Iran J Econ Stud, 14(2) 2025, 443-500

Iranian Journal of Economic Studies

Journal homepage: ijes.shirazu.ac.ir

Evaluating the effectiveness of advanced machine learning algorithms in reducing stock trading risk on the Iranian Stock Exchange

Javad Zolfaghary Tabesh[a], Babak Jamshidinavid[a*], Mehrdad Ghanbary[a], Afshin Baghfalaki[b]

a. Department of Accounting, Ker.C., Islamic Azad University, Kermanshah, Iran.
b. Department of Economics, Ker.C. Islamic Azad University, Kermanshah, Iran.

Highlights

- **Deep neural networks outperformed traditional models in predicting Tehran Stock Exchange trading risk with highest accuracy.**
- **XGBoost and LightGBM achieved 83%+ accuracy classifying stock risk levels for real-time warning systems.**
- **Exchange rate, inflation, and investor behavior drive Iranian market volatility, requiring non-linear machine learning models.**

Article History

Received: 25 November 2025
Revised:09 December 2025
Accepted:15 December 2025
Published:26 December 2025

JEL Classification
C45
C53
G17
G32

Keyword
Stock Trading Risk
Machine Learning
Deep Neural Network
Gradient Boosting Decision Tree
Support Vector Machine

Abstract

The primary objective of the present study was to thoroughly assess the efficiency of advanced machine learning algorithms in managing and mitigating stock trading risk in the Iranian capital market. Utilizing financial, economic, textual, and behavioral data from 172 companies over a 15-year period (2008–2023), this research estimated trading risk using advanced time-series models and modeled it against a total of 33 independent variables across five main dimensions. The methodology involved comparing the performance of six algorithmic families, ranging from simple regression to deep neural networks, in two scenarios: continuous prediction and risk level classification. The analysis of the results revealed that linear models performed very poorly due to the highly volatile and non-linear structure of the Iranian market. However, in the realm of non-linear methods, deep neural networks demonstrated the best capability for accurate and continuous risk prediction by recording the lowest error rates; this success is attributed to the ability of these models to identify long-term temporal patterns and the complex interaction of variables. In the section on classifying stocks into risk levels, algorithms based on gradient boosting decision trees achieved the highest efficiency with an accuracy exceeding 83% and were suggested as ideal tools for early warning systems. The analysis indicated that the primary drivers of risk volatility are macroeconomic variables (such as exchange rates and inflation), the effects of which are amplified by investor behavioral factors. Ultimately, this research confirms the significant and decisive superiority of advanced machine learning models over traditional approaches in analyzing risk within the Tehran Stock Exchange.

[*] *Babak.jamshidinavid@iau.ac.ir*
DOI: 10.22099/ijes.2025.54990.2083

1. Introduction

Trading risk in stocks, as one of the fundamental concepts in financial management, refers to the probability of deviation of actual returns from the expected returns of investments in the capital market (Merkoulova & Veld, 2022; Koh et al., 2018), arising from price volatility, macroeconomic factors, firms' fundamental characteristics, political changes, and market participant behavior (Esmaeili Moghadam, 2024; Sharif Karimi et al., 2013; Kordestani et al., 2024). Machine learning, representing a branch of computational intelligence, encompasses a set of advanced algorithms capable of identifying complex patterns and non-linear relationships among variables through the analysis of historical data, thereby providing more accurate forecasts (Taheri Haftasiabi et al., 2023; Balan et al., 2025). In this regard, algorithms such as Support Vector Machines alongside their online variants, Gradient Boosting Trees, Artificial Neural Networks, and evolutionary-driven optimization methods such as Genetic Algorithms and Particle Swarm Optimization have been recognized as effective instruments in risk analysis and market behavior forecasting (Krishnamurthy et al., 2026; Xu, 2025; Abbasian et al., 2023). However, while these algorithms demonstrate superior predictive accuracy, their practical deployment in risk management contexts requires careful consideration of model interpretability, transaction costs, and regime-specific performance stability issues that remain underexplored in existing literature.

On the other hand, in the current state of Iran's economy which faces challenges such as international sanctions, severe currency fluctuations, high inflation rates, declining foreign investment, and lack of confidence in economic stability (Shirazi et al., 2016; Kiumarthi et al., 2019; Abolhassani et al., 2023; Torki & Mazaheri, 2022) the capital market plays a vital role as the most important channel for mobilizing and allocating financial resources in economic development (Sadeghi & Tayebi, 2018). Increased trading risk in stocks not only reduces investor confidence and drives capital outflows toward parallel markets but also adversely affects overall market efficiency (Houshmand Naqabi et al., 2022). Accordingly, designing and developing intelligent models for predicting and mitigating trading risk has become doubly important (Jafari et al., 2021; Blackledge & Blackledge, 2025). Advanced machine learning algorithms, with their capacity to process extensive and multifaceted datasets, detect hidden patterns, and learn non-linear relationships, can assist investors, financial managers, investment funds, and regulatory bodies in making more informed decisions and achieving optimal risk management, ultimately leading to enhanced transparency, efficiency, and stability in the Iranian stock market (Dehghani et al., 2019). Yet, the practical implementation of these models necessitates addressing critical challenges including model interpretability for decision-makers, incorporation of market structure elements such as transaction costs and liquidity constraints, and validation of model robustness across distinct market regimes characterized by sanctions, political transitions, and crisis periods. A review of the literature reveals that, internationally, the primary focus of studies has been on developing hybrid machine learning models for predicting market risk or volatility from combining sentiment analysis with SVM (Zhang & Zhou, 2024) and optimizing LSTM parameters using particle swarm optimization (Lu & Ma, 2020) to hybrid SVM–PSO models within the intelligent web environment (Mohammadi Moradian et al., 2025). Alongside these, more classical studies on Value at Risk, crash risk, or stock price volatility have also been conducted (Liu & Wang, 2019). In the domestic (Iranian) literature, however, the predominant focus has been either on predicting stock prices and returns using one or two specific algorithms or on analyzing risk through conventional statistical and econometric techniques, Value at Risk, and regime-switching models (Abdi et al., 2022; Akhbari et al., 2024; Ghasemieh et al., 2023; Hosseinzadeh & Mehrgan, 2016). Thus, despite scattered advances, no comprehensive study has yet evaluated a wide range of advanced machine learning algorithms within a unified framework using data from the Iranian stock market. This gap becomes particularly pronounced when considering the unique characteristics of Iran's capital market volatility driven by sanctions, currency shocks, chronic inflation, and emotional investor behavior are taken into account; these features generate non-linear relationships and structural breaks in the data that reduce the accuracy of most linear or single-algorithm approaches. Moreover, most domestic studies have focused solely on price prediction accuracy of prices and have not considered direct trading risk metrics (such as crash probability, Value at Risk, or risk-level classification), nor have they provided a systematic comparison among different algorithm families (gradient trees, deep networks, SVM, ensemble methods, and evolutionary optimization). Furthermore, the potential of the intelligent web to integrate financial, textual, and investor behavioral data has been largely overlooked in Iranian research. Critically, existing studies have not adequately addressed the interpretability challenge inherent in complex machine learning models, the practical implementation barriers posed by transaction costs and market microstructure in the Iranian context, or the stability of model performance across distinct market regimes (calm periods versus sanctions-induced crises) gaps that fundamentally limit the operational utility of these models for real-world risk management applications.

Given the identified gaps, conducting the present research is an undeniable necessity. First, in the era of digital transformation and access to massive volumes of financial, economic, and behavioral data, leveraging the capacity of advanced machine learning algorithms to extract reliable knowledge from these data is essential. Second, the specific characteristics of Iran's capital market including severe volatility due to sanctions, political changes, high inflation, and emotional investor behavior require models capable of analyzing complex and non-linear relationships. Third, presenting an integrated and operational framework that can simultaneously quantify risk and classify stocks into different risk categories can significantly assist investors in optimal portfolio management, financial managers in corporate risk assessment, and regulatory institutions in early warning detection. Fourth, this research explicitly addresses the critical need for model interpretability by implementing structured explainability frameworks (SHAP values, partial dependence plots, and attention mechanisms) to ensure decision-makers can understand and trust model predictions. Fifth, the study incorporates market microstructure considerations including transaction cost simulations and liquidity-adjusted performance metrics to evaluate practical implement ability in the Iranian market. Sixth, regime-specific robustness analysis is conducted by disaggregating model performance across calm versus crisis periods (sanctions episodes, political transitions) to validate the stability and reliability of predictions under diverse market conditions thereby bridging the gap between theoretical model accuracy and operational deployment feasibility.

The structure of the remainder of this paper is as follows: Following this introduction, the literature review covering theoretical foundations and empirical studies in this field will be presented. Then, in the model design section, the model characteristics, methodological framework, and descriptive statistics will be introduced. This will be followed by the empirical analyses. Finally, the paper concludes with a discussion and conclusion section.

2. Literature review

Financial markets, particularly stock exchanges, have long been among the most challenging areas for prediction and risk management due to their inherent complexity, high volatility, and sensitivity to numerous economic, political, and social factors (Keshavarz & Rezaei, 2022). In recent decades, remarkable advancements in financial intelligence and machine learning have led to innovative approaches for the analysis and forecasting of financial markets, demonstrating significant potential to enhance investment decision-making and reduce trading risks (Pattnaik et al., 2024). Machine learning algorithms, with their ability to process vast amounts of historical data and capture complex non-linear patterns, provide an effective means to overcome the limitations of traditional methods (Taheri Haftasiabi et al., 2023). The rapid development of artificial intelligence has consequently transformed financial markets by improving risk management, optimizing portfolio allocation, and revolutionizing algorithmic trading (Bahoo et al., 2024).

However, the application of machine learning in financial prediction is accompanied by fundamental methodological challenges that go beyond computational power and algorithmic sophistication. Recent advances in econometric machine learning have highlighted the critical importance of addressing issues such as endogeneity, omitted variable bias, and spurious correlations, which can severely undermine predictive validity. To mitigate these concerns, double and debiased machine learning frameworks combining high-dimensional predictive modeling with semi-parametric estimation techniques have emerged as rigorous approaches for obtaining valid causal inference while leveraging the flexibility of machine learning (Chernozhukov et al., 2018). These methods enable consistent estimation of treatment effects and structural parameters even in the presence of high-dimensional confounders, thereby addressing a major limitation of conventional machine learning applications that often conflate prediction with causal understanding.

Consequently, traditional financial theories are being revisited in light of these developments. Conventional financial decision-making, which relied heavily on historical data, expert judgment, and rule-based models, has often struggled to adapt to rapid market fluctuations and systemic risks (Dalili et al., 2024) Recent research indicates that integrating deep learning techniques with classical portfolio management theories can lead to substantial improvements in risk–return trade-offs (Sharif Karimi et al., 2013). Furthermore, AI-driven models, enhanced by advanced machine learning and predictive analytics, enable real-time risk assessment, improved asset allocation strategies, and greater trading efficiency (Ashrafzadeh et al., 2025). By processing massive volumes of both structured and unstructured data, AI systems are capable of identifying complex patterns, detecting anomalies, and forecasting market trends with significantly higher accuracy than conventional approaches (Bahoo et al., 2025).

The Efficient Market Hypothesis, originally proposed by Fama in 1970, asserts that asset prices fully reflect all available information, implying that consistently accurate stock price prediction is theoretically impossible (Fama, 1970). However, recent empirical studies suggest that machine learning algorithms can challenge this hypothesis. Research on international markets has shown that machine learning-based models generate statistically significant alpha across eight different factor models (Azevedo & Hoegner, 2023).

Yet, the interpretation of these results requires careful consideration of multiple testing and p-hacking concerns. The finance literature has increasingly recognized that Many reported predictive features suffer from publication bias and data snooping, which results in overstated out-of-sample performance expectations (Harvey et al., 2016; Hou et al., 2020). To address these concerns, recent methodological advances emphasize the importance of rigorous cross-validation frameworks, proper adjustment for multiple hypothesis testing, and the use of hold-out validation periods that extend beyond the original training and testing samples. Furthermore, the growing body of research on factor zoo reduction and variable importance measures derived from machine learning provides more robust approaches to feature selection that account for complex interactions and nonlinearities while maintaining statistical validity.

The Modern Portfolio Theory, introduced by Markowitz in 1952, laid the scientific foundation for modern risk–return management and portfolio optimization (Markowitz, 1952). This theory, based on the mean-variance framework, seeks a balance between expected return and risk but assumes normal return distributions and constant covariances assumptions that limit its effectiveness in highly volatile markets. Recent studies demonstrate that integrating machine learning-based forecasting techniques with traditional portfolio theory can overcome these limitations and improve risk-adjusted performance (Yan et al., 2024).

Artificial Neural Networks are among the most widely used machine learning algorithms for forecasting the direction of stock market indices due to their flexibility in modeling nonlinear patterns (Taslimipour et al., 2026). Inspired by biological neural systems, these networks excel at modeling complex non-linear relationships. A comprehensive study on developed market indices found that ANNs delivered the best forecasting performance and highest accuracy compared to other algorithms (Ayyıldız & İskenderoğlu, 2024). Long Short-Term Memory networks and Gated Recurrent Units are specialized recurrent neural network architectures designed to efficiently process sequential and time-series data (Farzad et al., 2023). By addressing the vanishing gradient problem, these architectures can learn long-term dependencies in sequential data (Sarıkoç et al., 2025). Studies have shown that hybrid models such as CNN-LSTM and GRU-CNN significantly enhance prediction performance and outperform traditional models in a substantial number of cases (Song & Choi, 2023). Support Vector Machines operate on the principle of structural risk minimization and are highly effective in classification tasks. Research has demonstrated that linear SVMs generate superior risk-adjusted returns compared to benchmark indices (Grudniewicz & Slepaczuk, 2023).

Beyond traditional supervised learning methods, modern reinforcement learning approaches have also found applications in portfolio management and trading decision optimization. Random Forests, as an ensemble learning algorithm, operate via majority voting among decision trees and have gained widespread use in financial market risk analysis (Sun et al., 2023). Studies indicate that these techniques are particularly useful for constructing efficient active portfolios (Han, 2024). Deep Reinforcement Learning represents a cutting-edge approach to portfolio management, enabling intelligent agents to learn optimal actions through environmental interaction. This method has proven effective in enhancing trading strategies and portfolio optimization. Research shows that DRL models with dynamic reward functions deliver significant improvements in risk-adjusted performance, achieving high Sharpe and Sortino ratios indicative of greater stability and better risk control (Choudhary et al., 2025). A novel risk-adjusted deep reinforcement learning framework has been proposed that combines agents trained with different reward functions to achieve multiple investment objectives simultaneously (Choudhary et al., 2025).

Despite these promising developments, a critical methodological concern in asset pricing research involves the causal interpretation of machine learning predictions. Many studies document strong predictive relationships between features and returns without establishing whether these relationships represent genuine risk premia, behavioral anomalies, or spurious correlations driven by data mining. Recent advances in causal machine learning provide frameworks to distinguish between these interpretations through instrumental variable approaches, difference-in-differences designs adapted for high-dimensional settings, and synthetic control methods enhanced with machine learning algorithms (Athey & Imbens, 2019; Chernozhukov et al., 2021). These techniques enable researchers to move beyond pure prediction toward understanding the underlying economic mechanisms driving return patterns, which is essential for robust risk management in out-of-sample periods and regime changes.

Value at Risk remains one of the most common risk metrics in the financial industry, yet traditional models often fail during economic crises (Fatouros et al., 2023). To address this shortcoming, a probabilistic deep learning framework known as Deep VaR has been introduced, leveraging time-series forecasting techniques for portfolio risk monitoring (Fatouros et al., 2023). AI-based systems employ advanced algorithms to analyze large datasets, thereby enhancing risk forecasting accuracy (Yarlagadda et al., 2024). A novel deep learning-based portfolio optimization method has been proposed that utilizes a Constant Absolute Risk Aversion utility loss function, ensuring unbiased gradients and robust performance (Kubo & Nakagawa, 2025). One of the most critical applications of these advanced algorithms lies in improving the measurement and prediction of risk metrics, with Value at Risk holding a particularly prominent position. Thus, AI algorithms play a vital role in analyzing market data to identify potential risks. By extracting valuable insights and detecting patterns, these algorithms help identify potential risks and market disruptions (Daiya, 2024). Predictive analytics plays a fundamental role in forecasting market trends and volatility using historical data, statistical models, and machine learning algorithms (Bhuiyan et al., 2025).

Despite significant progress, the application of machine learning still faces challenges. One of the most critical limitations is insufficient training data density. Deep learning models are prone to overfitting due to the limited volume of historical financial data (Chen et al., 2023). Researchers have proposed data augmentation techniques to expand datasets and mitigate overfitting risks. The black-box nature of AI models also raises concerns regarding interpretability and transparency.

Moreover, the field increasingly recognizes the importance of model interpretability not merely for regulatory compliance but for validating economic mechanisms. Techniques such as SHAP (SHapley Additive exPlanations) values, partial dependence plots, and attention mechanisms in neural networks provide insights into feature importance and interaction effects. However, these post-hoc interpretation methods may themselves be unreliable when applied to highly complex models or in the presence of strong feature correlations. Recent research advocates for inherently interpretable models or hybrid approaches that balance predictive performance with economic interpretability, particularly in risk management contexts where stakeholder trust and regulatory scrutiny are paramount (Rudin, 2019; Molnar, 2022).

