



Applying metaheuristics and SVMs to forecast stock price crashes in Tehran Stock Exchange

Reza Raei ¹, Saeed Shirkavand ¹, Ali Jamali Neyshabour ^{1*}

¹ Department of Financial Management and Insurance, University of Tehran, Tehran, Iran.

Received: Oct 2023-17/ Revised: Dec 2024-10/ Accepted: Dec 2024-30

Abstract

Sudden and severe stock price crashes pose a significant challenge to stock markets. The substantial losses incurred from such events underscore the need for more effective forecasting tools. This study aims to enhance the predictive power of models for stock price crashes in Tehran Stock Exchange and commenced with a comprehensive literature review to identify key financial factors influencing stock price volatility. Given the high dimensionality of the dataset and the extended time period, metaheuristic algorithms were employed for feature selection. 10 algorithms, namely Ant Colony Optimization, Hill Climbing, Las Vegas, Whale Optimization, Simulated Annealing, Genetic Algorithm, Tabu Search, Particle Swarm Optimization (PSO), Honey Bee (HBA) and Firefly were utilized to reduce dimensionality and enhance model performance. Subsequently, Support Vector Machines were implemented to develop predictive models. The models were trained and evaluated using historical data from Tehran Stock Exchange spanning from 2001 to 2020. The findings of this research demonstrate that combining metaheuristic algorithms for model reduction and optimization, along with advanced machine learning techniques, yields results that can significantly improve investment decision-making.

Keywords: Stock Price Crash, Metaheuristic Algorithms, Support Vector Machine (SVM), Stock exchange

Paper Type: Original Research

1. Introduction

In today's turbulent world, forecasting the future is not only a competitive advantage but also a necessity for the survival of organizations and individuals. Rapid and unpredictable changes in the business environment have made data-driven decision-making more critical than ever. Governments, business managers, and investors are constantly seeking ways to predict future trends. From forecasting market fluctuations and stock prices to estimating consumer behavior and technological changes, all contribute to strategic decision-making. Accurate future forecasting is a powerful tool for planning, mitigating risk, and seizing new opportunities. Among the most significant challenges facing decision-makers is predicting unexpected events and severe market fluctuations. Events such as sudden stock price crashes can have a significant impact on the performance of organizations and shareholders' wealth. The sharp and sudden decline in stock prices, especially after the 2008 financial crisis, has raised significant concerns among investors. This event can lead to investment losses and market instability. Therefore, identifying the causes of this phenomenon and finding ways to predict and mitigate its effects is of paramount importance. A sudden stock market crash is a large and unusual change, typically defined as a negative return on a stock. This complex and recurring phenomenon in stock markets, unlike normal market fluctuations, often occurs without major economic shocks (Luo & Zhang, 2020) and results in significant losses of investor wealth. Other notable characteristics of stock market crashes include negative skewness in return distributions and their contagious nature (McLean & Pontiff, 2016). One of the primary variables associated with stock price crashes is firm-specific financial metrics. Studies have shown that characteristics such as return on assets, firm size, market-to-book value ratio, shareholders' equity, and the quick ratio exhibit a significant negative relationship with crash risk. High levels of debt are correlated with increased crash risk, as firms with significant leverage may struggle during downturns as evidenced by the work of Kangarlou et al (2024). Negative Tail Skewness (NCSKEW): This metric reflects the asymmetry in the return distribution, with negative values indicating a higher likelihood of crashes (Gillis, 2024). In addition, Various machine learning algorithms, including support vector machines (SVM) and random forests, have been employed to analyze complex relationships among these variables. These models can capture non-linear patterns that traditional methods might miss according to Kaya et al (2024). Numerous

