Predicting Audit Failure Using Metaheuristic Algorithms

Saeed Zarrin[1], Ahmad Mohammadi[2]*, Mehdi Zeynali[1]

1. Department of Accounting, Tabriz Branch, Islamic Azad University, Tabriz, Iran.
2. Department of Accounting, Tabriz Branch, Islamic Azad University, Tabriz, Iran (Corresponding author).

*** Corresponding author email address**: iacpa_a_mohammadi@yahoo.com

Received: 2024-05-11 **Reviewed:** 2024-06-16 **Revised:** 2024-08-01 **Accepted:** 2024-08-14 **Published:** 2024-09-10

Abstract

The aim of the present study is to predict audit failure using metaheuristic algorithms in companies listed on the Tehran Stock Exchange. To achieve this objective, 1,848 firm-year observations (154 companies over 12 years) were collected from the annual financial reports of companies listed on the Tehran Stock Exchange during the period from 2011 to 2022. In this study, four metaheuristic algorithms (including Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Bee Colony Optimization (BCO)) were utilized, as well as two methods for selecting the final research variables (the two-sample t-test and the forward stepwise selection method) to create the model. The results from the metaheuristic algorithms indicate that the overall accuracy of the GA, PSO, ACO, and BCO algorithms is 95.3%, 94.5%, 90.6%, and 92.8%, respectively, demonstrating the superiority of the Genetic Algorithm (GA) compared to other metaheuristic algorithms. Furthermore, the overall results from the variable selection methods indicate the efficiency of the stepwise method. Therefore, in companies listed on the Tehran Stock Exchange, the stepwise method and the Genetic Algorithm (GA) provide the most efficient model for predicting audit failure.

Keywords: *Audit failure, Prediction, Metaheuristic algorithms.*

How to cite this article:

Zarrin S, Mohammadi A, Zeynali M. (2024). Predicting Audit Failure Using Metaheuristic Algorithms. Management Strategies and Engineering Sciences, 6(3), 20-31.

1. Introduction

Predicting audit failure is an essential endeavor in the field of accounting and finance, given its implications for corporate governance, investor confidence, and the stability of financial markets. Over the years, audit failures have not only led to significant financial losses but have also tarnished the credibility of accounting firms and auditors [1]. As such, the development of effective models to predict audit failure is critical in mitigating these risks. This study seeks to explore the use of metaheuristic algorithms—such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Bee Colony Optimization (BCO)—to forecast audit failure, particularly in companies listed on the Tehran Stock Exchange. The choice of metaheuristic algorithms is motivated by their proven ability to handle complex optimization problems, which are often nonlinear and multidimensional, characteristics that are typical of financial prediction models [2].

The concept of audit failure arises when an auditor fails to identify material misstatements in the financial statements of an entity. This failure often results from negligence, insufficient audit procedures, or a lack of skepticism on the part of auditors. Previous studies have shown that audit failures can have widespread consequences, including bankruptcy, regulatory sanctions, and reputational damage [3, 4]. Given these consequences, predicting audit failure is a high-priority task for both regulatory bodies and the auditing profession.

Historically, various approaches have been used to predict audit failures, ranging from traditional statistical models to more contemporary machine learning methods [5]. For instance, Fafatas (2010) highlighted the conservative nature of auditors following audit failures, a behavior aimed at reducing the likelihood of future misstatements [4]. Furthermore, Jin et al. (2011) investigated the capacity of audit quality variables in predicting bank failures during the financial crisis, demonstrating the importance of incorporating audit-related variables into predictive models [6].

In recent years, the emergence of artificial intelligence (AI) and machine learning has revolutionized the predictive capabilities of audit models. Algorithms like neural networks and support vector machines have shown promise in identifying patterns and anomalies that may indicate potential audit failures [7, 8]. However, these methods often struggle with interpretability and may require large datasets to achieve accuracy, limiting their applicability in smaller firms or datasets [9].

Metaheuristic algorithms offer a promising alternative, as they are designed to optimize complex systems and can be applied even when the underlying relationships between variables are unknown or difficult to model directly [10]. These algorithms, which mimic natural processes such as evolution or swarm behavior, have been applied in various domains, including healthcare, engineering, and finance, with impressive results [11]. Their application in audit failure prediction, however, remains relatively unexplored. This study seeks to fill this gap by applying four metaheuristic algorithms—GA, PSO, ACO, and BCO—to predict audit failure in a sample of firms listed on the Tehran Stock Exchange. The choice of these algorithms is based on their demonstrated success in solving optimization problems in previous studies across diverse fields [12, 13].

The Tehran Stock Exchange provides a unique context for this study. Over the past decade, the Iranian economy has faced various challenges, including sanctions, inflation, and currency devaluation, all of which have had significant impacts on the financial health of firms listed on the stock exchange [14]. These economic pressures increase the likelihood of financial misstatements and, consequently, audit failures. Understanding and predicting these failures is essential not only for auditors but also for regulators and investors seeking to make informed decisions in a volatile market environment [15].

The metaheuristic algorithms used in this study have unique features that make them suitable for predicting audit failure. The Genetic Algorithm (GA), for example, is inspired by the process of natural selection and is widely used in optimization problems. It works by evolving a population of potential solutions over several generations, selecting the best-performing solutions and combining them to create new, potentially better solutions [16]. In previous studies, GA has been successfully applied to a range of problems, from healthcare diagnostics to cloud management [9, 17].