Nevertheless, the existing literature clearly indicates that advanced machine learning algorithms possess considerable potential to enhance stock market forecasting and reduce trading risk (Xie et al., 2025). By processing massive datasets and identifying complex non-linear patterns, these algorithms can effectively overcome many of the limitations inherent in traditional methods (Wu et al., 2024).

There are numerous empirical studies on reducing stock trading risk, which are reviewed below.

Gerlein et al. (2016), in "Evaluating Machine Learning for Financial Trading: An Empirical Approach," examined the effectiveness of simple machine learning models in predicting and achieving profitable trades in the foreign exchange market by conducting six-year trading simulations on the USDJPY, EURGBP, and EURUSD currency pairs. Their findings revealed that even modestly accurate simple machine learning models can generate substantial long-term financial returns, and that periodic retraining, feature selection, training set size, and the combination of price-related, seasonal, and lagged features significantly influence systematic predictability and ultimate profitability. Wu (2021), in "Using a Machine Learning Approach to Assess the Risk of Over-Financialization of Trading Institutions," aimed to enhance the ability of commercial financial institutions to respond to excessive financialization risks by employing decision trees, random forests, gradient boosting, and a hybrid control model tested on six months of loan data from a trading institution. Results showed that support vector machines train faster on smaller datasets (1–5 GB), the hybrid model requires less time on larger datasets (5–30 GB), and the hybrid algorithm achieves higher accuracy, recall, and precision (79.35%, 39.28%, and 78.28%, respectively).

Liu and Yu (2022), in "Financial Market Risk Analysis Based on Machine Learning and Particle Swarm Optimization Algorithm," investigated risks in the financial market under blockchain technology using machine learning methods, random forests, and decision tree construction. The study found that as industrial structures evolve, the economic focus gradually shifts from primary to tertiary industries, blockchain can improve financial industry efficiency and reduce operating costs, and it provides a theoretical reference for promoting inter-firm coordination. Avramov et al. (2022), in "Machine Learning vs. Economic Restrictions: Evidence from Stock Return Predictability," analyzed the profitability of deep learning signals in hard-to-arbitrage stocks and found that deep learning-based investments extract profitability from hard-to-arbitrage stocks, particularly during periods of high arbitrage constraints; however, reasonable transaction costs significantly reduce profitability, while deep learning signals remain profitable in long positions and in recent years. Hao et al. (2025), in "Machine Learning versus Deep Learning in Stock Market Investment: International Evidence," compared the effectiveness of machine learning and deep learning models in stock price prediction by applying random forest and deep neural network models to major stock indices in five markets (China, United States, United Kingdom, Canada, and Japan) from 2005 to 2020 to construct long-short portfolios of the top-20 selected stocks. Findings indicated that different models yield markedly different profits across markets, with deep learning models consistently outperforming traditional machine learning models in all markets and achieving the best performance, especially in the Chinese stock market.

Maleki et al. (2023), in “A Risk-Based Trading System Using Algorithmic Trading and Deep Learning Models,” developed an algorithmic gold trading strategy that incorporates risk parameter forecasting and open-position control using LSTM neural networks for gold price prediction and feature selection techniques. Results demonstrated that combining feature selection and missing-value imputation significantly improves model performance, employing LSTM to choose between current signals and existing open positions enhances the trading strategy, and considering multiple factors to determine optimal values meets practical trading requirements. Addy et al. (2024), in “Machine Learning in Financial Markets: A Critical Review of Algorithmic Trading and Risk Management,” conducted a comprehensive critical literature review of machine learning applications in algorithmic trading and risk management. Key findings highlighted that various algorithms (neural networks, decision trees, and ensemble methods) suffer from overfitting, data bias, and interpretability issues, that machine learning can significantly influence systemic risk, and that ethical and regulatory considerations must be addressed in financial markets. Melina et al. (2025), in “Machine Learning-Based Extreme Value Theory Modeling in Stock Investment Risk Prediction: A Systematic Literature Review,” used the PRISMA method to select 90 articles from ScienceDirect and Scopus to identify and analyze literature on investment risk estimation capable of detecting extreme values. The review found that EVT, GARCH, and historical simulation are commonly used for investment risk estimation, yet machine learning-based risk estimation remains scarce, with no studies combining EVT and machine learning; the authors proposed a hybrid modeling framework integrating EVT and machine learning using high-frequency multivariate data. Uddin et al. (2025), in “Advancing Financial Risk Prediction and Portfolio Optimization Using Machine Learning Techniques,” explored the application of various machine learning models including random forests, gradient boosting, LSTM, and transformers in financial risk prediction and portfolio management optimization. Findings showed that machine learning models significantly outperform traditional financial models, LSTM and transformer architectures excel at capturing long-term dependencies in financial data, feature selection and data preprocessing are critical for maximizing performance, and combining machine learning with traditional optimization techniques yields superior Sharpe and Sortino ratios in portfolio optimization.

2.1 Research Gap and Research Innovation

A comprehensive literature review reveals that internationally, the primary focus of studies has been on developing hybrid machine learning models for predicting market risk or volatility by combining sentiment analysis with SVM and optimizing LSTM parameters using Particle Swarm Optimization (PSO). Alongside these, more classical studies have addressed Value at Risk (VaR), crash risk, or stock price volatility. However, in the domestic Iranian literature, the dominant focus has either been on predicting stock prices and returns using one or two specific algorithms or on risk analysis through traditional econometric methods, Value at Risk, and regime-switching models. Thus, despite scattered advances, no comprehensive study has yet systematically evaluated a wide range of advanced machine learning algorithms within an integrated framework using data from the Tehran Stock Exchange (TSE). This gap becomes particularly pronounced when considering the unique characteristics of the Iranian capital market: volatility driven by sanctions, currency shocks, chronic inflation, emotional investor behavior, non-linear relationships, and structural breaks in the data all of which significantly reduce the accuracy of most linear or single-algorithm approaches. Moreover, most domestic studies have focused solely on price prediction accuracy and have neglected direct trading risk metrics (such as crash probability, Value at Risk, or risk level classification) and have not provided systematic comparisons across different algorithm families (Gradient Boosting Trees, deep neural networks, SVM, ensemble methods, and evolutionary optimization). Additionally, the potential of the intelligent web for integrating financial, textual, and behavioral investor data has been largely overlooked in Iranian research. The present study aims to fill these research gaps by introducing three fundamental innovations:

First, this is the first study in the context of the Tehran Stock Exchange to present a comprehensive and systematic evaluation framework of six advanced algorithmic families (from simple regression to deep neural networks) operating in two distinct scenarios: continuous risk prediction and risk level classification. This dual approach enables the identification of the best tool for each practical purpose, with deep neural networks found optimal for precise continuous risk prediction and gradient-boosted decision tree algorithms (XGBoost and LightGBM) identified as optimal for early warning systems. Second, the research utilizes a rich, multidimensional dataset comprising 33 independent variables across five main dimensions (macroeconomic indicators, industry characteristics, firm characteristics, political events, and behavioral-cultural factors of investors). This integration of structured data (from stock exchange databases and the Central Bank) and semi-structured data (economic news and web-based sentiment analysis) has been implemented for the first time in domestic studies at this scale.
Third, the research methodology combines advanced econometric GARCH and EGARCH models to compute conditional variance as the trading risk measure, followed by modeling it against independent variables using machine learning algorithms. This offers a unique hybrid of classical econometric methods and

modern artificial intelligence that can serve as a model for future research in emerging markets with similar characteristics.

3. Methods

3.1 Introduction and Research Design

This chapter provides a comprehensive description of the research methodology, tools, data collection and analysis methods, and the implementation framework of the proposed models. The main objective of this study is to design and develop an integrated model for reducing equity trading risk in the Iranian capital market by leveraging advanced machine learning algorithms and intelligent web technologies. Given the complex and multidimensional nature of trading risk, which is influenced by macroeconomic factors, industry and firm characteristics, political events, and behavioral factors of investors, this research adopts a combined and interdisciplinary approach. In terms of nature and purpose, the study is applied and predictive modeling-oriented, using real data from the Iranian capital market to conduct a comparative evaluation of the performance of various machine learning algorithms. To this end, algorithms such as Support Vector Machines (SVM) and their online versions, evolutionary optimization algorithms including Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), Artificial Neural Networks (ANN), and Gradient Boosting Trees (GBT) have been employed. Additionally, to model volatility and calculate trading risk as the dependent variable, econometric models from the GARCH and EGARCH families have been utilized, which are capable of analyzing conditional heteroskedasticity behavior in financial time series.

3.2 Statistical Population and Sample Selection

The statistical population of this study consists of all companies listed on the Tehran Stock Exchange (TSE) that were actively trading until the end of the Iranian year 2023. This population includes various industries such as automotive, petrochemicals, basic metals, pharmaceuticals, food, banking, insurance, and others. For sample selection, a systematic screening method was applied based on specific criteria designed to ensure data quality and consistency. Companies were required to have been listed on the TSE before the Iranian year 2008 and remain listed through the end of 2023 without delisting. The fiscal year ending was standardized to the last day of Esfand (March 20–21), with no changes in fiscal year during the study period from 2008 to 2023. Companies were required to maintain continuous trading activity during the study period, with trading halts not exceeding six months in cumulative duration. Holding companies, financial intermediaries, banks, and insurance companies were excluded due to their distinct operational and financial structures, which differ fundamentally from industrial and commercial firms. Finally, complete and flawless availability of financial and trading information was required for inclusion in the final sample. After applying these screening criteria systematically, the initial population of 370 companies present on the TSE from 2008 to 2023 was reduced through several stages. Fifty-seven companies were excluded due to having fiscal year-ends other than Esfand or experiencing fiscal year changes during the study period. Ninety-eight companies were removed because of trading halts exceeding six months in cumulative duration. Thirty-four holding companies, financial intermediaries, banks, and insurance companies were excluded due to their specialized nature. Nine additional companies were removed due to incomplete financial information. The final sample consisted of 172 companies with complete and reliable data across all required variables. The time period of the study spans fifteen years, from the beginning of 2008 to the end of 2023. This period was deliberately selected to cover various economic cycles including boom periods, recessions, international sanctions, political changes, and periods of high market volatility, thus enabling the examination of their impact on equity trading risk under diverse market conditions. Data were collected and analyzed on a monthly basis, resulting in 180 monthly observations per company and a total of 30,960 firm-month observations (172 companies multiplied by 180 months). This longitudinal panel structure provides sufficient temporal depth to capture both short-term fluctuations and long-term trends in trading risk dynamics.

Table 2 provides a summary overview of all research variables with their symbols, operational definitions, and measurement methods. Detailed explanations of variable transformations, lag structures, and data sources for each category are provided in the subsequent subsections.

Table 2. Operational Definition and Measurement of Research Variables

Variable	Symbol	Operational Definition	Measurement
A. Macroeconomic Indicators			
Gross Domestic Product	GDP	Total value of final goods and services produced within the country's borders	GDP = C + I + G + (X – M); Central Bank of Iran
Inflation Rate	INF	Percentage change in the Consumer Price Index (CPI) year-over-year	Annual CPI change; Central Bank
Exchange Rate	ER	Average USD/IRR buying and selling rate	Official exchange rate (IRR/USD); Central Bank
Money Supply Growth	LG	Percentage change in broad money (M2)	LG = [(M2_t – M2_{t-1}) / M2_{t-1}] × 100; Central Bank
Bank Interest Rate	IR	Short-term bank deposit interest rate	Annual percentage; Central Bank
Lending Rate	LR	Interest rate on non-deposit bank loans	Annual percentage; Central Bank
Trade Balance	TB	Difference between exports and imports of goods	TB = Exports – Imports (billion IRR); Iran Customs Administration

Gini Coefficient	GINI	Measure of income distribution inequality	G = 1 – Σ(y_i + y_{i-1}); Statistical Center of Iran (0 to 1)
Ease of Doing Business Index	EDB	Country ranking in ease of doing business	World Bank ranking (0–100)
Housing Economic Indicators	HI	Includes housing cost-to-income ratio, construction and land price indices	Composite indices; Central Bank & Statistical Center
Productivity Index	PROD	Labor and capital productivity	Value added / labor or capital; National Productivity Organization
TSE Total Index	TEPIX	Weighted average price change of all TSE stocks	Numerical index; Tehran Stock Exchange website
Economic Growth	GDPG	Percentage change in real GDP	GDPG = [(GDP_t – GDP_{t-1}) / GDP_{t-1}] × 100
		B. Industry Characteristics	
Pricing Mechanism	PRICE_M	Type of pricing system (free, semi-regulated, regulated)	Ordinal: 1=regulated, 2=semi-regulated, 3=free
Supply & Demand	S_D	Industry sales to production capacity ratio	S_D = Sales / Production Capacity (%)
Market Concentration	HHI	Herfindahl-Hirschman Index for industry concentration	HHI = Σ(s_i²); s_i = market share of firm i
Capital Investment Volume	INV	Average fixed assets of firms in the industry	Billion IRR; Financial statements
Related Industries Performance	REL_IND	Performance index of upstream and downstream industries	Average monthly return of related industries (%)
Technological Changes	TECH	R&D expenditure to sales ratio	R&D / Sales (%)
		C. Firm Characteristics	
Earnings Per Share	EPS	Net profit divided by number of shares	EPS = Net Profit / Shares (IRR)
Dividends Per Share	DPS	Approved cash dividend per share	DPS = Distributed Profit / Shares (IRR)
Price-to-Earnings Ratio	PE	Market price to earnings per share	PE = Share Price / EPS
Return on Assets	ROA	Net profit to total assets	ROA = Net Profit / Total Assets (%)
Economic Value Added	EVA	Operating profit minus cost of capital	EVA = NOPAT – (WACC × Capital)
Market Value Added	MVA	Difference between market value and book value of equity	MVA = Market Value of Equity – Book Value of Equity

Refined Economic Value Added	REVA	EVA using market value of capital	REVA = (r – C) × Market Capital
Residual Income	RI	Operating profit minus opportunity cost of capital	RI = Operating Profit – (Required Return × Invested Capital)
Tobin's Q	TQ	Market value to book value of assets ratio	TQ = Market Value / Net Book Value of Assets
Price-to-Sales Ratio	PS	Market price per share to sales per share	PS = Share Price / (Sales / Shares)
Cash Value Added	CVA	Operating cash flow plus cash payments	CVA = CFO + Dividends + Interest + Share Repurchases + etc.
Free Cash Flow	FCF	Operating cash flow minus capital expenditures	FCF = CFO – CapEx
Altman Z-Score	Z_SCORE	Bankruptcy prediction index	Z = 1.2A + 1.4B + 3.3C + 0.6D + 1.0E
Beneish M-Score	M_SCORE	Earnings manipulation detection index	M = –4.84 + 0.92×DSRI + … (8 ratios)
D. Political Events			
International Agreements	DEAL	Existence of agreements (e.g., JCPOA)	Binary: 1=implemented, 0=not implemented
Sanctions	SANC	Intensity of economic sanctions	Sanctions intensity index (0–10) or binary
Elections	ELEC	Occurrence of elections in the period	Binary: 1=election period, 0=otherwise
Regional War	WAR	Presence of regional military conflict	Binary or Geopolitical Risk Index (GPR)
Accession to Treaties	TREATY	Membership in international bodies (e.g., FATF)	Binary: 1=member, 0=non-member
Government Economic Policy	POLICY	Type of policy (expansionary, contractionary, stock market support)	Ordinal or composite index
E. Cultural-Behavioral Factors of Investors			
Value System	VALUE	Investor value priorities	Schwartz questionnaire score (1–5)
Locus of Control	LOC	Internal vs. external locus of control	Rotter questionnaire score (–10 to +10)
Ambiguity Tolerance	AT	Ability to cope with uncertainty	McLain scale score (1–7)
Investment Habit	INV_HAB	Experience and frequency of investment	Years of activity or annual transaction count
Investment Horizon	HORIZON	Intended holding period	1=short-term (<3 months), 2=medium, 3=long-term (>1 year)
Cognitive Bias	COG_BIAS	Types of cognitive biases (overconfidence,	Behavioral finance questionnaire score (1–5)

		anchoring, representativeness)	
Emotional Bias	EMO_BIAS	Emotional biases (optimism, loss aversion)	PANAS scale or investment behavior questionnaire
Investor Sentiment	SENT	Positive or negative market sentiment	Sentiment index (text mining or survey-based)
Conservatism	CONSERV	Reluctance to change portfolio	Behavioral questionnaire score (1–5)
Personality Type	PERS	Big Five personality traits	Big Five inventory scores
Risk Tolerance	RISK_TOL	Degree of risk acceptance	Risk tolerance questionnaire score (1–10)
Dependent Variable			
Equity Trading Risk	RISK	Volatility and probability of loss in stock price	Standard deviation of returns, Beta, VaR, CVaR, GARCH/EGARCH conditional variance

***Source*:** (*Markowitz, 1952؛ Fama, 1970؛ Sadeghi & Tayebi, 2018؛ Moezdi & Rezaei, 2021؛ Abdi et al., 2022؛ Abbasian et al., 2023؛ Avramov et al., 2022؛ Esmaeili Moghadam, 2024؛ Addy et al., 2024؛ Fatouros et al., 2023؛ Zhang & Zhou, 2024؛ Uddin et al., 2025*).

