Iranian Journal of Finance

Print ISSN 2676-6337
Online ISSN 2676-6345

Fraud Risk Prediction in Financial Statements through Comparative Analysis of Genetic Algorithm, Grey Wolf Optimization, and Particle Swarm Optimization

Zahra Nemati

Ph.D. Candidate, Department of Accounting, Zanjan Branch, Islamic Azad University, Zanjan, Iran. (Email: znemati5879@gmail.com)

Ali Mohammadi *

*Corresponding Author, Assistant Prof., Department of Accounting, Zanjan Branch, Islamic Azad University, Zanjan, Iran. (Email: ali_mohammedi93@yahoo.com)

Ali Bayat

Assistant Prof., Department of Accounting, Zanjan Branch, Islamic Azad University, Zanjan, Iran. (Email: ali.bayat22@yahoo.com)

Abbas Mirzaei

Assistant Prof., Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran. (Email: a.mirzaei@iauardabil.ac.ir)

Iranian Journal of Finance, 2024, Vol. 8, No.1, pp. 98-130.
Publisher: Iran Finance Association
doi: https://doi.org/ 10.30699/IJF.2024.384462.1401
Article Type: Original Article

Received: February 06, 2023
Received in revised form: August 30, 2023
Accepted: February 13, 2024
Published online: March 08, 2024

Abstract

Financial statements are critical to users, as the increasing fraud cases have left behind irreversible impacts. Hence, this study aims to identify the appropriate financial ratios for fraud risk prediction in the financial statements of companies listed on the Tehran Stock Exchange within the 2014–2021 period. The study is based on data from 180 companies listed on the Tehran Stock Exchange, encompassing a total of 1440 financial statements. To select the most appropriate ratios for fraud risk prediction, all financial ratios were tested by three metaheuristic algorithms, i.e., genetic algorithm, grey wolf optimization, and particle swarm optimization. Metaheuristic and data mining methods were employed for data analysis, and these analyses were conducted using MATLAB R2020a (MATLAB 9.8). According to the research results, the fitness function yielded 0.2708 in particle swarm optimization (PSO). With an accuracy of 72.92% after 19 iterations, PSO was more accurate and converged faster than the other algorithms. It also extracted 11 financial ratios: total debts to total assets, working capital to total assets, stock to current asset, accounts receivables to sales, accounts receivables to total assets, gross income to total assets, net income to gross income, current assets to current debt, cash balance to current debt, retained earnings and loss to equity, and long-term debt to equity. The support vector machine (SVM) classifier was then employed for fraud risk detection at companies through the ratios extracted by the proposed algorithms. The accuracy and precision of financial ratios extracted by PSO and SVM were reported at 80,60% and 71,20%, respectively, which indicates the superiority of the proposed model to other models. Considering that the results obtained from the performance evaluation of financial ratios provided by PSO-SVM demonstrate the capability of this method in predicting the likelihood of fraud in financial statements, it can assist financial statement users. By incorporating these ratios about the performance of the target companies and comparing them with those of other companies, users can make more informed decisions in economic decision-making, investments, credit assessments, and more, ultimately minimizing potential losses and risks.

Keywords: Financial ratios, Metaheuristic algorithm, Particle swarm optimization, and Support vector machine.

Introduction

We live in the information age when the capital market is the economy's driving force based on information. Accounting is an information system that produces financial statements for companies, considered the most important sources of information used in a capital market. If this information is accurate, decent, precise, and reliable, it can significantly help users to make investment decisions.

In recent decades, fraudulent financial reporting has been among the hottest topics law enforcers and legal institutions have raised worldwide (Rastatter et al., 2019). Associated with the bankruptcy of large-scale companies, fraud has caused serious concerns about the quality of financial reporting and increased risks and costs of businesses (Khani, 2005). Iran is no exception to the growth and harm of this economic phenomenon. In the past two decades, fraud has been the most frequent topic of discussion in financial markets and socioeconomic institutions in Iran (Raeisi et al., 2018). All over the world, legislators have passed different laws to support fraud prevention, such as the UK Public Interest Disclosure Act 1999, the Australia Corporations Act 2001, and the US Sarbanes Oxley Act 2002. In most developed countries, official organizations also report statistics on the occurrence of fraud and the introduction of fraudulent companies. For instance, the Association of Chief Police Officers exemplifies such organizations in the US. Conducting an analysis of fraud on a global scale twice a year, this association detects fraud and financial scandals. It publishes a comprehensive report of various types of fraud, frequency of fraud, and financial impacts.