Similarly, Particle Swarm Optimization (PSO) is inspired by the social behavior of birds flocking or fish schooling. Each particle in the swarm represents a potential solution, and particles communicate with each other to move towards the best solution found by the group. PSO has been used in various applications, including failure prediction in cloud data centers and financial forecasting [17, 18]. Its adaptability and robustness make it a strong candidate for predicting audit failure in the complex financial environment of the Tehran Stock Exchange.

Ant Colony Optimization (ACO), on the other hand, is inspired by the foraging behavior of ants. Ants leave pheromone trails that guide other ants to food sources, and this behavior is modeled in ACO to find optimal solutions to problems. ACO has been applied in diverse fields, including network optimization and fraud detection [19]. Its ability to explore multiple solutions simultaneously makes it a useful tool for predicting audit failure, where the relationships between variables are often complex and nonlinear.

Finally, Bee Colony Optimization (BCO) mimics the behavior of honeybees in finding food sources. Each bee in the colony represents a potential solution, and bees communicate with each other to find the best solution. BCO has been used in a variety of applications, including scheduling and resource allocation [17]. Its decentralized nature allows it to explore a wide range of potential solutions, making it a valuable addition to the suite of algorithms used in this study.

The use of these metaheuristic algorithms in audit failure prediction is expected to offer several advantages over traditional methods. First, these algorithms are well-suited to handling large, complex datasets, which are typical in financial analysis [3]. Second, metaheuristic algorithms are capable of optimizing models even when the underlying relationships between variables are poorly understood or highly nonlinear [8]. This is particularly important in audit failure prediction, where the causes of failure are often multifaceted and difficult to capture using traditional linear models [5]. Finally, metaheuristic algorithms are highly adaptable and can be tailored to specific contexts, such as the Tehran Stock Exchange, where economic conditions and regulatory frameworks may differ significantly from those in other markets [15].

In summary, this study seeks to apply four metaheuristic algorithms—GA, PSO, ACO, and BCO—to predict audit failure in companies listed on the Tehran Stock Exchange. The use of metaheuristic algorithms in this context is motivated by their proven success in optimizing complex systems and their potential to offer superior predictive accuracy compared to traditional methods. By developing models that can accurately predict audit failure, this study aims to contribute to the ongoing efforts to improve audit quality and reduce the risk of financial misstatements. The following sections will provide an overview of the methodology used in this study, as well as a detailed analysis of the results obtained from the application of the metaheuristic algorithms.

2. Methodology

This study is applied in nature and follows a quasi-experimental, ex-post facto research design, utilizing historical data. To collect the theoretical framework, the study utilized journals, books, and available databases. The necessary data for analysis were gathered from the *Rahavard Novin* software, as well as the websites of the "Research, Development, and Islamic Studies Management of the Securities and Exchange Organization," Codal, the Central Bank of Iran, and the Statistical Center of Iran.

The statistical population of this study includes all companies listed on the Tehran Stock Exchange. The time frame covers a 12-year period based on the financial statements of sample companies from 2011 to 2022.

For determining the sample, the study utilized a screening method. Companies that met the following conditions were selected as the sample, while the rest were excluded:

1. The fiscal year of the company ends on the last day of March each year.
2. The company did not change its fiscal year during the study period.
3. The companies under review are not investment firms, holding companies, financial intermediaries, or insurance companies.
4. The data and information of these companies are available.

Given these conditions and limitations, a total of 154 companies listed on the Tehran Stock Exchange were selected as the sample for the study.

Measurement of Research Variables

Dependent Variable (Response):

Audit Failure (Audit_Failure): The dependent variable in this study is audit failure, which, following previous research, is measured using audit errors (including both Type I and Type II audit errors). In other words, audit failure is a binary variable that equals 1 if either a Type I or Type II audit error occurred, and 0 otherwise. Audit errors are measured as follows:

- **Type I Audit Error**: If the auditor issues an adverse opinion, but the client's financial statements are not restated in the following year (i.e., no prior period adjustments are recognized), it is classified as a Type I error, and a value of 1 is assigned. If no error occurs, a value of 0 is assigned.

- **Type II Audit Error**: If the auditor issues a clean opinion in one year, but the client's financial statements are restated in the following year (i.e., prior period adjustments are recognized), it is classified as a Type II error, and a value of 1 is assigned. If no error occurs, a value of 0 is assigned.

Independent Variables (Predictors)

In this study, the factors influencing audit failure are categorized into three levels: auditor characteristics, client characteristics, and stock market and macroeconomic characteristics. These factors were determined based on theoretical discussions and prior empirical studies. One distinguishing feature and innovation of this study is that it examines the factors influencing audit failure across these three levels. The operational definitions of these variables are detailed below:

Auditor Characteristics:

- **Independent Auditor Size (AUDSIZE)**: A binary variable that equals 1 if the independent auditor is the Audit Organization or the *Mofid Rahbar* auditing firm, and 0 otherwise. Information related to this variable is extracted from the independent auditor's report or the board of directors' report.
- **Independent Auditor Tenure (AUDTEN)**: Represents the number of consecutive years that a particular auditing firm has audited a company's financial statements. This information is extracted from the independent auditor's report or the board of directors' report.
- **Audit Independence (DA)**: Following Chang et al. (2019), Chen et al. (2008), Chi et al. (2012), and Chen et al. (2010), audit independence is measured using discretionary accruals. The modified Jones model, introduced by Dechow et al. (1995), is used for this purpose. Discretionary accruals are estimated using the following formula:

$$TACC(it) / TA(it-1) = \alpha 0(1/ TA(it-1)) + \beta 1((\Delta REV(it) - \Delta REC(it)) / TA(it-1)) + \beta 2(PPE(it) / TA(it-1)) + \varepsilon(it)$$