*Corresponding Author: alijamali.ny@gmail.com

researchers have sought to explain the causes of sudden stock price crashes by examining market mechanisms and investor behavior. Theories such as price bubbles, and heterogeneous beliefs are among the most prominent (Chen, Hong, & Stein, 2001). However, despite the importance of the subject, fewer studies have explored the possibility of predicting this event. This research aims to contribute to the existing literature by developing a model to forecast stock price crash using a powerful machine learning model: support vector machines. By leveraging metaheuristic algorithms as optimizers and considering factors identified in previous studies as influencing sudden stock price crashes, this research seeks to fill a gap in the existing literature. The combination of artificial SVMs and metaheuristic algorithms for predicting stock price crashes offers significant advantages that can enhance the accuracy and efficiency of predictive models. Metaheuristic algorithms assist in identifying and selecting the most effective variables from complex data, which leads to dimensionality reduction and improved accuracy of neural networks. This combination enables the modeling of nonlinear relationships between variables, which is particularly effective in unstable market conditions. Additionally, the use of metaheuristic algorithms can reduce computational time and expedite the search process for the optimal combination of variables. Furthermore, SVMs, with their ability to learn from complex patterns in data, can provide more accurate predictions. Ultimately, this approach is highly flexible and capable of adapting to environmental changes, a feature that is especially important in today's volatile financial markets.

2. Literature Review

Sudden and severe fluctuations in stock prices have attracted significant attention from academics and professionals in recent years. Given the importance of stock returns to investors, stock market crashes, which lead to sharp declines in returns, have received more research attention compared to rallies. In the following sections, a brief overview of some of these studies will be provided. Abdi and Hosseini (2013) examined the impact of liquidity management on reducing the risk of stock price crashes. Their findings provided evidence of a high probability that liquidity management can reduce stock price crashes. Dang et al. (2016) demonstrated that firms with higher short-term debt ratios are less likely to experience severe price declines; in other words, debt maturity has a negative impact on crash risk. Dianati et al. (2012) demonstrated that working capital management significantly reduces the probability of stock price crashes. Ahmadpour et al. (2014) examined the impact of firm characteristics on the crash of listed companies on Tehran Stock Exchange. The variables of return on assets, firm size, market-to-book value ratio, shareholders' equity, and the quick ratio had a significant negative relationship with crash risk in listed companies on Tehran Stock Exchange. However, the variables of return on equity and leverage did not have a significant relationship with crash risk. Haghanifar (2015) investigated the impact of fundamental stock indices on the future crash of stocks in Tehran stock exchange. The study found a significant relationship between annual real stock returns, systematic risk, and stock price crash. However, there was no significant relationship between the price-to-earnings ratio, the previous year's average real stock returns, and future stock price crash.

3. Theoretical Foundations

This study uses stock price crash as the dependent variable. We define stock price crash as a weekly return that falls below a specified number of standard deviations from Firm-Specific Weekly Return of the year. This approach is consistent with previous studies (Hutten, 2009; Kim & Zhang, 2010). Firm-Specific Weekly Return of a company, W in equation 1, is calculated as the natural logarithm of one plus the simple return in week, plus an error term ε that comes from equation 2.

$$W_{j,\tau} = Ln(1 + \varepsilon_{j,\tau}) \quad \text{Equation 1}$$

$$r_{j,\tau} = a_j + B_{1j}r_{m,\tau-2} + B_{2j}r_{m,\tau-1} + B_{3j}r_{m,\tau} + B_{4j}r_{m,\tau+1} + B_{5j}r_{m,\tau+2} + \varepsilon_{j\tau} \quad \text{Equation 2}$$

In the above formula, $r_{j,t}$ represents the return of stock j in week τ , and r_m represents the market return (based on a market index). The returns of several weeks before and after are used in these calculations to demonstrate how the return of a stock changes over a specific period and to identify its return trend. Therefore, if w , the stock return in at least one week of the year, is less than a multiple of the standard deviation of the Firm-Specific Weekly Return for the entire year, then it can be said that the company has experienced a price crash. Due to various factors such as sanctions, exchange rate fluctuations, and continuous and unstable policy-making, the Iran's capital market often experiences severe fluctuations. Additionally, due to features such as price fluctuations limits and minimum trading volume, it requires considering a specific coefficient. The coefficient should be selected in such a way that only significant price declines that are noticeably different from the average are considered crashes. In this

research, through a trial-and-error approach based on analysis, the aforementioned coefficient is defined as 2. Selecting smaller coefficients would result in identifying insignificant fluctuations as crashes, while selecting greater coefficients would cause even very serious fluctuations to be overlooked as stock price crashes. In the first phase, through a literature review and utilizing library information, variables influencing stock price crashes were extracted, and relevant data was collected. In the second phase, the variables under consideration were optimally selected. Metaheuristic algorithms were employed for variable selection. The output of this phase will be models optimized based on training data and ready for validation after predicting stock price crashes (third phase) using a support vector machine.