The accurate prediction of fraud risk in financial statements will improve the ability to detect and prevent fraud and reduce the high fraud costs. Known as an auditor's responsibility for fraud and fault in auditing financial statements, Iran's Audit Standard 240 urges auditors to consider the concept of fraud in financial statements. However, according to Section 4 of this standard, even if an audit is planned and implemented correctly in accordance with relevant standards, it is complicated but essential to detect fraud, which is usually concealed.

Given the number of TSE-listed companies, membership in the Tehran Stock Exchange in the International Organization of Securities Commission will undoubtedly require improving the quality of financial information and predicting fraud risk in financial statements. Despite the importance of fraud risk detection in financial statements in Iran, no legal institutions can directly analyze and detect cases of financial fraud. Furthermore, there are no databases to disseminate the list of companies that commit fraud in financial statements. In fact, the cases of fraud analyzed in the Tehran Stock Exchange are announced privately but not publicly if judicial courts reach and issue verdicts (Etemadi & Zolfi, 2013).

Also, with advances in technology and high-speed communication networks, fraud methods have become so complicated that it is more accessible to commit fraud but more challenging to detect. In fact, fraudsters now act intelligently and quickly (Sadgali et al., 2019). Hence, detecting fraud is a challenging and complicated but essential task. Thus, studies have gradually started using artificial intelligence techniques rather than conventional methods and statistical analysis due to their reliance on restrictive hypotheses such as normal distribution and high classification error rates (Yao et al., 2019).

Financial statements published by publicly traded companies in the stock market are the most critical source of information in the capital market. This information should be transparent and reliable so that users such as investors, creditors, and managers can make informed economic decisions and allocate their capital effectively. Fraud in financial statements not only leads to resource wastage and the failure to achieve organizational goals but can also trigger economic crises on a macroeconomic scale, create concerns, and erode public trust in accounting, auditing professions, and financial reporting (Rezaei & Reilly, 2010). Additionally, due to the absence of an organization or platform dedicated to identifying companies engaged in fraud and the limitations of statistical methods in fraud detection, given the complexity of fraudulent practices, the question arises: Given the importance of predicting the likelihood of fraud, what are the best variables and methods for detecting fraud in financial statements?

Rarely has research been conducted on using artificial intelligence methods to extract suitable variables for predicting the likelihood of fraud and comparing the performance of these variables. Therefore, this study aims to identify appropriate financial ratios using heuristic methods as predictive variables for fraud risk prediction in financial statements.

Paper Organization

Section 1: Introduction, Objectives, and Research Innovations: In this section, we introduce the research problem, delineate the objectives, and highlight the key innovations driving this study. We establish the significance of our research within the broader context of financial fraud detection.

Section 2: Theoretical Foundations and Research Background This section delves into the theoretical underpinnings of our research. We review relevant literature that provides insights into the relationship between financial leverage and accounting practices, the potential impact of high leverage on fraud-related agreements and borrowing capabilities, and the influence of leverage on earnings management strategies. This section underscores the theoretical framework upon which our research is built.

Section 3: Research Hypothesis Here, we articulate our research hypotheses, which guide our investigation into the relationships between financial ratios, leverage, and financial fraud within company financial statements. We pose critical questions concerning the role of financial leverage in influencing accounting practices and potentially fraudulent activities.

Section 4: Research Methodology This section outlines our research methodology. We describe our data selection process, feature extraction, and feature selection techniques, including the application of metaheuristic algorithms such as Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Genetic Algorithm (GA). Additionally, we explain the classification methods used to predict financial fraud.

Section 5: Research Findings In this section, we present the empirical results of our research. We analyze the predictive performance of our model and provide insights into the key financial ratios that contribute significantly to fraud detection. We also discuss the implications of our findings for practitioners and policymakers.

Section 6: Discussion and Conclusion The final section encompasses the discussion of our results, highlighting their significance in financial fraud detection. We summarize vital takeaways and suggest avenues for future research in this critical domain.