Where:

- TACC(it) = total accruals of company i in year t
- TA(it-1) = total assets of company i in year t-1
- ΔREV(it) = change in revenue of company i between years t-1 and t
- ΔREC(it) = change in receivables of company i between years t-1 and t
- PPE(it) = gross amount of property, plant, and equipment of company i in year t
- ε(it) = the residual or discretionary accruals.
- **Auditor Change (Auditorchange)**: A binary variable that equals 1 if the independent auditor (audit firm) has changed compared to the previous year, and 0 otherwise.
- **Audit Firm Ranking (AUD_Rank)**: A binary variable that equals 1 if the independent auditor (audit firm) is ranked in Group A, and 0 otherwise.
- **Audit Opinion Type (AUDOPIN)**: A binary variable that equals 1 if the audit opinion is clean, and 0 otherwise. Information related to this variable is extracted from the independent auditor's report or the board of directors' report.
- **Audit Report Delay (AUD_Lag)**: Measured as the natural logarithm of the number of days between the balance sheet date and the audit report issuance date.
- **Auditor Specialization (Specialist)**: In this study, auditor specialization is measured using market share, similar to the approach of Etemadi et al. (2010). A higher market share indicates that the auditor has distinguished themselves from competitors in terms of audit quality.

Auditor Market Share = (Total assets of all clients in a specific industry audited by the audit firm) / (Total assets of all clients in the same industry)

Audit firms with market shares greater than (1 / number of firms in the industry) * (1 / 2) are considered industry specialists.

- **Audit Fee (AUDFEE)**: Measured as the natural logarithm of the audit fee paid to the auditor. Information related to this variable is extracted from the independent auditor's report or the board of directors' report.

Client Characteristics:

- **Company Size (SIZE)**: Measured as the natural logarithm of the company's total assets. Information related to this variable is extracted from the financial statements.
- **Liquidity (LIQID)**: Measured as the ratio of current assets to current liabilities. Information related to this variable is extracted from the financial statements.

- **Debt Ratio (DEBT)**: Measured as the ratio of total liabilities to total assets. Information related to this variable is extracted from the financial statements.
- **Profitability (PROF)**: Measured as the ratio of operating profit to total assets. Information related to this variable is extracted from the financial statements.
- **Company Age (AGE)**: Measured as the natural logarithm of the number of years since the company was established. Information related to this variable is extracted from the explanatory notes to the financial statements or the board of directors' report.
- **Growth Opportunity (GROWTH)**: Measured as the ratio of the market value of equity to the book value of equity. Information related to this variable is extracted from the financial statements.

Stock Market and Macroeconomic Characteristics:

- **Systematic Risk (SYSR)**: Estimated by calculating the beta coefficient for each company over the year using the CAPM model.
- **Unsystematic Risk (NSYSR)**: Measured as the standard deviation of the residuals from the daily CAPM model.
- **Inflation Rate (INF)**: The inflation rate (consumer price index) is obtained from the Central Bank of Iran or the Statistical Center of Iran.
- **Interest Rate (INT)**: The interest rate is obtained from the Central Bank of Iran or the Statistical Center of Iran.
- **Exchange Rate (CURR)**: Measured as the natural logarithm of the exchange rate (USD). Information related to the exchange rate is obtained from the Central Bank of Iran or the Statistical Center of Iran.

Economic Growth Rate (GGDP): Measured as the changes in gross domestic product (GDP). Information related to this variable is obtained from the Central Bank of Iran or the Statistical Center of Iran.

3. Findings

In Table 1, Panel A, some descriptive statistics of the variables, including the mean, median, minimum observations, maximum observations, and standard deviation, are presented. For example, the mean value of the company liquidity variable is 1.583, which, considering the standard deviation (0.803), indicates low volatility. The mean value of the company profitability variable is 0.156, which, considering the standard deviation (0.134), also shows low volatility. The mean value of the company's debt ratio is 0.559, indicating that 55.9% of the company's financial resources are financed through debt.

According to Panel B, the results of the t-test show that, at a 95% confidence level, the variables of auditor tenure, audit independence, audit report delay, audit fee, company size, company liquidity, company profitability, institutional ownership, major shareholder ownership, board independence, exchange rate, and economic growth rate differ significantly between companies with audit failure and those without audit failure. Moreover, the mean values of these variables are lower in companies with audit failure compared to companies without audit failure. The results also show that, at a 95% confidence level, the variables of the company's debt ratio, state ownership, unsystematic risk, and interest rate differ significantly between companies with audit failure and those without audit failure. Additionally, the mean values of these variables are higher in companies with audit failure compared to those without.

Finally, the results from Panel B show that, at a 95% confidence level, the variables of company age, company growth opportunity, client industry size, industry competition level, managerial ownership, audit committee size, audit committee independence, audit committee financial expertise, board size, CEO tenure, board financial expertise, systematic risk, and inflation rate do not differ significantly between companies with audit failure and those without audit failure.

According to Panel C, the results of the chi-square test show that, at a 95% confidence level, the variables of auditor size, auditor change, audit firm ranking, audit opinion type, and auditor specialization differ significantly between companies with audit failure and those without audit failure. In contrast, the type of client ownership (private or state-owned) does not differ significantly between companies with audit failure and those without.