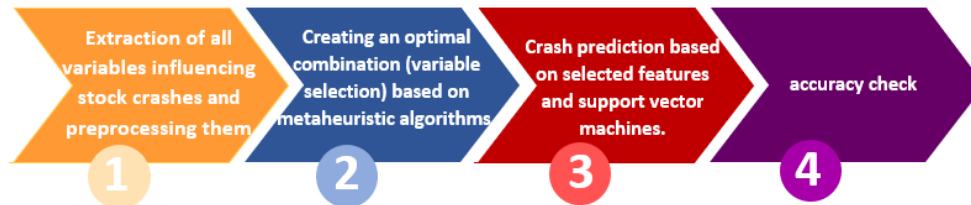


Figure 1. The research process and development of a stock price crash prediction model

The process begins with data collection, where relevant financial indicators and historical stock prices are gathered. Next, data preprocessing is performed to clean and normalize the dataset, ensuring accuracy in analysis. Following this, feature selection is conducted using metaheuristic algorithms to identify the most impactful variables influencing stock price movements. The selected features are then used to train various SVMs, optimizing their parameters through iterative testing. Finally, model validation is carried out using back testing techniques to assess predictive performance against unseen data. In the following, the theoretical foundations of variables affecting stock price crashes, metaheuristic algorithms and support vector machines will be discussed in order.

3.1 Financial variables affecting stock price crashes

A wide range of financial performance-related factors, including those arising from operational activities, liquidity status, profitability, and the company's debt status, significantly influence stock price crashes. A portion of the variables frequently used in the research literature is presented in the table on the following page. The large number of variables and the associated computational complexity justify the pursuit of optimality. Therefore, the application of metaheuristic algorithms for variable selection from the set of factors and the development of more agile models are considered. The subsequent section will provide a literature review and theoretical foundations of the aforementioned algorithms.

Table 1. Variables

#	Variables	#	Variables
1	Gross Profit Margin	21	FCFE to net income ratio
2	Operating Profit Margin	22	FCFE to net income ratio
3	Profit Margin Before Tax	23	Capital expenditure to revenue
4	Net profit margin	24	Capital expenditure to net income
5	Return on assets	25	Operating cash flow to revenue
6	Return on equity	26	Operating cash flow to total debt
7	EBIT on assets	27	Weekly market return (annual average)
8	Assets turnover ratio	28	Weekly stock return (annual average)
9	Fixed assets turnover ratio	29	Firm-Specific Weekly Return
10	Debt-to-equity ratio	30	Negative skewness of stock returns
11	Debt ratio	31	Market value-to-book value ratio
12	Short-term Debt ratio	32	Logarithm of sales
13	Long-term Debt ratio	33	Price-to-earnings ratio
14	Retained earnings on assets	34	Tobin's Q ratio
15	Interest coverage ratio	35	Price-to-sales ratio
16	Current ratio	36	Market value
17	Quick ratio	37	Book value
18	Cash ratio	38	Operating cash flow
19	Free cash flow to income ratio	39	Free cash flow to the firm
20	Free cash flow to net income	40	Free cash flow to equity

3.2 Feature selection based on metaheuristic algorithms

As real-world problems become increasingly complex, traditional optimization methods have reached their limits. These methods often require specific conditions such as continuity and differentiability, and frequently become trapped in local optima. Furthermore, the computational time required to solve problems increases dramatically

with the problem size. To overcome these limitations, metaheuristic algorithms have emerged as powerful tools. Inspired by natural phenomena such as evolution, ant colony behavior, or bird flocking, these algorithms provide a flexible and efficient approach to solving complex optimization problems. Most of these algorithms initiate the search process by generating a population of random solutions within the feasible region of the decision variables. This initial set of solutions is often referred to as a population, colony, swarm, or similar terms depending on the specific algorithm. Individual solutions within the population are commonly termed chromosomes, ants, particles, or analogous terms. Subsequently, a new set of solutions is generated using operators, typically relying on random number generation. In metaheuristic algorithms, the search for an optimal solution is conducted iteratively. In each iteration, the algorithm generates new solutions using various operators and then, through an intelligent selection mechanism, chooses the best solutions to form the next generation. This selection process, allows the algorithm to gradually converge towards a better solution. By focusing on high-quality solutions and discarding inferior ones, the algorithm progressively approaches a near-optimal solution. This iterative process continues until a predefined stopping criterion is met. Figure 2 provides a schematic representation of this process.