Our paper is structured as follows: In Section 1, we introduce the research problem, objectives, and research innovations, emphasizing the significance within the context of financial fraud detection. Section 2 provides the theoretical foundations and research background, establishing the theoretical framework of our study. Section 3 presents our research hypotheses, guiding our exploration of the relationships between financial ratios, leverage, and financial fraud. Section 4 outlines our research methodology, covering data selection, feature extraction, and using metaheuristic algorithms like Particle Swarm Optimization, Grey Wolf Optimizer, and Genetic Algorithm. We also explain our choice of classification methods for fraud prediction. Section 5 presents empirical findings, analyzing model performance and highlighting critical fraud-detecting financial ratios with implications for practitioners and policymakers. Finally, in Section 6, we discuss our results' significance in the context of financial fraud detection and conclude by summarizing key insights and suggesting future research directions.

Literature Review

Definition of Fraud

According to Aderibigbe and Dada (2007), fraud is intentional deception to deprive others of their properties and rights by planning and implementing fraudulent actions directly or indirectly. The Institute of Chartered Accountants of Nigeria (2006) defined fraud as a deliberate action taken by one or several individuals, such as managers, employees, or third parties, to present false information on financial statements.

The American Institute of Certified Public Accountants (AICPA) defined fraud as an extensive, illegal concept that could be employed to determine whether fraud was intentional or unintentional and to distinguish between fraud and mistake (Kaufmann et al., 2009).

Standard accounting definition of fraud: According to Section 24 of Audit Standards, distortion of financial statements can result from fraud or mistake. Based on this standard, "fraud" denotes any deliberate or deceptive actions taken by one or several individuals, such as managers, employees, or third parties, to gain an illegal advantage. Although fraud is a broad legal concept, what concerns an auditor include fraudulent actions leading to substantial distortion of financial statements (Audit Standards Committee, 2020).

Methods of Fraud

The Association of Certified Fraud Examiners (ACFE) classified cases of professional fraud into three categories: financial corruption, abuse of assets, and fraud in financial statements. The following figures report statistics for the frequency and financial effect of each fraud category:

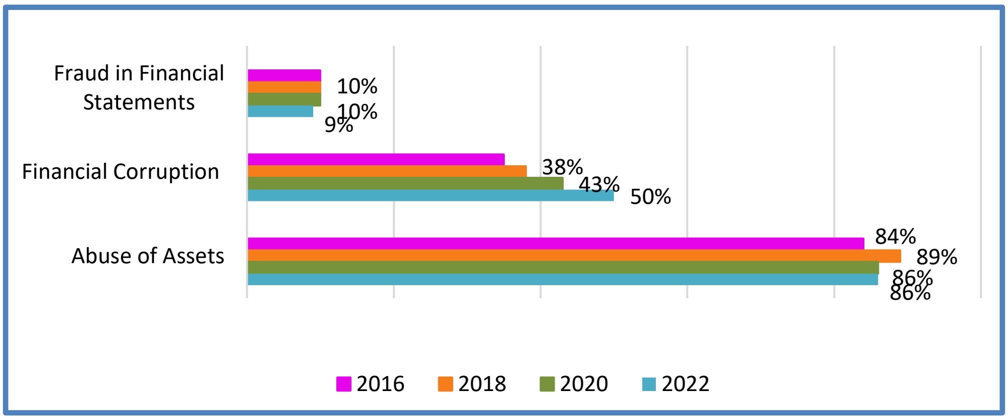

Figure 1. Frequency of fraud categories (Ref.: ACFE reports, 2022)

