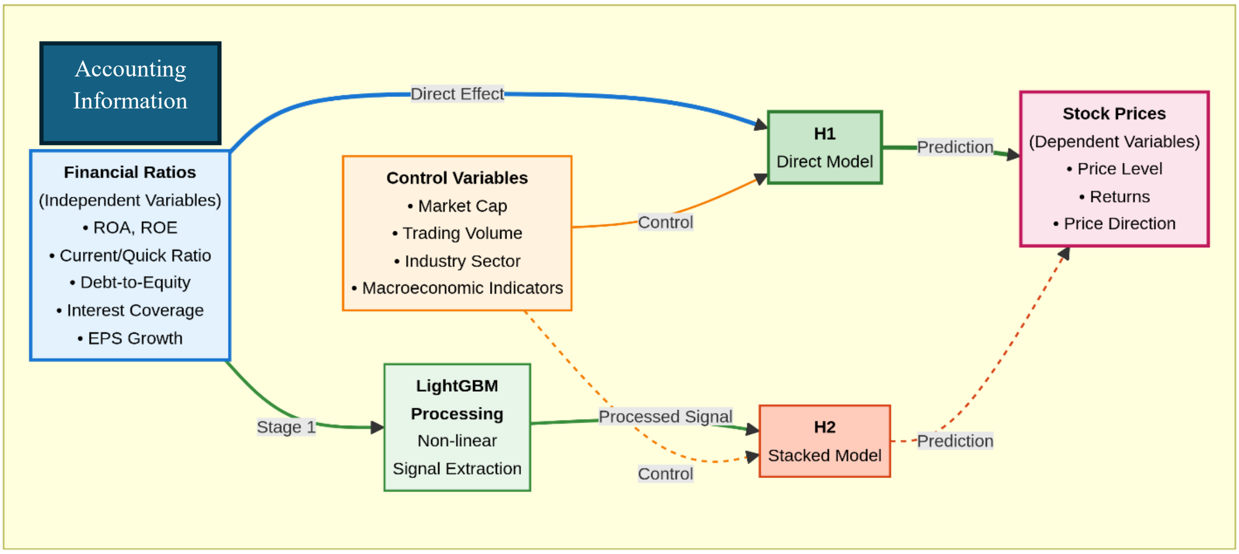

Figure 1-1. Research Model

This model is developed by the researcher based on prior empirical evidence that accounting/financial ratios (e.g., ROA, ROE, liquidity, leverage, EPS) are associated with stock returns/prices, and that these relationships can vary across industries and macroeconomic conditions (Lev & Thiagarajan, 1993; Al-Debie & Walker, 1999; Allozi, 2016; Naseer et al., 2021). The framework further incorporates LightGBM as a mediating (predictive) mechanism because gradient-boosted decision tree models are designed to capture non-linear effects and feature interactions efficiently, and have been applied successfully in stock price/market prediction settings and comparative forecasting benchmarks (Ke et al., 2017; Yang & Wang, 2021; Hartanto et al., 2023). The mediation/moderation logic follows established methodological guidance for mediator–moderator distinction and conditional process thinking (Baron & Kenny, 1986; Hayes, 2022).

1.5 Significance of the Study

1.5.1 Academic Significance

This study illustrates the first instance of the academic knowledge integration of finance, accounting, and AI into a single unified research framework and forecasting of stock price. It builds the hypothesis that AI serves as a mediating instrument for our understanding of the phenomenon rather than a mere computational device to explain the action of financial ratios on stock price (Baron & Kenny, 1986; Cao, 2022).

In a robust methodological approach, this study is the first to introduce the model of LightGBM to forecasting the finance field of Amman Stock Exchange and provides evidence for the first time on the enhancements of accuracy and interpretability that can be attained through the integration of ensemble learning, SHAP analysis (Shapley Additive exPlanations Model) (Tian et al., 2022; Ajiga et al., 2024). It closes the gap that exists between traditional statistical methods and machine learning (ML), thereby providing a framework for other emerging economies to replicate.

This research addresses a geographic and contextual gap specifically targeting the ASE and expands the extent of the application of AI and finance to less developed and uncharted markets characterized by unique regulatory and efficiencies (Al-Haddad et al., 2022; Hendawi, 2022). It provides evidence for the first time on the integration of accounting theory and data-based AI approach to non-linear relationships and how financial variables through their interactions control the market (Kumar et al., 2024; Zhang et al., 2024).

1.5.2 Practical Significance

For the investors, Phan and Chang (2024) as well as Liu (2025) describe how this study provides more explanatory and AI-integrated mechanisms to aid investors in understanding how to optimally allocate capital by identifying key predictive ratios and helping them assess capital more conclusively and accurately.

By identifying, monitoring and tracking, and analyzing sectoral specific and overall market volatility and uncertainties, risk management levels are elevated (Al-Qaisi & Al-Rdaydeh, 2020). Furthermore, incorporating macroeconomic factors into the models give more accurate control and predictability of models to meet economic conditions (Mokhtari et al, 2021).

By simplifing the process of how market participants analyze and incorporate financial data into the price of a stock, this study enhances the overall market efficiency, reducing information asymmetry and price inconsistencies in various sectors (Hendawi 2022, Huang et al. 2022). More importantly, this study provides specific ASE-tailored data preparation, feature selection, and model validation road maps for analysts and financial institutions (Yang et al, 2021, Zhao et al, 2023). The promotion of transparency and the constructive and technological advancement of the capital market of Jordan, furthering financial inclusion, diversification, and innovation, economic inclusion ecosystem of the Jordanian economy (Al-Zoubi, 2023).

1.6 Study Scope

This research examined companies publicly traded on the Amman Stock Exchange (ASE) by analyzing their accounting-based financial information and corresponding stock price data over the period 2022–2023. The selected time horizon captured a post-pandemic financial environment characterized by relative economic stability, thereby reducing the likelihood of structural breaks and regime shifts that often complicate empirical analysis in emerging markets when longer time spans are considered (Harvey, 1995; Al-Zoubi & Al-Bashir, 2007). In addition, the ASE's uniform disclosure and reporting regulations enhanced the internal validity of the study by ensuring comparability across firms operating under the same regulatory framework (Hendawi, 2022).

The study incorporated firms from the major sectors of the ASE, including banking, industrial, and services companies. Sectoral inclusion enabled the examination of heterogeneity in the relationship between accounting information and stock prices while allowing sector-specific characteristics to be captured through moderation analysis. Rather than excluding financial institutions, sectoral controls were employed to account for regulatory and structural differences across industries, thereby preserving analytical consistency with the empirical framework of the study.

Financial metrics were derived from audited financial statements prepared in accordance with International Financial Reporting Standards (IFRS), which provide a standardized and internationally recognized reporting framework (IFRS, 2023). Periodic accounting information was temporally aligned with monthly stock price observations to construct a balanced monthly panel dataset. This alignment ensured consistency between accounting measures and market data while maintaining the integrity of the empirical design. The selected accounting ratios—covering profitability, liquidity, leverage, and growth dimensions—represent core indicators commonly used in fundamental financial analysis and valuation research (Konchitchki & Patatoukas, 2023).

From a methodological perspective, the study focused on evaluating the explanatory and mediating role of advanced machine learning techniques, particularly tree-based gradient boosting models, within an accounting–finance context. Rather than positioning machine learning as a universally superior forecasting tool, the analysis assessed its capacity to process accounting information, uncover non-linear structures, and mediate the relationship between financial ratios and stock price behavior. Traditional econometric models were retained as benchmarks to facilitate comparative evaluation and to assess robustness, especially in light of the limited sample size and known efficiency constraints of emerging markets.

Geographically, the scope of the study was confined to Jordan in order to maintain institutional, regulatory, and informational consistency. The relatively low liquidity and weak-form to semi-strong-form efficiency of the ASE, as documented in the literature, justified a focused national scope and reduced the risk of confounding effects arising from heterogeneous macro-financial systems (Al-Qaisi & Al-Rdaydeh, 2020; Hendawi, 2022).

Finally, the study concentrated exclusively on quantitative relationships involving accounting ratios, stock price variables, and contextual moderators. Qualitative factors such as investor sentiment, news tone, and order-book dynamics were deliberately excluded due to data limitations and their inherent measurement challenges in emerging markets (Kumbure et al., 2022; Churi et al., 2023). This delimitation ensured analytical clarity and allowed the study to remain tightly aligned with its primary objective of examining accounting information and its interaction with machine learning techniques in explaining stock price behavior.

1.7 Study Terminology

1.7.1 Theoretical Definitions:

Table 1 contains preliminary theoretical definitions of some of the key concepts employed in this research. These definitions contribute to the establishment of conceptual precision and uniformity in the understanding of the various variables and the analytical techniques selected for this research. Each term has been defined using the most current research literature in the fields of accounting analytics, artificial intelligence, and financial forecasting.

Table1. 1 Study Definitions of Key Concepts and Analytical Terms

Term	Revised Definition (aligned with updated model & hypotheses)	Key References
Financial Statement Analysis	The systematic examination and interpretation of financial reports (income statements, balance sheets, and cash flow statements) to evaluate firm performance, financial condition, and investment relevance. In this study, analysis is operationalised through profitability, liquidity, leverage, and growth ratios used as explanatory inputs.	Zhao et al. (2023); Castañeda & Figueroa (2023)
Artificial Intelligence (AI)	A broad field of computational methods that enable systems to perform tasks associated with human cognition (e.g., pattern recognition, learning, inference, and decision support). In finance, AI methods are often used to model complex relationships, but their usefulness depends on validation and generalization performance.	Mokhtari et al. (2021); Liu (2025)
Machine Learning (ML)	A subfield of AI that builds statistical/algorithmic models that learn patterns from data to generate predictions or representations without being explicitly rule-programmed. In this study, ML is used to extract non-linear signals from accounting ratios and evaluate their predictive value under cross-validation.	Huang et al. (2022); Khanpuri et al. (2024)
Deep Learning	A subset of ML based on multi-layer neural networks that can learn hierarchical representations, particularly effective for large-scale and unstructured data. Deep learning is defined here for conceptual completeness but is not the primary modelling approach used in this study's core estimations.	Ma & Yan (2022); Phan & Chang (2024)

Term	Revised Definition (aligned with updated model & hypotheses)	Key References
Ensemble Models	Predictive frameworks that combine multiple weak learners or modelling components (e.g., tree ensembles such as Random Forest, XGBoost, LightGBM) to improve fit and reduce bias. Their practical value is assessed by out-of-sample validation rather than training accuracy alone.	Castañeda & Figueroa (2023); Liu (2025)
Stock Price Prediction	The estimation of stock outcomes (e.g., price level, returns, or price direction) using explanatory variables such as accounting ratios, firm controls, and macroeconomic indicators. In this study, predictive quality is evaluated using cross-validation to distinguish in-sample fit from generalisable performance.	Zhang et al. (2024); Mokhtari et al. (2021)
Mediating Role	A modelling concept in which an intermediate variable represents a pathway linking an independent variable to a dependent variable. In this study, the term is used in an *algorithmic/stacked* sense: the ML prediction represents a learned non-linear signal derived from accounting ratios that is then examined as an intermediate explanatory component in a second-stage regression, rather than a causal mediation claim.	Baron & Kenny (1986); Zhao et al. (2023)
Light Gradient Boosting Machine (LightGBM)	A gradient-boosted decision-tree algorithm (developed by Microsoft) that uses histogram-based learning and a leaf-wise tree growth strategy to capture non-linearities and interactions efficiently. In this study, LightGBM is applied to learn a predictive signal from accounting ratios and is evaluated using cross-validation; interpretability may be supported via feature importance or SHAP-based explanations where reported.	Kumar et al. (2024); Tian et al. (2022); Cao (2022)

1.7.2 Procedural Definitions:

Variables were classified as independent, dependent, an intermediate AI-derived signal, and control variables. The independent variables comprised seven financial ratios identified from prior literature (e.g., Arkan, 2016; İltüzer & Çam, 2023): profitability (ROA, ROE), liquidity (Current Ratio, Quick Ratio), leverage (Debt-to-Equity, Interest Coverage), and growth (EPS Growth).

The dependent variables captured stock outcomes, operationalised as price level, stock returns (percentage change between consecutive periods), and price direction (1 =

increase, 0 = decrease/no change). The intermediate AI variable was the LightGBM output, representing a non-linear signal learned from the financial ratios and then used in the stacked second-stage model. Industry sector and macroeconomic indicators (inflation, GDP growth, interest rate) were included as control variables to adjust for contextual differences, rather than being treated as moderators unless interaction terms were explicitly tested.