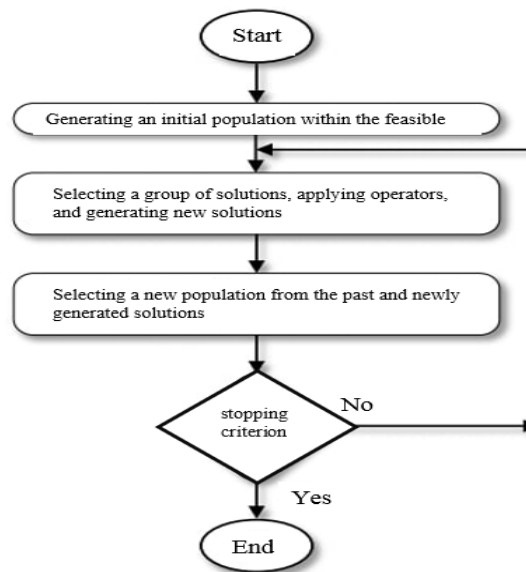


Figure 2. The overall process of metaheuristic algorithms in finding a solution

In this research, the following metaheuristic algorithms will be employed: Ant Colony Optimization, Hill Climbing, Las Vegas, Whale Optimization, Simulated Annealing, Genetic Algorithm, Tabu Search, Particle Swarm Optimization (PSO), Honey Bee (HBA) and Firefly. The advantages of each algorithm are detailed in the table below:

Table 2. Metaheuristic Algorithms Used in the Research and Their Advantages

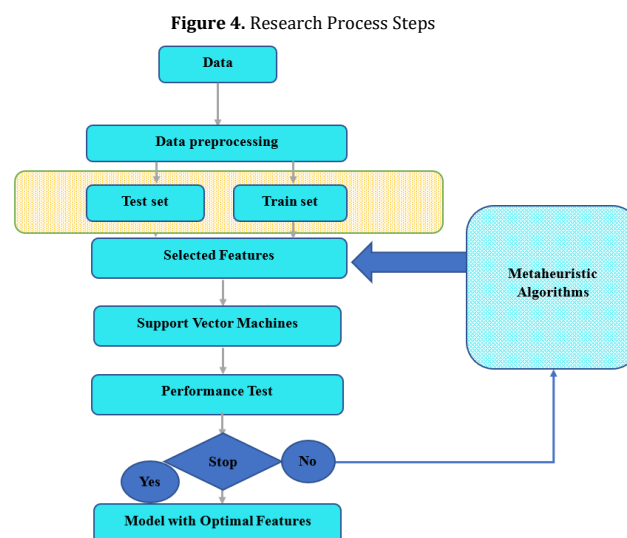
Algorithm	Innovators	Year of presentation	Benefits
Ant Colony Optimization	Marco Dorigo	1992	Guaranteed convergence to an optimal solution, suitable for dynamic applications requiring rapid adaptation to environmental changes.
Hill Climbing	--	--	Providing a satisfactory solution in a reasonable amount of time is a hallmark of this algorithm, making it suitable for problems with numerous variables, a vast search space, and numerous constraints.
Las Vegas	László Babai	1979	It employs a probabilistic approach that facilitates escape from local optima and convergence to better global solutions
Whale Optimization	Mirjalili	2016	The novel algorithm, which employs mechanisms for target encirclement and rotation, exhibits a high capability of finding solutions.
Simulated Annealing	Kirkpatrick	1983	It possesses simple and intuitive theoretical concepts, yet is highly effective in solving optimization problems in large search spaces.
Genetic Algorithm	--	--	Numerous advantages including parallel search, capability to solve high-dimensional and multi-objective complex problems, and a high probability of discovering high-quality solutions.
Tabu Search	Glover	1986	The algorithm builds a list of forbidden moves or points to avoid revisiting them in subsequent searches, thus facilitating escape from local optima.
Particle Swarm Optimization	Eberhart and Kennedy	1995	The algorithm possesses memory, enabling all particles to retain knowledge of good solutions. Each particle benefits from its own historical information.
Honey Bee	Karaboga	2005	Its ability to be parallelized and its suitable convergence rate in most problems allow it to approach optimal solutions in a relatively short time.
Firefly	Yang	2008	High flexibility in solving both continuous and discrete problems, as well as the ability to escape local optima and reach more desirable global solutions.