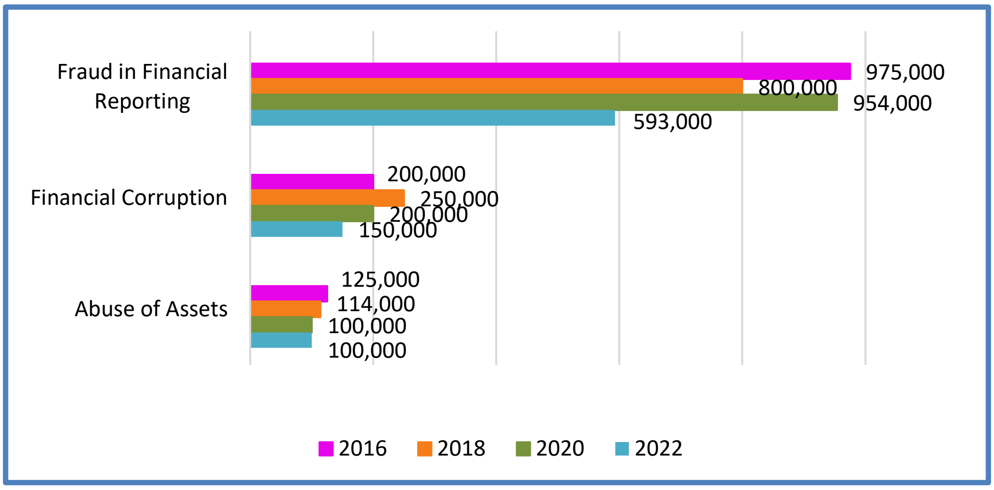

Figure 2. The financial effect of loss caused by fraud categories (Ref.: ACFE reports, 2022)

Financial Corruption: In this category of fraud, the employees of a department take advantage of their positions and influence to gain direct or indirect interests such as commission fees, involvement in the conflict of interests (plans of purchase and sales), and bribery (Sedighi Kamal, 2013).

Abuse of Assets: This category of fraud includes the theft or misuse of assets in an organization. The cases of this category include cash theft techniques, commodity theft, asset theft, and illegal use of assets as personal assets (Vakilifard et al. Ahmadi, 2009).

Fraud in Financial Statements: According to Section 24 of Audit Standards, this category of fraud includes deception attempts such as false documentation, manipulation or distortion of accounting records or backup documents of financial statements, false presentation or deliberate deletion of events in financial statements, and misuse of accounting standards for measurement, identification, classification, presentation, or intentional disclosure.

Financial Frauds Theories

Until now, numerous theories and perspectives regarding the factors contributing to fraud have been proposed, each attempting to elucidate the constituent elements of fraud. These theories span a multidisciplinary spectrum, ranging from financial and administrative sciences to law and psychology (Yazdani-Fazlabadi, 2016). Below, some of the most significant theories on fraud are presented.

Iceberg Theory: This theory emphasizes that detecting financial fraud is akin to spotting the tip of an iceberg, where most fraudulent activities remain concealed beneath the surface. It delves into structural factors such as organizational goals, hierarchy, financial resources, employee skills, and performance metrics. Additionally, it considers behavioral elements like attitudes, norms, and values (Yazdani Fazalabadi, 2016).

Fraud Triangle Theory: Schuessler and Cressey's Fraud Triangle Theory identifies three essential elements for fraud: motivation, opportunity, and rationalization. It posits that fraud is less likely if any of these elements is missing. Motivation arises from financial pressure, while weak internal controls and organizational policies facilitate opportunity. Rationalization is the final step in which the individual justifies fraudulent actions (Schuessler, Cressey, 1950).

Fraud Diamond Theory: Building upon the Fraud Triangle, this theory introduces a fourth element: capability. It considers an individual's capacity, intelligence, self-confidence, and persuasive abilities to engage in fraudulent activities. Economic downturns and organizational crises significantly enhance the capability component (Wolfe, Hermanson, 2004).

Fraud Pentagon Theory: Sorunke extends the Fraud Diamond Theory by introducing ethical characteristics as a fifth element. It argues that assessing an individual's ethical attributes, including their moral code, desires, and indifference to certain behaviors, is crucial for understanding their propensity for fraud (Sorunke, 2016).

Theory of Planned Behavior: This psychological theory focuses on intentions and attitudes as precursors to behavior. It suggests that individuals are more likely to engage in a specific behavior if they have a positive intention, a favorable attitude, and perceive control over the behavior. In the context of fraud, these factors can be instrumental in predicting fraudulent actions (Beck, Ajzen, 1991).