Table 1. Descriptive Statistics of Research Variables

Variable Name	Observations	Mean	Median	Maximum	Minimum	Standard Deviation
Auditor Tenure	1848	3.944	3.000	15.000	1.000	3.966

Auditor Independence	1848	0.049	0.038	0.290	-0.164	0.119
Audit Report Delay	1848	4.333	4.419	5.455	2.890	0.375
Audit Fee	1848	5.288	6.774	9.933	3.751	3.262
Company Size	1848	14.731	14.562	17.973	12.593	1.533
Company Liquidity	1848	1.583	1.366	8.107	0.599	0.803
Company Debt Ratio	1848	0.559	0.565	0.923	0.172	0.209
Company Profitability	1848	0.156	0.131	0.432	-0.059	0.134
Company Age	1848	3.685	3.784	4.263	2.565	0.345
Company Growth Opportunity	1848	6.618	4.034	27.804	0.624	6.977
Client Industry Size	1848	18.653	18.612	21.292	15.830	1.564
Industry Competition Level	1848	0.170	0.169	0.389	0.011	0.138
State Ownership	1848	0.453	0.550	0.898	0.000	0.338
Institutional Ownership	1848	0.545	0.645	0.921	0.054	0.322
Managerial Ownership	1848	0.217	0.229	0.306	0.045	0.068
Major Shareholder Ownership	1848	0.499	0.510	0.864	0.130	0.204
Audit Committee Size	1848	2.756	3.000	5.000	0.000	1.044
Audit Committee Independence	1848	0.634	0.667	1.000	0.000	0.291
Audit Committee Financial Expertise	1848	0.817	1.000	1.000	0.000	0.321
Board Size	1848	5.025	5.000	7.000	5.000	0.222
Board Independence	1848	0.653	0.600	1.000	0.000	0.183
CEO Tenure	1848	3.820	2.000	13.000	1.000	3.406
Board Financial Expertise	1848	0.644	0.600	0.857	0.200	0.163
Systematic Risk	1848	0.699	0.611	2.324	-0.594	0.773
Unsystematic Risk	1848	0.149	0.139	0.308	0.044	0.074
Inflation Rate	1848	27.858	30.850	46.500	9.000	13.339
Interest Rate	1848	17.233	18.000	23.140	12.500	3.568
Exchange Rate	1848	10.087	10.185	13.102	8.343	1.449
Economic Growth Rate	1848	0.700	3.000	8.300	-7.000	4.622

Table 2. t-test in Companies with and without Audit Failure

Variable Name	Audit Failure (Observations = 431)	No Audit Failure (Observations = 1417)	t-statistic
Auditor Tenure	3.433	4.038	-2.940***
Auditor Independence	0.038	0.053	-3.292***
Audit Report Delay	3.626	4.936	-2.488**
Audit Fee	4.705	5.582	-3.126***
Company Size	14.516	14.796	-3.176***
Company Liquidity	1.295	1.671	-9.544***
Company Debt Ratio	0.673	0.524	13.520***
Company Profitability	0.051	0.189	-22.855***
Company Age	3.661	3.692	-1.603
Company Growth Opportunity	6.167	6.842	0.762
Client Industry Size	18.661	18.651	0.111
Industry Competition Level	0.166	0.171	-0.615
State Ownership	0.509	0.410	3.576***
Institutional Ownership	0.508	0.592	-2.907***
Managerial Ownership	0.215	0.219	0.551
Major Shareholder Ownership	0.465	0.520	-4.591***
Audit Committee Size	2.722	2.767	-0.793
Audit Committee Independence	0.627	0.636	-0.575
Audit Committee Financial Expertise	0.805	0.821	-0.907
Board Size	5.028	5.024	0.315
Board Independence	0.602	0.677	-3.454***
CEO Tenure	3.566	3.897	-1.767*
Board Financial Expertise	0.647	0.643	0.442
Systematic Risk	0.719	0.689	1.810*
Unsystematic Risk	0.167	0.142	2.290**
Inflation Rate	26.920	28.144	-1.669*
Interest Rate	17.333	16.903	2.279**
Exchange Rate	9.865	10.155	-3.648**
Economic Growth Rate	0.554	0.745	-3.750**

Significance at 90%, ** Significance at 95%, *** Significance at 99%

Table 3. Chi-square Test in Companies with and without Audit Failure

Variable Name	Audit Failure (Observations = 431)	No Audit Failure (Observations = 1417)	Chi-square Statistic
Auditor Size	102	365	22.166***
Auditor Change	402	342	16.960***
Audit Firm Ranking	265	887	25.174***
Audit Opinion Type	229	781	18.525***
Auditor Specialization	204	751	4.251**
Client Ownership Type (Private or State-owned)	180	628	0.877

Significance at 90%, ** Significance at 95%, *** Significance at 99%

To address the research questions and achieve the study objectives, the intended models, which include those developed using Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Bee Colony Optimization (BCO), were created based on selected variables using the two-sample t-test and forward stepwise selection method. We will compare the results obtained from these models.

To determine the final variables for the study from the initial variables, both the two-sample t-test and forward stepwise selection methods were used. The difference between these two methods is that, in the two-sample t-test, the significance of the difference in means for each independent variable is examined without considering the relationship of other independent variables with the dependent variable. In contrast, the forward stepwise selection method starts with an empty set of features, and in each iteration, the best primary feature is selected and added to the previous set. The goal is to select variables that maximize the coefficient of determination (R^2) of the model. For conducting the above statistical tests, SPSS software was used.

To perform a two-sample t-test, the variances of the two samples must first be compared. In other words, the test of equal variances precedes the test of equal means. Levene's test, based on Fisher's statistic, is used to test the equality of variances. This test does not require the data distribution to be normal.