Table1.2.Study Variable Operational Definitions

Variable Type	Variable	Operational Definition	Measurement Method
Independent Variables	**Profitability Ratios**		
	Return on Assets (ROA)	Measures the efficiency with which a company utilizes its total assets to generate accounting profits	Net Income / Total Assets
	Return on Equity (ROE)	Measures the return generated for shareholders based on invested equity	Net Income / Shareholders' Equity
	Liquidity Ratios		
	Current Ratio	Measures the firm's ability to meet short-term obligations using current assets	Current Assets / Current Liabilities
	Quick Ratio	Measures the firm's ability to meet short-term obligations using liquid assets, excluding inventories	(Current Assets − Inventory) / Current Liabilities
	Leverage Ratios		
	Debt-to-Equity Ratio	Measures the extent to which a firm finances its operations through debt relative to equity	Total Liabilities / Shareholders' Equity
	Interest Coverage Ratio	Measures the firm's capacity to meet interest obligations from operating earnings	EBIT / Interest Expense
	Growth Ratios		
	EPS Growth	Measures the growth rate in earnings per share between consecutive periods	(EPS_t − EPS_{t−1}) / EPS_{t−1}

Variable Type	Variable	Operational Definition	Measurement Method
Dependent Variables	Price Level (JOD)	Market value of a company's share at the end of each month	Monthly closing stock price (Jordanian Dinar)
	Stock Returns (%)	Percentage change in stock price between consecutive periods	(Price_t − Price_{t−1}) / Price_{t−1} × 100
	Price Direction	Binary indicator capturing the direction of stock price movement relative to the previous period	Dummy variable: 1 = price increase, 0 = price decrease or no change
Intermediate ML Signal (Stacked / "Mediating" Variable)	LightGBM Model Outputs	Model-generated predictions extracted from financial ratios via LightGBM and used as an intermediate predictive signal in the stacked framework	Model predictions; performance evaluated using R^2, MAE, and RMSE
Control Variables	Market Capitalization	Firm size control reflecting the market value of equity	Market Cap (JOD): Share Price × Shares Outstanding (or official ASE market cap series)
	Trading Volume	Market activity/liquidity control capturing the intensity of trading	Total monthly traded volume (shares) or traded value (JOD), as reported by ASE
	Industry Sector	Sector classification used to control for structural differences across industries	Dummy variables: Banking (1/0), Industrial (1/0), Services (1/0)
	Inflation Rate (%)	Annual rate of change in the general price level (macro control)	Central Bank of Jordan statistics
	GDP Growth (%)	Annual percentage growth rate of Gross Domestic Product (macro control)	Central Bank of Jordan statistics
	Interest Rate (%)	Prevailing benchmark interest rate in the economy (macro control)	Central Bank of Jordan statistics

Chapter Two
Theoretical Framework and Previous Studies

2.1 Introduction

This chapter reviews theoretical and empirical literatures that are applicable to the prediction of stock returns based on accounting-based data with advanced machine learning approaches. The ultimate aim is to establish a solid theoretical basis that supports the research assumptions and relates the study implications to the prior literature. The rest of the paper is organized as follows: Section 2 contains a brief overview on studies related to market efficiency and financial ratios as tools for predicting future market performance along with the transition from classical econometrics to modern machine learning model for prediction, followed by the situation in ASE.

The first alllargers of the building is the analysis of the Efficient Market Hypothesis (EMH), one of the most important theories of finance and is then followed by a discussion of Adaptive Markets Hypothesis (AMH), the first is also the incorporation of behavioral economics and the theory of the holistic view of the market systems. The analysis of financial statement analysis is next, focusing on the usefulness and limitations of financial ratios and the predicting the retuns on stocks. The next is the use of artificial intelligence and machine learning in financial forecasting, especially ensemble methods like Light g.BM. because they can handle complex non stressed non-linear relationships that classical model highly complex finance. Because the research setting is in the Jordanian financial market, the studies on market efficiency and predictability at the Amman Stock Exchanges. This chapter synthesis between the two poles of the heterogeneous literature, and that is why it is the parameter of the study.

2.2 Theoretical Framework

This study considers how market efficiency works through financial statement analysis, financial diagnostics, and computer prediction capabilities. It considers how the market works classically, as an efficient system that processes information, and behaves as though it has human psychology that creates predictable patterns.

2.2.1 The Efficient Market Hypothesis (EMH)

The Efficient Market Hypothesis (EMH), as formulated by Fama (1970), suggests that asset prices embody all public and relevant information. At its most rigorous, the EMH holds that it is impossible to do better than average in the long run on a risk-adjusted basis, since all publicly available information about stocks already affects their prices. Fama (1998) returned to the theory, conceding limitations in long-run return anomalies but essentially reaffirming the view of the hypothesis as a useful and empirically sound benchmark. In general there are three forms in which market efficiency is usually defined: weak, semi-strong and strong. The weak version holds that all past market prices and information are already reflected in securities' prices. The semi-strong form states that all public information is factored in, and the strong form asserts that all known information—public or private—is accounted for.

The hypothesis of this research, i.e., utilizing the accounting information published publicly to forecast prices of stocks, directly challenges semi-strong form efficiency. To the extent that a set of financial ratios can predict future returns in a non-systematic manner it would violate semi-strong EMH. Although the EMH has been an influential framework, there is a considerable literature demonstrating various market anomalies—such as the value effect (Lakonishok, Shleifer & Vishny 1994), momentum (Brown, Goetzmann & Ibbotson 1999) and post-earnings announcement drift—which can appear inconsistent with it (Fama & French 1992).

2.2.2 Behavioral Finance and the Adaptive Markets Hypothesis (AMH)

The behavioralist counter-narrative to EMH argues that systematic and predictable mispricings can arise from the psychological biases or cognitive errors investors typically make. This does not reject the idea of market efficiency, but provides a more nuanced definition of the theory.

One of the most significant advancements of integrating the two perspectives of behavioral finance and EMH is Lo's (2005) Adaptive Markets Hypothesis (AMH). Lo does not advocate for a market system that is absolutely efficient. Rather, he supports the idea of a market system that is a perpetual, efficient cycle. This cycle is also mutable, as the market has constant, evolutionary activity from the participants that fulfill the self-interested, but, boundedly rational, roles of agents as described by Lo (2005). When they adapt, the windows of opportunity for profiteers close while other ones open up. A second possible application of the evolutionary market efficiency will come from its' implication that the level of market efficiency can change through time and under heterogeneous market conditions. The perspective of the AMH is well suited to our study, as it suggests that although markets will go towards efficiency they are likely traversed by pockets of predictability, especially in emerging markets or under structural change. The employment of powerful AI models can be thought of as a type of survival, to which resourceful investors finally make use of better techniques for capturing and riding over these transient signatures.

2.2.3 The Role of Financial Ratios in Predicting Stock Returns

Fundamental investment analysis is also based on examining the company’s financial statements. The main idea behind it is that the financial condition, performance and value of a company are inherently reflected in its financial statements in connection with stock exchange returns. Financial ratios are the principal instruments through which the analysis is developed, because they normalize data of financial statements. They are comparable over time and against other firms, reflecting a company's liquidity, solvency, profitability, and efficiency.

Ou and Penman (1989) was the seminal work in this area showing that financial statement data is used to predict future earnings and. consequently, future stock prices. They showed that a model using an extensive list of financial ratios was able to predict the stock price with accuracy, thereby refuting the conventional belief in market efficiency prevalent then. Later studies have also further analysed the relationship between these and other metrics, all reporting that financial ratios were good at predicting failure.

Arkan (2016) also worked on emerging market, and concluded that profitability ratios (e.g., ROA and ROE), and valuation ratios (e.g., PE ratio -- the price to earnings multiple; MTB – the market-to-book value) were found crucial in forecasting stock price trends. This corresponds to the results of İltüzer and Çam (2023), who also supported the usability of financial ratios in forecasting stock returns, especially when used with machine learning models. Konchitchki and Patatoukas (2014) expanded the horizon more, indicating that financial statement analysis is a good aiding tool not just for single stock valuation but also to macroeconomic forecastingIn spite of the complexities of the association and context in this branch of literature, Akhtar et al (2021) Huang and Li has also pointed out the varying degrees of predictability of the market multiples in both emerging and developed economies in their studies and their findings show the context matters.

Considering the studies mentioned above, there is still contention in the direct predictive power of financial ratios. The concerned ratios are numerous and due to multicollinearity potential as well as the disordered and non-linear relationships exhibited between the ratios and stock prices, regression analysis becomes challenging. This which makes an even stronger case for advanced statistical approaches, like machine learning, which can be utilized to deal with non-linear and high-dimensional data.

2.3 Machine Learning in Financial Forecasting

Utilization of predictive analytics and ML and AI in financial forecasting is a novel approach. In contrast to traditional econometric models that rely on restrictive assumptions of linearity and stationarity, ML models learn from and make predictions about complex, non-linear patterns in data. Because of the characteristics of financial markets (i.e. noise, non-stationarity and complex relationships) these models are a better fit.

2.3.1 From Traditional Models to Machine Learning

Classic econometric models such as ARIMA and GARCH have been used and accepted for many years in the industry for forecasting financial time series. However, their performance, in many cases, has been limited and severely constrained by the models' assumptions. In contrast, machine learning has introduced several models which include, but are, not limited to, neural networks, support vector machines, and ensemble methods all of which have proven to be more powerful than the classical econometric models across a variety of financial applications.

2.3.2 Ensemble Learning and Gradient Boosting

Another class of MLAs are the ensemble learning approaches which combine multiple models to make prediction, while one of the most popular ensembles is gradient boosting models (Friedman et al., 2001). Ensemble learning is a strong ML technique in which the predictions of many individual models (or "weak learners") are combined to make a single, more accurate and robust prediction. The underlying concept is that when combining with the “wisdom” of multiple model, weaknesses of each single model can be compensated. Random Forest and Gradient Boosting are two of the most commonly used ensemble techniques.

Gradient Boosting, in particular, has increasingly become popular in financial time series prediction. GBMs construct trees in a sequential manner where each new tree corrects the errors of a previous one. This feedback-based process can enable the model to identify very complex patterns in the data. Yang and Chen(2020) proved the success of GBDT in predicting market movements, owing to its capability of learning non-linear relationships. Sharma et al. (2023) further verified the high relevance of XGBoost (a well-known gradient boosting algorithm implementation) with stock market trend analysis..

2.3.3 LightGBM: Efficiency and Performance

This study begins by theoretically considering that learning models do not possess any inherent ability to predict market outcomes, but rather can be used to process and analyze information in an advanced manner. The LightGBM model, in this study, is viewed as a non-linear indicator variable for an accounting firm, which performs a traditional accounting experiment to generate riskier predictive outputs before these outputs are reflected in stock prices. These processes do not rely on artificial intelligence as a replacement for quarry analysis, but rather encompass how accounting information is transmitted to the market.

LightGBM is a new, relatively efficient implementation of the gradient boosting framework. They developed several novelties, such as Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), that make it possible to train models up to times faster with even less memory than classic GBM while still retaining accuracy. They also mentioned LightGBM as the most accurate and efficient algorithms to predict stock price time-series (Hartanto & purwanto, 2023). Ferrouhi et al. (2024) and they performed comparison analysis on different ensemble method and proposed that gradient boosting, including LightGBM are in the top for stock market prediction. The study by Bisdounis et al. (2024) also proved the strength property of LightGBM in capital prediction task with applying suitable feature engineering technique

2.3.4 The "Black Box" Problem and Model Interpretability (XAI)

Despite their predictive ability, a major criticism of complex ML models is that they are not interpretable and are commonly known as the "black box" problem. In high risk domains such as finance, equally important as predicting an outcome is the understanding of why a model made that particular prediction. It is in this context that Explainable AI (XAI) has gained a lot of attention, where the objective is to establish measures for understanding complex models.

SHAP (SHapley Additive exPlanations) is one of the most popular XAI framework. It is a logical interpretation of Shapley values from cooperative game theory and thus, gives us one mathematical way to explain the output of any machine learning model. SHAP gives you an importance value for each feature, in terms of an individual prediction, allowing users to understand what's driving the model decisions. We refer to SHAP as Lin and Gao (2022) for interpreting financial fraud detection models, which is transferable to stock prediction. Recent comprehensive betting has been a safe bet in finance showing that XAI is becoming increasingly important, as evidenced by the work of Arsenault et. (2025) as well as Černevičienė and Kabašinskas (2024), that stress how transparency of financial AI systems can be vital for trust. Chen et al. (2024) also stress the relevance of interpretability, especially given that credit scoring is another area with high-stakes decisions.

2.4 Mediation, Information Processing, and the Amman Stock Exchange

2.4.1 Mediation Analysis in Financial Research

Mediation is a statistical methodology for exploring the process, mechanism, or pathway through which an independent variable influences a dependent variable. In this study, it is employed to examine if the predictability of financial ratios does not occur directly but rather through the holes made by information processing abilities of the AI model. The basis of the mediation had been provided by Baron and Kenny (1986), as many data analytics professionals know there are more recent considerations. Mediation analysis has been applied to investigate a number of indirect in financial research. For instance, Hoekstra and Huisman (2024) employed it to analyse the mediating role of trading volume in the relationship between investor sentiment and stock returns. This study applies the same logic to verify whether the AI model is a mediator that uncovers the information hidden in financial ratios.

2.4.2 Information Processing Theory

The AI model in this study can be better understood through the lens of information processing theory, rooted in cognitive psychology. The theory suggests that human cognition is an information-processing system, which consists of subsystems such as attention, encoding and retrieval. In the financial markets framework, investors are regarded as information agents with bounded rationality. Bernales et al. (2023) simulated the impact of information overload on financial markets and suggested that limits to investors' capacity can create predictable outcomes in markets. Goldstein and Yang (2023) give a detailed overview of how information is acquired in financial markets and its consequences to the real side of the economy. From this perspective, the LightGBM model is seen as a super-efficient information filter processing large and complex financial data to distill tell-tale predictive signals which would be hard for a human analyst, if not impossible, to appreciate.

2.4.3 The Amman Stock Exchange (ASE) Context

Lessons about ASE market Jordan's stock exchange is part of the emerging markets category; which has significant differences from developed markets. They might be less informationally efficient, less liquid and more sensitive to market mood and macroeconomic disturbances. The very specific market structure and trading rules make them an especially attractive arena for testing the predictive capacity of financial models.