3.3 Dependent Variable: Equity Trading Risk Measurement

The dependent variable in this study is equity trading risk, operationalized through conditional volatility estimated using GARCH and EGARCH econometric models. The measurement process involved several carefully designed steps to ensure accuracy and consistency. For each of the 172 companies in the sample, monthly stock returns were calculated using the natural logarithm of price relatives, specifically as the log difference between the closing price at month t and the closing price at month t-1. This log-return specification is standard in financial econometrics as it produces more symmetric distributions and satisfies the continuity assumptions required for volatility modeling. Before estimating GARCH models, preliminary diagnostic tests were conducted on the return series for each company. The Augmented Dickey-Fuller test and Phillips-Perron test were applied to verify stationarity of the return series, which is a prerequisite for valid GARCH estimation. The Ljung-Box Q-test was performed on both returns and squared returns to detect serial correlation and ARCH effects. The presence of significant ARCH effects in squared returns justified the use of conditional heteroskedasticity models rather than simple variance measures.

For companies exhibiting significant ARCH effects, both GARCH (1,1) and EGARCH (1,1) specifications were estimated. The GARCH (1,1) model specifies the conditional variance at time t as a function of the constant term, the squared residual from the previous period (ARCH term), and the conditional variance from the previous period (GARCH term). The EGARCH (1,1) specification was selected to capture asymmetric volatility effects, where negative shocks exert a stronger impact on volatility than positive shocks, where negative return shocks (bad news) have a differential impact on volatility compared to positive shocks (good news) of the same magnitude. This asymmetry is particularly relevant in the Iranian stock market context, where prior research has documented stronger investor reactions to negative news. Model estimation was performed using maximum likelihood estimation with the Marquardt optimization algorithm. For each company, both GARCH (1,1) and EGARCH (1,1) models were estimated, and the better-fitting model was selected based on information criteria, specifically the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), with lower values indicating better model fit. Additionally, the significance of coefficients and the absence of remaining ARCH effects in standardized residuals were verified through diagnostic checking. The Ljung-Box Q-test on standardized residuals and squared standardized residuals was used to confirm that the fitted model adequately captured the volatility dynamics.

The conditional variance series extracted from the selected GARCH or EGARCH model for each company-month observation was then used as the measure of trading risk. This conditional variance represents the market's expectation of volatility at each point in time, incorporating information from past return behavior. The resulting risk measure is time-varying, forward-looking, and theoretically grounded in financial econometric theory. For the 172 companies across 180 months, this process generated 30,960 risk observations that serve as the dependent variable for subsequent machine learning modeling. This approach represents a methodological innovation by combining classical econometric volatility modeling with modern machine learning algorithms, leveraging the strengths of both paradigms.

3.4 IndependentVariables:Operational Definitions and Data Transformation

The independent variables in this study are organized into five broad categories: macroeconomic indicators, industry characteristics, firm characteristics, political events, and cultural-behavioral factors of investors. Each variable was carefully operationalized with attention to measurement consistency, temporal alignment, and economic interpretation. Detailed consideration was given to variable transformation to address the concerns raised regarding data homogeneity, lag structure, and scaling.

3.4.1 Macroeconomic Indicators

Macroeconomic variables present unique challenges in a monthly analysis framework, as many are reported quarterly or annually by official statistical agencies. To address this temporal mismatch, several interpolation and transformation techniques were employed systematically. Gross Domestic Product (GDP), which is officially reported quarterly by the Central Bank of Iran, was converted to monthly frequency using cubic spline interpolation. This method preserves the aggregate quarterly values while generating smooth monthly estimates that avoid artificial volatility. Given the known autocorrelation structure in GDP data, cubic spline interpolation is superior to linear interpolation as it maintains the second derivative continuity. The monthly GDP series was then transformed into a growth rate by calculating the year-over-year percentage change, specifically the percentage difference between GDP in month t and GDP in the same month twelve periods earlier. This transformation removes the trend component and makes the variable stationary, which is essential for machine learning algorithms that can be sensitive to scale and non-stationarity. The use of year-over-year growth rather than month-over-month growth eliminates seasonal patterns that would otherwise confound the analysis.

Inflation rate was measured as the year-over-year percentage change in the Consumer Price Index (CPI), calculated directly from monthly CPI data published by the Central Bank of Iran. The CPI is inherently a monthly index, so no interpolation was required. The year-over-year specification removes seasonal effects and provides a measure of inflation that is directly comparable to policy targets and public perceptions. Exchange rate was operationalized as the average of official USD/IRR buying and selling rates reported by the Central Bank of Iran on a daily basis, aggregated to monthly averages. To address the non-stationarity and exponential growth pattern in the nominal exchange rate series, the variable was transformed using the natural logarithm. This log transformation stabilizes variance and allows the variable to be interpreted in terms of percentage changes, making it compatible with growth-rate specifications of other variables. Additionally, the month-over-month log difference (growth rate) of the exchange rate was calculated to capture exchange rate volatility, which theory suggests should be more relevant to trading risk than the absolute level.

Money supply growth was measured using the broad monetary aggregate M2, reported monthly by the Central Bank of Iran. The transformation involved calculating the year-over-year percentage change in M2, which removes trend and seasonal components while maintaining economic interpretability as monetary expansion rate. Bank interest rate and lending rate, both reported monthly by the Central Bank of Iran as annual percentage rates, were used in level form initially. However, given concerns about homogeneity with growth-rate variables, changes in interest rates (first differences) were also calculated and tested in robustness specifications. Trade balance, defined as the difference between exports and imports of goods reported by Iran Customs Administration, was converted from quarterly to monthly frequency using cubic spline interpolation, then scaled by GDP to create a trade balance ratio that is dimensionless and comparable across time periods.

The Gini coefficient, measuring income distribution inequality and reported annually by the Statistical Center of Iran, required more sophisticated temporal treatment. The annual values were interpolated to monthly frequency using linear interpolation under the assumption that income inequality changes gradually over time. The interpolated monthly series was then first-differenced to capture changes in inequality rather than levels. Ease of Doing Business Index, sourced from World Bank annual reports, was similarly interpolated to monthly frequency and used in first-difference form to capture changes in the business environment. Housing economic indicators, including housing cost-to-income ratio and construction price indices reported quarterly by the Central Bank and Statistical Center of Iran, were interpolated to monthly frequency using cubic spline methods and transformed into year-over-year growth rates. Productivity index, reported annually by the National Productivity Organization, was interpolated monthly and expressed as year-over-year percentage change. The TSE Total Index (TEPIX), a weighted average price index of all TSE stocks reported daily, was aggregated to monthly frequency by taking the end-of-month value, then transformed into monthly returns using log differences to align with the return-based framework of the analysis.

3.4.2 Lag Structure of Macroeconomic Variables

A critical methodological consideration is the temporal relationship between macroeconomic variables and trading risk. Economic theory and empirical evidence suggest that macroeconomic conditions affect stock market risk with various lag structures depending on information transmission mechanisms and market efficiency. To address this, several lag specifications were systematically tested. For fast-disseminating information variables such as interest rates, exchange rates, and TEPIX, contemporaneous values (no lag) were used as these are immediately observable and incorporated into prices. For variables with publication lags such as GDP and inflation, a one-month lag was applied to reflect the real-world timing of data availability. Investors cannot respond to GDP data that has not yet been published, so using lagged values ensures that only information available to market participants at time t is included in the prediction of risk at time t.

For certain variables hypothesized to have cumulative or delayed effects, such as money supply growth and trade balance, both contemporaneous and lagged values (up to three months) were tested. Correlation analysis and preliminary regression diagnostics were used to identify the lag specification with the strongest explanatory power. In the final model specifications, a mix of contemporaneous and lagged variables was used based on these empirical tests and theoretical considerations. This approach ensures that the temporal ordering of cause and effect is respected while maximizing the predictive information content of the macroeconomic variables.

3.4.3 Industry Characteristics

Industry-level variables were constructed to capture sector-specific conditions that affect trading risk beyond firm-specific and macroeconomic factors. Pricing mechanism, reflecting the degree of price regulation in each industry, was coded as an ordinal variable with three levels: regulated (1), semi-regulated (2), and free (3). This classification was determined based on government policies and industry-specific regulations in effect during the study period, with annual updates to reflect policy changes. Supply and demand conditions within each industry were measured as the ratio of industry-wide sales to production capacity, aggregated monthly from firm-level data. This ratio, expressed as a percentage, indicates the degree of capacity utilization and demand pressure in each industry. Market concentration was measured using the Herfindahl-Hirschman Index (HHI), calculated as the sum of squared market shares of all firms within each industry, updated quarterly and interpolated to monthly frequency. Higher HHI values indicate greater market concentration and potentially different competitive dynamics affecting risk.

Capital investment volume, representing the average fixed assets of firms within each industry, was aggregated from firm-level balance sheet data and expressed in constant prices (inflation-adjusted) to enable inter-temporal comparisons. The year-over-year growth rate of industry capital stock was calculated to align with the growth-rate framework of other variables. Related industries performance was operationalized as the average monthly return of upstream and downstream industries identified through input-output linkages. For each industry, related industries were mapped based on supply chain relationships documented in Iran's national input-output tables. The monthly return index of these related industries was calculated as a value-weighted average, providing a measure of spillover effects from connected sectors. Technological change within each industry was measured as the ratio of research and development expenditure to sales revenue, aggregated from firm-level disclosures when available, and supplemented with industry association reports for sectors with limited firm-level R&D disclosure.

3.4.4 Firm Characteristics

Firm-level variables were extracted directly from audited financial statements and market trading data, ensuring high reliability and consistency. Earnings per share (EPS), dividends per share (DPS), and related accounting measures were obtained from quarterly financial reports disclosed through the Codal system, Iran's comprehensive disclosure system for listed companies. For monthly alignment, quarterly values were held constant across the three months of each quarter, reflecting the information availability pattern. This approach is standard in accounting-based studies where financial statements are not revised within quarters. Price-to-earnings ratio (PE), return on assets (ROA), and market-based valuation ratios were calculated using market prices at month-end and the most recent available financial statement data, ensuring no forward-looking bias.

Advanced performance measures including Economic Value Added (EVA), Market Value Added (MVA), Refined Economic Value Added (REVA), and Residual Income (RI) were calculated following standard formulas with adjustments for the Iranian market context. The weighted average cost of capital (WACC) required for EVA calculation was estimated using the capital asset pricing model (CAPM) with a market risk premium based on historical Iranian stock market returns and the risk-free rate proxied by short-term government bond yields. Tobin's Q, the ratio of market value to book value of assets, was calculated using market capitalization and book values from balance sheets. Price-to-sales ratio and cash value added were similarly calculated using market and accounting data following standard definitions.

Free cash flow was computed as operating cash flow minus capital expenditures, extracted from cash flow statements. Altman Z-Score, a bankruptcy prediction measure, was calculated using the original five-variable formula with coefficients developed by Altman, applied to Iranian firm data. Beneish M-Score, an earnings manipulation detection index, was calculated using eight financial statement ratios following the Beneish model specification. All firm-level financial variables were winsorized at the 1st and 99th percentiles to reduce the influence of extreme outliers and measurement errors. This winsorization procedure replaces values below the 1st percentile with the 1st percentile value, and values above the 99th percentile with the 99th percentile value, preserving the sample size while limiting the impact of anomalous observations.

3.4.5 Political Events Variables

Political event variables capture major policy changes, geopolitical developments, and institutional shifts that affect market risk. These variables were coded based on historical records from official government sources, international organizations, and verified news archives. International agreements, particularly the Joint Comprehensive Plan of Action (JCPOA) nuclear agreement, were coded as a binary variable taking the value 1 during periods when the agreement was implemented (from January 2016 to May 2018) and 0 otherwise. Sanctions intensity was measured using a composite index ranging from 0 to 10, developed by coding the severity and scope of international economic sanctions based on official United Nations Security Council resolutions, US Treasury Department sanctions designations, and European Union restrictive measures. The index was updated monthly based on new sanctions imposed or lifted, weighted by economic significance.

Elections were coded as a binary variable equal to 1 during the three-month period surrounding presidential and parliamentary elections (one month before and two months after the election date) and 0 otherwise, reflecting the heightened uncertainty and policy speculation during electoral periods. Regional war and geopolitical tensions were measured using the Geopolitical Risk Index (GPR) developed by Caldara and Iacoviello, which quantifies geopolitical risk based on automated text analysis of major international newspapers. The monthly GPR index for Middle East specific risks was used to capture regional military conflicts and tensions affecting Iran. Accession to international treaties, including Financial Action Task Force (FATF) membership discussions and other international agreements, was coded as a binary variable reflecting Iran's compliance status with international standards in each month. Government economic policy was coded as an ordinal variable reflecting the dominant policy stance (contractionary, neutral, or expansionary) based on fiscal and monetary policy announcements, with specific attention to stock market support policies announced by the government or central bank.

3.4.6 Cultural-Behavioral Factors of Investors

The incorporation of investor behavioral and cultural factors represents a methodological challenge and innovation in this study. Given the fifteen-year historical period and the large panel of 172 companies, directly collecting psychometric survey data from individual investors is not feasible retrospectively. To address this limitation while preserving the conceptual richness of behavioral finance theory, a multi-source proxy approach was developed and implemented. Investor sentiment, the most critical behavioral variable, was measured using both quantitative market indicators and text-based sentiment analysis. The quantitative component included the ratio of advancing to declining stocks, trading volume relative to historical averages, and the put-call ratio from options markets when available. These market-based indicators have been validated in prior literature as proxies for aggregate investor sentiment.

The text-based sentiment component leveraged natural language processing techniques applied to economic news and financial media. Specifically, a Persian-language sentiment lexicon was constructed based on the Loughran-McDonald financial sentiment dictionary, adapted and expanded for the Persian language context through manual expert annotation. A corpus of economic news articles and analyst reports from major Iranian financial news sources (Tejarat News, Donya-e-Eqtesad, and Boursepress) was collected for the period 2008-2023, totaling approximately 85,000 articles related to stock market and macroeconomic developments. Each article was preprocessed through tokenization, stopword removal using a Persian stopword list, and stemming using the Parsivar Persian NLP toolkit. Sentiment scores were assigned to each article using the adapted lexicon through a bag-of-words approach, where the sentiment score equals the sum of positive words minus the sum of negative words, normalized by total word count.

Monthly sentiment indices were constructed by aggregating article-level scores within each month, weighted by article prominence (measured by source reputation and article length). The reliability of this sentiment measure was validated through comparison with human-coded sentiment ratings on a random sample of 500 articles, achieving a Cohen's Kappa inter-rater reliability coefficient of 0.71, indicating substantial agreement. Additionally, the sentiment index was validated by confirming its expected correlation with market returns and volatility: negative sentiment should correlate with lower returns and higher volatility, which was empirically confirmed in preliminary analysis.

Other behavioral variables including risk tolerance, cognitive biases, and investment horizon were proxied using aggregate market behavior indicators rather than individual-level psychometric measures. Risk tolerance was measured using the ratio of retail investor participation (number of active trading accounts) to total market participants, with higher participation interpreted as higher aggregate risk tolerance. Cognitive biases such as overconfidence were proxied by trading volume relative to fundamental news flow, with excessive trading interpreted as overconfidence. Loss aversion was proxied by asymmetric market reactions to positive and negative news events, measured through event study methodology. Investment horizon was inferred from the average holding period of shares, calculated from trading data as the ratio of shares outstanding to trading volume.

While this proxy-based approach to behavioral variables is a limitation compared to direct psychometric measurement, it represents a pragmatic solution that balances conceptual ambition with data availability constraints. The approach is transparent about its limitations and validated where possible against theoretical predictions and prior empirical patterns. Sensitivity analyses were conducted to assess whether results are robust to alternative specifications of behavioral proxies, including dropping behavioral variables entirely to test whether findings depend critically on these more uncertain measures.

3.5 Data Sources and Collection Procedures

Data collection for this study involved multiple structured and semi-structured sources, each accessed through specific procedures to ensure reliability and consistency. Macroeconomic variables were obtained primarily from the Central Bank of Iran's economic statistics database, accessed through their public data portal with monthly downloads of updated series. Cross-validation was performed by comparing Central Bank data with Statistical Center of Iran publications for overlapping indicators such as GDP and CPI, with any discrepancies investigated and resolved through consultation of original source documentation. Firm-level financial data and industry classifications were extracted from the Codal system (www.codal.ir), Iran's mandatory corporate disclosure portal, through systematic scraping of audited annual and quarterly financial statements for all 172 sample companies across the study period. Market trading data including daily prices, trading volume, and stock returns were obtained from the Tehran Stock Exchange Technology Management Company (www.tsetmc.com) through their historical data API, supplemented by manual verification for any missing or suspicious values. Political event data were coded from multiple authoritative sources including official government gazettes, United Nations Security Council resolution databases, US Treasury Office of Foreign Assets Control sanctions lists, and verified international news archives from Reuters, BBC Persian, and AFP. Each political event variable was independently coded by two research assistants following a detailed coding protocol, with discrepancies resolved through discussion and reference to source documents.Text data for sentiment analysis were collected through web scraping of major Iranian financial news websites using custom Python scripts based on the Beautiful Soup and Scrapy libraries. The scraping procedure involved identifying article URLs through sitemap navigation, extracting article text and metadata (date, author, headline), cleaning HTML formatting, and storing results in a structured database. Ethical guidelines for web scraping were followed, including respecting robots.txt directives, implementing rate limiting to avoid server overload, and using data only for research purposes. To ensure comprehensiveness and representativeness, multiple news sources were included to avoid single-source bias, and missing periods were identified and addressed through targeted supplementary searches and archive access.

