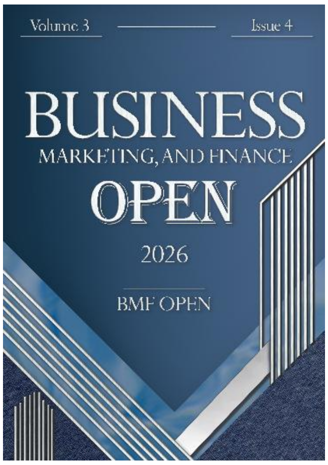
The Role of Fundamental Factors in Optimal Stock Portfolio Formation Using the Fama–French Five-Factor Model with an Emphasis on Data Mining: Evidence from Pharmaceutical Companies Listed on the Tehran Stock Exchange



Reza Khosravi[1], Jamshid Peikfalak[2,*] and Hassan Fattahi Nafchi[3]

[1] Department of Accounting, Shk.C., Islamic Azad University, Shahrekord, Iran;
[2] Department of Accounting, WT.C., Islamic Azad University, Tehran, Iran;
[3] Department of Accounting, University of Isfahan, Isfahan, Iran;

* Correspondence e-mail: peikfalak@iau.ac.ir

Abstract: The purpose of the present study was to determine the role of fundamental factors in optimal stock portfolio formation using the Fama–French five-factor model with an emphasis on data mining. The statistical population consisted of companies listed on the Tehran Stock Exchange during the 2012–2023 period. To achieve the research objectives and construct an optimal stock portfolio, dimensionality-reduction approaches, Data Envelopment Analysis (DEA), Support Vector Machine (SVM), and clustering algorithms were employed. Balance-sheet financial ratios, income-statement financial ratios, cash-flow-statement financial ratios, composite financial ratios, and risk and return measures based on the Fama–French five-factor model were used as model inputs for constructing four portfolios. The findings indicated that the support vector machine method and the fourth approach, which comprised the composite model, demonstrated superior performance in stock portfolio optimization. These findings can assist investors and stock analysts in identifying appropriate financial ratios for optimal portfolio formation while achieving an appropriate balance between return and risk.

Keywords: Fundamental Factors, Optimal Stock Portfolio, Fama–French Five-Factor Model

Citation: Khosravi, R., Peikfalak, J., & Fattahi Nafchi, H. (2026). The Role of Fundamental Factors in Optimal Stock Portfolio Formation Using the Fama–French Five-Factor Model with an Emphasis on Data Mining: Evidence from Pharmaceutical Companies Listed on the Tehran Stock Exchange. *Business, Marketing, and Finance Open*, 3(4), 1-20.

Received: 03 March 2026
Revised: 07 July 2026
Accepted: 15 July 2026
Published: 30 July 2026

1. Introduction

Portfolio selection is one of the central problems in financial management because investors must allocate limited capital among multiple securities while simultaneously considering expected return, risk, liquidity, firm-specific characteristics, market conditions, and their own investment constraints. The fundamental objective of portfolio optimization is not merely to identify stocks with high individual returns, but to construct a combination of assets capable of generating an acceptable level of return relative to the risk undertaken. In this context, portfolio formation represents a multidimensional decision-making problem in which the information content of financial statements, market-based indicators, risk-adjusted performance measures, and statistical or computational selection techniques can jointly influence the quality of investment decisions. Traditional financial management has therefore emphasized the systematic evaluation of return and risk in capital allocation decisions, while contemporary approaches increasingly integrate accounting information and computational methods into the portfolio-selection process [1]. Fundamental analysis is particularly important in this respect because it allows investors to evaluate the economic and financial position of firms through information derived from financial statements and other company-specific indicators before making investment decisions [2].

The selection of appropriate stocks becomes more complex when investors operate in emerging and relatively volatile capital markets such as the Tehran Stock Exchange. In such markets, stock prices may be influenced not only by corporate fundamentals but also by liquidity conditions, investor behavior, market sentiment, macroeconomic uncertainty, and changing expectations. Consequently, portfolio selection based solely on historical price behavior may provide an incomplete basis for investment decision-making. Earlier evidence from the Tehran Stock Exchange has shown that investors consider several interrelated dimensions when selecting stocks, suggesting that stock-selection decisions require a structured evaluation of multiple criteria rather than dependence on a single financial indicator [3]. Similarly, multicriteria approaches applied to leading companies listed on the Tehran Stock Exchange have demonstrated the usefulness of systematically ranking firms prior to portfolio construction [4]. These considerations support the view that firm fundamentals should be incorporated alongside market-based measures when designing an optimal portfolio.

Financial ratios are among the most widely used instruments for converting accounting information into comparable measures of corporate performance, liquidity, profitability, solvency, efficiency, and financial structure. Balance-sheet ratios can provide information regarding leverage, liquidity, and the relationship between assets and liabilities, whereas income-statement ratios reveal dimensions of profitability, operating performance, and the firm's capacity to generate earnings. Cash-flow ratios provide an additional perspective by focusing on the actual generation and use of cash, which may differ substantially from accounting earnings. Research examining traditional liquidity ratios and cash-flow-based measures has shown that information derived from the cash flow statement can complement conventional financial indicators in evaluating the ability of firms to sustain their activities [5]. Therefore, combining information from different financial statements may improve discrimination between financially stronger and weaker firms and may provide a more complete foundation for portfolio formation than reliance on one class of ratios alone.

Despite the usefulness of financial ratios, conventional ratio analysis has important methodological limitations. Ratios are often considered individually or within relatively narrow groups, while firm performance is inherently multidimensional. Moreover, different ratios may provide conflicting signals, making it difficult to reach an integrated assessment of corporate efficiency. Data Envelopment Analysis (DEA) provides an important solution to this problem because it permits simultaneous consideration of multiple inputs and outputs and evaluates the relative efficiency of comparable decision-making units. DEA has developed into a major nonparametric technique for efficiency measurement across a wide variety of organizational and financial applications [6]. In financial analysis, DEA can supplement traditional financial-ratio analysis by integrating several indicators into a unified efficiency framework, thereby identifying firms that achieve comparatively superior outcomes relative to the resources or financial conditions represented by their inputs [7].

The relevance of DEA to investment decision-making has been demonstrated in portfolio-selection research. An important advantage of DEA is that it can classify stocks or firms according to relative efficiency without requiring a predetermined functional relationship between financial inputs and performance outputs. This characteristic is particularly useful when investors simultaneously consider profitability, liquidity, risk, and other financial dimensions. Research has shown that DEA can be used as a portfolio-selection mechanism in which stock liquidity serves as a major criterion for identifying efficient investment opportunities [8]. Likewise, empirical research on equity portfolios has demonstrated that DEA-based stock screening can contribute to improved portfolio performance when efficiency information is incorporated into the investment-selection process [9]. These findings indicate that DEA can operate as an effective preprocessing and screening mechanism before the final portfolio is formed.

At the same time, portfolio analysis using DEA requires careful model specification. The designation of inputs and outputs, the selected efficiency model, and the treatment of risk and return can materially affect the set of stocks identified as efficient. Consequently, researchers have emphasized that portfolio analysis with DEA should be preceded by careful consideration of model structure and the economic meaning of the variables included in the efficiency framework [10]. More recent applications have further expanded DEA-based portfolio evaluation by combining efficiency measurement with fuzzy decision-making and ranking procedures. For example, DEA-based fuzzy portfolio models integrated with Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) have been employed to rank efficient portfolios under different risk indicators, demonstrating the potential value of combining efficiency analysis with multicriteria ranking tools [11]. In the Iranian capital-market context, DEA has also been used to form portfolios on the basis of efficiency derived from risk and distribution-based returns, further supporting its applicability to stock selection under multidimensional performance criteria [12].

However, identifying efficient stocks does not entirely resolve the portfolio-selection problem. Financial markets generate high-dimensional and often nonlinear datasets, and the relationships among accounting variables, market indicators, and future stock performance may be too complex for conventional analytical procedures. For this reason, machine-learning methods have increasingly entered portfolio-management research. These methods can detect structures and interactions in data that may not be captured through conventional linear techniques and can be used for stock screening, classification, return prediction, ranking, and allocation. Recent research on machine learning in portfolio decisions has emphasized the growing role of predictive algorithms in asset selection and portfolio management, particularly when investors face large and heterogeneous information sets [13]. Similarly, machine-learning-based portfolio analysis has been proposed as a mechanism for supporting more intelligent investment decisions by extracting useful patterns from financial and market data [14].