3.3 Prediction testing using Support Vector Machines

Support Vector Machine (SVM) is a powerful machine learning algorithm employed for data classification. SVM seeks to find the optimal hyperplane in the data space that best separates two classes. This hyperplane maximizes the margin between the closest data points of each class. These crucial points are termed as support vectors. SVM undergoes a two-phase training process: in the first phase, the optimal hyperplane is constructed using training data. In the second phase, the model's accuracy in classifying unknown data is evaluated using new data. Support Vector Machine (SVM) is a robust machine learning algorithm employed in a wide array of applications including pattern recognition, time series forecasting, market analysis, facial recognition, and medical diagnosis. By combining the strengths of traditional statistical methods and machine learning capabilities, SVM can identify intricate patterns within data and often outperforms other methods.

4. Methodology

The figure below presents a schematic representation of the research process:



At the outset of this research, an extensive literature review was conducted to identify key variables influencing the phenomenon of stock price crashes. These variables, selected based on previous studies and relevant theories, encompass economic, financial, and behavioral factors. Following the identification of variables, the necessary data was collected from reliable sources such as stock exchange databases, company financial reports, and other relevant information repositories. From the extensive set of identified variables, the most influential and relevant variables associated with stock price crashes were selected. To achieve this, advanced metaheuristic algorithms capable of optimally searching a vast space of variable combinations were employed. These algorithms include Ant Colony Optimization, Hill Climbing, Las Vegas, Whale Optimization, Simulated Annealing, Genetic Algorithm, Tabu Search, Particle Swarm Optimization (PSO), Honey Bee (HBA) and Firefly. By evaluating the performance of various models based on training data, the optimal combination of variables was identified. The output of this phase is a set of initial models trained on historical data and capable of predicting stock price crashes. In this research, variables selected by at least 8 out of the 10 metaheuristic algorithms were included in the Support Vector Machine models for predicting crash risk. To train the Support Vector Machine, 70% of the data is randomly selected, the remaining 30% of the variables are used to test the prediction accuracy.

5. Data Analysis and Findings

For this study, financial data from 150 companies on Tehran Stock Exchange spanning from 2001 to 2021 was collected. Subsequently, 38 variables, as detailed in Table 1, were computed based on this dataset.

Table 3. Characteristics of the data used in the research

Description	Characteristics
150	Number of firms
2001-2021	Time period
40	Independent variables
One variable	Dependent Variables

Therefore, the accuracy of the proposed model is 80%. Given that five variables, namely "return on equity", "debt ratio", "FCFE to revenue ratio", "negative skewness of stock returns", and "logarithm of sales", were selected based on the consensus of metaheuristic algorithms, a support vector machine model was run with randomly selected five variables to test the results. This test was repeated 300 times. The average results of these tests are presented in the table below (decimal places are rounded).

Table 6: Average results of stock crash prediction using Support Vector Machine with random variables, repeated 300 times Therefore, the accuracy of the prediction results using randomly selected variables is 71%.

No Crash	Crash	Total	Predictions
203	316	519	Correct predictions
133	79	212	Incorrect predictions
336	395	731	Total

6. Conclusion

This Study has shown that combining metaheuristic algorithms for dimensionality reduction and optimal structure selection, coupled with advanced machine learning techniques, can yield highly promising results in predicting stock price crashes. This combined approach enables the development of models capable of identifying complex and latent patterns in market data, and accurately forecasting the occurrence of sudden and severe price declines. The findings of this research clearly demonstrate that the developed models can provide relatively accurate predictions of stock price crash risk using available information. This enables investors, financial analysts, and economic decision-makers to optimize their investment strategies and mitigate portfolio risk by being aware of the likelihood of such events. Overall, the significance of this research lies in its provision of a comprehensive and effective framework for predicting stock price crashes, thereby contributing significantly to improved decision-making in financial markets. The findings of this study can serve as a tool for identifying investment opportunities, managing systemic risk, and mitigating losses arising from severe market fluctuations.