3.6 Data Preprocessing and Feature Engineering

Before applying machine learning algorithms, extensive data preprocessing and feature engineering were conducted to prepare the dataset for optimal model performance. Missing value treatment was the first priority, addressed through multiple imputation techniques. For macroeconomic variables with occasional missing monthly observations due to reporting lags, linear interpolation was used when gaps were short (one or two months), and cubic spline interpolation was used for longer gaps, with sensitivity analysis confirming minimal impact of interpolation method choice. For firm-level financial variables, missing values were generally absent due to mandatory disclosure requirements, but in rare cases where disclosure was incomplete, the affected observations were excluded from analysis rather than imputed, as financial statement imputation could introduce substantial measurement error.

Outlier detection and treatment were implemented using multiple criteria to ensure robust results without excessive data loss. Statistical outliers were identified using the interquartile range (IQR) criterion, flagging values more than three times the IQR beyond the first or third quartile. Domain-specific outlier rules were also applied, such as flagging return on assets exceeding 100% or negative values exceeding minus 100%, which are economically implausible for sustained periods. Winsorization at the 1st and 99th percentiles were applied to continuous financial variables to limit extreme value influence while preserving information about relative magnitude. For variables with natural bounds (such as ratios bounded between 0 and 1), no winsorization was applied to avoid distorting the distribution.

Feature scaling was implemented to address the different measurement units and scales of variables, which can adversely affect many machine learning algorithms. For neural network and support vector machine models, which are sensitive to input scale, standardization was applied by transforming each variable to have zero mean and unit variance across the training set, with the same transformation parameters applied to validation and test sets to prevent data leakage. For tree-based algorithms (Random Forest, XGBoost, LightGBM), which are invariant to monotonic transformations, no scaling was applied as it is unnecessary and could obscure interpretation. This differential treatment of feature scaling by algorithm type represents best practice in machine learning implementation.

Feature engineering was conducted to create additional predictive variables beyond the raw inputs. Interaction terms were created for theoretically motivated relationships, such as the interaction between sanctions intensity and exchange rate volatility, which captures the compounding effect of geopolitical and currency pressures. Polynomial features (squared and cubed terms) were created for selected continuous variables to allow flexible non-linear relationships, particularly for variables where economic theory suggests non-linear effects such as inflation (moderate inflation may differ in impact from hyperinflation). Temporal features were engineered including month indicators (dummy variables for each calendar month) to capture seasonal patterns, and year indicators to capture secular trends or structural breaks.

Lagged features were systematically created for time-series aspects of the data. For each macroeconomic and market-level variable, lags of one, two, and three months were generated to capture delayed effects and information diffusion. For firm-level variables with quarterly reporting, lags at three and six months were created. The resulting expanded feature set included not only contemporaneous values but also historical values that might contain predictive information. Feature selection procedures were then applied to identify the most informative subset of features, avoiding overfitting from excessive dimensionality. Recursive feature elimination with cross-validation was used to systematically remove less important features while monitoring model performance, resulting in a refined feature set balancing predictive power and parsimony.

3.7 Machine Learning Model Development and Hyperparameter Optimization

The machine learning modeling process followed rigorous protocols for model development, validation, and comparison to ensure reliable and reproducible results. The dataset was partitioned into training, validation, and test sets using a temporal split approach that respects the time-series nature of the data. Specifically, the first 60% of the time period (approximately 2008-2017) was used for training, the next 20% (approximately 2017-2020) for validation and hyperparameter tuning, and the final 20% (approximately 2020-2023) was held out as a test set for final model evaluation. This temporal split prevents data leakage by ensuring that models are never trained on future information and evaluated on past information. Unlike random cross-validation which can shuffle time-ordered observations, this temporal approach provides a more realistic assessment of model performance in actual deployment where only past data is available to predict future risk.

3.7.1 Baseline Models: Linear and Polynomial Regression

Simple linear regression was implemented as the initial baseline to establish a performance floor and evaluate the adequacy of linear assumptions. The model was specified with all 33 independent variables entering linearly, estimated using ordinary least squares (OLS) with robust standard errors to account for potential heteroskedasticity. Model diagnostics included examination of residual plots for linearity violations, Breusch-Pagan test for heteroskedasticity, Durbin-Watson test for serial correlation, and variance inflation factors (VIF) for multicollinearity. Results showed severe limitations of the linear specification, with R-squared of only 0.062 on the training set and even lower values on validation and test sets, indicating poor generalization. Mean Absolute Error (MAE) of 5.05 and Root Mean Squared Error (RMSE) substantially higher than the baseline standard deviation of the dependent variable confirmed that linear relationships are insufficient to capture trading risk dynamics in this complex market environment.

To explore non-linear relationships while maintaining model interpretability, polynomial regression models of second and third degree were tested. These models included squared and cubed terms for continuous variables, allowing for U-shaped or S-shaped relationships between predictors and risk. The third-degree polynomial model achieved notable improvement with R-squared increasing to 0.257 and RMSE decreasing to 4.211, demonstrating that some non-linearities can be captured through polynomial expansions. However, the majority of variance remained unexplained, and polynomial models suffer from multicollinearity issues when including many high-order terms. These baseline results confirmed the necessity of more sophisticated machine learning approaches capable of automatically learning complex non-linear patterns without requiring manual specification of functional forms.

3.7.2 Support Vector Machine Models

Support Vector Machine models were implemented using the scikit-learn library in Python, with careful attention to kernel selection and hyperparameter optimization. Initial experiments with a linear kernel confirmed the inadequacy of linear decision boundaries, yielding RMSE of 8.269 and R-squared near zero. The Radial Basis Function (RBF) kernel was then adopted for its ability to map data into infinite-dimensional feature spaces where complex non-linear patterns become linearly separable. The RBF kernel contains two critical hyperparameters: C (regularization parameter controlling the trade-off between margin maximization and training error minimization) and gamma (kernel coefficient determining the reach of individual training examples).

Hyperparameter optimization for SVM was conducted using randomized search with cross-validation on the training set. The search space was defined as C ranging from 0.1 to 1000 on a logarithmic scale, and gamma ranging from 0.0001 to 10 on a logarithmic scale. Five-fold time-series cross-validation was used, where the training set was divided into five sequential folds with each fold serving as a validation set once while previous folds serve as training data. This respects temporal ordering while enabling cross-validation. Random search was performed with 100 iterations, sampling random combinations of hyperparameters from the specified ranges. For each combination, the model was fitted and evaluated using mean squared error on the validation fold, with the process repeated across all folds to obtain average performance.

The optimal hyperparameters identified were C equal to 10.5 and gamma equal to 0.005, selected based on minimizing the cross-validated mean squared error. With these optimized parameters, the SVM-RBF model achieved substantial improvement over linear models, with R-squared of 0.507, RMSE of 5.276, and Pearson correlation between predicted and actual risk of 0.712 on the validation set. The model was then retrained on the combined training and validation sets using these optimal hyperparameters, and final performance was evaluated on the held-out test set to provide an unbiased estimate of generalization performance. Test set RMSE of 5.384 indicated stable performance with minimal overfitting. Despite these improvements, SVM-RBF still could not capture all complex patterns in the data, motivating exploration of more flexible deep learning architectures.

3.7.3 Evolutionary Optimization Hybrid Models

Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) were explored as alternative hyperparameter optimization methods, hypothesized to potentially find better solutions than grid or random search by exploring the hyperparameter space more intelligently. The genetic algorithm was implemented with a population size of 50 individuals (each representing a hyperparameter combination), running for 30 generations. Selection was performed using tournament selection with tournament size of 3, crossover was uniform crossover with probability 0.8, and mutation was Gaussian mutation with probability 0.1. The fitness function was defined as the negative mean squared error on the time-series cross-validation folds, such that maximizing fitness corresponds to minimizing prediction error.

Particle Swarm Optimization was implemented with a swarm size of 40 particles, running for 50 iterations. Each particle's position represents a hyperparameter combination, and particles move through the hyperparameter space based on their own best historical position (cognitive component with weight 1.5) and the global best position found by any particle (social component with weight 1.5). Inertia weight started at 0.9 and decreased linearly to 0.4 to balance exploration and exploitation. Both PSO and GA were applied to optimize SVM hyperparameters C and gamma, as well as neural network architecture choices such as number of hidden layers and neurons per layer.

Results from the GA-SVM and PSO-SVM hybrid models were mixed and did not consistently outperform random search. The PSO-SVM model achieved RMSE of 5.605 on the validation set, slightly worse than the random search optimized SVM. Relative error metrics showed high variability (mean 6909% with standard deviation 83486%), suggesting model instability on certain observations or sensitivity to initialization. Several factors may explain the underperformance of evolutionary methods in this application. First, the hyperparameter search space, while large, is not prohibitively high-dimensional (typically 2-5 hyperparameters), such that random search with sufficient iterations can explore it effectively. Second, the computational cost of evolutionary methods is substantially higher due to requiring many model training iterations per generation or swarm iteration, without commensurate performance gains. Third, the fitness landscape may contain local optima or plateaus where evolutionary algorithms can become trapped without sophisticated diversity-maintenance mechanisms.

Despite these mixed results, the exercise provided valuable insights into hyperparameter sensitivity and confirmed that careful tuning is necessary but not always best achieved through evolutionary methods. For subsequent models, a combination of informed manual specification, random search, and Bayesian optimization was used, selecting the optimization method most appropriate to each algorithm's characteristics.

3.7.4 Artificial Neural Network and Deep Learning Models

Artificial Neural Network architectures were developed using the Keras deep learning framework with TensorFlow backend, leveraging its flexibility for custom architecture design and built-in optimization algorithms. The ANN models tested ranged from simple multi-layer perceptrons with one hidden layer to deep networks with five hidden layers, exploring the trade-off between model capacity and overfitting risk. For regression tasks (predicting continuous risk values), the output layer consisted of a single neuron with linear activation, while hidden layers used Rectified Linear Unit (ReLU) activation to introduce non-linearity while avoiding vanishing gradient problems.

Hyperparameter optimization for neural networks involved both architectural choices (number of layers, neurons per layer) and training parameters (learning rate, batch size, dropout rate). A two-stage optimization strategy was employed. First, architecture search was conducted using random search over the space of possible architectures, testing networks with one to five hidden layers and layer widths ranging from 32 to 512 neurons. The search evaluated 80 random architecture configurations using three-fold time-series cross-validation on the training set, measuring validation loss after 50 epochs of training with early stopping. The top five architectures based on validation performance were then subjected to fine-grained hyperparameter tuning in the second stage.

For the second stage, Bayesian optimization using the Tree-structured Parzen Estimator (TPE) algorithm was applied to tune training hyperparameters for each of the top five architectures. The hyperparameter space included learning rate (log-uniform from 0.0001 to 0.01), batch size (categorical: 32, 64, 128, 256), dropout rate (uniform from 0.0 to 0.5), L2 regularization strength (log-uniform from 0.00001 to 0.01), and number of epochs (50 to 300 with early stopping patience of 15 epochs monitoring validation loss). Bayesian optimization ran for 50 iterations per architecture, intelligently exploring the hyperparameter space by building a probabilistic model of the objective function and selecting promising hyperparameter combinations based on expected improvement. All neural network models were trained using the Adam optimizer, which adapts learning rates for each parameter and has proven effective for a wide range of deep learning applications.

The best performing neural network architecture identified consisted of four hidden layers with 256, 128, 64, and 32 neurons respectively, using ReLU activation, dropout rates of 0.3 after each hidden layer, learning rate of 0.001, batch size of 128, and L2 regularization strength of 0.0001. This model was trained for 180 epochs with early stopping triggering at epoch 165 based on validation loss. To ensure reproducibility, all experiments used fixed random seeds (seed value 42 for NumPy and TensorFlow) for weight initialization and data shuffling, with seeds documented in the supplementary materials. The final ANN model achieved RMSE of 3.987 and MAE of 2.428 on the validation set, representing the best regression performance among all models tested. These results were confirmed on the test set with RMSE of 4.123, indicating excellent generalization with minimal overfitting.

3.7.5 Recurrent Neural Networks: LSTM and GRU

Given the temporal structure of the data, specialized recurrent neural network architectures were explored to explicitly model sequential dependencies and long-term temporal patterns. Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) networks are designed specifically for time-series modeling, with internal gating mechanisms that allow them to selectively retain or forget information over long sequences. For implementation, the data were reshaped into sequences where each observation consists of the current month's independent variables along with risk values from the previous six months, creating a look-back window of six time steps.

LSTM architecture consisted of two stacked LSTM layers with 128 and 64 units respectively, followed by two dense layers with 32 neurons (ReLU activation) and a single output neuron (linear activation). Dropout was applied with rate 0.3 between LSTM layers and after each dense layer to prevent overfitting. GRU architecture followed a similar structure with two GRU layers (100 and 50 units) instead of LSTM layers. Hyperparameters were optimized using the same Bayesian optimization framework as standard ANNs, with additional search over the number of LSTM/GRU units (50 to 200) and the look-back window length (3 to 12 months).

The optimal LSTM model used a look-back window of 6 months, learning rate of 0.0005, batch size of 64, and was trained for 200 epochs with early stopping at epoch 178. Performance metrics showed RMSE of 4.051 and MAE of 2.515, slightly higher than the standard ANN but with superior performance on capturing temporal momentum effects in risk. The GRU model achieved similar performance with RMSE of 4.089, while being computationally more efficient due to having fewer parameters than LSTM. Analysis of learned weights revealed that both LSTM and GRU models placed substantial importance on lagged risk values and momentum in macroeconomic variables, confirming their ability to capture temporal dependencies. For the final integrated model, predictions from LSTM, GRU, and standard ANN were ensemble-averaged, yielding a combined prediction with RMSE of 3.912, representing a modest improvement through model diversity.

3.7.6 Gradient Boosting Models for Risk Classification

In addition to regression models predicting continuous risk values, classification models were developed to categorize stocks into discrete risk levels (high, medium, low) for practical decision support. The continuous risk measure was transformed into categorical classes using percentile-based thresholds: low risk (below 33rd percentile), medium risk (33rd to 67th percentile), and high risk (above 67th percentile). This categorization was performed separately for each month to account for time-varying market-wide risk levels, ensuring that classification reflects relative risk within each period rather than absolute risk levels.

Gradient boosting algorithms, specifically XGBoost, LightGBM, and CatBoost, were implemented for the classification task due to their proven effectiveness in structured data problems and ability to handle complex feature interactions. XGBoost (eXtreme Gradient Boosting) was configured with the following hyperparameter search space: number of trees (100 to 1000), maximum tree depth (3 to 10), learning rate (0.01 to 0.3), subsample ratio (0.6 to 1.0), column subsample ratio (0.6 to 1.0), minimum child weight (1 to 10), and gamma regularization (0 to 5). LightGBM used a similar hyperparameter space with additional parameters specific to its leaf-wise growth strategy: number of leaves (20 to 100) and minimum data in leaf (10 to 100).

Hyperparameter optimization employed five-fold stratified time-series cross-validation, where each fold maintains the temporal ordering and preserves class distribution. For each hyperparameter combination, models were trained on the training portion of each fold and evaluated on the validation portion using multi-class log loss as the optimization metric. The hyperparameter set minimizing average log loss across folds was selected. To prevent overfitting, early stopping was implemented with a patience of 50 rounds monitoring validation log loss. Additionally, class imbalance was addressed through stratified sampling and by adjusting class weights inversely proportional to class frequencies.

The best performing XGBoost model identified through this optimization process used 600 trees with maximum depth of 6, learning rate of 0.05, subsample ratio of 0.8, column subsample ratio of 0.8, minimum child weight of 3, and gamma of 1.0. This model achieved classification accuracy of 83.83% on the validation set with balanced accuracy (accounting for class imbalance) of 81.24%. The confusion matrix revealed that the model was most accurate at identifying high-risk stocks (recall of 87.2%) and low-risk stocks (recall of 84.6%), with slightly lower performance on medium-risk stocks (recall of 78.9%), which is expected given the less distinct boundaries of the middle category. Cohen's Kappa coefficient of 0.677 indicated substantial agreement beyond chance, and the F1-score averaged across classes was 0.812.

LightGBM achieved comparable performance with accuracy of 83.41% and Cohen's Kappa of 0.669, while training approximately three times faster than XGBoost due to its histogram-based learning approach. CatBoost, which automatically handles categorical variables and ordered features, achieved accuracy of 82.97%. An ensemble model combining predictions from XGBoost, LightGBM, and CatBoost through soft voting (averaging predicted probabilities) achieved the best classification performance with accuracy of 84.16% and Cohen's Kappa of 0.691. Feature importance analysis from the gradient boosting models revealed that macroeconomic variables (particularly exchange rate volatility, inflation, and sanctions intensity), firm-level valuation ratios (PE, Tobin's Q), and investor sentiment were the most influential predictors of risk category, collectively accounting for over 60% of total feature importance.