Support Vector Machine (SVM) is especially relevant when the portfolio-selection problem is formulated as a classification task. SVM algorithms attempt to identify a separating hyperplane that maximizes the margin between observations belonging to different classes. In a portfolio-selection framework, stocks may first be assigned to favorable and unfavorable, efficient and inefficient, or otherwise defined categories, after which an SVM can learn the classification structure and predict the status of new observations. The distance of observations from the separating hyperplane can also provide a basis for ranking. Thus, SVM provides a bridge between classification and portfolio screening. Research in the Iranian market has already highlighted the usefulness of machine learning in the portfolio-optimization process and has proposed new methods for integrating learning algorithms into the selection and weighting of securities [15]. This development reflects the broader movement from static portfolio construction toward data-driven investment frameworks.

The integration of machine learning with conventional portfolio theory has also received considerable attention internationally. One line of research has combined predictive stock-selection models with Markowitz mean-variance optimization, showing that machine-learning-based forecasting can be used to improve the stock-selection stage prior to optimization [16]. More recent work has moved beyond simple prediction toward sequential and adaptive decision-making. Reinforcement-learning approaches, for example, have been developed to optimize portfolio selection through mechanisms that combine stock ranking and matching, enabling portfolio decisions to adapt dynamically to the information environment [17]. These developments demonstrate that the portfolio problem increasingly involves not only mathematical optimization but also intelligent information processing and data-driven ranking.

The growing complexity of investment opportunities has also encouraged the development of hybrid portfolio models. Portfolio optimization may involve financial, behavioral, ethical, or other decision criteria, and hybrid models can be designed to incorporate these dimensions simultaneously [18]. Related optimization frameworks have shown that portfolio selection can be formulated in knowledge-based systems in which financial considerations are integrated with additional investor objectives or constraints [18]. The importance of such hybrid approaches lies in their capacity to reflect the real-world structure of investment decisions, where portfolio quality is determined by several interdependent criteria rather than by a single objective function.

Another critical theoretical component of portfolio analysis concerns the modeling of expected returns. The Fama–French factor-model framework represents an important extension of single-factor asset-pricing models by recognizing that expected stock returns are associated with multiple systematic dimensions. The five-factor version incorporates the market factor together with size, value, profitability, and investment factors, thereby providing a richer structure for explaining cross-sectional return behavior. Within portfolio optimization, the relevance of the Fama–French five-factor framework lies in its capacity to integrate market-based risk and return information with firm characteristics that are fundamentally connected to corporate size, valuation, profitability, and investment behavior. This framework is therefore conceptually compatible with a research design that seeks to connect fundamental financial indicators to optimal stock selection.

The importance of more sophisticated risk assessment has been reinforced by recent portfolio-optimization research. Portfolio construction cannot be evaluated exclusively through absolute returns, because higher return may simply reflect greater exposure to risk. Accordingly, risk-adjusted measures such as the Sharpe ratio provide a more meaningful basis for comparing alternative portfolios. Recent evidence comparing active and passive strategies through Sharpe-ratio-based optimization has demonstrated the usefulness of assessing investment performance relative to the volatility borne by investors rather than considering return in isolation [19]. Similarly, portfolio-optimization research using adjusted measures of stock risk has emphasized that the quality of a portfolio depends substantially on how risk is measured and incorporated into the performance-evaluation framework [20]. These considerations are especially important when alternative classification and prediction models produce portfolios with different combinations of expected return and volatility.

Portfolio risk is also affected by interconnectedness across financial assets and markets. Evidence concerning return and volatility spillovers among technology stocks, decentralized finance instruments, and cryptocurrencies indicates that asset relationships can materially influence diversification opportunities and portfolio optimization [21]. Research combining selected stocks and cryptocurrencies has likewise shown that optimization results depend on the characteristics and interactions of the assets entering the portfolio [22]. Although the present study focuses on pharmaceutical stocks rather than cross-asset portfolios, such evidence reinforces the broader principle that portfolio formation should be sensitive to the risk-return structure of the selected investment universe and that optimization techniques must distinguish between raw performance and risk-adjusted performance.

Investor behavior provides another layer of complexity. Financial decisions are not always based exclusively on objective financial information, and psychological factors may alter security selection, risk tolerance, and portfolio composition. Recent evidence from the Tehran Stock Exchange has demonstrated that behavioral biases can influence portfolio optimization, indicating that actual investment behavior may diverge from purely rational allocation models [23]. Investor sentiment has likewise been shown to affect portfolio optimization in both the Tehran Stock Exchange and cryptocurrency markets, highlighting the relevance of market psychology to allocation decisions [24]. Furthermore, behavioral portfolio models incorporating investor preferences and memory have been proposed to represent more realistically how investors process prior experiences and make subsequent portfolio choices [25]. Although behavioral factors are not the direct focus of the present model, these studies illustrate why decision-support systems based on structured financial data may be valuable in reducing dependence on subjective judgments.

Taken together, the literature suggests that optimal portfolio construction requires an integrated framework. Fundamental analysis provides information concerning the underlying financial condition of firms; financial ratios translate this information into comparable indicators; DEA can identify efficient securities across multiple dimensions; machine-learning techniques can classify and rank stocks using complex data structures; the Fama–French five-factor framework can strengthen the modeling of risk and expected return; and the Sharpe ratio allows alternative portfolios to be compared on a risk-adjusted basis. Nevertheless, much of the existing literature has examined these elements separately. DEA-based studies have often focused on efficiency measurement, machine-learning studies have frequently concentrated on prediction, factor-model research has primarily addressed return explanation, and financial-ratio analysis has generally been used as a conventional screening device.

This fragmentation creates an important research opportunity, particularly within the pharmaceutical sector of the Tehran Stock Exchange. Pharmaceutical firms operate in an industry in which capital intensity, financing structure, profitability, working-capital management, cash-flow generation, and investment policies may differ considerably across companies. Consequently, examining only one group of financial ratios may fail to capture the full informational content relevant to stock selection. A comparative assessment of balance-sheet ratios, income-statement ratios, cash-flow-statement ratios, and composite ratios can therefore reveal whether different forms of fundamental information have different capacities to support portfolio formation. In addition, combining these categories with DEA and SVM makes it possible to move beyond descriptive fundamental analysis toward a predictive classification and ranking framework.

The methodological contribution of such an approach also lies in comparing portfolio performance across alternative information structures. If the addition of broader categories of financial information systematically improves classification accuracy, portfolio return, or the Sharpe ratio, this would indicate that the informational content of financial statements is cumulative rather than redundant. Conversely, if a narrower category performs equally well, then investors may be able to achieve efficient stock screening with a more parsimonious information set. Therefore, comparing the four financial-ratio approaches provides evidence not only about optimal portfolio construction but also about the relative investment relevance of different financial-statement components.

Accordingly, the aim of the present study was to determine the role of fundamental factors in forming an optimal stock portfolio among pharmaceutical companies listed on the Tehran Stock Exchange by integrating financial-ratio information, Data Envelopment Analysis, Support Vector Machine classification, and the Fama–French five-factor model within a data-mining framework.