The t-statistic for testing the equality of two-sample means, under the assumption of equal and unequal variances, is calculated. In the case of equal variances, equation (5) is used to calculate the t-statistic, where the degrees of freedom (df) are calculated as df = n1 + n2 - 2.

Equation (5): $t = ((\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)) / Sp \sqrt{(1/n_1 + 1/n_2)}$

Equation (6): $Sp = \sqrt{(((n_1 - 1) S_1^2 + (n_2 - 1) S_2^2) / (n_1 + n_2 - 2))}$

In the case of unequal variances, the t-statistic is calculated using equation (7) and the degrees of freedom using equation (8).

Equation (7): $t = ((\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)) / \sqrt{((S_2^2 / n_1) + (S_2^2 / n_2))}$

Equation (8): $df = ((S_1^2 / n_1) + (S_2^2 / n_2))^2 / (((S_1^2 / n_1)^2 / (n_1 - 1)) + ((S_2^2 / n_2)^2 / (n_2 - 1)))$

Where n_1 and n_2 represent the sample sizes of the first and second samples, and S_1 and S_2 are the standard deviations of the first and second samples, respectively.

The calculated t-statistic based on the above formulas is then compared to the critical t-value from the table, considering a 5% error level. If the calculated t-statistic is smaller than the critical t-value (p-value < 0.05), the difference in means between the two groups is considered significant; otherwise, it is rejected.

The results of the two-sample t-test are presented in Panels B and C of Table 1. Based on the results, it can be concluded that there is a significant difference in the means of the auditor size, auditor tenure, audit independence, auditor change, audit firm ranking, audit opinion type, audit report delay, auditor specialization, audit fee, company size, company liquidity, company debt ratio, company profitability, state ownership, institutional ownership, major shareholder ownership, board independence, unsystematic risk, interest rate, exchange rate, and economic growth rate between companies with and without audit failure at a 95% confidence level. Accordingly, based on the results of this test, 21 variables with significance levels below 5% are introduced as the final variables, as shown in Table 1.

In the forward stepwise selection method, the goal is to estimate a logistic regression model based on the initial research variables, with the aim of maximizing the model's coefficient of determination (R^2). The process starts with an empty set of features, and in each iteration, the best feature is selected and added to the previous set. The goal is to select variables that maximize the model's R^2. This process continues until no further improvement in R^2 is observed with the addition of any remaining variables.

Using the forward stepwise method, after 12 steps, the variables of auditor size, auditor tenure, audit independence, auditor change, audit firm ranking, audit opinion type, audit report delay, auditor specialization, audit fee, company size, company debt ratio, company profitability, state ownership, major shareholder ownership, board independence, unsystematic risk, and exchange rate—comprising 17 variables—were selected as the final variables.

To implement and evaluate the algorithms, the research data must be divided into two sets: training data and test data. The training data will be used to build the model, while the test data will be used to evaluate the model built in the training phase. Classification techniques can be compared based on criteria such as accuracy, speed, and robustness. The accuracy of a classification method depends on the number of correct predictions made by the model. Speed refers to the time required to build and use the model for classification, while robustness reflects the model's ability to handle unusual data or missing values.

To evaluate the accuracy and robustness of the models, the research data were randomly split into training and test sets 50 times. In each iteration, 75% of the data was used to train the model, and 25% was used to test and evaluate the model. The average results from 50 iterations of each model were considered the final outcome.

Table 4. Simulation Parameters

Simulation Parameter	Value
Total Dataset Sample Size	1848 firm-years (observations)
Number of Features per Sample (T-test set)	21 features
Number of Features per Sample (Stepwise set)	13 features
Number of Training Samples	1386
Number of Test Samples	462

The classification of a dataset containing two classes is expressed using four possible outcomes: true positive (TP), false negative (FN), true negative (TN), and false positive (FP). Based on these, the accuracy parameter can be defined. Accuracy is the most standard metric for summarizing classification performance across all classes, and it is calculated as follows:

Equation (9): accuracy = (TP + TN) / (TP + TN + FP + FN)

As shown above, to calculate the classification accuracy, the confusion matrix must first be computed. The values on the main diagonal represent the correct classifications by the method used.

Table 5. Confusion Matrix

False Positives (FP)	True Positives (TP)
True Negatives (TN)	False Negatives (FN)

Next, the confusion matrix for each classification method used to predict audit failure (with two different feature sets) is presented.

The results from the Genetic Algorithm (GA), presented in the form of a confusion matrix in Table 4, show that the model’s error rate in identifying companies with audit failure is approximately 2.9% based on the t-test and 1.1% based on the stepwise method. Therefore, the stepwise method is more effective for implementing the GA. The overall accuracy of the GA model shows that GA & Stepwise is more accurate than GA & T-test. Overall, the results from GA & Stepwise indicate that the accuracy of this model in predicting audit failure for companies listed on the Tehran Stock Exchange is approximately 95.3%. This indicates that, with the variables selected through the stepwise method and the GA, companies with audit failure can be identified with a 95.3% probability.