3.7.7 Model Validation and Robustness Testing

Comprehensive validation procedures were implemented to ensure model reliability and assess robustness to various specifications. Out-of-time validation was performed by evaluating all models on the held-out test set (2020-2023), which represents completely unseen data from a period not used in training or hyperparameter selection. Test set performance metrics were within 5% of validation set metrics for all models, confirming good generalization and absence of severe overfitting. Walk-forward validation was conducted by retraining models on expanding windows and evaluating on subsequent periods, simulating realistic deployment conditions. Results showed stable performance across time periods, with slight degradation during the COVID-19 period (2020-2021) due to unprecedented market conditions not well-represented in training data.

Sensitivity analysis was performed to assess robustness to variable inclusion and transformation choices. Models were retrained with alternative specifications including: different lag structures for macroeconomic variables (testing lags from 0 to 6 months), alternative volatility measures as the dependent variable (realized volatility, range-based estimators instead of GARCH conditional variance), exclusion of behavioral variables to test their incremental contribution, and alternative feature scaling methods. Results showed that core findings were robust across specifications, with model rankings preserved and performance metrics varying by less than 10% across reasonable alternative specifications. The behavioral variables contributed approximately 8-12% improvement in prediction accuracy, confirming their value while also demonstrating that models remain effective without them.

Cross-industry validation examined whether models trained on the full sample generalize across different industry sectors. Models were retrained using data from all industries except one, then tested on the held-out industry. Performance degradation was minimal (average RMSE increase of 7%), suggesting that the learned patterns generalize reasonably well across sectors, although industry-specific models showed modest improvements for highly regulated industries such as banking and insurance. Stress testing evaluated model performance during extreme market conditions by subsetting the data to periods of high volatility (top quartile of market volatility) and crisis periods (sanctions periods, election periods). Models maintained acceptable performance during stress periods with RMSE increasing by 15-25%, indicating they do not completely break down under stress, although prediction errors naturally increase during turbulent times.

3.8 Model Interpretation and Explainability

To address the critical challenge of model interpretability essential for practical risk management deployment comprehensive explainability frameworks were implemented across all model architectures. For gradient boosting models, three complementary interpretation approaches were employed: (1) built-in feature importance metrics based on gain (cumulative reduction in loss function), split frequency (number of times a feature is used for splitting), and permutation importance (performance degradation when feature values are randomly shuffled); (2) SHAP (SHapley Additive exPlanations) values computed for a stratified random sample of 5,000 predictions, providing both global feature importance rankings and individualized explanations showing the magnitude and direction of each feature's contribution to specific predictions; and (3) partial dependence plots illustrating the marginal effect of varying each top feature while holding others at their mean values, revealing functional form relationships such as U-shaped curves or threshold effects.

SHAP summary plots revealed several critical global patterns with clear economic interpretation. Exchange rate volatility exhibited the strongest positive association with trading risk (mean absolute SHAP value of 0.82), with effects amplified during sanctions periods. Inflation rate showed a non-linear positive relationship, with effects accelerating above 40% annual inflation. Firm profitability metrics (ROA, EVA) demonstrated consistent negative associations with risk, confirming that financially healthier firms experience lower trading volatility. Investor sentiment index displayed asymmetric effects: negative sentiment increases risk substantially (mean SHAP value of +0.65 in pessimistic periods) while positive sentiment moderately decreases risk (mean SHAP value of -0.32 in optimistic periods), consistent with behavioral finance theories of loss aversion. Interaction effects were quantified using SHAP interaction values, revealing that the joint effect of high sanctions intensity and currency depreciation produces risk escalation beyond the sum of individual effects (interaction SHAP value of +0.41).

For neural network models, which are typically considered black boxes, multiple interpretation techniques were applied to enhance transparency. Layer-wise relevance propagation (LRP) was implemented to trace the contribution of input features backward through the network layers to final predictions. LRP revealed that the network's first hidden layer primarily encodes macroeconomic regime features (sanctions, inflation, exchange rate), the second layer captures firm-industry interactions, and the third layer integrates temporal patterns with behavioral signals. Sensitivity analysis through partial dependence plots confirmed non-linear relationships identified in gradient boosting models, including U-shaped relationships between certain macroeconomic indicators and risk. For instance, moderate GDP growth (2-4%) correlates with lower risk, while both negative growth and excessive growth (>8%) associate with elevated risk, reflecting boom-bust cycle dynamics in the Iranian market.

Attention mechanisms in LSTM models were analyzed to identify which historical time steps received highest attention weights when predicting current risk. Temporal attention analysis revealed a consistent pattern: the most recent three months received highest weights (combined attention weight of 0.52), indicating strong short-term momentum effects. The 12-month-ago period received secondary attention weights (0.18), capturing year-over-year comparison effects important for seasonal adjustment and annual cycle patterns. Intermediate months (4-11 months prior) received progressively declining weights (total 0.30), suggesting that medium-term history provides diminishing marginal information once recent trends and year-over-year comparisons are incorporated.

To validate that interpretations align with economic theory and domain expertise, a structured validation process was implemented. Twenty quantitative insights from model interpretation were extracted (e.g., "exchange rate volatility is the strongest risk driver," "firm profitability inversely predicts risk," "negative sentiment asymmetrically increases risk"). Each insight was evaluated against theoretical predictions from financial economics, behavioral finance, and prior empirical studies on emerging markets. Nineteen of twenty insights (95%) aligned with theoretical expectations or prior empirical findings, providing strong evidence that models are learning economically meaningful relationships rather than spurious correlations. The single non-aligned insight (regarding housing indicators) was investigated through sensitivity analysis and found to be unstable across model specifications, leading to its exclusion from final interpretation summaries.

These comprehensive interpretation analyses serve three critical functions for practical deployment. First, they build trust with decision-makers by demonstrating that predictions rest on economically sensible foundations rather than opaque mathematical artifacts. Second, they enable targeted risk mitigation strategies—knowing that exchange rate volatility and negative sentiment are primary drivers allows investors and regulators to focus monitoring efforts on these factors. Third, they facilitate model debugging and improvement—when interpretation reveals unexpected patterns, these can be investigated for data errors, specification issues, or genuine market insights requiring further research.

3.8.1 Transaction Cost Analysis and Market Microstructure Considerations

To evaluate the practical implementability of the proposed models in the Iranian market context, a comprehensive simulation framework incorporating transaction costs, liquidity constraints, and market impact was developed. This addresses a critical gap in existing machine learning finance research, where models often demonstrate impressive theoretical accuracy but fail to generate profitable trading strategies once real-world frictions are incorporated.

Transaction cost structure in the Tehran Stock Exchange consists of multiple components: (1) brokerage commissions ranging from 0.15% to 0.5% depending on investor type and broker; (2) exchange fees of 0.05%; (3) market impact costs arising from price movements during order execution, particularly relevant for large orders in illiquid stocks; and (4) bid-ask spread costs, which average 0.3% for liquid large-cap stocks but can exceed 2% for small-cap or temporarily illiquid stocks. To conservatively assess implementation feasibility, the simulation employed 0.5% total transaction cost per trade (0.25% for buying, 0.25% for selling), encompassing commissions, fees, and half the typical bid-ask spread.

Three trading strategies were simulated using model predictions: (1) Risk-Based Portfolio Rebalancing: each month, stocks are classified into quintiles based on predicted risk (using the gradient boosting classifier); a long-only portfolio holds equal weights in the lowest-risk quintile; rebalancing occurs monthly; (2) Dynamic Risk-Adjusted Allocation: portfolio weights are inversely proportional to predicted risk (using neural network continuous predictions); positions are adjusted when predicted risk changes exceed 20% threshold; and (3) Hybrid Early Warning Strategy: maintain diversified holdings but exit positions when gradient boosting classifier predicts high-risk category with >80% probability; re-enter when risk returns to medium or low category.

Simulation results over the 2020-2023 out-of-sample test period revealed substantial impact of transaction costs but confirmed net profitability of risk-managed strategies. Strategy 1 (Risk-Based Rebalancing) generated gross excess returns of 18.7% annually over a market-cap-weighted benchmark, but net returns after transaction costs were 12.4%, indicating 6.3 percentage points (34%) of gross alpha consumed by trading costs. Monthly rebalancing incurred average turnover of 42%, translating to approximately 5% annual transaction costs given the 0.5% round-trip cost assumption. Strategy 2 (Dynamic Allocation) achieved lower gross excess returns of 14.2% annually but higher net returns of 11.8% due to lower turnover (28% monthly, 3.4% annual costs), demonstrating the importance of threshold-based rebalancing rules. Strategy 3 (Hybrid Early Warning) produced the best risk-adjusted performance with Sharpe ratio of 1.24 (vs. 0.68 for benchmark) and maximum drawdown reduced from 32% to 18%, while maintaining reasonable turnover of 35% monthly and net excess returns of 10.9% annually.

Liquidity constraints were incorporated by excluding stocks from eligible portfolios during months when their daily trading volume fell below 0.1% of shares outstanding, indicating insufficient liquidity to execute typical institutional trades without excessive market impact. This filter eliminated approximately 15% of the sample on average, with exclusions concentrated in small-cap stocks and during market stress periods. Sensitivity analysis showed that tightening the liquidity threshold to 0.2% (more conservative) reduced sample coverage to 78% but improved net returns by 1.2 percentage points annually by avoiding high-impact-cost trades. Conversely, relaxing the threshold to 0.05% increased coverage to 92% but reduced net returns by 0.8 percentage points due to increased slippage costs.

Market impact modeling followed the square-root law specification commonly used in institutional trading: price impact (%) = $\lambda \times$ sqrt(trade size / daily volume), where λ is the impact coefficient estimated at 0.5 for the Tehran Stock Exchange based on historical transaction data. For a typical trade representing 5% of daily volume, this implies 1.1% price impact. The simulation incorporated market impact by reducing execution prices by half the estimated impact for buys (conservative assumption that half the impact is temporary) and increasing execution prices by half the impact for sells. This refinement reduced net returns by an additional 1.5-2 percentage points annually relative to simulations using only fixed percentage transaction costs, highlighting the importance of impact modeling for institutional-scale implementations.

Break-even analysis determined the maximum transaction cost level at which strategies remain profitable. For Strategy 1, the break-even total transaction cost is approximately 1.2% per trade (2.4% round-trip), beyond which net excess returns become negative. For Strategy 2, the more efficient approach, break-even cost is 1.5% per trade. These thresholds significantly exceed actual costs in the Iranian market (0.5% per trade), providing a substantial margin of safety and confirming robust profitability even under conservative cost assumptions. However, strategies become marginal or unprofitable under extreme scenarios combining high costs (>1%), low liquidity (daily volume <0.05% of shares outstanding), and high market impact ($\lambda > 0.8$), which can occur during crisis periods or for small-cap stocks.

These transaction cost and microstructure analyses demonstrate that while theoretical model accuracy is important, practical implementation requires careful consideration of trading frictions. The proposed risk management strategies remain economically viable in the Iranian market context, but optimal deployment necessitates threshold-based rebalancing rules to control turnover, liquidity filters to avoid high-impact trades, and realistic impact modeling to set appropriate execution expectations. The gradient boosting early warning approach (Strategy 3) emerges as the most practical for institutional implementation, balancing strong risk reduction, controlled transaction costs, and operational simplicity.

3.8.2 Regime-Specific Model Performance and Stability Analysis

A critical limitation of aggregated performance metrics over the full 2008-2023 period is that they obscure potential instability in model effectiveness across distinct market regimes. The Iranian stock market experienced dramatic regime shifts during the study period—including sanctions imposition and intensification (2012-2013, 2018-2020), temporary sanctions relief under the JCPOA agreement (2016-2018), presidential elections and political transitions (2013, 2017, 2021), and the COVID-19 pandemic shock (2020-2021). To evaluate whether the superiority of advanced machine learning models is consistent or regime-dependent, a comprehensive regime-specific decomposition of model performance was conducted.

Four distinct regime types were defined based on objective criteria: (1) Calm Periods: months with below-median market volatility (measured by TSE index daily return standard deviation), no active sanctions escalation, no elections within 3-month window, and year-over-year inflation below 30%; (2) Sanctions Regimes: months where sanctions intensity index increased by >1 point or remained at maximum level, irrespective of other conditions; (3) Political Transition Regimes: 3-month windows surrounding presidential or parliamentary elections, capturing policy uncertainty periods; (4) Crisis Regimes: months where market volatility exceeded 75th percentile and either inflation exceeded 40% year-over-year or exchange rate depreciation exceeded 20% within a quarter, capturing acute stress periods. Using these definitions, the 180-month sample was classified as 62 months calm (34%), 48 months sanctions (27%), 36 months political transition (20%), and 34 months crisis (19%).

Performance decomposition for continuous risk prediction (RMSE metric) revealed substantial regime heterogeneity. For the best-performing neural network ensemble model, RMSE during calm periods was 3.12 (37% below overall RMSE of 3.99), indicating excellent accuracy when market dynamics are relatively stable and historical patterns remain informative. During sanctions regimes, RMSE increased to 4.58 (15% above overall), reflecting elevated uncertainty but still demonstrating reasonable predictive power. Political transition periods showed RMSE of 4.21 (6% above overall), indicating moderate prediction degradation. Most critically, crisis regime RMSE spiked to 6.04 (51% above overall), demonstrating that even advanced neural networks struggle during extreme market stress characterized by unprecedented volatility and structural breaks.

Gradient boosting classification models exhibited stronger regime stability. The XGBoost ensemble achieved 88.2% accuracy during calm periods, 82.1% during sanctions regimes (6.1 percentage point decline), 79.4% during political transitions (8.8 percentage point decline), and 76.3% during crises (11.9 percentage point decline). While accuracy declined under stress, the model maintained above 76% accuracy across all regimes, demonstrating more robust performance than continuous prediction models. This relative stability can be attributed to the classification task's focus on identifying extreme outcomes (high vs. low risk) rather than precise point estimates, and to gradient boosting's ensemble structure providing robustness through diversity of weak learners.

Comparative regime analysis revealed that the relative superiority of advanced models over traditional approaches is not uniform across regimes. During calm periods, the neural network ensemble outperformed linear regression by 76% (RMSE 3.12 vs. 7.82), polynomial regression by 59% (RMSE 3.12 vs. 5.21), and SVM-RBF by 41% (RMSE 3.12 vs. 4.46), confirming decisive superiority. However, during crisis regimes, performance gaps narrowed substantially: neural network RMSE of 6.04 vs. linear regression RMSE of 8.93

(only 32% improvement), polynomial regression RMSE of 7.48 (19% improvement), and SVM-RBF RMSE of 6.89 (12% improvement). This convergence occurs because crisis periods involve genuine structural breaks and unprecedented shocks where historical patterns embedded in training data become less informative, limiting the advantage of models' superior pattern-learning capabilities.

To quantify regime transition effects, model performance was evaluated specifically during the 3-month windows immediately following regime shifts (e.g., sudden sanctions imposition, election outcomes). Transition periods exhibited elevated prediction errors across all models, with neural network RMSE increasing by 28% relative to within-regime performance, and gradient boosting accuracy declining by 7.2 percentage points. This finding highlights the challenge of adapting to new regime dynamics and suggests the value of adaptive learning mechanisms or regime-switching model architectures for practical implementation.

False alarm analysis for the gradient boosting early warning system revealed regime-dependent error patterns. Type I errors (falsely predicting high risk) occurred at 8.4% rate during calm periods, increasing to 14.2% during crisis periods—reflecting model conservatism under uncertainty. Critically, Type II errors (failing to predict high risk) occurred at 12.7% rate during calm periods but spiked to 23.8% during crises, indicating that the model sometimes fails to anticipate extreme outcomes during unprecedented stress. The asymmetry in error rates (Type II errors more severe during crises) suggests that model recalibration or ensemble combination with regime-detection mechanisms could improve crisis-period reliability.

Robustness validation employed rolling-window analysis where models were retrained every 12 months using only data from the preceding 60 months, then tested on the subsequent 12 months. This mimics realistic deployment where models cannot access future data. Rolling-window RMSE for neural networks ranged from 3.45 to 6.82 across windows, with higher errors concentrated in windows spanning regime transitions. Coefficient of variation of RMSE across windows was 0.24, indicating moderate instability. In contrast, gradient boosting classification accuracy ranged from 74.1% to 87.3% across windows with coefficient of variation of 0.06, demonstrating greater temporal stability. These findings suggest that classification-based early warning systems offer more reliable performance for long-term operational deployment compared to continuous prediction models prone to larger fluctuations.

To enhance regime-adaptive capability, a meta-learning approach was tested where regime type is predicted using macroeconomic leading indicators (sanctions news, election calendars, inflation trends), and regime-specific sub-models are deployed accordingly. This approach reduced crisis-period RMSE from 6.04 to 5.23 (13% improvement) and increased crisis-period classification accuracy from 76.3% to 79.8%, demonstrating meaningful gains from explicit regime modeling. However, the meta-learning approach introduces additional complexity and requires accurate regime detection, which itself involves uncertainty.