2. Methods and Materials

In terms of purpose, the present study is classified as applied research. The required data were collected from audited financial statements and the Rahavard Novin financial database/software. To estimate the results, the data were initially prepared in Microsoft Excel 2017 and subsequently analyzed using MATLAB 2022. In this study, a training dataset was constructed using financial ratios extracted from companies' financial statements. The values obtained for return, risk, liquidity, and expected return for each stock were then used in conjunction with Data Envelopment Analysis (DEA) to classify the stocks. Efficient Decision-Making Units (DMUs) were classified into the favorable category, whereas inefficient units were classified into the unfavorable category. Subsequently, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) technique was employed to eliminate redundant or anomalous observations. Among the clusters within each category, the cluster containing the largest number of observations was selected. In this manner, redundant data were removed, after which the model was solved using a Support Vector Machine (SVM). Following this stage, unclassified data were entered into the SVM to predict their corresponding categories. After determining the characteristics of the hyperplanes, the Euclidean distance of each observation from the hyperplanes was calculated. Finally, the classified stocks were ranked according to these distances. In the final stage, investment portfolios were constructed based on the proposed approach, and the portfolio demonstrating the best performance was selected as the portfolio for the subsequent year. The required data were collected from audited financial statements and the Rahavard Novin financial database/software. For estimation purposes, the data were prepared in Microsoft Excel 2017 and analyzed using MATLAB 2022. The statistical population consisted of all pharmaceutical companies listed on the Tehran Stock Exchange during the 2012–2023 period, with the objective of constructing a stock portfolio for the forecast period of 2023–2024. Given that the research model included five input variables and four output variables, the minimum required number of stocks was 27. After determining the number of stocks required to solve the model, the pharmaceutical products industry was selected from among the various industries represented on the Tehran Stock Exchange. It should also be noted that the input variables of the Banker–Charnes–Cooper (BCC) model must be nonnegative. Therefore, stocks with negative input values, stocks for which complete information was unavailable for 2022–2023, and stocks that had been delisted from the stock exchange were excluded from the sample. Ultimately, 27 stocks remained. Accordingly, the final dataset comprised information related to 27 pharmaceutical companies listed on the Tehran Stock Exchange.

The financial ratios employed in the present study were classified into four categories based on the financial statements: balance-sheet financial ratios, income-statement financial ratios, cash-flow-statement financial ratios, and composite financial ratios. Subsequently, using the values obtained for return, risk, liquidity, and expected return for each stock, together with Data Envelopment Analysis (DEA), the stocks were classified into efficient and inefficient groups. After removing outliers using the DBSCAN clustering method, the unclassified observations were entered into the Support Vector Machine to predict their respective categories. Given that financial ratios were divided into four categories and optimal stock portfolio selection was evaluated for each category using the return and risk dimensions of the Fama–French five-factor model, four optimal portfolios were obtained:

1. Optimal stock portfolio based on the analysis of balance-sheet financial ratios and the Fama–French five-factor model.
2. Optimal stock portfolio based on the analysis of income-statement financial ratios and the Fama–French five-factor model.
3. Optimal stock portfolio based on the analysis of cash-flow-statement financial ratios and the Fama–French five-factor model.
4. Optimal stock portfolio based on the analysis of composite financial ratios and the Fama–French five-factor model.

Using the Sharpe ratio, these portfolios were compared with one another, and the informational content of each category of financial statements was evaluated based on the results of these comparisons. To provide a clearer understanding of the proposed model, a schematic representation of the overall model and its implementation stages on the Tehran Stock Exchange is presented in this section.

The criteria employed in the model were summarized into four categories: balance-sheet financial ratios, income-statement financial ratios, cash-flow-statement financial ratios, and composite financial ratios. Subsequently, the values obtained for return, risk, liquidity, and expected return for each stock were used together with Data Envelopment Analysis to classify stocks into efficient and inefficient groups.

Table 1. Classification of Financial Ratios

Financial Ratios	Classification
Current ratio; quick ratio; total debt-to-equity ratio; total assets-to-total liabilities ratio; debt-to-assets ratio; fixed assets-to-current assets ratio; total liabilities-to-total assets ratio	Balance-sheet financial ratios
Gross profit-to-sales ratio; net profit-to-sales ratio; interest coverage ratio; profit margin; earnings per share; degree of operating leverage	Income-statement financial ratios
Cash return on assets; cash return on equity; cash coverage of financing costs; cash quality of earnings; quality of sales; cash conversion cycle; cash conversion efficiency index; net cash balance index	Cash-flow-statement financial ratios
Sales-to-total assets ratio; accounts receivable turnover ratio; return on equity; return on investment; inventory turnover; operating cash flow-to-total debt ratio; operating cash flow-to-long-term debt ratio; operating cash flow-to-current liabilities ratio	Composite financial ratios

To calculate the efficiency of companies listed on the Tehran Stock Exchange, a framework was developed in which the variables associated with each ratio were designated as inputs or outputs according to their underlying characteristics. Among these ratios, those that were more commonly used and for which no missing observations existed in the database were selected. Based on the nature of each ratio, the selected variables were entered into an input-oriented BCC model as either inputs or outputs in relation to each of the three specified criteria.

Table 2. Stock Efficiency Models

Model	Inputs	Outputs
Model 1: Return-Based Efficiency	Beta; debt-to-net-worth ratio (positive coefficient); current ratio; total asset turnover (negative coefficient)	Average monthly return over one year
Model 2: Risk-Based Efficiency	Monthly variance of changes up to the year under consideration, representing annual risk (positive coefficient)	Net earnings per share (EPS); net profit-to-sales ratio; return on assets; earnings before tax-to-current liabilities ratio (all with positive coefficients)
Model 3: Liquidity-Based Efficiency	Beta; debt-to-net-worth ratio (positive coefficient); current ratio; total asset turnover (negative coefficient)	Liquidity ranking score

Following implementation of the DEA algorithm, companies whose efficiency score equaled 1 for each of the aforementioned models were designated as **efficient** and assigned a score of 1. Companies with efficiency scores below 1 were designated as **inefficient** and assigned a class label of −1.

Under the classification approach based on Data Envelopment Analysis, each of the three criteria measured above—return, risk, and liquidity—was entered into the input-oriented BCC model as either an input or an output, together with other selected financial ratios identified through the literature review, in order to determine the class label of each stock (record). Based on the model outputs, units or stocks identified as efficient for each criterion were separately assigned to the favorable class, whereas inefficient stocks were assigned to the unfavorable class. Thus, upon completion of the model, each record received three labels of either 1 or −1, representing the efficient and inefficient categories, respectively. The input-oriented BCC model was implemented through programming in MATLAB. It should be noted that this procedure was conducted separately for stocks in each year. The BCC model was selected instead of the Charnes–Cooper–Rhodes (CCR) model because the BCC specification generally identifies a larger number of units as efficient. Since the number of inefficient observations was typically considerably larger, the input-oriented BCC model was employed to generate a relatively larger and more balanced set of efficient stocks.

In this section, predictions were performed separately for each efficiency model. Specifically, using the Fama–French five-factor modeling approach, variables associated with each of the four approaches—balance-sheet ratios, income-statement ratios, cash-flow-statement ratios, and composite financial ratios—were entered into the model using data from the in-sample period. The efficiency associated with each model was considered the output or dependent variable. After the models had been estimated using the specified variables, efficiency values for each model were predicted separately for the out-of-sample period and for each of the three efficiency models. Subsequently, the efficiency vector was determined under the four specified approaches, and the efficiency status of each observation was assigned to one of the eight possible states reported in Table 4.1. In effect, prediction was performed using the Fama–French five-factor approach, and the predictive performance of the model was subsequently tested against out-of-sample data.

For the classification of observations previously categorized according to this approach, the same proposed procedure was applied. A hierarchical linear Support Vector Machine was again employed, with Euclidean distance from the separating hyperplane serving as the ranking criterion. The differences between this section and the preceding section concerned two aspects. First, the underlying content of the classification differed; second, a two-level tree was employed instead of a three-level tree. All other procedures related to implementation of the Support Vector Machine remained unchanged. More specifically, the stocks were initially divided into two categories based on expected return. Subsequently, based on liquidity, each of the categories generated in the previous stage was divided into two further categories. Following stock ranking, the investment portfolio was constructed. After this stage, unclassified observations were entered into the Support Vector Machine to predict their categories. Given that the financial ratios were classified into four categories and that optimal stock portfolio selection based on each category was evaluated using the return and risk dimensions of the Fama–French five-factor model, four optimal portfolios were obtained:

1. Optimal stock portfolio based on the analysis of balance-sheet financial ratios and the Fama–French five-factor model.
2. Optimal stock portfolio based on the analysis of income-statement financial ratios and the Fama–French five-factor model.
3. Optimal stock portfolio based on the analysis of cash-flow-statement financial ratios and the Fama–French five-factor model.
4. Optimal stock portfolio based on the analysis of composite financial ratios and the Fama–French five-factor model.