Financial Ratios

In human societies, the necessity of financial information has always caused intellectual challenges in providing decision-makers with helpful information. In this case, attempts have been made to find the correlation and proportion of different types of financial information. With the massive increase in the amount and flow of financial information in the 1890s, financial statements were presented first based on demands and necessary items. After that, they were presented in comparative columns. Since the 19th century, they have finally been extracted and presented as financial ratios to acquire significant information from financial statements. In fact, financial ratios are used in practical areas to analyze corporate performance, compare yields in different years, predict a company's future success and performance, and compare different companies in research studies to design and propose various models (Zare Bahmanmiri & Malekian, 2016). According to many Iranian and non-Iranian studies (e.g., Kamrani and Abedini (2022), Cheng, Kao and Lin (2021), Rezaei, Nazemi Ardakani, and Naser Sadr Abadi (2020), Tashdidi, Sepasi, Etemadi, and Azar (2019), Omidi, Min, Moradinaftchali, and Piri (2019), Jan (2018), Omar, Johari, and Smith (2017), Ebrahimi and Khajavi (2017), Zare Bahmanmiri and Malekian (2016)), financial ratios are capable of describing the importance of corporate features concerning significant events such as fraud. This study used financial ratios as the independent variables or fraud predictors of financial statements.

Financial ratios are classified into the following categories. To classify financial ratios into the categories mentioned in this paper, we have drawn upon established literature in financial management. The categorization process has been informed by authoritative texts, including "Modern Financial Management," translated by Dr. Ali Jahan Khani and Mojtaba Shoori. Additionally, "Financial Management," authored by Dr. Reza Tehrani. Moreover, in "Financial Management," authored by Dr. Ahmad Ahmadpour and Dr. Mahmoud Yahyazadeh. These sources provide comprehensive insights into the classification and analysis of financial ratios, contributing to the development of our research.

Ratios of Short-Term Affordability Evaluation or Liquidity Ratios: As their name suggests, these ratios provide information regarding a company's liquidity status and indicate its ability to fulfill short-term financial commitments. According to St. Piero Anderson (1984), current assets of fraudulent companies result incredibly from their inventories and accounts receivable (Persons, 2011).

Ratios of Long-Term Affordability Evaluation or Leverage Ratios: Measuring the ability of a company to fulfill long-term commitments, these ratios indicate how much debt was used in financing and capital structure. Christie (1990) demonstrated a positive correlation between leverage and income-increasing accounting behaviors. Pearson's (2011) perspective suggests that higher leverage may sometimes be associated with a higher potential for covenant violations and a reduced ability to obtain additional capital through borrowing. Lazim and Jalani (2017) asserted that increased financial leverage positively influences income management behaviors based on accruals. Additionally, due to the potential wealth transfer from debt holders to managers associated with higher financial leverage, some managers may be motivated to manipulate financial statements to meet debt covenants. Charalambous (2002) pointed out that higher debt levels may increase the likelihood of fraud. Brazil and colleagues (2009) and Alden et al. (2012) believed in the impact of financial leverage on fraud in financial statements."

Ratios of Asset Management or Ratios of Working Capitals (Efficiency): These ratios can be used to measure the efficiency of a company in employing assets (Dalnial, 2014). Kirkos, Spathis, and Manolopoulos (2007) reported significant differences between fraudulent and non-fraudulent companies in the mean ratio of working capital.

Profitability Ratios: These ratios are employed to measure the affordability of a company to generate income and acquire earnings (Jahankhani & Shoori, 2013). According to Kreutzfeldt & Wallace (1986), businesses with lower earnings are more significantly likely to commit fraud in financial statements than businesses with higher earnings.

Metaheuristic Algorithms

Metaheuristic algorithms try to find the best solution of all the possible solutions to an optimization problem. They are nature-inspired algorithmic frameworks designed to offer a suboptimal solution to an optimization problem. In particular, these algorithms are preferable when the available information is incomplete and the computing capacity is limited (Bianchi et al., 2009).

Genetic Algorithm: The genetic algorithm (GA) is a method of search and optimization developed in selection principles and natural genetics (Xu & Wunsch, 2005). This algorithm benefits from biological evolution techniques (e.g., inheritance, biological mutation, and Darwinian selection principles) to find an optimal model prediction or adaptation formula. According to Darwin's theory of gradual evolution, only individuals with superior genes can survive and reproduce new offspring. In contrast, individuals with inferior genes will be eliminated in the survival of the fittest. This process continues until the acquisition of an optimal solution or the end of runtime (Taghavi & Nobari, 2006).

Grey Wolf Optimization: GWO is a metaheuristic algorithm inspired by grey wolves' hierarchical structure and social behavior while hunting. This population-based algorithm has a simple process that can easily be generalized to problems with higher dimensions.