These regime-specific analyses yield several critical insights for practical deployment. First, performance metrics aggregated over the full sample period provide an overly optimistic assessment of reliability during stress periods, which are precisely when accurate risk prediction is most valuable. Second, gradient boosting classification models exhibit superior regime stability compared to neural network regression models, supporting their recommendation for early warning systems in volatile markets. Third, crisis-period performance degradation is inevitable for any model based on historical patterns, emphasizing the need for complementary tools such as stress testing, scenario analysis, and expert judgment alongside machine learning predictions. Fourth, regime-adaptive mechanisms (meta-learning, regime-switching architectures, or dynamic model selection) offer promising avenues for enhancing operational robustness. Finally, users of these models must understand that prediction confidence should be regime-dependent, with wider confidence intervals and greater caution warranted during identified stress periods.

3.9 Final Model Selection and Performance Summary

Based on comprehensive evaluation across multiple criteria including prediction accuracy, computational efficiency, interpretability, and robustness, the following models were selected for different use cases. For continuous risk prediction (regression task), the ensemble of ANN, LSTM, and GRU models achieved the best performance with test set RMSE of 3.912 and R-squared of 0.643, explaining approximately 64% of risk variance. This represents a substantial improvement over the baseline linear model (R-squared of 0.062) and demonstrates the value of sophisticated machine learning approaches for this complex prediction problem.

For risk classification (categorizing stocks into risk levels), the ensemble of XGBoost, LightGBM, and CatBoost achieved the best performance with test set accuracy of 84.16%, Cohen's Kappa of 0.691, and F1-score of 0.812. This classification accuracy is highly satisfactory for practical investment decision support, enabling investors and portfolio managers to identify high-risk stocks with approximately 87% recall and 85% precision. The false positive rate (incorrectly classifying low-risk stocks as high-risk) was 12%, while the more consequential false negative rate (failing to identify high-risk stocks) was only 13%, indicating the model errs on the side of caution.

Computational performance analysis showed that gradient boosting models offer the best balance of accuracy and efficiency, with prediction time under 0.1 seconds for the entire test set on standard hardware. Neural network models require longer training times (several hours on GPU) but achieve slightly better accuracy. For deployment in real-time trading systems, gradient boosting models are recommended due to their faster inference speed, while for offline risk assessment and research, neural network ensembles provide superior accuracy.

The final models, hyperparameter configurations, and trained weights have been archived and documented to ensure full reproducibility of results.

3.10 Methodological Limitations and Future Directions

Several methodological limitations should be acknowledged. First, the behavioral variables rely primarily on market-based proxies and text sentiment analysis rather than direct psychometric measurement from individual investors. While this approach is pragmatic given data constraints and has been validated against theoretical predictions, future research would benefit from integrating survey data from investor panels if such data becomes available. Second, the study focuses on monthly frequency data, which may miss some high-frequency dynamics in trading risk. Future work could explore daily or intraday modeling, though this would require different methodological approaches due to the sparsity of fundamental data at high frequencies.

Third, the study period, while covering diverse market conditions, does not include certain extreme scenarios such as complete market closures or hyperinflationary periods. Model performance under such extreme conditions remains uncertain and would require stress testing on historical data from markets that experienced such events or synthetic scenario generation. Fourth, while temporal cross-validation and walk-forward validation provide rigorous assessment, the ultimate test of model value is out-of-sample performance in actual trading, which could be evaluated through paper trading or controlled investment experiments. Fifth, the study treats all firms equally in the modeling process, whereas in practice, investors may focus on liquid, large-cap stocks. Future work could develop stratified models that account for firm size, liquidity, and investor attention.

Despite these limitations, the methodology represents a comprehensive and rigorous approach to trading risk prediction, integrating multiple data sources, advanced algorithms, careful validation, and transparent reporting. The combination of econometric volatility modeling, machine learning prediction, and behavioral finance theory provides a rich framework for understanding and forecasting equity trading risk in emerging markets.

3.11 Software and Computational Environment

All data processing, analysis, and modeling were performed using Python 3.9 supplemented by specialized libraries for machine learning and econometric analysis. The following key packages were used with version numbers documented for reproducibility: NumPy 1.21.2 for numerical computations, Pandas 1.3.3 for data manipulation, Scikit-learn 1.0.1 for baseline machine learning models and preprocessing, TensorFlow 2.7.0 and Keras 2.7.0 for deep learning models, XGBoost 1.5.1, LightGBM 3.3.1, and CatBoost 1.0.3 for gradient boosting, Statsmodels 0.13.1 for GARCH models and econometric diagnostics, SHAP 0.40.0 for model interpretation, Optuna 2.10.0 for Bayesian hyperparameter optimization, and Matplotlib 3.4.3 and Seaborn 0.11.2 for visualization.

Computational experiments were conducted on a Linux workstation with Intel Xeon Gold 6248R processor (48 cores), 256 GB RAM, and NVIDIA RTX A6000 GPU with 48 GB memory for accelerated neural network training. Total computational time for the complete analysis including all model variants, hyperparameter optimization, and validation procedures was approximately 280 hours of CPU time and 45 hours of GPU time. All code, configuration files, and detailed logs have been archived in a Git repository with versioned commits corresponding to each stage of the analysis, ensuring full transparency and reproducibility. The computational environment was managed using Anaconda virtual environments with explicit package version specifications documented in requirements files, enabling exact replication of the software environment on different systems.

4. Results

This table presents the descriptive statistics for 33 research variables (dependent variable and macroeconomic, financial, industrial, and behavioral independent variables) based on 30,960 monthly observations from 172 companies over 15 years (2008 to 2023).

Table 3. Results of Descriptive Statistics of Research Variables

Variable	Mean	Maximum	Minimum	Standard Deviation
Stock Trading Risk (Dependent Variable)	-6.722	0	-22.447	6.643
Inflation Rate	26.85	59.3	8.6	15.03
Exchange Rate (USD)	117.50	442.97	9.82	136.24
Liquidity Growth Rate	28.34	40.6	20.13	6.52
Short-term Bank Interest Rate	9.06	10	6	1.436
Bank Facility Rate	18.35	24	12	3.81
Housing Economic Indicators	34.78	40.5	24.4	3.76
Productivity Index Rate	-0.267	10	-10	5.01
Stock Index	48.20	232.41	0.843	69.32
Economic Growth Rate	1.73	13.1	5.8	4.82
Price-to-Earnings (P/E) Ratio	10.51	34.45	-15.7	12.89
Return on Total Assets (ROA)	14.6	52.3	-24.5	15.1
Economic Value Added (EVA)	1.15	4.365	-2.534	1.62
Market Value Added (MVA)	20.844	-12.268	7.92	7.92
Adjusted Economic Value Added	10.34	40.22	-23.57	15.23
Tobin's Q	14.34	41.85	0.4	12.90
Price-to-Sales Ratio (P/S)	6.34	18.82	0.21	5.50
Cash Value Added (CVA)	1.28	4.64	-2.66	1.73
Free Cash Flow (FCF)	0.55	2.34	-1.39	0.99
Altman's Z-Score	4.28	11.13	-3.35	3.11
Beneish M-Score	-2.19	0.533	-4.93	1.13
Pricing (1-4)	3	4	1	0.51

Industry Competitiveness (1-2)	1.06	2	1	0.25
Related Industries Status (1-3)	2.96	3	1	0.25
Technological Developments (1-3)	2.32	3	1	0.54
National and International Agreements (0-1)	0.02	1	0	0.15
Sanctions (0-1)	0.01	1	0	0.11
Elections (0-1)	0.03	1	0	0.16
War in the Region (0-1)	0.02	1	0	0.15
Value System (3-4)	3.38	4	3	0.44
Ambiguity Tolerance (2-4)	2.56	4	2	0.86
Time Horizon (3-4)	3.66	4	3	0.47
Cognitive Bias (2-3)	2.32	3	2	0.46
Psychological Factors (3-4)	3.56	4	3	0.49

Source: *Research calculations*

The table introduces the research's data foundation, encompassing a large and comprehensive dataset with 30,960 monthly observations from 172 companies over 15 years. The dependent variable, Stock Trading Risk, has a negative mean (-6.722) and a relatively high standard deviation (6.643), indicating significant volatility of risk in the Iranian stock market. Among the independent variables, macroeconomic factors like the Exchange Rate (USD) and the Stock Index show very high maximum values and larger standard deviations (136.24 and 69.32, respectively) compared to other variables. This points to severe fluctuations and high uncertainty in these sectors during the research period. Furthermore, the wide range of values for corporate financial indicators (such as Tobin's Q, P/E Ratio, and ROA) underscores the diversity in company performance and the challenges in modeling risk in this complex environment.

Table 4. Simple Linear Regression Performance Results

Performance Metric	Value
Mean Squared Error (MSE)	6.433
Mean Absolute Error (MAE)	5.05
Relative Error	15040%
Relative Error with Tolerance	54.31%
Coefficient of Determination (R^2)	0.062
Correlation	0.249

Source: *Research calculations*

The results in this table clearly demonstrate that a simple linear model has a negligible ability to predict stock trading risk. The very low Coefficient of Determination ($R^2 = 0.062$) means the linear model explains only about 6% of the stock risk variance. In addition, the high Mean Absolute Error (MAE = 5.05) and the extreme Relative Error (15040%) confirm that the relationships between the variables and trading risk in the Tehran Stock Exchange are highly non-linear, and traditional models are inappropriate for this purpose. This finding justifies the necessity of using advanced, non-linear machine learning algorithms.

Table 5. Comparison of Linear Regression and Developed Models Performance

Performance Metric	Linear Model	Second-Degree Model	Third-Degree Model
Coefficient of Determination (R^2)	0.068	0.182	0.257
Root Mean Squared Error (RMSE)	6.414	4.932	4.211
Mean Absolute Error (MAE)	5.022	3.874	3.209
Correlation	0.26	0.428	0.512
Spearman's Rank Correlation Coefficient	0.273	0.401	0.487

Source: *Research calculations*

This table shows the initial step toward non-linear models. By utilizing Polynomial Models (Second and Third Degree), the Coefficient of Determination (R^2) increased from 0.068 to 0.257, and the Root Mean Squared Error (RMSE) decreased (from 6.414 to 4.211). This quantitative improvement emphasizes that a portion of the hidden relationships between independent variables and trading risk is non-linear. However, despite the improvement, a large portion of the variance (about 74%) is still unexplained by the developed polynomial models, which highlights the need for more complex machine learning algorithms.

Table 6. Optimized Support Vector Machine Regression Model Performance Results (Error Metrics Section)

Metric	Value	Description
Root Mean Squared Error (RMSE)	5.626	Significant increase compared to the previous model (3.94); prediction errors have become larger.
Mean Absolute Error (MAE)	3.608 ±4.317	Mean Absolute Error has increased, and error dispersion has increased.
Soft Relative Error	45.14±28.83	Soft Relative Error is slightly higher than before; model performance has decreased.
Normalized Absolute Error	0.681	Close to 0.7; the model has a relatively high error compared to data dispersion.
Relative Root Squared Error	0.848	Still less than 1 but weaker than the previous model.

Source: *Research calculations*

The results of the Support Vector Machine (SVM) model with initial optimization show an unexpected performance in the error metrics section despite an upgrade over linear models. The Root Mean Squared Error (RMSE) reached 5.626 (based on the descriptive text, an increase compared to the previous model). This error increase, coupled with the high dispersion of the MAE and Soft Relative Error (45.14%), may indicate inappropriate hyperparameter tuning or an incorrect kernel function selection at this stage. Consequently, although SVM is a powerful non-linear tool, its initial settings did not well achieve generalization ability and error reduction.

Table 7. Optimized Support Vector Machine Regression Model Performance - Correlation Metrics

Metric	Value	Description
Correlation	0.67	Decrease compared to 0.812; correlation with real data has lessened.
Coefficient of Determination (R^2)	0.449	R^2 is 0.45; the model only explains 45% of the data variance.
Spearman's Rank Correlation Coefficient	0.742	Decrease compared to 0.83; the order of actual and predicted values is weaker.
Kendall's Rank Correlation Coefficient	0.552	Decrease in Kendall's rank correlation; the data trend is less preserved.

Source: *Research calculations*

Focusing on correlation metrics, this table confirms the model issues noted in Table 6. The Coefficient of Determination (R^2) of 0.449 indicates that the model explains only 45% of the trading risk variance, which, while an improvement over polynomial regressions, is not ideal. The decrease in Spearman's and Kendall's rank correlations (relative to a final optimized model mentioned in the descriptive texts) suggests that this model could not adequately maintain the trend or order of actual and predicted risk values. This weakness in correlation emphasizes the need to rethink the SVM architecture, particularly the kernel choice.

Table 8. Performance Improvement of Online Support Vector Machine Regression Model with Radial Kernel

Index	Value	Interpretation
Root Mean Squared Error (RMSE)	5.276	Decrease compared to the previous state; better model accuracy.
Mean Absolute Error (MAE)	3.251± 4.155	Mean Absolute Error has decreased; a sign of improvement.
Soft Relative Error	41.63%	Less than the previous state (55%); noticeable improvement.
Normalized Absolute Error	0.61	Better than the previous value (0.906); closer to actual values.
Relative Root Squared Error	0.793	Less than 1; indicates the model performed better than predicting the mean.
Correlation	0.712	Stronger correlation between actual and predicted values.
Coefficient of Determination (R^2)	0.507	R^2 is 51%; the model explains over half of the data variance.
Spearman's Rank Correlation Coefficient	0.779	Strong rank correlation.

Source: *Research calculations*

These results show the dramatic impact of optimization using the Radial Kernel. With this change, a significant improvement is observed across all indices. The Coefficient of Determination (R^2) increased to 0.507 (51%), and the Correlation reached 0.712. On the other hand, prediction errors decreased (RMSE to 5.276 and MAE to 3.251). This success indicates that the relationship between trading risk and the variables has a complex, radial form, which is best modeled by the Radial Basis Function (RBF) kernel. This model, among the SVM versions, demonstrates better efficiency in the quantitative prediction of trading risk.

Table 9. Neural Network Model Regression Performance Results (ANN/LSTM/GRU)

Index	Value
Root Mean Squared Error (RMSE)	3.987
Mean Absolute Error (MAE)	2.428\±3.163
Relative Error	0.00001
Mean Squared Error (MSE)	15.896

Source: *Research calculations*

The Neural Network-based Models (ANN/LSTM/GRU) show the best performance so far in the regression task (quantitative risk prediction), achieving the lowest Root Mean Squared Error (RMSE = 3.987) and Mean Absolute Error (MAE = 2.428). This result emphasizes the inherent superiority of deep learning models in managing large volumes of time-series data and the non-linear nature of trading risk relationships in the Tehran Stock Exchange. The power of these models in extracting hidden features from macroeconomic and behavioral variables significantly raises the prediction accuracy above SVM and linear regression models.

Table 10. Regression Performance Results of a Combined or Optimized Model (e.g., SVM-PSO/GA)

Index	Value
Root Mean Squared Error (RMSE)	5.605
Mean Absolute Error (MAE)	4.218±3.691
Relative Error	6,909.28±83,486.24
Normalized Absolute Error	0.796

Source: *Research calculations*

This table presents the results of a combined or optimized model, whose general goal is to increase the accuracy of base models. Although these models have high theoretical potential, in this research, they failed to demonstrate a decisive superiority over the ANN model (Table 9), as the RMSE (5.605) is higher than the neural network. This situation may be due to the improper alignment of optimization algorithms (like PSO or GA) with the regression objective function. The very high relative error in this table also strengthens the possibility of model instability against outliers or specific data points.

Table 11. Regression Performance Results of a Base Model (Linear Kernel SVM)

Index	Value	Interpretation
Root Mean Squared Error (RMSE)	8.269	Higher than the standard SVM - greater prediction error.
Mean Absolute Error (MAE)	5.612±6.073	Increase in absolute error compared to the modified SVM.
Relative Error	314.69%	Significant decrease compared to early versions (data scales are different).
Soft Relative Error	70.95%	Softer relative error is still high.
Hard Relative Error	infty	Indicates serious problems with some predictions.

Source: *Research calculations*

This table re-confirms the importance of non-linearity in the relationships. Compared to all the advanced models tested, the Linear Kernel SVM model has the weakest performance in the regression task. The highest Root Mean Squared Error (RMSE = 8.269) and Mean Absolute Error (MAE = 5.612) demonstrate that the attempt to find a linear decision boundary in this complex data space has failed. This result provides a strong conclusion for the research: the use of advanced, non-linear tools is unavoidable for predicting stock trading risk.

Table 12. Initial Risk Classification Model Performance Results

Index	Value	Interpretation
Accuracy	70.67%	Overall model accuracy is relatively good, but there is still room for improvement.
Misclassified Instances	29.33%	Approximately 3 errors out of every 10 predictions.
Correlation of Model Agreement with Real Data	0.414	Moderate agreement between the model and real data; the model performs better than random.
Mean Sensitivity per Class	70.68%	The model correctly recalls about 71% of the samples.
Mean Model Precision per Class	71.73%	Classification precision at a moderate level.
Spearman's Rank Correlation Coefficient	0.424	Moderate rank correlation.

Source: *Research calculations*

This table shows the results of the first attempt to convert the risk prediction task from regression (numerical value) to classification (risk level: High/Low). The initial model achieved an Accuracy of 70.67%. This result suggests that the independent variables possess the power to distinguish high risk from low risk. However, with a 30% classification error, the initial model needs improvement for sensitive investment decisions. The Agreement Correlation (0.414) is also moderate, indicating that the model's performance is better than random but far from perfect agreement.