Using the Sharpe ratio, the four portfolios were compared with one another, and the informational content of each set of financial statements was examined based on the comparative results. To provide a clearer overview of the proposed model, a schematic representation of the complete model and the stages of its implementation on the Tehran Stock Exchange is presented in this section.

3. Findings and Results

The first stage of the analysis evaluated the classification performance of the Support Vector Machine (SVM) under the four financial-ratio approaches. The classification results for both the training and testing samples are presented in Table 3.

Table 3. Accuracy of the Support Vector Machine Model Across the Four Research Approaches

Approach	Partition	Correct, n	Correct, %	Incorrect, n	Incorrect, %	Total, n
First approach (S)	Testing	425	68	185	32	620
	Training	412	72	160	28	572
Second approach (S1)	Testing	483	77	137	23	620
	Training	450	79	119	21	569
Third approach (S2)	Testing	510	82	110	18	620
	Training	470	83	95	17	565
Fourth approach (S3)	Testing	517	83	103	17	620
	Training	470	84	89	16	559

As shown in Table 3, classification accuracy progressively improved from the first to the fourth approach. In the testing sample, the percentage of correctly classified observations increased from 68% under the first approach to 77% under the second approach, 82% under the third approach, and 83% under the fourth approach. A similar pattern was observed for the training data, for which classification accuracy increased from 72% to 79%, 83%, and 84%, respectively. Thus, the fourth approach produced the highest classification accuracy, followed by the third, second, and first approaches.

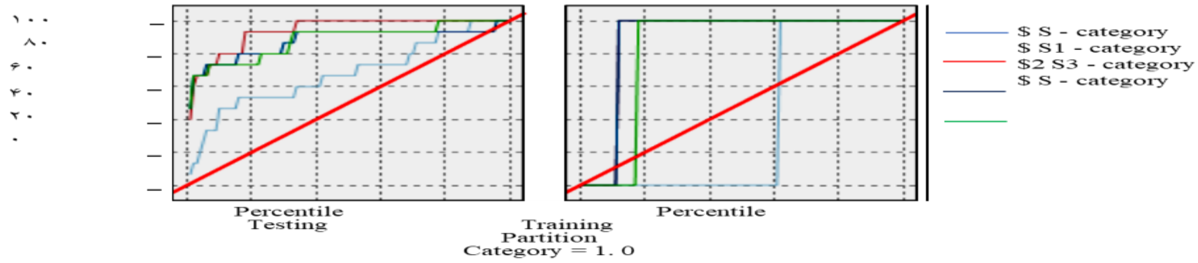


Figure 1. Gain Index Adjustment Rate for the Accuracy of Support Vector Machine Models Across the Four Research Approaches

The Gain index presented in Figure 1 confirmed the pattern observed in Table 3. The predictive accuracy curve associated with the fourth approach converged more rapidly toward the 100% explanatory threshold than did the curves associated with the preceding approaches. This finding provided further evidence of the superior predictive performance of the fourth approach.

Following classification, optimal portfolio weights were determined for each of the four SVM-based prediction models. Portfolio return was calculated as the sum of the products of each stock's annual return and its corresponding portfolio weight. The stock-level composition of the four optimal portfolios is summarized in Table 4.

Table 4. Stock-Level Composition of the Optimal Portfolios Based on the Support Vector Machine

Approach	Stock Symbol	Coefficient	Stock Weight	Annual Stock Return	Weighted Return
First	Delqma	.334	.03948	.74298	.029333
First	Dpars	.522	.061702	.23374	.014422
First	Shtehran	.440	.052009	.33539	.017443

First	Dkosar	.318	.037589	.08423	.003166
First	Dtemad	.541	.063948	1.21057	.077414
First	Dfara	.492	.058156	.08995	.00523
First	Ddam	.529	.06253	.59347	.037109
First	Dshimi	.333	.039362	.38304	.015077
First	Vpakhsh	.462	.05461	.89151	.048685
First	Dabor	.334	.03948	1.34037	.052918
First	Dasveh	.412	.04870	.22156	.01079
First	Drazak	.310	.036643	.43591	.015973
First	Diran	.504	.059574	1.96690	.11718
First	Pakhsh	.378	.044681	.32579	.014557
First	Sobhan	.342	.040426	.29593	.011963
First	Dalbor	.472	.055792	.71847	.040085
First	Darou	.391	.046217	.48957	.022627
First	Dkimi	.460	.054374	.69973	.038047
First	Djaber	.446	.052719	.42574	.022444
First	Dsina	.440	.052009	.23603	.012276
Second	Delqma	.471	.031621	.74298	.023494
Second	Dpars	.988	.066331	.23374	.015504
Second	Shtehran	.627	.042095	.33539	.014118
Second	Dkosar	.998	.067002	1.21057	.081111
Second	Dtemad	.857	.057536	.08995	.00518
Second	Dfara	.911	.061161	.59347	.036297
Second	Ddam	.715	.048003	1.53003	.073446
Second	Dshimi	.489	.03283	.38304	.012575
Second	Vpakhsh	.956	.064183	.89151	.057219
Second	Dabor	.449	.030144	1.34037	.040405
Second	Dasveh	.636	.042699	.22156	.00946
Second	Drazak	.434	.029137	.43591	.012701
Second	Diran	.960	.064451	.19669	.012677
Second	Pakhsh	.662	.044444	.32579	.01448
Second	Sobhan	.524	.03518	.29593	.010411
Second	Dalbor	.980	.065794	.71847	.047271
Second	Darou	.677	.045451	.48957	.022252
Second	Dkimi	.875	.058745	.69973	.041105
Second	Djaber	.818	.054918	.42574	.023381
Second	Dsina	.868	.058275	.23603	.013755
Third	Delqma	.514	.042876	.23347	.010022
Third	Dpars	.402	.033534	.33523	.011247
Third	Shtehran	.928	.077411	1.21081	.093711
Third	Dkosar	.760	.063397	.00899	.00057
Third	Dtemad	.412	.034368	.59326	.020296
Third	Dfara	.680	.056723	1.53020	.086788
Third	Ddam	.443	.036954	.38335	.014155
Third	Dshimi	.928	.077411	.89181	.069012
Third	Vpakhsh	.548	.045712	1.34081	.061272
Third	Dabor	.574	.047881	.22148	.010609
Third	Dasveh	.649	.054137	.43593	.023599
Third	Drazak	.627	.052302	.32570	.01704
Third	Diran	.477	.03979	.29579	.011775
Third	Pakhsh	.955	.079663	.71856	.057235

Third	Sobhan	.637	.053136	.48973	.026014
Third	Dalbor	.836	.069736	.69963	.048797
Third	Darou	.747	.062313	.42571	.026529
Third	Dkimi	.871	.072656	.23615	.017149
Fourth	Delqma	.514	.042876	.23374	.010022
Fourth	Dpars	.402	.033534	.33539	.011247
Fourth	Shtehran	.928	.077411	1.21057	.093711
Fourth	Dkosar	.760	.063397	−.08995	−.0057
Fourth	Dtemad	.412	.034368	.59347	.020296
Fourth	Dfara	.680	.056723	1.53003	.086788
Fourth	Ddam	.443	.036954	.38304	.014155
Fourth	Dshimi	.928	.077411	.89151	.069012
Fourth	Vpakhsh	.548	.045712	1.34037	.061272
Fourth	Dabor	.574	.047881	.22156	.010609
Fourth	Dasveh	.649	.054137	.43591	.023599
Fourth	Drazak	.627	.052302	.32579	.01704
Fourth	Diran	.477	.03979	.29593	.011775
Fourth	Pakhsh	.955	.079663	.71847	.057235
Fourth	Sobhan	.637	.053136	.48957	.026014
Fourth	Dalbor	.836	.069736	.69973	.048797
Fourth	Darou	.747	.062313	.42574	.026529
Fourth	Dkimi	.871	.072656	.23603	.017149

The first and second SVM-based portfolios each contained 20 selected stocks, whereas the third and fourth portfolios each contained 18 stocks. The weighted return of an individual stock reflected the contribution of that security to overall portfolio performance after accounting for its portfolio weight. The total portfolio return was consequently obtained by summing the weighted returns of all included stocks.