Table 6. Confusion Matrix Results for Predicting Audit Failure (Genetic Algorithm)

GA & Stepwise	Negative	Positive	GA & T-test	Negative	Positive
Positive	0.084	0.989	Positive	0.105	0.971
Negative	0.916	0.011	Negative	0.895	0.029
Accuracy	0.953		Accuracy	0.933	

The results from the Particle Swarm Optimization (PSO) algorithm, presented in the confusion matrix, show that the model’s error rate in identifying companies with audit failure is approximately 3.6% based on the t-test and 2.5% based on the stepwise method. Therefore, the stepwise method is more effective for implementing the PSO algorithm. The overall accuracy of the PSO model shows that PSO & Stepwise is more accurate than PSO & T-test. Overall, the results from PSO & Stepwise indicate that the accuracy of this model in predicting audit failure for companies listed on the Tehran Stock Exchange is approximately 94.5%. This indicates that, with the variables selected through the stepwise method and the PSO algorithm, companies with audit failure can be identified with a 94.5% probability.

Table 7. Confusion Matrix Results for Predicting Audit Failure (Particle Swarm Optimization)

PSO & Stepwise	Negative	Positive	PSO & T-test	Negative	Positive
Positive	0.086	0.975	Positive	0.093	0.964
Negative	0.914	0.025	Negative	0.907	0.036
Accuracy	0.945		Accuracy	0.936	

The results from the Ant Colony Optimization (ACO) algorithm, presented in the confusion matrix, indicate that the model's error rate in identifying companies with audit failure, based on the T-test and Stepwise methods, is approximately 8.1% and 9.6%, respectively. Therefore, considering the variable selection method for implementing the ACO algorithm, the T-test method is more efficient. The overall accuracy results show that CO & T-test is more accurate than CO & Stepwise. In general, the results of CO & T-test indicate that the accuracy of this model in identifying and predicting audit failure in companies listed on the Tehran Stock Exchange is approximately 90.6%. This suggests that, with the variables selected through the T-test method and the ACO algorithm, companies with audit failure can be identified with a probability of 90.6%.

Table 8. Confusion Matrix Results for Identifying Audit Failure (Ant Colony Optimization)

CO & Stepwise	Negative	Positive	CO & T-test	Negative	Positive
Positive	0.102	0.904	Positive	0.108	0.919
Negative	0.898	0.096	Negative	0.892	0.081
Accuracy	0.901		Accuracy	0.906	

The results from the Bee Colony Optimization (BCO) algorithm, presented in the confusion matrix, show that the model's error rate in identifying companies with audit failure, based on the T-test and Stepwise methods, is approximately 5.6% and 8.5%, respectively. Therefore, considering the variable selection method for implementing the BCO algorithm, the T-test method is more efficient. The overall accuracy results show that BCO & T-test is more accurate than BCO & Stepwise. In general, the results of BCO & T-test indicate that the accuracy of this model in identifying and predicting audit failure in companies listed on the Tehran Stock Exchange is approximately 92.8%. This suggests that, with the variables selected through the T-test method and the BCO algorithm, companies with audit failure can be identified with a probability of 92.8%.

Table 9. Confusion Matrix Results for Identifying Audit Failure (Bee Colony Optimization)

BCO & Stepwise	Negative	Positive	BCO & T-test	Negative	Positive
Positive	0.091	0.915	Positive	0.088	0.944
Negative	0.909	0.085	Negative	0.912	0.056
Accuracy	0.912		Accuracy	0.928	

Table 10. Summary of Audit Failure Prediction Results Using Metaheuristic Algorithms

Algorithm	Accuracy (T-test)	T-test FN	Accuracy (Stepwise)	Stepwise FN
GA	0.933	0.029	0.953	0.011
PSO	0.936	0.036	0.945	0.025
CO	0.906	0.081	0.901	0.096
BCO	0.928	0.056	0.912	0.085

The summary results from the metaheuristic algorithms show that the overall accuracy of the GA, PSO, CO, and BCO algorithms is 95.3%, 94.5%, 90.6%, and 92.8%, respectively. This indicates that the Genetic Algorithm (GA) is more effective than other metaheuristic algorithms. Additionally, the overall results from the variable selection methods suggest that the Stepwise method is more efficient. Therefore, in companies listed on the Tehran Stock Exchange, the Stepwise method combined with the Genetic Algorithm (GA) provides the most effective model for predicting audit failure.

4. Discussion and Conclusion

The results of this study demonstrate the effectiveness of metaheuristic algorithms in predicting audit failures in companies listed on the Tehran Stock Exchange. Specifically, the Genetic Algorithm (GA) was shown to have the highest predictive accuracy at 95.3%, followed closely by Particle Swarm Optimization (PSO) at 94.5%, Bee Colony Optimization (BCO) at 92.8%, and Ant Colony Optimization (ACO) at 90.6%. These findings align with previous research indicating the efficacy of metaheuristic approaches in handling complex optimization tasks and predicting financial outcomes (Li et al., 2018; Jiang & Jones, 2018). Additionally, the stepwise method of variable selection was proven to be the most effective in refining the model, further confirming its applicability in predictive modeling contexts [3].

The superiority of the Genetic Algorithm (GA) over other metaheuristic algorithms can be attributed to its evolutionary optimization technique, which iteratively improves solutions based on natural selection processes. Similar results were observed in other fields, such as cloud management and medical diagnostics, where GA outperformed other algorithms in terms of accuracy and reliability [16]. This suggests that GA's adaptability to various data types and its ability to explore multiple potential solutions in parallel makes it an excellent candidate for financial prediction tasks, particularly in audit failure forecasting.

The performance of PSO, which achieved a 94.5% accuracy rate, demonstrates its strength as an optimization technique, particularly in scenarios where the relationships between variables are complex and nonlinear. PSO has been widely recognized for its effectiveness in fields such as financial modeling and data center management, as it can quickly converge on optimal solutions by simulating social behaviors [17]. The success of PSO in this study reaffirms its utility in audit failure prediction, a field where the causes of failure are often interconnected and difficult to model using traditional linear approaches (Chen et al., 2021).