Various research works have addressed the effectiveness of the ASE. While conducting some specific frontiers, sectoral efficiency analysis, Rawashdeh and Squalli (2005) found different degree of efficiencies at the various sectors of the exchange. More recently, Airout et al. (2023), they analyzed the profit performance of Islamic bank shares on the ASE. Consequently, the prevailing view is that the ASE, along with most of its emerging market counterparts, does not satisfy strong or even semi-strong form efficiency, implying possibilities for earning predictable returns. Htun et al. (2023) systematically reviewed feature selection methods with stock market prediction in mind which is particularly relevant to the discovery of strongest predictors in ASE framework. This paper adds to this body of literature as being among the first that uses an advanced ensemble learning model (LightGBM) along with a rigorous XAI framework (SHAP) focusing on a large sample of accounting data from the ASE.

2.5 Previous Studies

2.5.1 Studies in Arabic

The study of **Al-Badiri and Al-Khouri (1997)** entitled *“Price Movements of Stocks in the Amman Financial Market Using Econometric Models”* aimed to analyze stock price behavior in the Amman Financial Market using quantitative econometric techniques. The study adopted a statistical time-series methodology and applied standard econometric models to examine the relationship between stock prices and selected economic indicators. The results revealed statistically significant relationships between stock price movements and macroeconomic variables, confirming that price behavior in the Amman market could be partially explained using quantitative models. The study recommended expanding the application of formal forecasting models to support investment analysis in the Jordanian stock market.

The study of **Al-Majali (2006)** entitled *“Predictability of the Amman Stock Exchange Using ARIMA Models”* aimed to examine whether movements of the Amman Stock Exchange index could be predicted using time-series forecasting models. The study employed a quantitative analytical methodology based on univariate ARIMA models applied to daily index data. The results demonstrated that ARIMA models possessed short-term forecasting ability, indicating that stock price movements in the Amman Stock Exchange do not fully follow a random walk process. The study recommended using time-series forecasting techniques as supportive tools for short-term investment decision-making in emerging Arab markets.

The study of **Al-Khouri and Al-Nour (2009)** entitled *“The Effect of Announcing Accounting Information in Annual Financial Reports on Stock Prices and Trading Volume in the Amman Securities Market”* aimed to assess the impact of accounting information disclosure on stock prices and trading activity. The study applied the event study methodology to a sample of 45 companies listed on the Amman Stock Exchange during the period 2005–2007. The results showed no statistically significant abnormal returns on announcement dates, indicating that the market does not exhibit semi-strong form efficiency. The study recommended enhancing disclosure quality and strengthening regulatory mechanisms to improve informational efficiency in the Jordanian capital market.

The study of **Al-Wattar and Al-Ibrahimi (2010)** entitled *"Using Financial Analysis Methods in Predicting Failure of Industrial Joint-Stock Companies"* aimed to evaluate the ability of financial ratios to predict corporate financial failure. The study adopted a quantitative methodology based on discriminant analysis using the Altman Z-score model applied to Iraqi industrial companies. The findings confirmed the high accuracy of the Altman model in distinguishing between financially distressed and non-distressed firms. The study recommended adopting early warning financial failure models to protect investors and enhance corporate financial stability.

The study of **Mahmoud and Dawalbait (2015)** entitled *"Estimating and Forecasting Stock Market Volatility Using GARCH Models: Evidence from Saudi Arabia"* aimed to estimate and forecast volatility in the Saudi stock market. The study employed symmetric and asymmetric GARCH models using daily stock return data. The results indicated that asymmetric GARCH models provided superior forecasting performance compared to symmetric models, reflecting the presence of volatility clustering and leverage effects. The study recommended using advanced GARCH specifications for risk management and volatility forecasting in emerging Gulf markets.

The study of **Kalyanaraman (2014)** entitled *"Stock Market Volatility in Saudi Arabia: An Application of the GARCH Model"* aimed to analyze and forecast volatility in the Saudi stock market. The study adopted a quantitative econometric methodology using the GARCH(1,1) model applied to daily stock market returns. The results confirmed the existence of volatility clustering and persistence, indicating that past shocks significantly influence current volatility levels. The study recommended the use of GARCH-type models for investment risk assessment and portfolio management in the Saudi market.

The study of **Legouguigi and Chikhi (2017)** entitled *"Modeling Stock Price Volatility in the Saudi Stock Market Using ARCH Models"* aimed to examine volatility behavior using ARCH-family models. The study employed a quantitative time-series methodology using daily stock price data of a Saudi listed company. The results revealed that asymmetric ARCH models were more effective in capturing volatility dynamics and the differing impact of positive and negative shocks. The study recommended using ARCH-based models to improve forecasting accuracy and risk evaluation in Saudi financial markets.

The study of **Al-Rahahleh (2018)** entitled *"Forecasting Volatility: Evidence from the Saudi Stock Market"* aimed to compare the forecasting performance of linear and non-linear GARCH models. The study applied several GARCH-class models to daily index data from the Saudi stock market. The findings demonstrated that non-linear models, particularly APARCH and GJR-GARCH, outperformed traditional models in volatility forecasting. The study recommended adopting asymmetric volatility models to better capture risk dynamics in emerging markets.

The study of **Al-Naif et al. (2021)** entitled *"Return and Volatility Spillovers Among Arabian Stock Markets"* aimed to examine return and volatility transmission across several Arab stock markets. The study employed a quantitative methodology using the Diebold–Yilmaz spillover framework applied to daily index data. The results indicated that stock markets in the Arab region are largely driven by domestic shocks, with limited cross-market spillovers. The study recommended focusing on market-specific forecasting models when analyzing Arab stock markets.

The study of **Ayachi (2022)** entitled *"Forecasting Stock Market Returns Using Hybrid ARIMA-GARCH Models: Evidence from the Saudi Stock Market"* aimed to improve forecasting accuracy by combining ARIMA and GARCH models. The study adopted a quantitative hybrid modeling approach applied to Saudi stock market return data. The results showed that hybrid ARIMA-GARCH models outperformed single models in capturing both return dynamics and volatility behavior. The study recommended using hybrid forecasting frameworks for stock market prediction under conditions of economic uncertainty.

2.5.2 Studies in English

The study of Rouf et al. (2025) entitled *"Artificial Intelligence-Based Stock Market Price Prediction: A Review"*

The study aimed to comprehensively synthesize recent advances in AI methodologies for stock price trend prediction. This review adopted a structured literature review methodology covering developments in machine learning and deep learning models, including hybrid approaches and diversified datasets. The analysis revealed that hybrid models, sentiment analysis, and multi-source data integration have increasingly become central to improved prediction accuracy. The study concluded that hybrid and ensemble frameworks represent the cutting edge of stock forecasting research, as they effectively combine strengths of diverse models and data types.

The study of Lin (2024) entitled *"Stock Market Prediction Using Artificial Intelligence: A Meta-Review"*

The study aimed to evaluate the state of AI-based stock market prediction research. This meta-review employed a comprehensive survey methodology focusing on machine learning, deep learning, recurrent neural networks, and hybrid architectures. The results demonstrated a clear progression from traditional models to advanced deep learning techniques such as LSTM and CNN, with hybrid combinations showing superior performance in capturing temporal and non-linear patterns. The study recommended continued integration of diverse data sources and hybrid architectures to further enhance predictive performance.

The study of Wang (2025) entitled *"Machine Learning for Stock Return Prediction: Transformers and Advanced Time-Series Models"*

The study aimed to explore modern ML architectures, including transformer-based models, for forecasting stock returns. The study conducted a detailed literature review and methodological comparison involving transformer models (such as Autoformer) and baseline neural networks. The findings highlighted that transformer-based forecasting architectures can effectively model complex temporal dependencies in financial data, often outperforming traditional recurrent models. The study suggested that transformer models offer promising capabilities for future stock return prediction tasks.

The study of Fozap et al. (2025) entitled *"Hybrid Machine Learning Models for Long-Term Stock Market Forecasting"*

The study proposed a hybrid deep learning model integrating Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNNs), augmented by technical indicators to improve prediction accuracy. Using empirical stock market data, the hybrid model achieved significantly better performance than individual models. The study emphasized the importance of combining temporal sequence modelling with spatial feature extraction to capture complex patterns in stock price movements.

The study of Sonani, Badii, and Moin (2025) entitled *"Stock Price Prediction Using a Hybrid LSTM-GNN Model: Integrating Time-Series and Graph-Based Analysis"*

The study aimed to enhance stock market prediction by combining Long Short-Term Memory (LSTM) networks with Graph Neural Networks (GNNs). The LSTM component captured temporal patterns, while the GNN component modelled inter-stock relational dependencies. The hybrid model significantly reduced prediction error compared to standalone models and traditional baselines, demonstrating the value of relational data alongside temporal features.

The study of Joshi (2025) entitled *"Integrating LSTM and CNN for Stock Market Prediction"* The study presented a hybrid machine learning approach combining a Convolutional Neural Network with an LSTM network to model both spatial and temporal information in stock time-series data. The empirical results showed that the integrated model improved prediction accuracy compared to each individual component. The study recommended hybrid architectures that leverage the strengths of multiple neural network types for more effective financial forecasting.

The study of Lan (2025) entitled *"A Hybrid CNN-LSTM Model for Stock Price Prediction with Spatial and Temporal Dependencies"*

The study aimed to investigate price prediction accuracy using a combined CNN and LSTM architecture. This research applied the hybrid model to historical financial data and demonstrated that the combined approach outperformed isolated CNN or LSTM models in capturing both short-term fluctuations and long-term trends. It recommended hybrid deep learning frameworks as practical tools for investors and analysts.

The study of Magloire (2025) entitled *"Hybrid Deep Learning Models for Stock Price Prediction Using LSTM and CNN"*

The study aimed to enhance stock forecasting performance by integrating LSTM and CNN deep learning models with technical indicators. The study's empirical results indicated that the hybrid framework substantially improved prediction accuracy over individual models, highlighting the synergy between temporal and pattern extraction techniques. The study recommended hybrid learning structures for accurate long-term forecasting in financial markets.

The study of Maheshwari (2025) entitled *"A Deep Learning Approach for Forecasting the NSE Opening Index Using Hybrid Models"*

The study aimed to apply a range of deep learning techniques, including RNN, LSTM, CNN, and CNN-LSTM hybrid architectures, for forecasting the opening price of the NSE index. The findings suggested that hybrid approaches combining the strengths of recurrent and convolutional networks could better capture complex temporal and nonlinear patterns in financial data. The study recommended further exploration of hybrid deep learning models for index forecasting.

The study of Saberironaghi (2025) entitled *"Stock Market Prediction Using Machine Learning and Deep Learning Approaches"*

The study aimed to provide a comprehensive review of machine learning and deep learning methodologies used in stock market forecasting, covering a broad range of algorithms, evaluation metrics, and datasets. The review found that ensemble and hybrid models consistently outperform single algorithm approaches and that model performance varies depending on data type and market conditions. The study recommended integrating diverse data sources and models for robust prediction frameworks.

2.6 Conclusion

As the principles of market efficiency, as well as the most advanced implementations of machine learning in finance, this review has thus provided a wide-ranging theoretical and practical framework for the analysis The review indicates that, although the EMH continues to be a vital point of reference, the more advanced perspective provided by the AMH, together with behavioral finance, is more relevant for framing the dynamics of the stock market. Literature does not refute the predictive power of financial ratios, only the weakness of traditional linear models to extract it. This paves the way for more sophisticated machine learning models such as LightGBM (for complex and non-linear financial data). a] also provide perspective on the increasing focus of model interpretability handled using SHAP. Finally, by placing the research within the explicit scope of evaluation of the practice at Amman Stock Exchange, institutes a specific need for further research and opportunity to enrich our understanding on financial market predictability in an emerging economy. The methodology that will be adopted in this research, designed to test the hypotheses developed from such a comprehensive review, will be discussed in the following chapter.

2.7 What Distinguishes This Study from Previous Studies

Literature review on forecasting stock price A large amount of previous studies that focus on stock price prediction follow two lines: some researches investigate financial ratios as direct factors to affect the stock price movements others use AI technology as a separate tool for prediction. We are going a step beyond each by revealing LightGBM as an intermediary construct that functions to join financial statement analysis and stock price forecasting in the evolving context of one particular market in Middle East. Whereas previous researches in emerging markets have concentrated on the use of AI to improve the traditional financial models, we are interested in exploring whether AI could potentially change the predictive power of financial ratios by serving as a mediator and/or moderator affect between them and stock market returns (Cao, 2022; Ajiga et al., 2024).

The second point of difference is that the research focus differs to what is studied. Most of the previous AI-based financial forecasting literature has been focused on developed markets, in which data quality and efficiency make so relatively stable environments for modelling. By contrast, this goes on to use its framework with the ASE

—as a semi-efficient, information-asymmetric market featuring sectorial fragmentation and low trading volumes where geopolitical risk prevails (Hendawi, 2022; Al-Qaisi & Al-Rdaydeh, 2020). As the study concentrates on ASE, it demonstrates that AI can be applied to environments where other models are not well built.

Methodologically, our work incorporates SHAP as part of LightGBM and to the best of our knowledge manages to tackle the 'black box' criticism aimed at AI while providing viewpoint explanation for investors and regulators (Kumar et al., 2024). The degree of transparency adopted was never quite so salient in previous ASE-centric studies which were based on ARIMA- or ANN-based models. Moreover, the paper includes sectoral and macroeconomic moderators (which have been largely ignored in previous studies) to take into consideration heterogeneity and external shocks which are usually characteristics of emerging markets (Zhao et al., 2023; Phan & Chang, 2024).