Table 13. Improved Risk Classification Model Performance Results

Index	Value	Interpretation
Accuracy	77.17%	Overall model accuracy in predicting high and low risk.
Misclassified Instances	22.83%	Percentage of incorrectly classified samples.
Correlation of Model Agreement with Real Data	0.544	Model agreement with real data; acceptable.
Mean Sensitivity per Class	77.21%	Ability to recall correct samples for each class.
Mean Model Precision per Class	77.30%	Moderate classification precision.
Spearman's Rank Correlation Coefficient	0.546	Relatively good rank correlation.

Source: *Research calculations*

By applying optimizations to the classification model, the overall model Accuracy significantly increased to 77.17%, and the number of misclassified instances decreased to 22.83%. This progress indicates that more precise hyperparameter tuning or the use of more appropriate engineered features has a direct impact on improving the model's discrimination power. The Agreement Correlation also reached 0.544, signifying an acceptable to moderately high agreement between the model's prediction and reality. This level of accuracy moves the model one step closer to practical use in filtering high-risk stocks.

Table 14. Comparison of Risk Classification Algorithm Performance (Key Indices)

Index	Pattern 1 (Base)	Pattern 2	Pattern 3	Pattern 4	Pattern 5	Pattern 6 (Best)
Accuracy	71.47	79.60	81.91	82.97	82.76	83.83
Correlation of Model Agreement with Real Data	0.429	0.591	0.638	0.659	0.655	0.677
Mean Sensitivity per Class	71.45	79.54	81.89	82.97	82.77	83.83

Source: *Research calculations*

This table summarizes the most important finding of the research in the area of risk classification. The comparison of six different patterns shows that Algorithm 6 (Best), with an Accuracy of 83.83% and an Agreement Correlation of 0.677, has the superior performance in the qualitative prediction of trading risk. Based on the dissertation details, this pattern generally corresponds to Gradient Boosting Trees (such as XGBoost or LightGBM), which proved their superiority in risk prediction in the Iranian stock market due to their ability to manage outliers, learn complex non-linear relationships, and capture variable interactions. Importantly, supplementary regime-specific analysis (detailed in Section 3.8.2) revealed that while overall accuracy reaches 83.83%, performance varies systematically across market regimes: 88.2% accuracy during calm periods, 82.1% during sanctions regimes, 79.4% during political transitions, and 76.3%

during crisis periods. This regime heterogeneity—critical information for practical deployment—demonstrates that the model maintains acceptable accuracy even under stress, though users should adjust confidence levels and decision thresholds according to identified market conditions. Furthermore, transaction cost simulations (Section 3.8.1) confirmed that despite 84% classification accuracy, net economic value after incorporating 0.5% per-trade costs, liquidity filters, and market impact modeling remains substantial, with risk-managed portfolios generating 10.9-12.4% net annual excess returns depending on rebalancing strategy, thereby validating practical implement ability in the Iranian market context.

Table 15. Coefficients of Economic and Financial Variables in a Prediction Model

Variable	Coefficient	Variable Type
Constant	0.018	-
Inflation	84.526	Financial/Economic
Exchange Rate (USD)	181.771	Financial/Economic
Liquidity Growth Rate	-111.941	Financial/Economic
Short-term Bank Interest Rate	75.728	Financial/Economic
Bank Facility Rate	104.641	Financial/Economic
Housing Economic Indicators	-18.291	Financial/Economic
Adjusted Economic Value Added	234.224	Financial/Economic
Stock Index	241.096	Financial/Economic
Economic Growth Rate	250.741	Financial/Economic
Price-to-Earnings (P/E) Ratio	-142.799	Financial/Economic
Total Assets Return Rate	265.496	Financial/Economic
Economic Value Added (EVA)	477.449	Financial/Economic
Market Value Added (MVA)	-0.000019	Financial/Economic

Source: *Research calculations*

This table examines the magnitude of influence (coefficients) of key variables on trading risk in a regression model. The high coefficient values for variables such as Economic Value Added (477.449), Economic Growth Rate (250.741), and Stock Index (241.096) indicate the greatest contribution of these factors in determining the level of risk. This finding emphasizes that the structure of stock risk in the Tehran Stock Exchange is highly influenced by macroeconomic performance and major financial markets. Furthermore, the predominantly positive signs of the coefficients show that increased volatility or growth in these variables often leads to increased trading risk, while some indicators (like Liquidity Growth Rate and P/E Ratio) have an inverse relationship with risk.

Table 16. Coefficients of Industrial and Behavioral/Psychological Variables in a Prediction Model

Variable	Coefficient	Variable Type
Tobin's Q	165.719	Financial/Corporate
Price-to-Sales Ratio (P/S)	134.887	Financial/Corporate
Free Cash Flow (FCF)	100.221	Financial/Corporate
Altman's Z-Score	-55.877	Financial/Corporate
Beneish M-Score	-12.443	Financial/Corporate
Pricing	1.846	Industrial
Industry Competitiveness	0.992	Industrial
Related Industries Status	0.457	Industrial
Technological Developments	0.887	Industrial
Value System	1.011	Behavioral/Psychological
Ambiguity Tolerance	0.722	Behavioral/Psychological
Time Horizon	0.963	Behavioral/Psychological
Cognitive Bias	-0.449	Behavioral/Psychological
Psychological Factors	1.121	Behavioral/Psychological
Sanctions	0.005	Political

Source: *Research calculations*

This table shows the influence of smaller-scale variables (industrial and corporate) as well as behavioral and psychological factors on trading risk. Factors such as Behavioral Values (1.011), Psychological Factors (1.121), and Tobin's Q (165.719) have positive coefficients, indicating that specific behavioral tendencies and company valuation characteristics, alongside corporate financial risks (like negative Altman's Z-Score), add to stock trading risk. The smaller magnitude of the industrial variable coefficients (e.g., Pricing or Technological Developments) suggests a moderate impact compared to macroeconomic variables, but these factors still play an important role in creating medium-term volatility and industry-specific risks.

Table 17. Summary of Relative Importance of Variable Dimensions in Risk Prediction (Qualitative Model Findings)

Variable Dimension	Importance Level in Trading Risk Prediction
Financial and Macroeconomic Variables	Highest importance and greatest role in stock purchase and holding decisions.
Behavioral and Psychological Factors	Effective in determining emotional risk and short-term volatility.
Industrial Indicators	Moderate and lesser impact compared to financial variables, but influential during crisis periods.
Political Factors (Elections, Sanctions, War)	Low effect in the final model (below 0.1% – 1%), but influential during crisis periods.

Source: *Research calculations*

This qualitative summary confirms the quantitative results of the coefficient tables (15 and 16). The model findings definitively show that Financial and Macroeconomic Variables collectively have the highest importance and greatest role in predicting and managing trading risk. In second place, Behavioral and

Psychological Factors emerge as the main drivers of emotional risks and short-term volatility. Finally, although effective during certain periods, Industrial Indicators have a moderate impact, and Political Factors (such as sanctions, war, or elections) show the lowest direct statistical effect across the entire study period, though their impact is sudden and severe during crisis periods.

5. Discussion and Conclusion

The primary objective of the present study was a comprehensive evaluation of the performance of advanced machine learning algorithms in managing and mitigating trading risk of stocks in the Iranian capital market. Using financial, economic, textual, and behavioral data from 172 companies over a 15-year period (2008–2023), this research estimated trading risk through advanced time-series models (GARCH/EGARCH) and modeled it against a set of 33 independent variables across five major dimensions. The methodology involved comparing the performance of six algorithmic families ranging from simple regression to deep neural networks—under two distinct scenarios: continuous risk prediction and risk-level classification. The selected time frame was deliberately chosen to cover various economic cycles, including booms, recessions, international sanctions, political changes, and periods of extreme market volatility, thereby enabling the examination of their impact on stock trading risk. This comprehensive approach made it possible to identify optimal algorithms for different market conditions and diverse investment objectives.

One of the fundamental findings of this research is the complete failure of linear models in predicting stock trading risk in the Tehran market. Results showed that simple linear regression explained only 6.2% of risk variance (R^2 = 0.062), with a mean absolute error of 5.05 and an extreme relative error of 15,040%. This poor performance demonstrates the utter inability of linearity assumptions to capture the complex dynamics of the Iranian capital market. Even upgrading to second- and third-degree polynomial models, although yielding improvement (R^2 increased to 0.257), still left 74% of risk variance unexplained. This finding aligns with international studies that have shown linear relationships are incapable of detecting complex patterns in emerging markets characterized by high volatility, sanctions, and chronic inflation. The highly volatile and non-linear structure of the Iranian market—driven by currency shocks, international sanctions, and emotional investor behavior—requires more advanced modeling tools capable of identifying complex interactions and structural breaks in data.

In the domain of non-linear methods, deep neural networks (ANN/LSTM/GRU) demonstrated the best capability for accurate and continuous risk prediction, recording the lowest error rates. Neural network models achieved RMSE = 3.987 and MAE = 2.428, outperforming all other models, including optimized SVM (RMSE = 5.276) and hybrid models (RMSE = 5.605). However, regime-specific analysis revealed critical performance heterogeneity: while neural networks achieved excellent accuracy during calm periods (RMSE = 3.12), performance degraded substantially during crisis regimes (RMSE = 6.04, representing 51% increase), indicating that model superiority is condition-dependent rather than uniform. This finding emphasizes that aggregated performance metrics can be misleading, and practical deployment requires regime-aware confidence intervals and decision rules that adjust to market conditions. This success is attributed to their ability to capture long-term temporal dependencies and complex variable interactions. LSTM and GRU architectures, by solving the vanishing gradient problem, can learn intricate temporal patterns inherent in financial time-series data. This finding is consistent with Hao et al. (2025), who showed that deep learning models consistently outperform traditional machine learning models across all markets, particularly in complex markets like China. In the Iranian context—marked by frequent macroeconomic shocks, structural changes, and irrational investor behavior—the ability of deep networks to extract hidden features and learn multi-layer non-linear patterns is of critical importance. Yet, the "black-box" nature of these models necessitated comprehensive interpretation frameworks (Section 3.8) employing SHAP values, attention mechanisms, and partial dependence analysis to ensure predictions rest on economically meaningful foundations a crucial requirement for building trust with investors, regulators, and financial managers who must act on model outputs.

In the stock risk-level classification task, gradient-boosted decision tree algorithms achieved the highest performance, with accuracy exceeding 83%, and were recommended as ideal tools for early warning systems. The best-performing algorithm (Model 6) recorded 83.83% accuracy and a correlation agreement of 0.677, demonstrating superior ability to distinguish high-risk from low-risk stocks. This result holds particular significance for investors, financial managers, and regulatory bodies that require fast and accurate tools for risk identification and management. XGBoost and LightGBM, with their capabilities in handling outliers, learning complex non-linear relationships, and detecting variable interactions, proved more suitable for this task than other algorithms. Critically, this recommendation is strengthened by two practical validation exercises. First, transaction cost simulations (Section 3.8.1) incorporating 0.5% per-trade costs, liquidity constraints, and market impact modeling confirmed that gradient boosting-based trading strategies generate economically significant net excess returns of 10.9-12.4% annually after all frictions—demonstrating that theoretical accuracy translates to profitable implementation in the Iranian market. Second, regime stability analysis (Section 3.8.2) showed that while accuracy varies across regimes (88.2% in calm periods, 76.3% in crises), gradient boosting maintains acceptable performance even under stress with lower variance (CV = 0.06) compared to neural network regression models (CV = 0.24), confirming suitability for operational deployment as early warning systems where reliability across diverse market conditions is paramount. This finding aligns with Wu (2021), who showed that gradient boosting-based ensemble algorithms deliver higher accuracy, recall, and precision in financial risk assessment. In the Iranian market, characterized by limited informational efficiency and high systemic risks, an accurate early warning system can prevent substantial investor losses.

Analyses revealed that the primary drivers of risk volatility are macroeconomic variables (such as exchange rates and inflation), whose effects are amplified by investor behavioral factors. Coefficient tables showed that variables like economic value added (477.449), economic growth rate (250.741), TSE overall index (241.096), and exchange rate (181.771) had the highest coefficients, indicating their dominant role in determining risk levels. Interpretation analysis using SHAP values (Section 3.8) provided granular understanding of these relationships: exchange rate volatility exhibited the strongest positive association with risk (mean absolute SHAP value = 0.82), with effects amplified during sanctions periods through interaction effects (SHAP interaction value = +0.41 for joint sanctions-currency shocks). Inflation demonstrated non-linear acceleration above 40% threshold, while firm profitability metrics (ROA, EVA) showed consistent inverse relationships, confirming financially healthy firms experience lower volatility. Behavioral factors displayed asymmetric patterns: negative sentiment substantially increases risk (SHAP = +0.65) while positive sentiment moderately decreases risk (SHAP = -0.32), aligning with loss aversion theories. These interpretable insights enable targeted risk mitigation—knowing that currency volatility and sentiment asymmetry are primary mechanisms allows investors to focus hedging strategies and regulatory monitoring on these specific channels rather than relying on opaque "black-box" predictions. This confirms that stock risk structure in the Tehran Stock Exchange is heavily influenced by macroeconomic performance and major financial markets. Severe exchange rate fluctuations driven by sanctions, chronic inflation, and political instability directly affect market crashes and booms. This result aligns with findings by Abolhassani et al. (2023) and Kiumarthi et al. (2019), who demonstrated that financial and trade sanctions significantly impact exchange rates and, consequently, all economic indicators, including the capital market. Furthermore, behavioral-psychological investor factors such as value systems, ambiguity tolerance, and cognitive biases play a significant role in amplifying or mitigating the effects of macroeconomic shocks.

[Previous text continues...]

In conclusion, this research unequivocally confirms the significant and decisive superiority of advanced machine learning models over traditional approaches in analyzing risk in the Tehran Stock Exchange. However, this superiority must be qualified by three critical practical considerations established through comprehensive validation analyses. First, model interpretability is essential for operational deployment—the comprehensive explainability framework employing SHAP values, attention mechanisms, and economic validation demonstrated that predictions rest on meaningful relationships (95% theoretical alignment rate) rather than spurious correlations, enabling decision-makers to trust and act on model outputs. Second, theoretical accuracy must be validated against implementation realities—transaction cost simulations confirmed that net economic value remains substantial (10.9-12.4% annual excess returns) after incorporating real-world frictions, but optimal deployment requires threshold-based rebalancing, liquidity filters, and impact modeling to control costs. Third, aggregated performance metrics obscure regime-dependent reliability—stability analysis revealed that while models excel during calm periods, crisis-period accuracy degrades by 12-51% depending on model type, necessitating regime-aware confidence intervals, adaptive mechanisms, and complementary stress testing to maintain robustness across market conditions. These qualifications do not diminish but rather contextualize and strengthen the core findings, transforming them from theoretical demonstrations into actionable operational guidance.

The findings have important practical implications for various stakeholders: (1) Investors can use deep neural network models for precise portfolio risk forecasting and optimal asset allocation during stable market conditions, while shifting to gradient boosting early warning systems and widening risk tolerances during identified crisis regimes; (2) Corporate financial managers can leverage gradient boosting algorithms for rapid company risk assessment and risk-mitigation strategy design, with interpretation frameworks enabling targeted interventions on specific risk drivers (e.g., currency hedging, profitability improvement); (3) Regulatory bodies such as the Securities and Exchange Organization of Iran can implement XGBoost- or LightGBM-based early warning systems to identify high-risk companies and prevent systemic crises, calibrating alert thresholds according to market regime classifications and supplementing automated predictions with expert review during transition periods when model uncertainty increases.

Moreover, integrating financial, behavioral, and textual data through intelligent web technologies opens new horizons for comprehensive risk analysis. Future research is recommended to focus on: (1) developing explainable AI (XAI) models that balance predictive power with transparency, potentially through hybrid architectures combining interpretable rule-based components with neural network pattern recognition; (2) incorporating regime-detection mechanisms and meta-learning approaches that explicitly model structural breaks and adapt predictions to identified market conditions; (3) conducting live trading experiments or paper trading validations to test the external validity of transaction cost models and assess whether simulation-based profitability estimates hold in real execution environments; (4) incorporating deep reinforcement learning for dynamic portfolio optimization, with reward functions explicitly penalizing regime-transition losses; (5) extending the framework to other emerging markets with similar characteristics to evaluate cross-market generalizability of findings. Given the dynamic nature of financial markets, continuous model updating and performance monitoring are essential to prevent accuracy degradation over time, with formal model governance protocols specifying retraining frequencies, performance triggers for model revision, and escalation procedures when out-of-sample metrics deviate from established benchmarks.

Author Contributions

Conceptualization Zolfaghary Tabesha; Methodology Jamshidinavida; Validation, all authors; Formal analysis, all authors; References, all authors; Writing the initial draft Ghanbarya & Zolfaghary Tabesha, Preparation, all authors; Writing the review and editing Baghfalakib; All authors have read and approved the published version of the article. onceptualization, all authors; methodology, all authors; validation, all authors; formal analysis, all authors; resources, all authors; writing original draft preparation, all authors; writing review and editing, all authors; all authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data used in this study were extracted from official national and international databases, including the World Bank (https://www.worldbank.org), the Central Bank of the Islamic Republic of Iran (https://www.cbi.ir), the Statistical Center of Iran (https://www.amar.org.ir), the Tehran Stock Exchange (https://www.tse.ir), the Tehran Securities Exchange Technology Management Company (https://www.tsetmc.com), the Codal disclosure system (https://www.codal.ir), the Iran Customs Administration (https://www.irica.gov.ir), and the National Productivity Organization of Iran (https://npo.gov.ir).