To evaluate the overall results of the four SVM approaches, portfolio return, portfolio risk, and the Sharpe ratio were compared with the corresponding market values. These results are presented in Table 5.

Table 5. Performance of the Support Vector Machine Portfolios Compared with the Market

Performance Measure	First Approach	Second Approach	Third Approach	Fourth Approach	Market
Return	.65012	.68523	.71203	.81256	.52369
Risk	.01225	.03147	.03452	.04173	.01154
Sharpe ratio	17.2543	19.1442	20.1746	26.5022	9.27699

Table 5 indicates that all four SVM-based portfolios generated returns exceeding the market return of .52369. Portfolio return increased systematically across the four approaches, from .65012 under the first approach to .81256 under the fourth approach. Although portfolio risk also increased across the approaches, risk-adjusted performance improved. In particular, the fourth approach generated the highest Sharpe ratio (26.5022), indicating the strongest risk-adjusted portfolio performance.

The results therefore showed that incorporating progressively more comprehensive financial information improved portfolio performance. In particular, the fourth approach, based on composite financial ratios, generated the highest return and Sharpe ratio. In other words, simultaneous consideration of the broader set of financial criteria resulted in the most favorable SVM-based optimal portfolio.

In the second stage of the analysis, classification accuracy was assessed using the Fama–French five-factor model under the same four financial-ratio approaches. Table 6 presents the training and testing results.

Table 6. Accuracy of the Fama–French Five-Factor Classification Model Across the Four Research Approaches

Approach	Partition	Correct, n	Correct, %	Incorrect, n	Incorrect, %	Total, n
First approach (N)	Testing	485	76	150	24	635
	Training	468	76	140	24	608
Second approach (N1)	Testing	485	77	150	23	635
	Training	464	77	144	23	608
Third approach (N2)	Testing	500	78	135	22	635
	Training	490	80	118	20	608
Fourth approach (N3)	Testing	510	82	125	20	635
	Training	498	82	110	18	608

The classification results were consistent with theoretical expectations. The fourth approach exhibited the highest classification accuracy among the four approaches. It was followed by the third approach, while the second approach generally performed better than the first approach. Specifically, testing accuracy increased from approximately 76%–77% under the first two approaches to 78% under the third approach and 82% under the fourth approach. Training accuracy similarly reached 82% under the fourth approach.

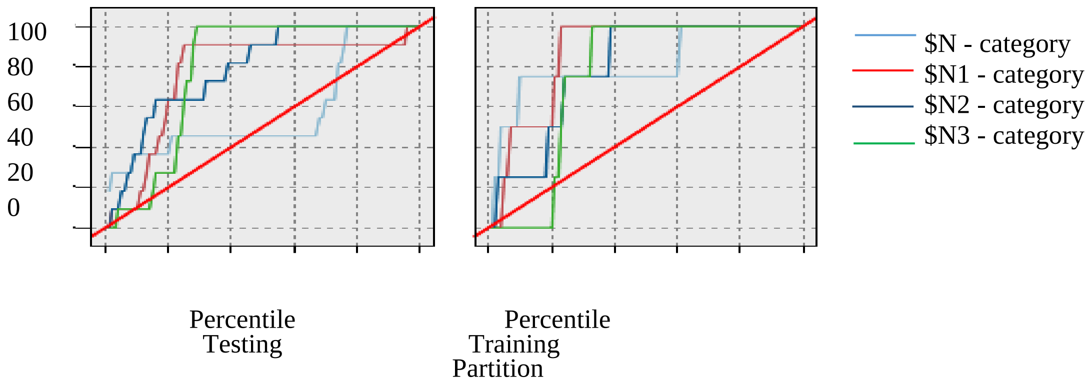


Figure 2. Gain Index Adjustment Rate for the Accuracy of the Fama–French Five-Factor Model Across the Four Research Approaches

The Gain index depicted in Figure 2 was consistent with the classification results and indicated stronger predictive performance for the more comprehensive approaches.

For the Fama–French portfolios, the weight assigned to each efficient stock was calculated by dividing its individual importance index by the sum of the importance indices of all efficient stocks. After portfolio formation, total portfolio return was calculated as the sum of the weighted returns of the constituent stocks and subsequently compared with market performance.

The first Fama–French portfolio consisted of 22 stocks. Its total coefficient was 13.19, and its reported total weighted return was .65. The selected stocks and their respective coefficients, portfolio weights, annual returns, and weighted returns were as follows: Delqma (.63, .05, .23, .01), Dpars (.60, .05, .34, .02), Shtehran (.45, .03, .64, .02), Dkosar (.64, .05, .66, .03), Dtemad (.55, .04, 1.21, .05), Dfara (.58, .04, −.09, .00), Ddam (.57, .04, .59, .03), Dshimi (.73, .06, 1.53, .08), Vpakhsh (.42, .03, .38, .01), Dabor (.75, .06, .89, .05), Dasveh (.60, .05, 1.34, .06), Drazak (.58, .04, 1.41, .06), Diran (.56, .04, .22, .01), Pakhsh (.61, .05, .44, .02), Sobhan (.47, .04, 1.66, .06), Dalbor (.69, .05, .33, .02), Darou (.77, .06, .72, .04), Dkimi (.46, .03, −.26, −.01), Djaber (.45, .03, .49, .02), Dsina (.73, .06, .70, .04), Dlor (.65, .05, .43, .02), and Damin (.72, .05, .24, .01).

Under the second approach, the Fama–French five-factor model did not predict any stock as simultaneously efficient according to the two specified criteria. Consequently, no investable portfolio could be proposed under the second approach. Nevertheless, when additional input or predictor variables were incorporated into the subsequent models, stocks predicted to be efficient reappeared in the candidate set.

The third approach produced a portfolio consisting of 23 stocks, with a total coefficient of 14.41 and a reported total weighted return of .54. The constituent stocks and their respective coefficients, portfolio weights, annual returns, and weighted returns were as follows: Delqma (.43, .03, .74, .02), Dpars (.88, .06, .23, .01), Shtehran (.59, .04, .34, .01), Dkosar (.51, .04, −.09, .00), Dtemad (.68, .05, 1.53, .07), Dfara (.42, .03, .38, .01), Ddam (.80, .06, .89, .05), Dshimi (.48, .03, 1.34, .04), Vpakhsh (.86, .06, .15, .01), Dabor (.53, .04, 1.40, .05), Dasveh (.74, .05, .44, .02), Drazak (.39, .03, 1.66, .05), Diran (.53, .04, .04, .00), Pakhsh (.53, .04, .48, .02), Sobhan (.47, .03, .02, .01), Dalbor (.71, .05, .33, .02), Darou (.58, .04, .30, .01), Dkimi (.88, .06, .72, .04), Djaber (.41, .03, −.26, −.01), Dsina (.56, .04, .49, .02), Dlor (.82, .06, .70, .04), Damin (.80, .06, .43, .02), and Drouz (.81, .06, .24, .01).

The fourth approach resulted in a more concentrated portfolio consisting of 17 stocks, with a total coefficient of 12.38 and a reported total weighted return of .55. The selected stocks and their respective coefficients, portfolio weights, annual returns, and weighted returns were as follows: Delqma (.49, .04, .74, .03), Dpars (.74, .06, .23, .01), Shtehran (.92, .07, .34, .02), Dkosar (.88, .07, 1.21, .09), Dtemad (.94, .08, −.09, −.01), Dfara (.77, .06, .59, .04), Ddam (.67, .05, 1.53, .08), Dshimi (.91, .07, .89, .07), Vpakhsh (.55, .04, .15, .01), Dabor (.92, .07, .22, .02), Dasveh (.54, .04, .20, .01), Drazak (.92, .07, .72, .05), Diran (.57, .05, .84, .04), Pakhsh (.41, .03, .49, .02), Sobhan (.77, .06, .70, .04), Dalbor (.77, .06, .43, .03), and Darou (.62, .05, .24, .01).

The overall return, risk, and risk-adjusted performance of the Fama–French portfolios are presented in Table 7.