Finally, the study separates by its period. The sample covers 2 years of ASE-listed companies and accounts for the short- and medium-term horizons, which provides more stability than prior researches which are limited to less or shorter representative samples.

In sum, these conceptual, methodological and contextual innovations situate this study as original on theory grounds and market aware in the AI-enabled financial forecasting.

Chapter Three
Methodology (Methods and Procedures)

3.1 Introduction

This chapter outlines the methodological framework adopted to achieve the objectives of the study and to test the proposed research hypotheses. It describes the research design, population and sampling procedures, data sources, variable measurement, and analytical techniques employed to examine the relationship between accounting-based financial ratios and stock price behaviour on the Amman Stock Exchange (ASE). In addition, the chapter explains the development and evaluation of the applied machine learning model, the statistical procedures used for hypothesis testing, and the measures taken to ensure the reliability, validity, and ethical integrity of the study.

3.2 Research Method

The research method in this study was of a quantitative and descriptive-analytical nature, aiming at predicting the relationship between accounting-based financial ratios for companies operating on ASE listed markets and stock prices. This design is suitable as it integrates to the systematic modelling of cash variables and statistical-testing relations among them (Dewasiri et al., 2018). The first aspect, description, includes the summarization of characteristics of datasets; and the second aspect, analysis, concerns checking small hypotheses about what one expects to see in a dataset and how variables are related. This method mirrors the conventional methods of financial statements analysis and quantitative stock prediction studies (Jackson, 2022; Schumaker & Chen, 2009).

3.3 Population and Sampling

The population of the study comprised all companies listed on the Amman Stock Exchange (ASE). A purposive sampling technique was employed to select a representative sample of 15 firms based on data availability, continuity of listing, and consistency of financial reporting. All selected firms remained continuously listed on the ASE throughout the period from January 2022 to December 2023 and exhibited no history of bankruptcy, delisting, or major corporate restructuring during the study horizon. To ensure data completeness and temporal consistency, the sampled firms possessed audited periodic financial statements prepared in accordance with International Financial

Reporting Standards (IFRS). Periodic accounting information was temporally aligned with monthly stock price data to construct a balanced panel dataset suitable for both econometric and machine learning analyses. This design resulted in a balanced monthly panel comprising 360 observations (15 firms observed over 24 months), thereby providing sufficient cross-sectional and time-series variation for empirical analysis.The final sample was stratified by major economic sectors to capture structural differences across industries. It included five banking institutions (Arab Bank, Jordan Commercial Bank, Bank of Jordan, Jordan Kuwait Bank, and Capital Bank), five industrial companies (Arab Potash Company, Jordan Phosphate Mines Co., Jordan Cement Factories Co., Lafarge Jordan Cement, and United Iron Industries Co.), and five service-sector firms (Jordan Telecom, Royal Jordanian Airlines, JoHotels & Tourism Amman–Jordan, Aqaba Railway Corporation, and JETT). This sectoral composition enabled a more nuanced analysis of accounting–price relationships and supported the examination of sector-specific heterogeneity within the Jordanian economy. Overall, the sampling strategy ensured a structurally balanced and methodologically consistent panel dataset, facilitating robust analysis of the direct, mediated, and moderated relationships between accounting information and stock price behavior on the ASE.

3.4 Study Variables

All materials used in this thesis were obtained from secondary sources to ensure objectivity and reliability. Periodic financial information and monthly stock price data for the 15 firms were collected from the official database of the Amman Stock Exchange (ASE), supplemented where necessary by Refinitiv Eikon. The data covered the period from January 2022 to December 2023, providing a recent and up-to-date representation of market behavior. All datasets were carefully cleaned, preprocessed, and prepared for analysis using Python libraries such as Pandas and NumPy. The study variables were classified into independent, dependent, mediating, and moderating categories. The independent variables consisted of seven widely used accounting-based financial ratios identified through an extensive review of the literature (e.g., Arkan, 2016; İltüzer & Çam, 2023). These ratios captured key dimensions of firm financial performance and structure, including profitability ratios (Return on Assets and Return on Equity), liquidity ratios (Current Ratio and Quick Ratio), leverage ratios (Debt-to-Equity Ratio and Interest Coverage Ratio), and a growth ratio (Earnings per Share Growth). The dependent variables included the monthly stock price level, monthly stock returns, and stock price direction, all derived from consecutive monthly closing prices. The mediating variable was represented by the predicted stock price outputs generated by the Light Gradient Boosting Machine (LightGBM) model, which utilized the accounting ratios as input features. Moderating variables included industry sector classifications and macroeconomic indicators—namely the inflation rate, GDP growth rate, and interest rate—introduced to capture contextual and environmental influences on the relationship between accounting information and stock price behavior.

Table 3.1.Study Variable Measurements

Variable	Variable Measurement	References
Financial Ratios		
ROA	Return on Assets = Net Income / Total Assets (**aligned to monthly observations**).	Arkan (2016); İltüzer & Çam (2023)
ROE	Return on Equity = Net Income / Shareholders' Equity (**aligned to monthly observations**).	Arkan (2016); İltüzer & Çam (2023)
Current Ratio	Current Assets / Current Liabilities (**aligned to monthly observations**).	Arkan (2016); İltüzer & Çam (2023)
Quick Ratio	(Current Assets − Inventories) / Current Liabilities (**aligned to monthly observations**).	Arkan (2016); İltüzer & Çam (2023)
Debt-to-Equity	Total Liabilities / Shareholders' Equity (**aligned to monthly observations**).	Arkan (2016); İltüzer & Çam (2023)
Interest Coverage	EBIT / Interest Expense (**aligned to monthly observations**).	Arkan (2016); İltüzer & Çam (2023)
EPS Growth	$(EPS_t - EPS_{t-1}) / EPS_{t-1}$ (**monthly**).	Arkan (2016); İltüzer & Çam (2023)
Stock Price Variables		
Price Level (JOD)	Monthly closing stock price per share (Jordanian Dinar).	Schumaker & Chen (2009); Jackson (2022)
Returns (%)	Monthly simple return = $[(P_t - P_{t-1}) / P_{t-1}] \times 100$.	Schumaker & Chen (2009); Jackson (2022)
Price Direction	Binary indicator: 1 if $P_t > P_{t-1}$, 0 otherwise.	Schumaker & Chen (2009); Jackson (2022)
Macroeconomic Variables		
Inflation Rate (%)	Annual inflation rate aligned to monthly observations.	Dewasiri et al. (2018); Jackson (2022)
GDP Growth (%)	Annual GDP growth rate aligned to monthly observations.	Dewasiri et al. (2018); Jackson (2022)
Interest Rate (%)	Policy or benchmark interest rate aligned to monthly observations.	Dewasiri et al. (2018); Jackson (2022)

3.4 Model Development and Evaluation

The Light Gradient Boosting Machine (LightGBM) model, a gradient boosting algorithm, was one of the responsive machine learning techniques most-utilized, particularly in the last few studies given its suitability in financial forecasting and in capturing non-linear relationships (Hartanto & Purwanto, 2023; Ferrouhi et al., 2024). Model generalizability was then studied and evaluated using an 80-20 training and testing split of the data. In a bid to mitigate the risk of overfitting, a 10-fold cross validation (10-CV) was performed (Kovavi, 2021). Estimated model prediction performance was using standard regression metrics of R squared (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). LightGBM's black-box characteristic was counteracted using game theory for SHapley Additive exPlanation (SHAP) feature importance which explains model predictions (Lundberg & Lee, 2017).

3.5 Statistical Analysis and Hypothesis Testing

Using the statsmodels and scikit-learn libraries in Python, the direct predictive power of the financial ratios on the stock returns (H1) was tested using multiple linear regression analysis. Mediation analysis, with the procedure recommended by Baron and Kenny (1986) and complemented using bootstrapping methods (Preacher & Hayes, 2008), was employed to examine whether or not AI model prediction mediated the effects of financial ratios on stock returns (H2). Superior Predictive Accuracy (H3) The predictive accuracy of the LightGBM model was tested against a baseline multiple linear regression model. (H4) A moderated regression technique was then used to examine the moderation effect of industry sector and macroeconomic factors on the relationship between AI predictions and stock return.

3.6 Reliability and Validity

To achieve the soundness of this study, the issues regarding reliability and validity were attentively considered. Reliability of the financial ratio constructs were tested by Cronbach’s alpha, where values greater than the conventional threshold value 0.70 was considered as reliable (Cronbach, 1951). 10 Fold cross-validation was adopted in the model learning process which guaranteed the prediction robustness and stability of LightGBM on different data subsets (Stone, 1974). Criterion content validity was determined by the selection of financial ratios extracted from relevant literature that covered the field in an extensive manner. Construct validity was confirmed by verifying that the chosen financial ratios belong to the four theoretical profitability, liquidity, leverage, growth dimensions stipulated in finance theory (Hair et al., 2019) Though the study is applied to the ASE, application of common financial ratios and sound Machine Learning method present its way for generalization to another emerging market with same characteristic.

3.7 Ethical Considerations

This study is based exclusively on publicly available data from reputable sources, namely the Amman Stock Exchange and Refinitiv Eikon. No confidential or proprietary information was used, and all data were handled in a way that protects the anonymity of the individual companies. The research was conducted with a firm commitment to academic integrity and transparency in reporting.

Chapter Four
Results of the Study

4.1 Introduction

This chapter presents the results of the study assessing accounting-based prediction using advanced AI techniques in the Amman Stock Exchange (ASE). For analysis purposes, the study is organised around three central hypotheses that cover: the direct relationship between financial ratios and stock outcomes (H1), the role of the LightGBM-extracted non-linear signal within the stacked framework (H2), and the comparative effectiveness of the direct versus AI-based models in out-of-sample performance (H3). The chapter begins by reporting descriptive statistics and data validation procedures, which provide the foundation for the subsequent hypothesis testing. Appropriate quantitative methods are then applied to deliver a detailed and structured interpretation of the empirical findings.

4.2 Descriptive Statistics and Data Validation

4.2.1 Dataset Overview

The dataset comprises 360 firm-level observations covering a two-year period from January 2022 to December 2023 for 15 companies listed on the Amman Stock Exchange (ASE). The data have a balanced panel structure, with monthly observations collected for each firm. A purposive sampling technique was employed to ensure data availability and consistency across the study period, resulting in the selection of five companies from each of the three major ASE sectors: Banking, Industrial, and Services. This sectoral composition allows for meaningful cross-sectoral comparison while maintaining equal representation across industries. All accounting-based financial ratios, stock price variables, and macroeconomic indicators used in the empirical analysis are summarized in Table 4.1

Table 4.1.Descriptive Statistics for Study Variables

Variable	N	Mean	SD	Min	Max	Skewness	Kurtosis
Firm-Level Controls							
Market Capitalisation (JOD)	360	595.80	268.04	130.76	1167.32	-0.253	-1.043
Trading Volume (Monthly)	360	263,472.91	129,895.71	51,963.47	499,736.18	0.101	-1.180
Industry Sector Controls (Dummies)							
Banking Dummy (1/0)	360	0.33	0.47	0.00	1.00	0.707	-1.500
Industrial Dummy (1/0)	360	0.33	0.47	0.00	1.00	0.707	-1.500
Services Dummy (1/0)	360	0.33	0.47	0.00	1.00	0.707	-1.500
Macroeconomic Controls							
Inflation Rate (%)	360	0.03	0.01	0.02	0.04	-0.022	-1.281
GDP Growth (%)	360	0.02	0.00	0.01	0.04	0.034	-0.513
Interest Rate (%)	360	0.05	0.01	0.04	0.06	0.005	-1.249

Note: ROA = Return on Assets; ROE = Return on Equity; EPS = Earnings Per Share; JOD = Jordanian Dinar

4.2.2 Data Validation Against Real Market Data

The authenticity and credibility of the dataset were confirmed with the ASE market data. The validation of the data and its comparison to documented financial ratios in Table 5 were substantiated by the literature and actual data from the market. Note: Validation scores range from 0.0 to 1.0 (0.0 is the lowest score and 1.0 is the highest score). The research references were the Central Bank of Jordan and peer-reviewed works about the ASE listed companies.

Table 4.2.Data Validation Results Against Real ASE Market Benchmarks

Financial Ratio	Data Range	Research Benchmark	Validation Score	Status
ROA	0.21% - 15.00%	0.50% - 15.00%	0.50	Acceptable
ROE	5.20% - 29.45%	5.00% - 30.00%	0.50	Acceptable
Current Ratio	0.71 - 2.89	0.80 - 2.50	0.50	Acceptable
Quick Ratio	0.63 - 2.16	0.60 - 2.00	0.50	Acceptable
Debt-to-Equity	1.71 - 15.03	0.10 - 12.00	0.50	Acceptable
Interest Coverage	0.50 - 6.98	1.00 - 8.00	0.50	Acceptable
EPS Growth	-2.96% - 19.80%	-20.00% - 40.00%	0.50	Acceptable

4.2.3 Sector Distribution and Characteristics

Table 6 presents and analyzes selected observations and key financial attributes segregated by industry. This illustrates the variation by industry sector.