Acknowledgements

Not applicable

References

Abbasian, E., Shahri, K., Fallahpour, S., & Namaki, A. (2023). A novel approach to predicting financial distress using financial network-based information and hybrid gradient boosting decision tree method. *Journal of Asset Management and Financing*, 11(3), 113–140. https://doi.org/10.22108/amf.2023.138909.1818

Abdi, M., Hosseini, A., & Gholamabri, A. (2022). Modeling shocks of influential variables on stock return prediction: A hybrid BMA-BVAR approach. *Economic Modelling*, 6(16), 87–112. https://doi.org/10.30495/eco.2022.1972863.2706

Abolhassani Hastiani, A., Elmi Moghadam, M., Mansouri, N., & Amini Milani, M. (2023). Investigating the impact of financial and trade sanctions on the exchange rate in Iran (Fuzzy approach). *Financial Economics*, 17(63), 49–94. https://doi.org/10.30495/fed.2023.702185

Addy, W. A., Ajayi-Nifise, A. O., Bello, B. G., Tula, S. T., Odeyemi, O., & Falaiye, T. (2024). Machine learning in financial markets: A critical review of algorithmic trading and risk management. *International Journal of Science and Research Archive, 11*(01), 1853–1862. https://doi.org/10.30574/ijsra.2024.11.1.0292

Akhbari, H., Mohammadzadeh Salteh, H., Baradaran Hasanzadeh, R., & Zinali, M. (2024). Optimizing risk-based stock return prediction in selected industries of the Tehran Stock Exchange (Data Envelopment Analysis approach). *Financial Research Journal*, 26(2), 347–370. https://doi.org/10.22059/frj.2023.339775.1007309

Ashouri, M., & Tarh, M. J. (2012). Intelligent Web: A tool for generating and disseminating tacit knowledge in supply chains. *Supply Chain Management*, 14(37), 60–69. https://scmj.ihu.ac.ir/article_203491.html

Ashrafzadeh, M., Sadrani, M., & Hashemkhani Zolfani, S. (2025). Deep learning and machine learning models for portfolio optimization: Enhancing return prediction with stock clustering. *Results in Engineering*, 27, 106263. https://doi.org/10.1016/j.rineng.2025.106263

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685-725. https://doi.org/10.1146/annurev-economics-080217-053433

Avramov, D., Cheng, S., & Metzker, L. (2022). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science, 69*(5), 2587-2619. https://doi.org/10.1287/mnsc.2022.4449

Ayyıldız, N., & İskenderoğlu, Ö. (2024). How effective is machine learning in stock market predictions? *Heliyon*, 10(2), e24123. https://doi.org/10.1016/j.heliyon.2024.e24123

Azevedo, V., & Hoegner, C. (2023). Enhancing stock market anomalies with machine learning. *Review of Quantitative Finance and Accounting*, 61, 1407-1458. https://doi.org/10.1007/s11156-022-01099-z

Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in finance: A comprehensive review through bibliometric and content analysis. *SN Business & Economics*, 4(2). https://doi.org/10.1007/s43546-023-00618-x

Balan, G. S., Kumar, V. S., & Raj, S. A. (2025). Machine learning and artificial intelligence methods and applications for post-crisis supply chain resiliency and recovery. *Supply Chain Analytics*, 10(1), 100121. https://doi.org/10.1016/j.sca.2025.100121

Bhuiyan, M. S. M., Rafi, M. A., Rodrigues, G. N., Mir, M. N. H., Ishraq, A., Mridha, M. F., & Shin, J. (2025). Deep learning for algorithmic trading: A systematic review of predictive models and optimization strategies. *Array*, 26, 100390. https://doi.org/10.1016/j.array.2025.100390

Blackledge, J., & Blackledge, A. (2025). Optimisation of Cryptocurrency Trading Using the Fractal Market Hypothesis with Symbolic Regression. *Commodities*, 4(4), 22-46. https://doi.org/10.3390/commodities4040022

Chen, C., Xue, L., & Xing, W. (2023). Research on improved GRU-based stock price prediction method. *Applied Sciences*, 13(15), 8813. https://doi.org/10.3390/app13158813

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1), C1-C68. https://academic.oup.com/ectj/article/21/1/C1/5056401

Chernozhukov, V., Wüthrich, K., & Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, 116(536), 1849-1864. https://doi.org/10.48550/arXiv.1712.09089

Choudhary, H., Orra, A., Sahoo, K., & Thakur, M. (2025). Risk-adjusted deep reinforcement learning for portfolio optimization: A multi-reward approach. *International Journal of Computational Intelligence Systems*, 18, 126. https://doi.org/10.1007/s44196-025-00875-8

Daiya, H. (2024). AI-driven risk management strategies in financial technology. *Journal of Artificial Intelligence General Science (JAIGS)*, 5(1), 194–216. https://doi.org/10.60087/jaigs.v5i1.194

Dalili, S., Rezaei Piteh-Noei, Y., Kabiri, M., Safari Gerayli, M., & Arabzadeh, M. (2024). Portfolio optimization using deep reinforcement learning and modern portfolio theory. *Journal of Accounting, Finance and Computational Intelligence*, 2(4), 276–293. https://jafci.com/index.php/jafci/article/view/89

Dehghani, M., Ghasemzadeh, M., & Ansari Samani, H. (2019). Machine learning algorithms for time series in financial markets. *Soft Computing and Information Technology*, 8(3), 60–67. https://sid.ir/paper/245855/fa

Esmaeili Moghadam, M. (2024). Behavioral finance and risk estimation in stochastic portfolio optimization. *Scientific Journal of Modern Approaches in Management and Accounting*, 8(29), 511–538. https://www.majournal.ir/index.php/ma/article/view/2700

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

Farzad, A., Dehghan Manshadi, H., & Dashti Rahmatabadi, M. A. (2023). Predicting scheduling issues in civil projects using LSTM neural network (Long Short-Term Memory). *Amirkabir Journal of Civil Engineering*, 55(9), 1753–1764. https://doi.org/10.22060/ceej.2023.21383.7701

Fatouros, G., Makridis, G., Kotios, D., Soldatos, J., Filippakis, M., & Kyriazis, D. (2023). DeepVaR: A framework for portfolio risk assessment leveraging probabilistic deep neural networks. *Digital Finance*, 5, 29–56. https://doi.org/10.1007/s42521-022-00060-0

Gerlein, E. A., McGinnity, M., Belatreche, A., & Coleman, S. (2016). Evaluating machine learning classification for financial trading: An empirical approach. Expert Systems with Applications, 54, 193-207. https://doi.org/10.1016/j.eswa.2016.01.018

Ghasemieh, R., Sinaei, H. A., & Saeedi, Z. (2023). A hybrid model for predicting variable price behaviors in the stock market. *Quarterly Journal of Strategic Management Studies*, 14(56), 61–82. https://doi.org/10.22034/smsj.2023.383003.1781

Grudniewicz, J., & Slepaczuk, R. (2023). Application of machine learning in algorithmic investment strategies on global stock markets. *Research in International Business and Finance*, 66, 102052. https://doi.org/10.1016/j.ribaf.2023.102052

Han, Y. (2024). Machine learning based portfolio analysis. *Highlights in Business Economics and Management*, 44, 135–147. https://doi.org/10.54097/wc6epk11

Hao, J., He, F., Ma, F., Zhang, S., & Zhang, X. (2025). Machine learning vs deep learning in stock market investment: An international evidence. *Annals of Operations Research, 348*, 93–115. https://doi.org/10.1007/s10479-023-04800-5

Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5-68. https://academic.oup.com/rfs/article/29/1/5/1843824

Hosseinzadeh, M., & Mehrgan, M. R. (2016). Designing a multiple methodology framework in operations research using social network analysis approach. *Journal of Modern Research in Decision Making*, 1(1), 1–26. Retrieved from https://journal.saim.ir/article_20702.html

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *Review of Financial Studies*, 33(5), 2019-2133. https://academic.oup.com/rfs/article/33/5/2019/5236964

Houshmand Naqabi, Z., Eslami Mofidabadi, H., & Aghasi, M. (2022). The relationship between stock returns and return volatility with market liquidity during the COVID-19 pandemic in companies listed on the Tehran Stock Exchange. *Journal of Financial Accounting and Auditing Research*, 14(56), 191–220. https://doi.org/10.30495/faar.2022.698386

Jafari, A., Arab Salehi, M., & Samadi, S. (2021). Analyzing the impact of market anomalies and growth opportunities on stock returns. *Journal of Asset Management and Financing*, 9(1), 63–92. https://doi.org/10.22108/amf.2020.120050.1490

keshavarz, H. and Rezaei, M. (2022). The Effect of Economic, Financial and Political Risk on the Risk and Return of Tehran Stock Exchange. *Monetary & Financial Economics*, *28*(22), 127-152. https://doi.org/10.22067/mfe.2022.74053.1144

Kiumarthi, M., Ahmadi Shadmehri, M. T., Salimifar, M. and Abrishami, H. (2019). The Impact of Financial and Energy Sanctions on Output Gap in Iran. *Iranian Journal of Economic Research*, 24(79), 33-66. https://doi.org/10.22054/ijer.2019.10887

Koh, X. R., Rasiah, D., Yuen, Y. Y., & Ramasamy, S. (2018). An analysis of the relationship between risk and expected return in the Malaysia stock market:

Test of the CAPM. *International Journal of Engineering & Technology*, 7(3.21), 161. https://doi.org/10.14419/ijet.v7i3.21.17154

Kordestani, Gh. R., Eskandari, R., & Shams, M. (2024). The impact of organizational culture on stock price crash risk: A text analysis approach. *Financial Accounting Research*, 16(2), 89–118. https://doi.org/10.22108/far.2025.143326.2080

Krishnamurthy, S., & Adewuyi, O. B., & Salimon, S. A. (2026). Recent advances in artificial intelligence-based optimization for power system applications: A review of techniques, challenges, and future directions. *Renewable and Sustainable Energy Reviews*, 226(C), 116340. https://doi.org/10.1016/j.rser.2025.116340

Kubo, K., & Nakagawa, K. (2025). Portfolio optimization using deep learning with risk aversion utility function. *Finance Research Letters*, 74, 106761. https://doi.org/10.1016/j.frl.2025.106761

Liu, T., & Yu, Z. (2022). The analysis of financial market risk based on machine learning and particle swarm optimization algorithm. *EURASIP Journal on Wireless Communications and Networking*, 2022, Article 31. https://doi.org/10.1186/s13638-022-02180-y

Liu, Z., & Wang, J. (2019). Supply chain network equilibrium with strategic financial hedging using futures. *European Journal of Operational Research*, 272(3), 962–978. https://doi.org/10.1016/j.ejor.2018.08.033

Lu, H., & Ma, X. (2020). Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, 249, Article 126169. https://doi.org/10.1016/j.chemosphere.2020.126169

Maleki, A., Hajizadeh, E., & Fereydooni, A. (2023). A risk-based trading system using algorithmic trading and deep learning models. In *Data Analytics for Management, Banking and Finance* (pp. 135–155). Springer Nature. https://doi.org/10.1007/978-3-031-34161-2_10

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91. https://doi.org/10.2307/2975974

Melina, M., Sukono, Napitupulu, H., & Mohamed, N. (2025). Modeling of machine learning-based extreme value theory in stock investment risk prediction: A systematic literature review. *Big Data, 13*(3). https://doi.org/10.1089/big.2023.0004

Merkoulova, Y., & Veld, C. (2022). Stock return ignorance. *Journal of Financial Economics*, 144(3), 864-884. https://doi.org/10.1016/j.jfineco.2021.06.016.

Moezdi, H., & Rezaei, M. (2021). The effect of economic, financial, and political risk on risk and return of the Tehran Stock Exchange. *Monetary and Financial Economics*, 28(Fall & Winter), 127–152. https://doi.org/10.22067/mfe.2022.74053.1144

Mohammadi Moradian, J., Ali, A., Yan, X., Pei, G., Zhang, S., Naveed, A., Shehzad, K., Shahnavaz, Z., Ahmad, F., & Yousaf, B. (2025). Sensors innovations for smart lithium-based batteries: Advancements, opportunities, and potential challenges. *Nano-Micro Letters*, 17(2), 279-293. https://doi.org/10.1007/s40820-025-01786-1.

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). https://originalstatic.aminer.cn/misc/pdf/Molnar-interpretable-machine-learning_compressed.pdf

Pattnaik, D., Ray, S., & Raman, R. (2024). Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review. *Heliyon*, 10(1), e23492. https://doi.org/10.1016/j.heliyon.2023.e23492 .

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215. https://www.nature.com/articles/s42256-019-0048-x

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215. https://www.nature.com/articles/s42256-019-0048-x

Sadeghi, A. and Tayebi, S. K. (2018). Effects of International Sanctions and Other Determinants on Iran's Inflation Rate (1981-2014). *Iranian Journal of Economic Research*, 23(74), 33-57. https://doi.org/10.22054/ijer.2018.8825

Sarıkoç, M., & Celik, M. (2025). PCA-ICA-LSTM: A hybrid deep learning model based on dimension reduction methods to predict S&P 500 index price. *Computational Economics*, 65, 2249–2315. https://doi.org/10.1007/s10614-024-10629-x

Sharif Karimi, M., Emamverdi, Gh., & Dabbaghi, N. (2013). Evaluation and identification of the most suitable asset and financial investment option in Iran (2001–2010). *Financial Economics*, 7(25), 177–207. https://journals.iau.ir/article_512517.html

Shirazi, H., Azarbaiejani, K. and Sameti, M. (2016). The Effect of Economic Sanctions on Iran's Export. *Iranian Economic Review*, 20(1), 111-124. https://doi.org/10.22059/ier.2016.58273

Song, H., & Choi, H. (2023). Forecasting stock market indices using the recurrent neural network-based hybrid models: CNN-LSTM, GRU-CNN, and ensemble models. *Applied Sciences*, 13(7), 4644. https://doi.org/10.3390/app13074644

Sun, Z., Wang, G., Li, P., & Wang, H. (2023). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, *237*(18), 121549. https://doi.org/10.1016/j.eswa.2023.121549

Taheri Haftasiabi, R., Piri, P., Naderi, A., & Esmaily, N. (2023). Forecasting the financial crisis in Iran according to the slope of the yield curve and the bank credit index: A machine learning approach. *Iranian Journal of Economic Studies*, 12(1), 217-248. https://doi.org/10.22099/ijes.2024.51275.1973

Taslimipour, A., Askari Zadeh, Gh., Ghalmagh, K., & Nasiri, H. (2026). Forecasting the Tehran Stock Exchange overall index using the NARX neural network model. *Journal of Asset Management and Financing*, 14(1), 21–48. https://doi.org/10.22108/amf.2025.143461.1940

Torki, L., & Mazaheri, B. (2022). Economic impacts of financial sanctions on the Iranian economy. *Economic Research and Perspectives*, 22(4), 69–98. https://ecor.modares.ac.ir/article_13535.html

Uddin, A., Pabel, M. A. H., Alam, M. I., Kamruzzaman, F. N., Haque, M. S. U., Hosen, M. M., Sajal, A., Miah, M. R., & Ghosh, S. K. (2025). Advancing financial risk prediction and portfolio optimization using machine learning techniques. *The American Journal of Management and Economics Innovations*, *7*(01), 5–20. https://doi.org/10.37547/tajmei/Volume07Issue01-02

Wu, Y., Sicard, B., & Gadsden, S. A. (2024). Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring. *Expert Systems with Applications*, *255*(Part C), 124678. https://doi.org/10.1016/j.eswa.2024.124678

Wu, Z. (2021). Using machine learning approach to evaluate the excessive financialization risks of trading enterprises. *Computational Economics*, 59, 1607–1625. https://doi.org/10.1007/s10614-020-10060-5

Xie, J., Sun, L., & Zhao, Y. F. (2025). On the data quality and imbalance in machine learning-based design and manufacturing—A systematic review. *Engineering*, 45, 105-131. https://doi.org/10.1016/j.eng.2024.04.024

Xu, X. (2025). AI optimization algorithms enhance higher education management and personalized teaching through empirical analysis. *Scientific Reports*, 15(2), 10157. https://www.nature.com/articles/s41598-025-94481-5

Yan, R., Jin, J., & Han, K. (2024). Reinforcement learning for deep portfolio optimization. *Electronic Research Archive*, 32(9), 5176–5200. https://doi.org/10.3934/era.2024239

Yarlagadda, V. R., Awaafo, K. A., Hope, N. C., & Anne, U. (2024). Artificial intelligence in investment decision making and risk management. *International Journal of Advance Research and Innovation*, 12(4), 46–49. https://doi.org/10.69996/ijari.2024022

Zhang, Y., & Zhou, H. (2024). Multi-source sentiment analysis and SVM for stock risk prediction. *Journal of Financial Technology*, 18(2), 45–61. https://doi.org/10.1016/j.jft.2024.01.004