Table 7. Performance of the Fama–French Five-Factor Portfolios Compared with the Market

Performance Measure	First Approach	Second Approach	Third Approach	Fourth Approach	Market
Return	.8596	—	.6495	.6789	.3265
Risk	.04256	—	.04526	.04685	.02635
Sharpe ratio	16.8536	—	19.1248	21.5211	9.6674

As shown in Table 7, the Sharpe ratios of all portfolios that could be constructed using the Fama–French five-factor approach exceeded the market Sharpe ratio of 9.6674. No portfolio was available for the second approach because no stocks satisfied the required efficiency criteria.

Although the first approach showed the highest raw return (.8596), the fourth approach generated the highest risk-adjusted performance, with a Sharpe ratio of 21.5211. The third approach produced a Sharpe ratio of 19.1248, while the first approach yielded 16.8536. The results therefore indicated that incorporating composite financial information in the fourth approach enhanced risk-adjusted portfolio performance. Accordingly, the use of information derived jointly from the different components of the financial statements produced the strongest Fama–French portfolio from a risk-adjusted perspective.

To summarize the portfolio-selection results produced by the two analytical methods, Table 8 compares the Sharpe ratios across the four approaches.

Table 8. Comparison of Sharpe Ratios Across the Support Vector Machine and Fama–French Five-Factor Approaches

Method	First Approach	Second Approach	Third Approach	Fourth Approach
Support Vector Machine	17.2543	19.1442	20.1746	26.5022
Fama–French five-factor model	16.8536	—	19.1248	21.5211

The comparison demonstrates that the highest Sharpe ratio across all estimated portfolios was 26.5022, obtained from the Support Vector Machine under the fourth approach. The second-highest Sharpe ratio was 21.5211, obtained from the Fama–French five-factor model under the fourth approach. Thus, both modeling strategies identified the fourth approach as the most favorable in terms of risk-adjusted portfolio performance, while the SVM produced the superior overall result.

For completeness, the principal comparative findings of the two methods are summarized in Table 9.

Table 9. Summary of Comparative Portfolio-Optimization Findings

Criterion	Support Vector Machine	Fama–French Five-Factor Model	Best Result
Highest classification accuracy—testing	83%	82%	SVM, fourth approach
Highest classification accuracy—training	84%	82%	SVM, fourth approach
Highest available Sharpe ratio—first approach	17.2543	16.8536	SVM
Second approach Sharpe ratio	19.1442	No portfolio formed	SVM
Third approach Sharpe ratio	20.1746	19.1248	SVM
Fourth approach Sharpe ratio	26.5022	21.5211	SVM
Best-performing financial-ratio specification	Composite ratios	Composite ratios	Fourth approach
Overall best portfolio	Fourth approach	Fourth approach	SVM, fourth approach

Overall, the findings revealed a consistent improvement in classification and portfolio performance as the financial-ratio specification became more comprehensive. Under the Support Vector Machine, testing accuracy increased from 68% in the first approach to 83% in the fourth approach, while training accuracy increased from 72% to 84%. The same general pattern was observed under the Fama–French classification framework, where the fourth approach achieved an accuracy of 82% in both the testing and training samples.

The portfolio-performance results further demonstrated the value of the fourth approach. The highest Sharpe ratio obtained in the entire analysis was 26.5022 for the SVM-based fourth portfolio. The corresponding Fama–French portfolio also achieved its method-specific maximum Sharpe ratio under the fourth approach, equal to 21.5211. Consequently, the results indicate that the Support Vector Machine combined with the fourth approach, which incorporates composite financial information, provided the strongest performance for optimal stock portfolio formation.

4. Discussion and Conclusion

The present study examined the role of fundamental factors in optimal stock portfolio formation among pharmaceutical companies listed on the Tehran Stock Exchange by integrating financial-ratio information, Data Envelopment Analysis (DEA), Support Vector Machine (SVM) classification, and the Fama–French five-factor framework. The findings demonstrated that the use of progressively more comprehensive financial information improved both classification accuracy and portfolio performance. Under the SVM approach, testing accuracy increased from 68% in the first approach to 77% in the second approach, 82% in the third approach, and 83% in the fourth approach. A similar pattern emerged in the training sample, where accuracy increased from 72% to 79%, 83%, and 84%, respectively. The Fama–French-based classification results also showed the strongest performance for the fourth approach, with testing and training accuracies reaching 82%. Therefore, the general pattern of findings indicates that the combined use of fundamental information from multiple financial statements improves the capacity of the models to distinguish between relatively efficient and inefficient stocks.

This result can be explained by the multidimensional nature of firm financial performance. Stock selection based on only one category of accounting information may capture a limited dimension of corporate quality, whereas a combined financial-ratio framework simultaneously reflects liquidity, profitability, leverage, operating efficiency, cash-generating capacity, and capital utilization. Fundamental analysis is premised on the assumption that the financial and economic characteristics of a firm contain relevant information for determining investment attractiveness, and the present findings support this assumption by showing that broader fundamental information produced more accurate stock classification [2]. This finding is also consistent with the view that portfolio-selection decisions in the Tehran Stock Exchange require simultaneous consideration of several criteria rather than reliance on a single factor [3]. Previous multicriteria stock-ranking research in the Iranian capital market has similarly indicated that structured evaluation of multiple corporate characteristics can improve the selection of firms for inclusion in investment portfolios [4].

The superior performance of the fourth approach also highlights the informational complementarity of the different components of financial statements. Balance-sheet indicators describe financial structure and short- and long-term solvency, income-statement ratios reflect profitability and operating performance, and cash-flow ratios provide information concerning the capacity of the firm to convert operations and earnings into actual cash flows. The finding that the composite-ratio approach outperformed the more restricted approaches suggests that these information sources are complementary rather than substitutable. This interpretation is compatible with previous evidence indicating that cash-flow-based ratios can complement traditional liquidity ratios and provide additional insight into the ability of firms to sustain their operations [5]. The present study extends that logic by demonstrating that combining multiple types of financial ratios can also contribute to superior stock classification and portfolio construction.

The results related to DEA further support the usefulness of multidimensional efficiency analysis in portfolio formation. In the present study, DEA was used to classify stocks as efficient or inefficient based on the simultaneous consideration of selected inputs and outputs related to return, risk, liquidity, and financial ratios. This approach is consistent with the fundamental logic of DEA, in which relative efficiency is evaluated by considering multiple inputs and outputs without requiring the specification of a predetermined functional form [6]. The application of DEA in the present study enabled the researcher to move beyond isolated financial ratios and generate an integrated efficiency assessment for each stock. This finding aligns with previous research showing that DEA can supplement conventional financial-ratio analysis by integrating several dimensions of corporate performance into a unified efficiency framework [7].

The usefulness of DEA for portfolio screening observed in the present study is also consistent with earlier portfolio-selection research. Previous Iranian evidence has shown that DEA can be used to identify optimal stocks while explicitly considering liquidity as an important investment criterion [8]. Likewise, international research has demonstrated that DEA-based screening can enhance equity portfolio performance when efficiency measures are incorporated into the stock-selection process [9]. The present findings also correspond with more recent research in which portfolio formation based on efficiency measures derived from risk and distribution-based returns produced a structured mechanism for distinguishing among candidate securities [12]. At the same time, the results reinforce the importance of appropriate DEA model specification because the selection of inputs, outputs, and efficiency criteria can materially influence which securities are classified as efficient [10]. The use of the input-oriented BCC model in the present study was therefore important because it produced a workable set of efficient observations for subsequent classification.

Another major finding was the high classification performance of the SVM, particularly when the composite financial-ratio approach was used. The fourth SVM specification achieved the highest overall accuracy, with 83% correct classification in the testing sample and 84% in the training sample. Moreover, the Gain index confirmed that this approach converged more rapidly toward complete explanatory power than the earlier approaches. These findings indicate that SVM can effectively learn the classification structure generated by the multidimensional efficiency model and can distinguish favorable from unfavorable stocks with relatively high predictive accuracy. The result is consistent with the growing literature on the use of machine learning for portfolio decision-making, which emphasizes the capacity of learning algorithms to identify complex relationships in large and heterogeneous financial datasets [13].