This algorithm consists of three significant steps:

1) Observing, tracking, and pursuing a prey
2) Approaching, enclosing (encircling), and confusing the prey until it stops moving
3) Attacking the prey (Mirjalili & Lewis, 2014)

Particle Swarm Optimization: In the early 1990s, many studies were conducted on the social behavior of animal groups. Inspired by those studies, Eberhart and Kennedy (1995) introduced the particle swarm optimization (PSO) algorithm to optimize nonlinear continuous functions. This algorithm was inspired by simulating the behavior of a flock of flying birds. It can solve a wide range of optimization problems (El-Shorbagy & Hassanien, 2018). Consider a flock of birds hunting randomly for food in a space with only one piece of food. The birds are still determining where the food is. Pursuing the bird that is in the shortest distance from food can be one of the best strategies. This strategy is the central theme of this algorithm. Referred to as a particle, each solution in this algorithm represents a bird in the flock of flying birds. A particle has a fitness value calculated by a fitness function. It also has a velocity responsible for directing this particle's moves. Each particle continues moving in the problem space by following optimal particles in the current state. In other words, some particles are first generated randomly. They then try to find an optimal solution by updating generations (Nath et al., 2014).

The Particle Swarm Optimization (PSO) algorithm is one of swarm intelligence's most essential optimization methods. This algorithm is inspired by simulating the collective behavior of birds in flight. PSO can be used to solve a wide range of optimization problems effectively.

In our approach, the Particle Swarm Optimization (PSO) algorithm selection was driven by its inherent features, such as its ability to explore solution spaces efficiently and converge quickly, making it a superior choice compared to alternative metaheuristic algorithms.

Particle Swarm Optimization (PSO) is considered an evolutionary computational algorithm that relies on swarm intelligence and is a population-based optimizer. The optimization process in PSO begins with randomly initializing a set of potential solutions. Then, it iteratively searches for the optimum. The algorithm finds the best position by following the best particles. Compared to other evolutionary algorithms, PSO has intelligent characteristics.

Due to the advantages of Particle Swarm Optimization, it is suitable for scientific research. In this group, each member has a velocity vector and a position vector in the search space. In each iteration, the new positions of particles are updated based on their current velocity vectors, the best position found by that particle, and the best position found by the best particle in the group. This updating process allows the particles to explore the search space and converge toward the optimal solution over time.

Let us assume that the search space is a d-dimensional space. The position and velocity of particle i in the search space can be represented as follows:

$X_i = (x_{i_1}, x_{i_2}, x_{i_3}, \dots, x_{i_d})$ Position Vector X_i (1)

$V_i = (v_{i_1}, v_{i_2}, v_{i_3}, \cdots, v_{i_d})$ The velocity of ith particle with the velocity V_i is as velocity vector (2)

To describe the best position that particle $P_{i.best}$ finds, you can use equation 3 and denote it as

$$P_{i.best} = (p_{i_1}, p_{i_2}, p_{i_3}, \dots, p_{i_d}) \quad (3)$$

The best position found by the best particle among all particles in the swarm, denoted as $P_{i.best}$, can be defined as follows in equation 4:

$$P_{g.best} = (p_{g_1}, p_{g_2}, p_{g_3}, \dots, p_{g_d}) \quad (4)$$

In each iteration, the velocity and position of each particle will be updated based on their own previous best position and the best position among the entire population. In the PSO algorithm, there are three functions for updates: (V) velocity, (X) position, and (W) inertia weight, which are updated according to the following equations.

$$V_{i,j}(t+1) = w.v_{i,j}(t) + C_1.R_1.\left(pbest_{i,j} - x_{i,j}(t)\right) + C_2.R_2.\left(gbest_{i,j}(t) - x_{i,j}(t)\right) \quad (5)$$