The slightly lower performance of ACO and BCO, while still impressive, may be explained by their inherent nature of slower convergence rates and a greater reliance on exploration rather than exploitation in optimization processes. While these algorithms are well-suited to certain domains such as fraud detection and resource allocation [19], their ability to predict audit failure may be slightly hindered by the need for rapid convergence in highly dynamic financial environments like the Tehran Stock Exchange [15]. Nevertheless, their accuracy rates of over 90% indicate that they are still effective tools for predictive modeling, especially when used in conjunction with other algorithms or methods.

When comparing the results of this study with prior research on audit failure prediction, it is evident that the use of advanced algorithms significantly enhances predictive accuracy compared to traditional methods. For instance, Jin et al. (2011) found that the inclusion of audit quality variables in predictive models improved the ability to forecast bank failures, but their accuracy rates were considerably lower than those achieved by the metaheuristic algorithms used in this study [6]. This difference may be attributed to the algorithms' ability to handle complex, nonlinear relationships, which are typical in financial data, more effectively than traditional statistical models [5].

Moreover, the results of this study align with those from other sectors where metaheuristic algorithms have been applied for failure prediction. For example, in healthcare, metaheuristic methods have been employed to predict diagnostic errors with high accuracy [20]. The success of these algorithms in both healthcare and finance demonstrates their versatility and reinforces the idea that metaheuristic approaches are particularly well-suited to predicting outcomes where the relationships between variables are not straightforward [9].

Despite the promising results, this study is not without its limitations. First, the dataset used in this research is limited to firms listed on the Tehran Stock Exchange, which may restrict the generalizability of the findings. Economic conditions in Iran, such as sanctions and inflation, may have unique impacts on corporate financial health, and these factors may not be present in other contexts (Dong et al., 2019). Therefore, the findings may not necessarily apply to firms in different economic or regulatory environments. Second, while the metaheuristic algorithms demonstrated high accuracy rates, they are computationally intensive and require significant processing power, which could limit their applicability in smaller firms or settings with limited access to computational resources [17].

Additionally, the study did not explore the interpretability of the algorithms. While the accuracy of metaheuristic algorithms is impressive, their results are often difficult to interpret compared to traditional models. This “black box” nature of AI algorithms can be problematic for auditors and regulators who need to understand the rationale behind the predictions to make informed decisions [9]. Finally, while the stepwise method for variable selection proved effective, other methods of feature selection, such as LASSO or Ridge regression, were not explored, potentially limiting the comprehensiveness of the variable selection process [3].

Future research could address the limitations of this study by expanding the dataset to include firms from multiple stock exchanges and countries. A comparative analysis across different economic regions could provide a deeper understanding of how audit failure prediction models perform under varying regulatory, economic, and financial conditions [15]. Moreover, future studies could explore hybrid models that combine the strengths of multiple metaheuristic algorithms, potentially increasing predictive accuracy and reliability. For instance, combining GA with PSO could yield a model that benefits from the exploration capabilities of PSO and the exploitation capabilities of GA [17].

Another area for future research is the interpretability of metaheuristic algorithms. While these algorithms are effective at predicting audit failures, their black-box nature limits their transparency. Incorporating explainable AI techniques could help bridge this gap by making the outputs of these algorithms more understandable to auditors and regulators [9]. Additionally, future studies should investigate the use of alternative variable selection methods, such as LASSO or Ridge regression, to determine whether these techniques could improve the efficiency and accuracy of audit failure prediction models [3].

Finally, future research could explore the application of metaheuristic algorithms in other areas of audit, such as detecting material misstatements or assessing audit quality. These applications would further validate the use of advanced AI methods in auditing and could potentially lead to new insights into the factors that contribute to audit failures [8]. Moreover, exploring the integration of metaheuristic approaches with traditional auditing practices could yield practical tools for auditors, helping them to identify and mitigate risks more effectively [1].

The findings of this study have several practical implications for auditors, regulators, and companies. First, the high accuracy of the GA and PSO algorithms suggests that these methods could be adopted by auditing firms as part of their risk assessment procedures. By integrating these algorithms into their audit processes, firms can improve their ability to detect potential audit failures early, reducing the risk of financial misstatements and enhancing the overall quality of their audits [9]. However, firms should also be aware of the computational demands of these algorithms and ensure they have the necessary infrastructure to support their implementation [17].

Regulators could also benefit from the use of metaheuristic algorithms in audit oversight. Given the high accuracy rates demonstrated in this study, regulatory bodies could use these algorithms to identify firms at high risk of audit failure, enabling more targeted and efficient interventions [3]. This would allow regulators to allocate resources more effectively and focus their efforts on companies that pose the greatest risks to financial markets.

Finally, companies themselves could use these predictive models to assess their own financial health and audit risks. By adopting metaheuristic algorithms, companies can proactively identify areas of financial weakness and take corrective actions before an audit failure occurs. This not only helps to improve the accuracy of financial reporting but also enhances corporate governance and investor confidence [15]. However, companies should also invest in training their staff to understand and interpret the results of these algorithms, ensuring that the outputs are used effectively in decision-making processes [9].

In conclusion, this study demonstrates the significant potential of metaheuristic algorithms in predicting audit failures. By leveraging the strengths of algorithms such as GA and PSO, auditors and companies can improve their risk assessment processes and enhance the overall quality of financial reporting. However, further research is needed to address the limitations of these methods and explore their broader applicability in auditing and financial analysis.