The superior performance of the SVM is also in line with evidence showing that machine-learning algorithms can improve stock analysis and support more intelligent investment decisions [14]. In the Iranian context, machine-learning methods have already been proposed as useful tools for portfolio optimization, particularly because they can improve the prediction and selection stages of the investment process [15]. The results of the present study extend these findings by showing that the value of machine learning is enhanced when the input data themselves are constructed through an economically meaningful fundamental-analysis framework. In other words, SVM performance appears to depend not only on the classification algorithm but also on the informational richness of the financial variables supplied to the model.

The results are further consistent with research combining machine-learning-based stock selection with conventional portfolio optimization. Previous work has demonstrated that predictive stock selection can be integrated with mean-variance optimization to improve portfolio construction [16]. More recent reinforcement-learning research has likewise shown that stock ranking and matching procedures can improve portfolio-selection decisions when algorithms adapt to complex investment information [17]. Although the present study used SVM rather than reinforcement learning, the common implication is that the quality of portfolio construction can be improved by introducing intelligent classification and ranking mechanisms prior to the final allocation stage.

The portfolio-performance results provide additional support for the superiority of the fourth approach. Under the SVM framework, portfolio return increased from .65012 in the first approach to .68523 in the second, .71203 in the third, and .81256 in the fourth, compared with a market return of .52369. More importantly, the fourth approach achieved the highest Sharpe ratio, equal to 26.5022. This indicates that its superior return was not merely a consequence of higher risk, but that it also provided the most favorable risk-adjusted performance. The result supports the view that portfolio evaluation should be based on the relationship between return and risk rather than on return alone. Recent Sharpe-ratio-based portfolio research has similarly emphasized the importance of comparing portfolio performance relative to the volatility accepted by investors [19].

The significance of adjusted risk measures in portfolio formation has also been emphasized in recent Iranian research. Portfolio optimization based on adjusted measures of stock risk has shown that the quality of an investment portfolio depends on the way in which risk is incorporated into the performance-measurement framework [20]. The present study supports this position because the ranking of the portfolios becomes more meaningful when the Sharpe ratio is considered alongside absolute return. Although the SVM portfolios generally displayed increasing risk across the four approaches, the increase in return more than compensated for the increase in risk, resulting in a substantial improvement in risk-adjusted performance under the fourth approach.

The Fama–French-based findings showed a similar overall pattern. Although the first Fama–French portfolio produced the highest raw return among the available Fama–French portfolios, the fourth approach generated the highest Sharpe ratio, equal to 21.5211. Therefore, the composite financial-ratio approach remained the best option when performance was evaluated on a risk-adjusted basis. This finding is theoretically meaningful because the Fama–French five-factor model incorporates multiple systematic dimensions associated with market exposure, size, value, profitability, and investment. The alignment between the factor-model results and the fundamental-ratio results suggests that stock-selection quality improves when accounting-based information and systematic return-related characteristics are considered together.

The inability of the second Fama–French approach to generate a proposed portfolio is also informative. In that specification, no stocks were simultaneously predicted to be efficient according to the relevant criteria. This result indicates that a restricted predictor set may be insufficient to identify an investable group of securities when the selection criteria are relatively demanding. However, once additional variables were incorporated into the third and fourth approaches, efficient candidate stocks again emerged. This pattern provides further support for the argument that portfolio selection benefits from richer information structures. Hybrid optimization research has similarly emphasized the usefulness of integrating multiple financial and decision criteria when constructing portfolios under complex constraints [18].

The comparison between the SVM and Fama–French models produced one of the most important findings of the study. Across comparable approaches, the SVM generally produced higher Sharpe ratios than the Fama–French model. The largest difference appeared under the fourth approach, where the SVM achieved a Sharpe ratio of 26.5022 compared with 21.5211 for the Fama–French model. Thus, although both models identified the fourth approach as the best-performing specification, the SVM generated the superior portfolio overall. This suggests that the nonlinear or high-dimensional classification capability of machine-learning models may provide an advantage over factor-based modeling when the purpose is portfolio screening and ranking rather than merely explaining expected returns.

The result is compatible with contemporary evidence emphasizing the growing importance of machine learning in portfolio decisions [13]. It also supports research suggesting that machine-learning approaches can outperform more traditional investment-selection procedures by exploiting relationships that may not be sufficiently captured by conventional models [14]. Nevertheless, the strong performance of the Fama–French model indicates that traditional financial theory remains highly relevant. Rather than implying that machine learning should replace factor models, the findings suggest that the two approaches may be complementary: factor models provide economically interpretable structures for understanding expected returns, whereas machine-learning models provide flexible tools for prediction, classification, and ranking.

The findings also have implications for diversification and portfolio risk. Contemporary markets are characterized by increasing interconnectedness, and return and volatility spillovers can affect the diversification benefits available to investors. Previous research on traditional technology stocks, decentralized finance instruments, and cryptocurrencies has shown that spillover structures can materially alter portfolio-optimization decisions [21]. Research combining stocks and cryptocurrencies has likewise demonstrated that optimal allocation is sensitive to the risk-return characteristics of the asset universe [22]. Although the present study examined only pharmaceutical stocks, the same principle applies within an industry-specific portfolio: investors must identify securities that collectively produce a favorable balance between expected return and risk rather than selecting stocks solely according to individual profitability.

The observed advantage of systematic, data-driven portfolio selection can also be interpreted in light of behavioral-finance research. Investors do not always process financial information rationally, and behavioral biases may alter portfolio composition and reduce decision quality. Recent evidence from the Tehran Stock Exchange has shown that behavioral biases can influence portfolio optimization [23]. Investor sentiment has similarly been found to affect portfolio decisions in both stock and cryptocurrency markets [24]. Moreover, portfolio models incorporating behavioral preferences and investor memory suggest that previous experiences can influence present allocation choices [25]. Against this background, the use of a structured DEA-SVM framework may be particularly valuable because it provides a systematic decision-support mechanism that reduces reliance on subjective judgments and emotional responses.

The ranking and weighting procedure used in the present study also corresponds with the broader literature on integrated portfolio-evaluation systems. DEA-based fuzzy portfolio models combined with TOPSIS have demonstrated that efficient portfolios can be further ranked according to different risk indicators [11]. Such findings support the general methodological principle used in the present research: efficiency identification alone is insufficient, and additional ranking mechanisms are required to transform an efficient set of stocks into a practical investment portfolio. In the current study, SVM classification, Euclidean distance from separating hyperplanes, stock ranking, and subsequent portfolio weighting collectively provided this additional decision structure.

Overall, the findings suggest that financial-statement information has significant incremental value in optimal portfolio formation. The progression from narrower financial-ratio categories to the composite approach improved classification accuracy and risk-adjusted portfolio performance under both analytical frameworks. The fourth approach consistently produced the strongest results, while the SVM-based fourth approach generated the highest overall Sharpe ratio. These results support an integrated view of portfolio optimization in which fundamental analysis, DEA-based efficiency measurement, machine-learning classification, and factor-based return modeling are combined rather than treated as competing alternatives. The findings also reinforce the broader financial-management principle that investment decisions should be based on a comprehensive evaluation of both corporate fundamentals and market-based risk-return characteristics rather than on isolated accounting ratios or price-based measures alone [1].

A number of limitations should be considered when interpreting the findings. First, the study was restricted to pharmaceutical companies listed on the Tehran Stock Exchange, which limits the direct generalizability of the results to other industries and capital markets. Second, the final sample included a relatively limited number of firms because companies with incomplete information, negative BCC inputs, or delisting events were excluded. Third, the results are dependent on the selected financial ratios, DEA specification, classification rules, and the particular SVM structure employed. Fourth, the analysis did not explicitly incorporate transaction costs, taxes, bid-ask spreads, short-selling restrictions, portfolio turnover, or other real-world implementation frictions. Fifth, macroeconomic variables, political uncertainty, market-regime changes, and behavioral indicators were not directly included in the model. Finally, some financial ratios may be affected by accounting policies and reporting quality, which may influence their comparability across firms and periods.