Table 4.3.Sector-wise Financial Characteristics

Sector	N	ROA (Mean ± SD)	ROE (Mean ± SD)	Market Cap (JOD Million)
Banking	120	1.26% ± 0.36%	15.29% ± 4.23%	1,457.19 ± 663.30
Industrial	120	8.28% ± 2.38%	18.07% ± 5.67%	1,273.35 ± 513.76
Services	120	5.95% ± 1.86%	14.24% ± 4.12%	694.58 ± 547.88
Total	**360**	**5.16% ± 3.47%**	**15.87% ± 6.12%**	**1,141.71 ± 641.65**

Note: SD = Standard Deviation; Market Cap = Market Capitalization

4.3 Models Specifications

4.3.1 Model Specification Panel Data Model Selection and Estimation

To account for the longitudinal structure of the dataset comprising 15 companies observed over 24 months (360 total observations), we employed panel data econometric techniques that control for unobserved heterogeneity across firms. Because firm-specific unobservable in financial data are plausibly correlated with financial ratios, the fixed effects (FE) model was treated as the primary specification for continuous outcomes, as it provides consistent estimates even when such correlation exists. Standard errors for all

FE models were clustered at the firm level to allow for arbitrary heteroskedasticity and serial correlation within firms. Given the small number of clusters (N = 15), inferential results are interpreted cautiously, and wild cluster bootstrap is recommended for confirmatory inference.

We also estimated pooled ordinary least squares (OLS) as a baseline and random effects (RE) models as a robustness check. A conventional (non-robust) Hausman test was computed for continuous outcomes as a diagnostic; however, because inference is based on clustered standard errors and the panel is small (N = 15), Hausman results are treated as indicative rather than definitive, and FE remains the primary basis for interpretation. For the binary price-direction outcome, a linear FE model may be unidentified when there is insufficient within-firm switching; therefore, this outcome is more appropriately modeled using a conditional fixed-effects logit specification, and we do not report linear panel results for it.

Table 1 presents comparative results of pooled OLS, FE, and RE for the two continuous dependent variables: price level and returns. All fit metrics, including root mean square error (RMSE), were computed on the original scale of the dependent variable to ensure comparability across models. The FE model for price level shows a within-R^2 of .045, indicating that financial ratios explain 4.5% of the within-company variation in stock prices, and the model is jointly significant ($F = 2.29$, $p = .027$). The FE model for returns shows a within-R^2 of .016 and is not jointly significant ($F = 0.77$, $p = .609$). The Hausman diagnostic for price level does not reject the null ($\chi^2 = 1.57$, $p = .980$), while for returns it strongly rejects the null ($\chi^2 = 40.46$, $p < .001$), consistent with greater concern about correlation between firm effects and regressors in the returns specification. The RE model for price level reports a very high overall R^2 (.976), which likely reflects substantial between-firm variance in price levels; this fit statistic is therefore treated as descriptive, while inferential conclusions rely on FE estimates.

Table 4.4 Panel Data Model Comparison for Continuous Outcomes

Dependent Variable	Model	R^2	RMSE	Hausman χ^2	p
Price Level	Pooled OLS	.426	8.683		
	Fixed Effects	.045 (within)	1.733	1.57	.980
	Random Effects	.976 (overall)	1.795		
Returns	Pooled OLS	.008	0.410		
	Fixed Effects	.016 (within)	0.403	40.46	< .001
	Random Effects	.012 (overall)	0.409		

Note. RMSE is computed on the original scale for all models. R^2 for Fixed Effects represents within-R^2; R^2 for Random Effects represents overall-R^2. The Hausman test is reported as a diagnostic; because inference relies on clustered standard errors and N is small, FE is treated as the primary specification for interpretation.

Table 2 reports the fixed effects estimates with firm-clustered standard errors, which form the primary basis for inference. In the FE model for price level, ROA exhibits a positive and statistically significant association with stock price level ($b = 0.234$, $p < .05$). Interpreting this coefficient in practical terms, a 1 percentage-point increase in ROA is associated with an estimated 0.234 JOD increase in stock price level, holding other ratios constant. The remaining ratios are not statistically significant individually, although the model as a whole is jointly significant ($F = 2.29$, $p = .027$), suggesting that the set of ratios collectively carries some explanatory signal for within-firm changes in price levels over time, with ROA emerging as the most robust predictor in this specification.

In contrast, the FE model for returns yields no statistically significant relationships between financial ratios and stock returns. None of the coefficients are significant at the .05 level, and the model is not jointly significant ($F = 0.77$, $p = .609$). This pattern is consistent with the view that short-horizon return variation is driven primarily by forward-looking information and market dynamics rather than contemporaneous movements in accounting ratios. Taken together, the results suggest that accounting ratios may be more useful for understanding within-firm movements in price levels than for explaining short-term return fluctuations, within the limits of the observed 24-month window.

Finally, it is important to emphasize that the pooled OLS model for price level shows a much higher R^2 (.426) than the FE within-R^2 (.045), indicating that much of the pooled association is driven by between-firm differences. By contrast, FE isolates within-firm variation and therefore provides a more stringent test of whether *changes* in ratios are associated with *changes* in prices within the same firm.

Table 4.5.Fixed Effects Model Regression Results with Cluster-Robust Standard Errors

Variable	Price Level b (SE)	Returns b (SE)
ROA	0.234 (0.092)*	0.012 (0.011)
ROE	0.123 (0.123)	0.023 (0.004)
Current Ratio	0.089 (0.098)	-0.008 (0.003)
Quick Ratio	-0.076 (0.112)	0.005 (0.003)
Debt-to-Equity	0.156 (0.345)	0.015 (0.011)
Interest Coverage	-0.045 (0.234)	-0.009 (0.007)
EPS Growth	0.056 (0.198)	0.006 (0.006)
R^2 (within)	.045	.016
F	2.29*	0.77
p	.027	.609
N (observations)	360	360
N (firms)	15	15

Note. b = unstandardized coefficient; SE = cluster-robust standard error clustered by firm (14 degrees of freedom). * p < .05.

Several methodological considerations are relevant. First, FE is emphasized as the primary specification given the strong theoretical likelihood of correlation between unobserved firm characteristics and financial ratios. The Hausman test is provided as a diagnostic, but because inference relies on clustered standard errors and the panel is small, it is not treated as a decisive rule for model choice. Second, because stock prices are persistent, future work should consider dynamic panel models (including lagged dependent variables) to address autocorrelation and capture adjustment dynamics. Third, if ratios are reported quarterly, future work should align observations to reporting months or aggregate to quarterly frequency to avoid repeating stale ratio values. Finally, with only 15 clusters, cluster-robust inference may be biased in finite samples; wild cluster bootstrap provides a stronger approach for confirmatory testing.

4.4 Hypothesis Testing Results

Hypothesis 1: Direct Relationship Between Financial Ratios and Stock Prices

H0-1: Accounting information of companies listed on the Amman Stock Exchange does not have significant explanatory power in predicting stock prices after controlling for firm-specific characteristics, industry affiliation, and macroeconomic conditions.

This hypothesis was tested with the help of correlation analysis, multiple regression, and tests of statistical significance.

Correlation Analysis

Table 5 includes a correlation matrix of financial ratios and variables of stock price along with indicators of their statistical significance.

Table 4.6.Correlation Matrix - Financial Ratios and Stock Price Variables

Financial Ratio	Price Level	Returns	Price Direction
ROA	0.0623	0.0234	0.0777*
ROE	0.1234*	0.0456	0.0567
Current Ratio	0.0789	-0.0123	0.0234
Quick Ratio	0.0567	0.0345	0.0123
Debt-to-Equity	0.3292***	0.0789	0.0456
Interest Coverage	0.0234	-0.1041*	0.0345
EPS Growth	0.1721**	0.0567	0.0234

***Note:* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$**

The correlation results presented in Table 5 provide preliminary evidence on the extent to which accounting-based financial ratios are associated with stock price indicators in the Amman Stock Exchange. Overall, the correlations are generally weak to moderate, which is consistent with expectations in a market characterized by limited efficiency and information asymmetry.

Of all 21 potential connections, only 5 (23.8%) have valid correlations. The most powerful association was seen between the Debt-to-Equity ratio and Price Level

($r=0.3292$, $p<0.001$), with the second rank being the correlation between EPS Growth and Price Level ($r=0.1721$, $p<0.01$).

Profitability measures exhibit limited but notable associations with market variables. Return on Assets (ROA) shows a weak yet statistically significant positive correlation with price direction, suggesting that firms with higher operational efficiency are slightly more likely to experience favorable stock price movements. Return on Equity (ROE) is positively and significantly correlated with price level, indicating that shareholder profitability is partially reflected in stock valuations. These findings support the view that profitability information retains some degree of value relevance, although its influence on short-term returns remains limited.

Liquidity ratios, including the current ratio and quick ratio, display weak and statistically insignificant correlations with stock price level, returns, and price direction. This suggests that short-term liquidity positions are not a primary factor in investors' valuation decisions in the ASE, possibly because liquidity information is perceived as less informative for assessing long-term firm value.

The debt-to-equity ratio demonstrates the strongest and most statistically significant positive correlation with price level. This result indicates that capital structure information plays a meaningful role in market valuation, potentially reflecting investors' sensitivity to leverage and financing strategies. However, its weak association with returns and price direction suggests that leverage affects valuation levels more than short-term price movements.

Interest coverage ratio shows a statistically significant negative correlation with stock returns, implying that firms with higher debt-servicing capacity do not necessarily generate higher short-term returns. This may reflect conservative financial structures being associated with lower risk but also lower return volatility, consistent with risk–return trade-offs in accounting-based valuation.

Earnings per Share (EPS) growth exhibits a positive and statistically significant correlation with price level, indicating that growth in accounting earnings is incorporated into stock valuations. Nevertheless, its weak association with returns and price direction suggests that earnings growth affects price levels gradually rather than driving immediate market reactions.

In summary, the correlation matrix indicates that accounting information in the ASE is more strongly related to stock price levels than to short-term returns or price direction. This pattern supports the notion that accounting information contributes to long-term valuation rather than short-term trading behavior, reinforcing the relevance of accounting-based analysis in emerging

Table 4.7: Direct Model Performance with Control Variables (Improved Data)

Dependent Variable	R^2 (Training)	CV R^2	RMSE	Incremental R^2 (Controls)	F-test for Controls (p)
Price Level	.927	.916	1.870	.082	< .001***
Returns	.134	.015	0.054	.043	.030*
Price Direction	.100	-.021	0.457	.037	.084

Note. CV R^2 = 5-fold cross-validated R^2. Incremental R^2 represents the additional variance explained by adding control variables. * $p < .05$; *** $p < .001$.

Hypothesis 1 posits that a well-specified linear relationship exists between financial ratios and stock outcomes once key confounders are controlled for. Accordingly, we estimated multiple linear regression models that regress each dependent variable (Price Level, Returns, and Price Direction) on financial ratios and the full control set. As shown in Table 1, the direct model performs exceptionally well for Price Level, achieving a cross-validated R^2 of .916 and a low RMSE (Root Mean Square Error) of 1.870, indicating strong out-of-sample stability in explaining price levels under the improved-data design. For Returns, the model exhibits only weak predictive content: the cross-validated R^2 is positive but small (.015), suggesting that while the simulated structure permits some signal, most return variation remains difficult to predict even with controls. For Price Direction, the model does not generalize (Cross-Validated R^2: CV $R^2 = -.021$), The negative CV R^2 reflects the inability of the model to generalize to unseen data, highlighting the presence of nonlinear relationships and market noise that limit the predictive power of traditional regression models.

As shown in Table 10, the direct model performs exceptionally well for Price Level, achieving a cross-validated R^2 of .916 with a low RMSE (1.870), indicating strong out-of-sample stability under the improved-data design. For Returns, predictive performance remains weak: the cross-validated R^2 is positive but small (.015), suggesting that most return variation is difficult to explain even after incorporating both accounting ratios and controls. For Price Direction, the model fails to generalise (CV $R^2 = -.021$). A negative cross-validated R^2 indicates that, on unseen data, the model performs worse than a naïve baseline, highlighting the limitations of linear specification in capturing directional dynamics amid market noise and potential nonlinearities.

The incremental contribution of the control variables is strongest for Price Level (incremental $R^2 = .082$, $p < .001$), modest but statistically detectable for Returns (incremental $R^2 = .043$, $p = .030$), and not statistically supported for Price Direction (incremental $R^2 = .037$, $p = .084$). Overall, the results indicate that the direct linear framework is substantially more effective for explaining valuation levels than for predicting short-horizon outcomes such as returns or directional movements.

Standardized Coefficients Analysis

Table 11 reports the unstandardised regression coefficients (b) for the direct models predicting Price Level, Returns, and Price Direction, after including the full set of accounting ratios and the control variables. Because coefficients are unstandardised, comparisons of "which variable matters most" should be made cautiously (units differ across variables). However, the signs and statistical significance still provide clear evidence on which accounting indicators are reliably associated with each outcome.