Authors’ Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

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

Ethical Considerations

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

References

[1] D. Barr-Pulliam, H. L. Brown-Liburd, and K.-A. Sanderson, "The Effects of the Internal Control Opinion and Use of Audit Data Analytics on Perceptions of Audit Quality, Assurance, and Auditor Negligence," *Auditing a Journal of Practice & Theory,* vol. 41, no. 1, pp. 25-48, 2021, doi: 10.2308/ajpt-19-064.

[2] T. Bosman, "The Measurement of Audit Quality in the Netherlands: A Practical Note," *Maandblad Voor Accountancy en Bedrijfseconomie,* vol. 95, no. 1/2, pp. 17-31, 2021, doi: 10.5117/mab.95.56820.

[3] M. Elmarzouky, K. Hussainey, and T. Abdelfattah, "Do Key Audit Matters Signal Corporate Bankruptcy?," *Journal of Accounting and Management Information Systems,* vol. 21, no. 3, 2022, doi: 10.24818/jamis.2022.03001.

[4] S. Fafatas, "Auditor Conservatism Following Audit Failures," *Managerial Auditing Journal,* vol. 25, no. 7, pp. 639-658, 2010, doi: 10.1108/02686901011061333.

[5] Y. Jiang and S. Jones, "Corporate Distress Prediction in China: A Machine Learning Approach," *Accounting and Finance,* vol. 58, no. 4, pp. 1063-1109, 2018, doi: 10.1111/acfi.12432.

[6] J. Y. Jin, K. Kanagaretnam, and G. J. Lobo, "Ability of Accounting and Audit Quality Variables to Predict Bank Failure During the Financial Crisis," *SSRN Electronic Journal,* 2011, doi: 10.2139/ssrn.1738483.

[7] F. Adolfi, J. S. Bowers, and D. Poeppel, "Successes and Critical Failures of Neural Networks in Capturing Human-Like Speech Recognition," 2022, doi: 10.48550/arxiv.2204.03740.

[8] D. G. Whiting, J. V. Hansen, J. B. McDonald, C. Albrecht, and W. S. Albrecht, "Machine Learning Methods for Detecting Patterns of Management Fraud," *Computational Intelligence,* vol. 28, no. 4, pp. 505-527, 2012, doi: 10.1111/j.1467-8640.2012.00425.x.

[9] R. J. Chen *et al.*, "Algorithm Fairness in AI for Medicine and Healthcare," 2021, doi: 10.48550/arxiv.2110.00603.

[10] S. Kilic, K. Lopcu, and S. Paksoy, "Artificial Neural Network Models to Build an Early Warning System for Turkish Commercial Banks Before and After the 2001 Financial Crisis," 2014, doi: 10.36880/c05.00963.

[11] F. Salfner, M. Lenk, and M. Malek, "A Survey of Online Failure Prediction Methods," *Acm Computing Surveys,* vol. 42, no. 3, pp. 1-42, 2010, doi: 10.1145/1670679.1670680.

[12] W. Bekker *et al.*, "Isolated Free Fluid on Computed Tomography for Blunt Abdominal Trauma," *Annals of the Royal College of Surgeons of England,* vol. 101, no. 8, pp. 552-557, 2019, doi: 10.1308/rcsann.2019.0078.

[13] T. Cook and S. MacDougall-Davis, "Complications and Failure of Airway Management," *British Journal of Anaesthesia,* vol. 109, pp. i68-i85, 2012, doi: 10.1093/bja/aes393.

[14] L. Dong, R. Sarikas, and A. Djatej, "Partner Rotation or Extended Rotations? The Effect of Confirmation Bias and Motivated Reasoning Bias on Objectivity and Independence: A Framework," *Journal of Global Business Insights,* vol. 4, no. 2, pp. 156-168, 2019, doi: 10.5038/2640-6489.4.2.1070.

[15] A. I. Nour, "The Impact of Corporate Governance Mechanisms on Corporate Failure: An Empirical Evidence From Palestine Exchange," *Journal of Accounting in Emerging Economies,* vol. 14, no. 4, pp. 771-790, 2023, doi: 10.1108/jaee-10-2022-0283.

[16] L. Zhao and K. Sakurai, "On Revenue Driven Server Management in Cloud," 2012, doi: 10.5220/0003901002950305.

[17] X. Li, X. Jiang, P. Garraghan, and Z. Wu, "Holistic Energy and Failure Aware Workload Scheduling in Cloud Datacenters," *Future Generation Computer Systems,* vol. 78, pp. 887-900, 2018, doi: 10.1016/j.future.2017.07.044.

[18] N. Salins, "Utility of Clinical Variables for Deciding Palliative Care in Paraquat Poisoning: A Retrospective Study," *Indian Journal of Critical Care Medicine,* vol. 28, no. 5, pp. 453-460, 2024, doi: 10.5005/jp-journals-10071-24708.

[19] A. F. E. Wuerges and J. A. Borba, "Accounting Fraud: An Estimation of Detection Probability," *SSRN Electronic Journal,* 2011, doi: 10.2139/ssrn.1954783.

[20] J. Abimanyi-Ochom, S. B. Mudiyanselage, M. Catchpool, M. Firipis, S. W. A. Dona, and J. J. Watts, "Strategies to Reduce Diagnostic Errors: A Systematic Review," *BMC Medical Informatics and Decision Making,* vol. 19, no. 1, 2019, doi: 10.1186/s12911-019-0901-1.