Future research should replicate the proposed framework in other industries and markets to evaluate the robustness and external validity of the findings. Comparative studies could examine whether the fourth composite-ratio approach continues to outperform narrower information structures in banking, petrochemical, automotive, technology, or multi-industry samples. Future studies could also compare SVM with alternative machine-learning and deep-learning algorithms, including random forests, gradient boosting, neural networks, long short-term memory models, and reinforcement-learning approaches. Incorporating macroeconomic variables, investor sentiment, market-regime indicators, ESG measures, and textual information from corporate disclosures could further strengthen the predictive framework. Researchers may also examine dynamic or rolling portfolio rebalancing, transaction costs, downside-risk measures, conditional value at risk, and alternative performance metrics. Finally, ensemble frameworks combining DEA, the Fama–French model, machine learning, and multicriteria decision-making methods may provide an even more comprehensive approach to portfolio optimization.

From a practical perspective, investors, portfolio managers, brokerage firms, and financial analysts can use the findings to improve stock-screening and portfolio-construction procedures. Rather than relying on a limited set of profitability or liquidity ratios, decision-makers should consider integrated information from the balance sheet, income statement, and cash flow statement. A preliminary DEA-based screening stage can be used to identify relatively efficient stocks, after which machine-learning tools such as SVM can classify and rank investment candidates. Portfolio managers should evaluate the resulting allocations using risk-adjusted measures such as the Sharpe ratio rather than focusing only on absolute return. Financial institutions can also incorporate such models into decision-support systems and analytical platforms to provide more systematic portfolio recommendations. For investors in the pharmaceutical sector, the results specifically suggest that composite financial information combined with SVM-based classification can provide a useful framework for constructing portfolios with superior risk-adjusted performance.

Authors' Contributions

Authors equally contributed to this article.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all participants who participate in this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

[1] H. Vakilifard and M. Vakilifard, *Financial Management*, 10th ed. Jangal Publications, 2015.

[2] S. M. A. Shohadai, *Fundamental Analysis in the Capital Market*, 2nd ed. Tehran: Chalesh Publishing House, 2008.

[3] A. Pahlavan, E. Ramezanpour, and M. H. Gholizadeh, "Prioritizing Factors Affecting Stock Selection in Tehran Stock Exchange Using Fuzzy Network Analysis Process," presented at the Third Conference on Financial Mathematics and Applications, 2012.

[4] A. R. Keyghabadi, Z. Lashgari, and A. Shabani, "Ranking the Top Listed Companies for Portfolio Selection Using AHP Technique: Case Study of the Top 50 Companies of Tehran Stock Exchange," presented at the Second National Conference on Applied Research in Management and Accounting Sciences, 2014.

[5] S. Mehrani, F. Ghior, and T. Bahramfar, "Examining the Relationship between Traditional Liquidity Ratios and the Ratios Obtained from the Cash Flow Statement in Order to Evaluate the Continuity of Companies' Activity," *Accounting and Auditing Reviews,* vol. 12, no. 2, pp. 3-17, 2014.

[6] W. D. Cook and L. M. Seiford, "Data Envelopment Analysis (DEA) - Thirty Years On," *European Journal of Operational Research,* vol. 192, pp. 1-17, 2014, doi: 10.1016/j.ejor.2008.01.032.

[7] S. Khajawi, A. Ghiori Moghadam, and M. Ghafari, "The Data Envelopment Analysis Technique Is a Supplement to the Traditional Analysis of Financial Ratios," *Financial Accounting and Audit Reviews,* vol. 17, no. 6, pp. 41-56, 2019.

[8] S. Khajawi and A. Ghiori Moqdeh, "Data Envelopment Analysis, a Method for Choosing the Optimal Portfolio According to the Amount of Stock Liquidity," *Accounting Developments,* vol. 4, no. 2, pp. 27-52, 2013.

[9] E. Pätäri, T. Leivo, and S. Honkapuro, "Enhancement of Equity Portfolio Performance Using Data Envelopment Analysis," *European Journal of Operational Research,* vol. 220, no. 2, pp. 786-797, 2012, doi: 10.1016/j.ejor.2012.02.006.

[10] A. C. Tarnaud and H. Leleu, "Portfolio Analysis with DEA: Prior to Choosing a Model," *Omega,* vol. 75, pp. 57-76, 2017, doi: 10.1016/j.omega.2017.02.003.

[11] A. Aggarwal, A. Gupta, R. Verma, and R. Kumari, "DEA Based Fuzzy Portfolio Evaluation Models Integrated with TOPSIS Techniques to Rank the Efficient Portfolios under Different Risk Indicators," 2023.

[12] H. A. Heydarzadeh, F. Rahnamee Roudposhti, A. Rashidi, and S. E. Najafi, "Portfolio Formation Based on Efficiency Derived from Risk and Distribution-Based Returns with Data Envelopment Analysis Approach," *Decision Making and Operations Research,* vol. 9, no. 2, pp. 289-304, 2024.

[13] M. Guidolin, G. Panzeri, and M. Pedio, "Machine Learning in Portfolio Decisions," in "BAFFI CAREFIN Centre Research Paper," 2024. [Online]. Available: https://ssrn.com/abstract=4988124

[14] M. Munawir and U. S. Sulistyawati, "Stock Portfolio Analysis With Machine Learning Algorithmic Approach for Smart Investment Decisions," *International Journal Software Engineering and Computer Science (Ijsecs),* vol. 4, no. 3, pp. 860-870, 2024, doi: 10.35870/ijsecs.v4i3.2606.

[15] S. Heratizadeh and F. Rezaei, "Presenting a New Method for Utilizing Machine Learning in the Process of Portfolio Optimization," *Decision Making and Operations Research,* vol. 9, no. 4, pp. 968-984, 2023.

[16] A. Chaweewanchon and R. Chaysiri, "Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning," *International Journal of Financial Studies,* vol. 10, no. 3, p. 64, 2022, doi: 10.3390/ijfs10030064.

[17] A. Chaher, "Optimizing portfolio selection through stock ranking and matching: A reinforcement learning approach," *Expert Systems with Applications,* vol. 269, p. 126430, 2025, doi: 10.1016/j.eswa.2025.126430.

[18] P. Gupta, M. K. Mehlawat, and A. Saxena, "Hybrid Optimization Models of Portfolio Selection Involving Financial and Ethical Considerations," *Knowledge-Based Systems,* vol. 37, pp. 318-333, 2013, doi: 10.1016/j.knosys.2012.08.014.

[19] Z. Wang, "Active vs. Passive Investment in the Post-Pandemic U.S. Stock Market: A Sharpe Ratio-Based Portfolio Optimization Compared to the S&P 500," *Highlights in Business Economics and Management,* vol. 61, pp. 41-45, 2025, doi: 10.54097/j6kd1r14.

[20] A. Mousizadeh, J. Ramazani, M. Ali Akbari, M. Safarigrayeli, and R. Rezaiyan, "Portfolio optimization with adjusted risk degree of stocks based on performance measurement model," *Investment Knowledge,* vol. 15, no. 58, pp. 23-42, 2024, doi: 10.30495/jik.2025.23818.

[21] R. J. Oben, M. Seraj, and Ş. Z. Eyüpoğlu, "Volatility and Return Spillovers Among US Traditional Technology Stocks, Decentralized Finance Instruments And conventional Cryptocurrencies: Implications for Portfolio Optimization," *Review of Behavioral Finance,* vol. 17, no. 5, pp. 807-834, 2025, doi: 10.1108/rbf-12-2024-0377.

[22] N. Shahbazi and S. Barkhordari, "Optimizing a portfolio comprising selected stocks and cryptocurrencies," *Budget and Finance Strategic Research,* vol. 6, no. 2, pp. 11-35, 2025.

[23] M. Eskandari, A. Mohseni, and M. Ghasemi, "The Effect of Behavioral Biases on Portfolio Optimization of Investors in the Tehran Stock Exchange: An Approach Based on Modern Behavioral Finance," *Accounting, Finance, and Computational Intelligence,* pp. 1-17, 2026.

[24] S. Nourollahi, A. Jafari, and A. Pakmaram, "The Impact of Investor Sentiment on Portfolio Optimization in the Tehran Stock Exchange and Cryptocurrency Markets," *Business, Marketing, and Finance Open,* pp. 1-21, 2025.

[25] V. Mousavi Kakhki and S. Khatabi, "Presenting a stock portfolio optimization model based on behavioral preferences and investor memory," 2024. [Online]. Available: https://civilica.com/doc/1961759.