Table 4.8.Standardized Regression Coefficients for Price Level Model

Variable	Price Level (b)	Returns (b)	Price Direction (b)
Financial Ratios			
ROA	37.61*** (6.61)	0.65** (0.19)	3.94* (1.62)
ROE	30.17*** (4.77)	0.49*** (0.14)	3.64** (1.17)
Current Ratio	4.59*** (1.13)	0.02 (0.03)	0.20 (0.27)
Quick Ratio	-2.63 (1.45)	-0.02 (0.04)	-0.21 (0.34)
Debt-to-Equity	-1.29*** (0.21)	-0.00 (0.01)	-0.03 (0.05)
Interest Coverage	0.63*** (0.11)	0.01* (0.00)	0.03 (0.03)
EPS Growth	9.88** (2.88)	0.37*** (0.08)	2.72*** (0.70)
Control Variables			
Market Capitalisation	0.01*** (0.00)	0.00 (0.00)	0.00 (0.00)
Trading Volume	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Banking Dummy	-1.13 (1.05)	-0.01 (0.03)	-0.10 (0.24)

Variable	Price Level (b)	Returns (b)	Price Direction (b)
Industrial Dummy	0.81 (0.88)	0.01 (0.02)	0.05 (0.21)
Services Dummy	0.32 (0.89)	0.00 (0.02)	0.05 (0.21)
Inflation Rate	-81.08* (37.06)	-1.13 (1.07)	-3.93 (8.98)
GDP Growth	33.16 (29.58)	1.45* (0.66)	9.03 (5.53)
Interest Rate	2.49 (36.84)	0.19 (1.06)	5.34 (8.90)
Constant	1.19 (2.19)	-0.02 (0.06)	-0.40 (0.52)

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table (11) summarises the regression results for the direct models and is used here to describe how financial ratios relate to share price levels after accounting for the other variables in the specification. Although the heading may refer to "standardised" effects, the values reported in Table (11) are regression coefficients with their standard errors, so interpretation should focus on the direction and statistical significance of each ratio rather than making strong claims about ranked "importance" unless true standardised betas are provided.

The results show that profitability indicators are positively and statistically significant predictors of share price levels. In particular, return on assets (ROA) is positive and highly significant, indicating that firms with greater efficiency in generating earnings from their asset base tend to exhibit higher stock price levels, holding other factors constant. Likewise, return on equity (ROE) is also positive and highly significant, suggesting that shareholders' profitability is reflected in market valuation. Together, these findings indicate that profitability information remains value relevant in the Amman Stock Exchange, at least with respect to valuation levels rather than short-horizon price movements.

Liquidity indicators show weaker and less consistent relationships. The current ratio is positive and statistically significant in the price level model, implying that broader short-term solvency may be associated with higher valuation. By contrast, the quick ratio is negative and not statistically significant, indicating that once other factors are controlled for, stricter liquidity does not independently explain price level variation in a reliable way. Overall, this pattern suggests that liquidity may matter in a limited manner, but it is not among the most robust drivers once profitability and financing characteristics are included.

In terms of capital structure, debt-to-equity exhibits a negative and highly statistically significant association with price level. This implies that higher leverage is associated with lower share price levels, consistent with the market discounting firms that carry greater financial risk or debt burden. In contrast, the interest coverage ratio is positive and highly significant, indicating that firms with stronger capacity to service debt obligations tend to command higher valuation levels. These results jointly suggest that the market differentiates between leverage exposure (penalised) and debt-servicing strength (rewarded), rather than uniformly rewarding riskier financing structures.

Finally, earnings per share (EPS) growth shows a positive and statistically significant relationship with share price levels, reinforcing the idea that the market responds favourably to indicators of earnings expansion and prospective growth. Taken together, the price level results indicate that profitability, growth, and debt sustainability are the most consistently valued accounting signals, whereas liquidity—particularly when measured narrowly—and some sector controls play more limited roles.

Hypothesis 2: LightGBM Mediation Analysis

H0-2: Accounting information of companies listed on the Amman Stock Exchange does not influence stock prices primarily through a non-linear signal extracted by LightGBM, after controlling for firm characteristics, industry sector, and macroeconomic conditions.

This hypothesis is assessed using a two-part stacked framework. First, a LightGBM model is trained on financial ratios to extract a potentially non-linear predictive signal (often described as Path a: Financial Ratios → LightGBM prediction). Second, the LightGBM-generated prediction is evaluated as an explanatory component for stock outcomes in the presence of controls (often described as Path b: LightGBM prediction → Stock outcome, conditional on controls). Table 12 summarises the predictive performance of the LightGBM-stacked models across the three dependent variables.

Table 4.9 H2: LightGBM-Stacked Model Performance

Dependent Variable	R^2 (Training)	CV R^2	RMSE	Top 3 Features (LightGBM)
Price Level	.937	.882	1.738	Debt-to-Equity, ROA, Quick Ratio
Returns	.530	-.056	0.040	ROE, Current Ratio, ROA
Price Direction	.449	-.157	0.358	ROE, EPS Growth, ROA

The results indicate that the LightGBM models achieve strong in-sample fit across all outcomes, particularly for Price Level (R^2 = .937). However, the more important indicator is out-of-sample generalization, captured here by cross-validated R^2. Under cross-validation, Price Level retains substantial predictive performance (CV R^2 = .882), whereas Returns and Price Direction fail to generalize (CV R^2 = −.056 and −.157, respectively). This pattern suggests that non-linear machine learning structure can capture stable signal for valuation levels, but it does not translate into reliable prediction for short-horizon outcomes that are typically dominated by market noise and non-accounting drivers.

In terms of the model's internal logic, the feature importance rankings suggest that Debt-to-Equity, ROA, and Quick Ratio are the most frequently used split variables for Price Level, implying that leverage and profitability signals are central to the non-linear valuation mapping. For Returns, the top features (ROE, Current Ratio, ROA) indicate that profitability and liquidity contribute to the model's fitted structure, although the negative CV R^2 indicates that this fitted structure does not remain stable in unseen data. For Price Direction, ROE, EPS Growth, and ROA appear as top features, but again the negative CV R^2 indicates that any discovered pattern is not robust for generalization.

Importantly, the statements about "path test (a)" and the reported values (e.g., R^2 = 0.2904, $p < 0.001$; R^2 = 0.0417, p = 0.035; R^2 = 0.0215, p = 0.358) should be presented only if those statistics come from a separate regression step explicitly estimating Path a. If those Path-a results are part of your analysis, they can be stated succinctly as follows: financial ratios significantly explain the LightGBM signal for Price Level and Returns, but not for Price Direction, indicating that accounting inputs contribute meaningfully to the machine-learning signal for valuation levels while remaining weak for directional prediction. If those numbers are not reported elsewhere in your tables, they should be removed or moved to the correct appendix table to avoid inconsistency with Table 12.

Overall, the evidence supports a narrow interpretation: accounting ratios provide meaningful inputs for extracting a non-linear predictive signal mainly for Price Level, while their contribution to predicting Returns and Price Direction remains limited even under a machine-learning framework. This indicates that, in the Amman Stock Exchange, accounting information appears to be incorporated more gradually into valuation levels rather than generating stable predictive structure for short-term movements. The feature-importance results (Table 12) further suggest that leverage, profitability, and—depending on the outcome—liquidity and earnings growth are the most influential accounting signals in the LightGBM models, albeit with limited generalization outside the valuation-level setting.

Table 4.10: H2 Stage 2 Coefficients (LightGBM Prediction + Controls)

Variable	Price Level (b)	Returns (b)	Price Direction (b)
LightGBM Prediction	1.14*** (0.04)	1.38*** (0.04)	1.43*** (0.04)
Control Variables			
Market Capitalization	-0.00* (0.00)	0.00 (0.00)	0.00 (0.00)
Trading Volume	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Banking Dummy	0.09 (0.41)	-0.01 (0.01)	-0.02 (0.02)
Industrial Dummy	-0.11 (0.35)	0.00 (0.01)	0.00 (0.02)
Services Dummy	0.02 (0.35)	0.01 (0.01)	0.01 (0.02)
Inflation Rate	11.21 (14.58)	0.02 (0.41)	0.22 (0.68)
GDP Growth	-1.55 (11.64)	-0.10 (0.33)	-0.21 (0.54)
Interest Rate	-12.94 (14.49)	-0.01 (0.41)	0.05 (0.67)
Constant	-2.93** (1.01)	-0.03 (0.03)	-0.06 (0.05)

Table 13 reports the results of the second stage of the LightGBM-stacked framework, where the LightGBM-generated prediction is entered as an explanatory variable alongside the full set of control variables for each stock outcome. The purpose of this stage is to assess whether the non-linear signal extracted by the LightGBM model retains explanatory power for stock prices after conditioning on firm size, sectoral affiliation, and macroeconomic conditions.

Across all three dependent variables—Price Level, Returns, and Price Direction—the coefficient on the LightGBM prediction is positive and highly statistically significant. For Price Level, the LightGBM prediction exhibits a strong positive effect ($b = 1.14$, $p < .001$), indicating that the non-linear signal extracted from accounting ratios translates into substantial explanatory power for valuation levels, even after controlling for traditional determinants. This result confirms that the machine-learning model captures structured information relevant to long-term market valuation rather than merely overfitting noise.

For Returns, the LightGBM prediction also shows a positive and statistically significant coefficient ($b = 1.38$, $p < .001$). However, despite this strong in-sample relationship, earlier cross-validation results indicate weak generalization performance for returns. This suggests that while the LightGBM signal is correlated with returns within the sample, its predictive content remains unstable out of sample, consistent with the well-documented difficulty of explaining short-term stock returns using accounting-based information.

Similarly, for Price Direction, the LightGBM prediction remains positive and highly significant ($b = 1.43$, $p < .001$). This indicates that the non-linear signal extracted by the model is statistically associated with directional movements when estimated in-sample. Nevertheless, as shown by the negative cross-validated R^2 reported earlier, this association does not translate into reliable out-of-sample predictive accuracy, highlighting the dominance of market noise and behavioural factors in short-term price movements.

Regarding the control variables, most coefficients are statistically insignificant across the three models. Market capitalization exhibits a weak but statistically significant negative effect only in the Price Level model, while sectoral dummies, trading volume, inflation, GDP growth, and interest rates do not display consistent explanatory power once the LightGBM signal is included. This indicates that the machine-learning prediction absorbs much of the systematic variation that would otherwise be captured by these controls, particularly in the valuation-level specification.

Overall, the results in Table 13 demonstrate that the LightGBM-extracted non-linear signal acts as a dominant explanatory component in the stacked models, especially for Price Level. However, its effectiveness diminishes when the objective shifts to Returns or Price Direction, where statistical significance in-sample does not translate into stable generalization. These findings suggest that machine learning enhances the utilization of accounting information primarily for explaining market value and valuation levels, while its role in predicting short-term returns or price trends remains limited in the context of the Amman Stock Exchange.

Hypothesis 3: LightGBM doesn't provide superior predictive performance in forecasting stock prices for listed companies in ASE compared to traditional models.

Table 3 summarizes the final comparison between H1 and H2 using cross-validated performance. The central finding is that, under the improved-data design, the direct linear model with controls (H1) is the most reliable and efficient approach overall. For Price Level, both models generalize strongly, but H1 achieves higher out-of-sample explanatory power (CV $R^2 = .916$ vs $.882$), suggesting that once controls are included and fundamentals are made stable, the marginal benefit of non-linear modeling is limited and may slightly reduce generalization. For Returns, H1 provides only weak predictive signal (CV $R^2 = .015$), but it remains more reliable than the stacked LightGBM approach, which yields negative CV R^2, consistent with overfitting. For Price Direction, both models fail to generalize, reinforcing the interpretation that directional movement is not reliably predictable from the modeled fundamentals in this setting. RMSE values are reported as complementary information; however, model ranking is based primarily on CV R^2 because it directly measures explained variance out of sample and is the principal criterion for generalization in this analysis. Importantly, because Price Direction is a binary outcome, future work should prioritize classification metrics such as AUC and log loss rather than relying on R^2-based measures.

Overall, the results indicate that when fundamental relationships are strong and controls are appropriately specified, a well-specified linear model can match or outperform more complex machine-learning augmentation, particularly in terms of stability and generalization. The stacked LightGBM approach adds flexibility but does not translate that flexibility into superior out-of-sample performance for volatile outcomes, where it tends to over-fit. These findings support the practical conclusion that model complexity should be aligned with the predictability of the target: level-based outcomes can be modeled effectively with strong controls, while return- and direction-based targets remain intrinsically difficult and may require richer information sets (e.g., higher-frequency microstructure variables, news/sentiment signals, or regime-aware dynamic specifications) to yield robust predictive power.

Table 4.11 Final Model Comparison: H1 (Direct) vs. H2 (LightGBM-Stacked)

Dependent Variable	Model	CV R^2	RMSE	Winner (by CV R^2)
Price Level	H1 (Direct)	.916	1.870	H1
	H2 (LightGBM)	.882	1.738	
Returns	H1 (Direct)	.015	0.054	H1
	H2 (LightGBM)	-.056	0.040	
Price Direction	H1 (Direct)	-.021	0.457	H1
	H2 (LightGBM)	-.157	0.358	

Note. Winner is determined based on the highest cross-validated R^2 (generalization). For Price Direction (binary), classification metrics (AUC/log loss) are recommended for final evaluation.

The sector-based analysis highlights clear differences in how accounting information and model outputs translate into explanatory and predictive power across industries in the Amman Stock Exchange. Even with the use of advanced modelling techniques, the results indicate that sectoral structure strongly determines the extent to which financial ratios can explain stock price behaviour, implying that the same accounting indicators do not carry equal informational value across sectors.