References

- Abdi, Mohamad, Hosseini, & Maqsood (2013). The Effect of Liquidity Management on Reducing the Stock Price Crash Risk. *International Research Journal of Applied and Basic Sciences*, (6) 333-338.
- Ahmadpour, A., Zare Behnamiri, M. J., & Heidari Rastami, K. (2014). The impact of firm characteristics on the risk of stock price decline (companies listed on the Tehran Stock Exchange). *The Securities Market*, 7(28), 29-45. (In Persian)
- Chen, Joseph, Hong, Harrison and Stein, Jeremy, (2001), Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices, *Journal of Financial Economics*, 61, issue 3, p. 345-381.
- Dang, V.A., Lee, E., Liu, Y., & Zeng, C.H. (2016). Does debt maturity affect stock price crash risk? www.ssrn.com.
- Dianati, Z., Moradzadeh, M., & Mahmoodi, S. (2012). The impact of institutional investors on reducing the risk of stock price decline. *Investment Knowledge*, 1(2), 1-18. (In Persian)
- Eberhart, R.C., & Kennedy, J. (1995). A new optimizer using particle swarm theory. In *Proceedings of the 6th international symposium on micro machine and human science*, (Nagoya, Japan) (pp. 39-43).
- Gillis, G. (2024). Predictability of firm-level stock crash risk. *The American Finance Association Journal*, 78(1), 1-25.
- Glover F., *Future paths for integer programming and links to artificial intelligence*, *Computers and Operations Research*, 1986, 13, pp. 533-549.
- Haghanifar, A., Shamakhi, H., & Hajji, G. A. (2015). The impact of investment criteria on the future stock price crash risk in the Iranian capital market. *International Conference on New Research in Management, Economics, and Accounting*. (In Persian)
- Hutton, A.P., Marcus, A.J., Tehranian, H., (2009). Opaque financial reports, R2, and crash risk. *Journal of Financial Economics*, 94, 67-86.
- Karaboga, D. (2005) *An Idea Based on Honey Bee Swarm for Numerical Optimization*. Technical Report-TR06, Erciyes University, Engineering Faculty, Department of Computer Engineering.
- Kaya, D., Reichmann, D., & Reichmann, M. (2023). Out-of-sample predictability of firm-specific stock price crashes: A machine learning approach. *Journal of Business Finance & Accounting*.
- Kim, J. B., Zhang, L., 2010. Does accounting conservatism reduce stock price crash risk? Available at: (www.ssrn.com).
- Kirkpatrick, S., Gelatt, Jr., C.D. and Vecchi, M.P. (1983) *Optimization by Simulated Annealing*. *Science*, 220, 671- 680.

- László Babai (1979). "Monte-Carlo algorithms in graph isomorphism testing". Université de Montréal Technical Report, DMS 79-10, p. 42.
- Luo, Y., & Zhang, C. (2020). Economic Policy Uncertainty and Stock Price Crash Risk. *Research in International Business and Finance*, 51, 101112.
- McLean, R.D., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *The Journal of Finance* 71 (1): 5-32.
- M. Dorigo, *Optimization, Learning and Natural Algorithms*, PhD thesis, Politecnico di Milano, Italy, 1992.
- Mirjalili, S., & Lewis, A. (2016). The Whale Optimization Algorithm. *Advances in Engineering Software*, 95, 51-67.
- Najafi Kangarlou, N., Jebel Ameli, F. , Mehrara, M. , & Fatehi, M. B. (2024). Probability of stock price crash and the expected return of stock under sanctions. *Iranian Journal of Finance*, 8(1), 131-145.
- YANG Xin-she. *Nature-inspired metaheuristic algorithms*[M]. Luniver Press, 2008: 83-96.