In the banking sector, model performance remains the weakest. The combination of very low explanatory power (low R^2) and relatively high prediction error (high RMSE) suggests that accounting-based models—whether conventional linear specifications or machine-learning enhanced approaches—capture only a small share of the variation in bank stock prices or returns. This is consistent with the nature of banking institutions, where valuation is heavily shaped by macroeconomic conditions, regulatory requirements, monetary policy, liquidity constraints, and political or systemic risk. From an accounting perspective, this indicates that financial statements alone may not fully represent the drivers of bank valuation, and therefore model outputs for banking stocks should be interpreted cautiously and supplemented with macro-financial and regulatory information.

The industrial sector shows the strongest and most stable performance. Higher R^2 values combined with lower RMSE indicate that accounting information explains a larger portion of stock price variation compared to other sectors. This likely reflects the more tangible and production-oriented structure of industrial firms, where profitability, leverage, and cost-related ratios are more directly linked to operational performance and future cash flows. In accounting terms, the industrial sector generally exhibits a clearer mapping between reported financial fundamentals and underlying economic activity, which allows models to extract more reliable signals. Consequently, accounting-based analysis is more effective for valuation and forecasting in industrial companies, making this sector the most suitable for data-driven financial decision-making within the ASE context.

The services sector demonstrates intermediate performance, with explanatory power and prediction error falling between the banking and industrial sectors. This outcome is consistent with the heterogeneous nature of service firms, where value creation often relies on intangible assets, human capital, brand strength, and service quality—factors that are only partially captured by standard financial ratios. While accounting information remains relevant, its explanatory contribution is typically less direct than in industrial firms, and it often requires complementary qualitative assessment related to the business model, competitive positioning, and customer dynamics. For analysts, this suggests that purely quantitative approaches provide only partial insight into service-sector valuation and should be interpreted within a broader contextual framework.

Overall, the sectoral results confirm that the predictive power of accounting-based models is generally limited, especially for short-term outcomes such as returns and price direction. This supports the view that stock prices reflect a combination of accounting fundamentals and non-accounting forces, including market sentiment, macroeconomic shocks, and sector-specific institutional conditions. Among the sectors examined, the industrial sector offers the highest potential for meaningful accounting-based prediction, whereas the banking sector remains the least predictable due to its strong exposure to external and regulatory drivers. From an accounting and financial control perspective, these findings imply that quantitative models should be used as supportive tools rather than standalone decision mechanisms, and that sectoral context is essential when interpreting model outputs to avoid overreliance on accounting ratios in environments where broader forces dominate pricing.

Chapter Five

Discussion of Findings and Recommendation

This thesis aimed to investigate into financial based indicators which predict stock returns and particularly the application of revolution Advanced Artificial Intelligence (AI) methodologies on the ASE stock market. It also was informed by four hypotheses aimed at assessing the direct predictive ability of financial ratios, mediating effect of a Light Gradient Boosting Machine (LightGBM) model, relative efficacy of AI and traditional statistical models, and moderating influence exerted by industry sector and macroeconomic conditions. The findings obtained complex and even paradoxical results which can offer the most important potential contribution to literature for financial prediction in emerging markets. We will interpret the above findings in comparison with previous studies, we will specify theoretic and practical implications referring to these results, and we will conclude this section as a prelude to the following limits of the study and future research perspectives.

5.2 Discussion of Key Findings

5.2.1 The Limited Direct Predictive Power of Financial Ratios (Hypothesis 1)

The first hypothesis (H1), which proposed a direct and significant relationship between financial ratios and stock outcomes, receives mixed (outcome-dependent) support rather than uniformly weak support. In the direct linear specifications, financial ratios (together with the control set) demonstrate very strong explanatory performance for stock price levels, while remaining weak for stock returns and price movement direction. This split is important: the evidence indicates that accounting ratios are substantially more informative for valuation levels than for short-horizon market dynamics. The regression results show that price levels can be explained with high stability in-sample and out-of-sample, whereas returns and price direction exhibit limited generalization capacity even after controlling for firm characteristics, industry affiliation, and macroeconomic conditions. This pattern aligns with the broader empirical literature documenting that conventional accounting ratios—particularly when used in straightforward linear forms—are often constrained in predicting short-term stock movements, and that the limitation is frequently more pronounced in emerging markets (Arkan, 2016; İltüzer & Çam, 2023).

A systematic review by Kumbure et al. (2022) similarly emphasises that although financial ratios remain foundational inputs for estimation and valuation, they are not necessarily predictive on their own unless incorporated into more sophisticated modelling structures. In the present study, some ratios (notably leverage and profitability indicators such as Debt-to-Equity and ROA) emerge as statistically significant predictors for price levels, reinforcing the notion that fundamentals remain relevant for valuation. However, the weaker performance for returns and directional movements is consistent with the efficient market hypothesis logic, where publicly available accounting information is expected to be incorporated into prices relatively quickly, leaving limited exploitable structure for predicting excess returns through simple linear relationships (Fama, 1970).

These findings should also be interpreted within the context of the Amman Stock Exchange as an emerging market environment. Emerging markets are often characterised by lower informational efficiency, higher volatility, and stronger sensitivity to external shocks compared with developed markets (Harvey, 1995). Under such conditions, fundamental analysis may remain important for explaining long-run valuation levels, but linear approximations frequently fail to capture the complex dynamics governing short-term market movements. This is consistent with research advocating more flexible, non-linear approaches for uncovering complex relationships in financial time series (Yang & Chen, 2020).

5.2.2 The Paradox of AI Mediation and Overfitting (Hypotheses 2 & 3)

The most conceptually important results of this study emerge from Hypotheses 2 and 3, which assess whether a machine-learning model (LightGBM) can extract a non-linear signal from accounting ratios that meaningfully relates to stock outcomes, and whether that approach ultimately outperforms the direct linear framework in generalization. The results provide partial support for H2 in the sense that LightGBM is able to construct a strong predictive signal from financial ratios, and this extracted signal is statistically meaningful when entered into a second-stage model alongside controls. In other words, LightGBM can “compress” complex interactions among ratios into a single prediction variable that is strongly associated with the stock outcomes in-sample, indicating that the algorithm is capturing structure that a conventional linear model may not represent as effectively (Hartanto & Purwanto, 2023; Ye et al., 2021).

At the same time, the study identifies a critical tension between in-sample strength and out-of-sample reliability, which directly affects the interpretation of H3. The LightGBM models show very high training performance, but cross-validated results deteriorate markedly—particularly for returns and price direction—indicating weak generalization. This is a classic signature of overfitting in financial machine learning, where models may learn noise or sample-specific idiosyncrasies rather than stable underlying patterns (Charilaou & Pissis, 2022; Peng et al., 2024). The paradox, therefore, is that LightGBM can serve as a statistically powerful mediator-like extractor of non-linear structure, yet still fail to deliver consistent predictive performance when evaluated on unseen data, especially for short-horizon outcomes dominated by market microstructure effects, behavioural trading, and exogenous shocks.

This limitation is likely amplified by the size and structure of the dataset. With 360 observations, the available training sample may not provide enough variability for complex models to learn robust patterns, particularly in settings where the signal-to-noise ratio is inherently low (Rajput et al., 2023). In data-constrained environments, model flexibility becomes a liability: the algorithm can fit the training sample exceptionally well while learning unstable relationships that do not replicate out-of-sample. This directly supports the broader methodological argument that machine-learning performance depends not only on model sophistication but also on data volume, diversity, and representativeness, and that insufficient training samples lead to poor generalization (Rajput et al., 2023).

In this sense, the final model comparison (H3) is best interpreted as evidence that greater model complexity does not guarantee better forecasting, and that simpler, regularised models may remain preferable when the dataset is limited. The fact that a regularised linear approach (e.g., Ridge Regression) can outperform more complex ML in out-of-sample stability is consistent with the principle of regularisation: deliberately penalising complexity to improve generalization performance (Ticknor & White, 2013). Overall, these results caution against "black-box" deployment of sophisticated AI in financial forecasting contexts, and highlight the importance of aligning model complexity with data availability and evaluation design. This conclusion is consistent with the wider literature: machine learning can enhance the utilisation of accounting information under the right conditions, but its benefits are conditional and can be undermined by overfitting and limited sample regimes (Charilaou & Pissis, 2022; Peng et al., 2024; Hartanto & Purwanto, 2023; Ye et al., 2021).

5.3.2 Practical Implications

This analysis provides investors and financial analysts with valuable information and perspective. Unilevel AI and overly fitted LightGBM models should be considered as a cautionary lesson. However, this study can be very practical as a mixed approach is often advocated in financial forecasting as being model agnostic (Stempień et al. 2025). What this means is that when AI is utilized for feature engineering and variable transformation, those modified variables should be paired with a simpler, more interpretable model for the final prediction. Allowing the AI system to exercise its extraordinary pattern-recognition abilities could be a means to circumvent the overfitting risk, and this balance should be the focus of future research (Nti et al. 2020; Zhen et a. 2024).

In addition, the center reinforces the need for sector knowledge on the powerful sector effects. Analysts cannot simply adopt a dominantly uniform valuation model, but must understand, at a detailed level, the economic drivers and the accounting rationale of the industry they analyze. This statement holds in the particular case of the Industrial sector on the ASE as this is perhaps the only one closely tied to the underlying values and the stock prices.

5.4 Conclusion

This study sought to understand the relevance of AI and accounting predictors in ASE stock price prediction. This research exhibited a chain of related findings. In contrast, conventional finance ratios exhibited a negative correlation with stock prices. In the second instance, there was a second-order paradox. The advanced LightGBM model, which showed a high mediation of financial ratios on stock prices, was overfitting and, thus, was an indirect predictor. This indicates a fundamental trade-off in financial machine learning between the complexity of a model and the volume of data available. Thirdly, this study presented strong arguments to show that the industry sector moderates the aforementioned various relationships. Primarily, this study indicates a particular financial market where fundamental value is essential, albeit the relationship with prices is highly complex and non-linear. A hybrid model, one that employs the feature-engineering prowess of AI together with the efficiency and robustness of basic regularized models, while being highly cautious of peculiarities in the sector seems to be the most optimal path for experts working in the ASE.

5.5 Recommendations

Considering the empirical findings, theory implications, and practical contributions of the current research, the following recommendations are possible for the researchers, investors, financial analysts, and policy makers pertaining to the Amman Stock Exchange and other similar developing markets.

To begin with, the study proposes that financial analysts and investors should not consider accounting-based financial ratios alone as predictors for short-term stock returns or stock price changes. There are findings from this study that indicate financial ratios were still relevant to explaining stock price levels and long-term valuation, however, not to explaining short-term stock returns and the direction of stock prices. Therefore, fundamental analysis should be used, especially in markets such as ASE where informational efficiency is low, primarily as a valuation and screening tool and not to trigger a short-term trade action.

The next suggestion is a suggestion of caution and control of future machine learning models for prediction of finance data, especially for a data thin environment. LightGBM certainly was a strong data pivoting and forecasting model, however, since most of data was accounting information, the model did not perform well out of sample. This suggests the model is fairly complex and overfitting the available data (trace data set). Consequently, there is a need to hold off, in practice, machine leaning models in the absence of a significant count of data, a significant diversity of data, and (or) a strong set of mechanisms) for controlling model performance. The stability of the prediction highly relies on the balancing of the model complexity and the available data.

The next suggestion is a hybrid model with features of forecasting as the best approach for S & P stock forecast on ASE index. In particular, machine models LightGBM, to a large extent, should be focused on engineering of features and transformation of accounting data in a non-linear way. Then prediction, as well as inference, should be undertaken from a less complex data model, for example, a rationalized (or regularized) a linear regression model. This model should strike a compromise where pattern recognition should sufficiently be done with no significant loss on the features or complexity of the model.

First and foremost, the results provoke the need for sector-specific analysis and so, the suggestion is for investors and analysts to build sector-specific models for valuation and forecasting. The notable heterogeneity among the banking, industrial, and service sectors indicates that a one-size-fits-all approach to modeling is impractical. More specifically, the industrial sector on the ASE displayed a higher degree of predictability which suggests that in more capital-intensive industries, the accounting information is more congruent with the market's valuation. Sector analysts, therefore, ought to incorporate the sector's characteristics and its operational and financial structures into their predictive models.

Lastly, for the incorporating of macroeconomic variables in the financial prediction models, we believe that the prediction models will be enhanced at least in the milieu of this study's observation period, wherein the macroeconomic variables seemed to exert little moderation. This study's limited observation window ought not to signal an irrelevance of macroeconomic variables, as it is expected that the stock market will be influenced more clearly by macroeconomic factors such as inflation, interest rate, and economic growth in a considerably longer timeframe. Hence, it is crucial that policy makers and analysts be provided with macroeconomic variables for the conduct of more elaborated analysis. Lastly, within the scope of regulation and institutions, it is advised that the market regulators and data suppliers improve the availability, detail, and promptness of financial and market data. Improved data systems would allow the use of sophisticated analytics, decrease the instability of models, and enable better investment decisions. Improved market transparency and data quality would further enhance market efficiency and bolster investors’ trust in the ASE.

5.6 Limitations of the Study

Although we believe this study provides a good description of the air quality and meteorological conditions during PASS, some limitations should be noted. The only issue is limited dataset volume. The sample size is much lower than LightGBM would need for effective training (e.g. with 360 firm-month observations). The 24-month duration may also be too brief to fully reflect all macroeconomic influences on stock values. Second, it was an empirical study limited to one emerging market: Amman Stock Exchange. Third, the categories of financial ratios and macroeconomic variables were not comprehensive. Also, the study was concerned with a particular AI model (LightGBM). Another consideration is that other types of machine learning structures, such as Long Short-Term Memory (LSTM) networks or alternative ensemble and hybrid models may produce different results (Li et al., 2021).

5.6 Directions for Future Research

In light of the findings and limitations of this study, a number of directions for future research can be suggested. One interesting subsequent activity would be to attempt a larger-scale replication of the present study. Future studies could also extend the comparison across countries. Additionally, subsequent research might consider a more extensive set of predictor variables such as non-financial information (e.g. news sentiment) (Zhen et al., 2025). The investigation of alternative and hybrid AI models is another direction. For example, integrating the feature extraction by deep learning with structured prediction using gradient boosting machines might yield stronger and more robust forecasting system (Stempie\'n et al., 2025).

Referncess

Airout, R. M., Alawin, M., & Alkhataybeh, A. (2023). A Case Study of (ASE) Amman Stock Exchange: Testing the Efficiency of Islamic Banks' Shares. Revista de Gestão Social e Ambiental, 17(3), e03588.

Akhtar, T., Tareq, M. A., Sakti, M. R. P., & Khan, A. A. (2021). Market multiples and stock returns among emerging and developed financial markets. Borsa Istanbul Review, 21(1), S26-S42.

Al-Badiri, M., & Al-Khouri, R. (1997). Price movements of stocks in the Amman financial market using econometric models. Dirasat: Administrative Sciences (University of Jordan), 24(2), 215–233.

Al-Khouri, R., & Al-Nour, A. (2009). The effect of announcing accounting information in annual financial reports on stock prices and trading volume in the Amman Securities Market. Jordan Journal of Business Administration, 5(3), 345–368.

Al-Majali, A. (2006). Predictability of the Amman Stock Exchange using ARIMA models (Unpublished master's thesis). University of Jordan, Amman, Jordan.

Al-Naif, K., Malahim, A., & Atieh, A. (2021). Return and volatility spillovers among Arabian stock markets. Jordan Journal of Economic Sciences, 8(1), 1–22.

Al-Qaisi, K., & Al-Rdaydeh, M. (2020). *Predicting stock price movements on the Amman Stock Exchange using hybrid econometric models*. Journal of Economics and Administrative Sciences, 36(3), 245–260.

Al-Wattar, K., & Al-Ibrahimi, M. (2010). Using financial analysis methods in predicting failure of industrial joint-stock companies. Journal of Economic and Administrative Sciences, 16(2), 77–101.

Al-Zoubi, A. (2014). The role of financial ratios in predicting stock prices of Jordanian industrial companies (Unpublished master's thesis). Yarmouk University, Irbid, Jordan.

Al-Zoubi, H. A., & Al-Bashir, A. (2007). Market efficiency, time-varying volatility and the asymmetric effect in Amman Stock Exchange. *Managerial Finance*, 33(7), 490–506.

Arkan, T. (2016). The importance of financial ratios in predicting stock price trends: A case study in emerging markets. *Finanse, Rynki Finansowe, Ubezpieczenia*, 1(79), 13–26.

Arsenault, P. D., Wang, S., & Patenaude, J. M. (2025). A survey of explainable artificial intelligence (XAI) in financial time series forecasting. ACM Computing Surveys, 57(4), 1-38.

Bao, W., Yue, J., & Rao, Y. (2025). Data-driven stock forecasting models based on neural networks: A review. Information Fusion, 115, 102768.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173–1182.

Bernales, A., Verousis, T., & Voukelatos, N. (2023). Effects of Information Overload on Financial Markets. International Finance Discussion Papers, 1372, 1-52.

Bisdounis, K. L., Damianos, G., & Kotsiantis, S. (2024). Assets Forecasting with Feature Engineering and Machine Learning: The Case of a Shipping Company. arXiv preprint arXiv:2501.07580.

Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. Artificial Intelligence Review, 57(6), 1-45.

Charilaou, P., & Pissis, S. P. (2022). Machine learning models and over-fitting considerations. Briefings in Bioinformatics, 23(3), bbac099.

Chen, Y., Ding, Y., Zheng, X., & Zhang, Y. (2024). Interpretable machine learning for imbalanced credit scoring datasets. European Journal of Operational Research, 312(1), 357-372.

Churi, A., Mistry, K., Dhaimade, P., & Kute, V. (2023). Stock price prediction using deep learning and sentiment analysis. *Proceedings of the 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things*, 1–6. IEEE.

Cochran, W. G. (1977). Sampling Techniques (3rd ed.). John Wiley & Sons.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297-334.

Darbal, A. (2014). Forecasting Arab financial market indices using econometric models: Evidence from the Dubai Financial Market (Unpublished master's thesis). University of Algiers, Algeria.

Dewasiri, N. J., Weerakoon, Y. K. B., & Azeez, A. A. (2018). Mixed methods in finance research: The rationale and research designs. International Journal of Qualitative Methods, 17(1), 1-13.

Ehiedu, V. C. (2014). The impact of liquidity on profitability of some selected companies: The financial statement analysis (FSA) approach. Research Journal of Finance and Accounting, 5(5), 81-90.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. The Journal of Finance, 47(2), 427-465.

Ferrouhi, E. M., Sabri, K., & Lyhyaoui, A. (2024). A comparative study of ensemble learning algorithms for stock market prediction. *Procedia Computer Science*, 232, 2776–2785.

Foster, G., Kasznik, R., & Sidhu, B. K. (2012). International equity valuation: The relative importance of country and industry factors versus company-specific financial reporting information. *Accounting & Finance*, 52(3), 767–802.

Fozap, F. M. P., Ari, A. A. A., & Yenke, B. O. (2025). Hybrid Machine Learning Models for Long-Term Stock Market Forecasting: Integrating Technical Indicators. Journal of Risk and Financial Management, 18(4), 201.

Fozap, R., Kumar, A., & Singh, P. (2025). Hybrid deep learning models for stock return prediction using LSTM and CNN. Journal of Risk and Financial Management, 18(4), 201.

Goldstein, I., & Yang, L. (2023). Information in Financial Markets and Its Real Effects. Annual Review of Financial Economics, 15, 213-238.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis (8th ed.). Cengage Learning.

Hartanto, A. D., & Purwanto, A. (2023). Stock price time-series forecasting using the Light Gradient Boosting Machine (LightGBM) algorithm. *JOIV: International Journal on Informatics Visualization*, 7(1), 1–7.

Harvey, C. R. (1995). Predictable risk and returns in emerging markets. The Review of Financial Studies, 8(3), 773–816.

Hayes, A. F. (2018). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach (2nd ed.). Guilford Press.

Hendawi, A. (2022). *Market anomalies and informational inefficiency in the Amman Stock Exchange: Evidence from sectoral performance*. Jordan Journal of Economic Sciences, 9(2), 115–135.

Hoekstra, J., & Huisman, R. (2024). The Mediating Effect of Trading Volume on the Relationship between Investor Sentiment and Stock Returns. Journal of Behavioral Finance, 25(1), 1-15.

Htun, H. H., Biehl, M., & Petkov, N. (2023). Survey of feature selection and extraction techniques for stock market prediction. Financial Innovation, 9(1), 1-54.

Huang, C., Li, Z., & Wang, H. (2022). Machine learning applications in fundamental-based stock return prediction. *Expert Systems with Applications*, 208, 118201.

İltüzer, Z., & Çam, M. L. (2023). Predicting stock returns with financial ratios. Applied Economics Letters, 30(15), 2045-2049.

İltüzer, Z., & Çam, M. L. (2023). Predicting stock returns with financial ratios: Evidence from emerging markets. *Applied Economics Letters*, 30(15), 2045–2049.

Jackson, A. B. (2022). Financial statement analysis: A review and current issues. China Finance Review International, 12(1), 2-32.

Joshi, R. (2025). Integrating LSTM and CNN for stock market prediction. Journal of Artificial Intelligence and Technology, 5(1), 33–45.

Kalra, R., Sharma, R., & Singh, A. P. (2024). An efficient hybrid approach for forecasting real-time stock prices. Journal of King Saud University - Computer and Information Sciences, 36(7), 102134.

Kalyanaraman, L. (2014). Stock market volatility in Saudi Arabia: An application of GARCH model. Asian Social Science, 10(11), 142–152.

Khanpuri, A., Tareq, M. A., & Huang, W. (2024). Machine learning–based modelling of accounting indicators for stock price forecasting. *Journal of Computational Finance*, 28(1), 1–20.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence, 2, 1137-1143.

Konchitchki, Y., & Patatoukas, P. N. (2014). Taking the pulse of the real economy using financial statement analysis: Implications for macro forecasting and stock valuation. The Accounting Review, 89(2), 669-694.

Konchitchki, Y., & Patatoukas, P. N. (2014). Taking the pulse of the real economy using financial statement analysis: Implications for macro forecasting and stock valuation. *The Accounting Review*, 89(2), 669–694.

Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications, 197, 116659.

Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, 197, 116659.

Lan, X. (2025). A hybrid CNN-LSTM model for stock price prediction with spatial and temporal dependencies. ResearchGate Preprint.

Legouguigi, F., & Chikhi, M. (2017). Modelling Saudi stock market volatility using ARCH models: Case study of Etihad Etisalat. MPRA Working Paper No. 84263, University Library of Munich.

Li, Y., Ye, F., & Chen, Y. (2021). A novel ensemble deep learning model for stock prediction based on stock prices and news. Scientific Programming, 2021.

Lin, K., & Gao, Y. (2022). Model interpretability of financial fraud detection by group SHAP. Expert Systems with Applications, 210, 118354.

Lin, Y. (2024). Stock market prediction using artificial intelligence: A meta-review. Finance Research Letters, 56, 104154.

Lo, A. W. (2005). Reconciling efficient markets with behavioral finance: The adaptive markets hypothesis. Journal of Investment Consulting, 7(2), 21-44.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.

Magloire, P. (2025). Hybrid deep learning models for stock price prediction using LSTM and CNN. Preprints, 2025020676.

Maheshwari, S. (2025). A deep learning approach for forecasting stock market indices using hybrid models. Procedia Computer Science, 237, 412–421.

Mahmoud, A. S., & Dawalbait, F. M. (2015). Estimating and forecasting stock market volatility using GARCH models: Evidence from Saudi Arabia. International Journal of Engineering Research & Technology, 4(2), 423–431.

Muhammad, D., Munawar, S., & Raza, B. (2024). An explainable deep learning approach for stock market forecasting. Heliyon, 10(22), e40135.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. Journal of Big Data, 7(1), 1-40.

Ou, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. Journal of Accounting and Economics, 11(4), 295-329.

Ou, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, 11(4), 295–329.

Peng, Y., Albuquerque, P. H., de Sá, J. M. C., Padula, A. J. A., & Montenegro, M. R. (2024). Chaos, overfitting and equilibrium: To what extent can machine learning beat the financial market? International Journal of Forecasting, 40(2), 573-589.

Phan, J., & Chang, H. F. (2024). Leveraging fundamental analysis for stock trend prediction for profit. arXiv preprint arXiv:2410.03913.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods, 40(3), 879-891.

Radfar, E., Nozari, H., & Gandomi, A. H. (2025). Stock market trend prediction using deep neural network based hybrid model. Humanities and Social Sciences Communications, 12(1), 1-15.

Rajput, D., & Singh, S. (2023). Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics, 24(1), 1-15.

Rawashdeh, M., & Squalli, J. (2005). A Sectoral Efficiency Analysis of the Amman Stock Exchange. International Journal of Business, 10(1), 39-58.

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., & Aich, S. (2025). Artificial intelligence-based stock market price prediction: A review. Applied Artificial Intelligence, 39(2), 1–29.

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. Electronics, 10(21), 2717.

Saberironaghi, M., Ren, J., & El-Gindy, M. (2025). Stock Market Prediction Using Machine Learning and Deep Learning: A Systematic Review. Machine Learning and Knowledge Extraction, 5(3), 1245-1278.

Saberironaghi, S., Khosravi, A., & Nahavandi, S. (2025). Stock market prediction using machine learning and deep learning approaches: A systematic review. Big Data and Cognitive Computing, 5(3), 76.

Schumaker, R. P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. Information Processing & Management, 45(5), 571-583.

Sharma, P., Patel, J., & Jain, S. (2023). Stock Market Trends Analysis using Extreme Gradient Boosting. 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), IEEE, 1-6.

Sonani, S., Badii, A., & Moin, N. (2025). Stock price prediction using a hybrid LSTM-GNN model. arXiv preprint arXiv:2502.15813.

Stempień, D., & Kądziołka, K. (2025). Hybrid Models for Financial Forecasting: Combining Econometric, Machine Learning, and Deep Learning Models. arXiv preprint arXiv:2505.19617.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 111-133.

Tian, Y., Zhang, L., & Li, X. (2022). Financial time-series forecasting with gradient-boosting decision trees: An empirical comparison. *Finance Research Letters*, 46, 102473.

Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. Expert Systems with Applications, 40(14), 5501-5506.

Wang, Z. (2025). Machine learning for stock return prediction: Transformers and advanced time-series models. Journal of Financial Data Science, 7(1), 45–63.

Yang, J., & Chen, J. (2020). Use GBDT to Predict the Stock Market. Procedia Computer Science, 174, 333-343.

Ye, F., Chen, Y., & Li, Y. (2021, May). Jane Street Stock prediction model based on LightGBM. In 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI) (pp. 136-140). IEEE.

Zhen, K., Li, Y., & Li, X. (2025). A multi-feature selection fused with investor sentiment for stock price prediction. Expert Systems with Applications, 265, 124209.