$$X_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1) \quad (6)$$

$$W(t+1) = wmax - \left(\frac{wmax - wmin}{t_{max}}\right).(t+1) \quad (7)$$

t represents the number of iterations. $v_{i,j}$ is the velocity of the i-th particle, and $x_{i,j}$ is the position of the i-th particle. R_1 and R_2 are two random numbers from 0 to 1. C_1 is a constant acceleration coefficient (towards the best path found by the particle), and C_2 is another constant acceleration coefficient (towards the best path found among the entire population). W is a positive inertia weight coefficient (indicating the influence of the velocity vector from the previous iteration, $V_i(t)$, on the velocity vector in the current iteration, $V_i(t+1)$. To prevent excessive speed during the movement of a particle from one location to another (avoiding vector velocity divergence), velocity changes are limited to the range between V_min and V_max, where V_min≤V ≤ V_max. The upper and lower speed limits will be determined based on the type of problem. The convergence test in this algorithm is as follows: By default, a predefined number of iterations is set at the beginning, and each stage, it checks if the number of iterations has reached the specified value. If the number of iterations is less than the initial setting, it needs to return to stage 2; otherwise, the algorithm terminates. Additionally, if there has not been any change in the cost of the best particle over several consecutive iterations, for example, 20 iterations in a row, then the algorithm terminates. In this case, you must return to stage 2 if the termination condition is unmet.

The algorithm assumes that each solution to the problem is considered a particle, and at each stage, a set of solutions is represented as particles. The figure below illustrates the movement of a particle in the vector space under the influence of the displacement vector and particle velocity.

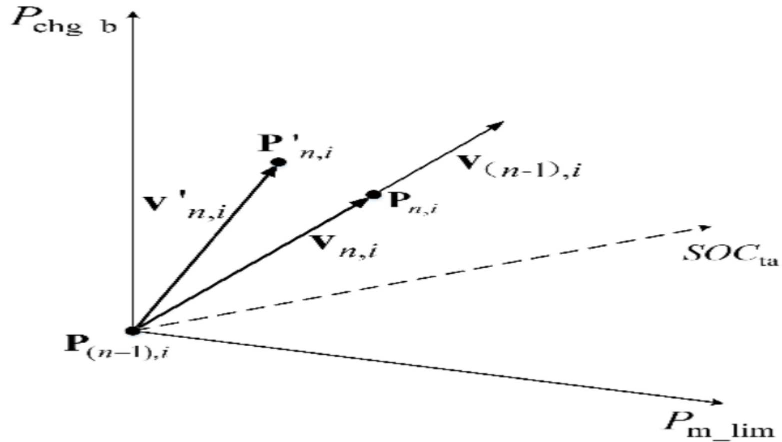

Figure.3 The Effect of Velocity on Particle Position (Zero and Banirostam, 2018)

As shown in the figure above, the new velocity vector of a particle is updated based on its current velocity vector, the desired position of the best member of the population, and the previous desired position of the particle.

The flowchart of the Particle Swarm Optimization algorithm is depicted in the figure below.

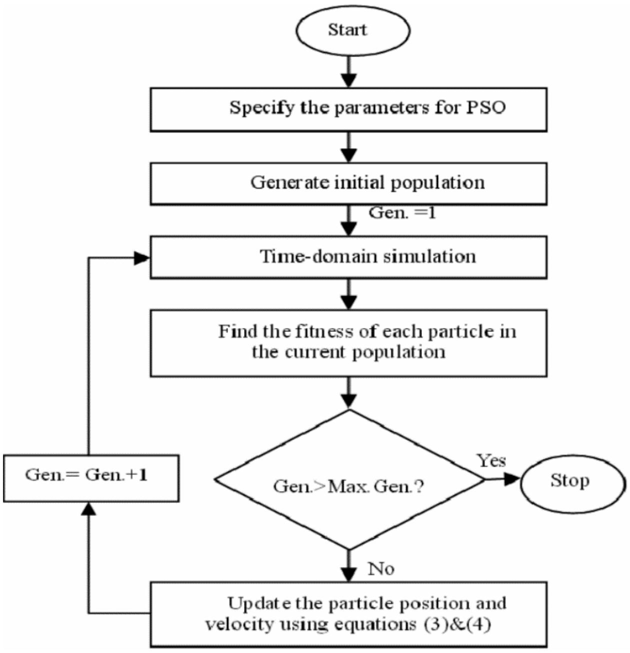

Figure 4 Flowchart of the Particle Swarm Optimization Algorithm (Clerc, 2010)

Support Vector Machine

The support vector machine (SVM) is a two-class classification method with supervised learning. Introduced by Vapnik (1995) based on the statistical learning theory, it can be employed to solve classification or regression problems. This algorithm draws some hyperplanes in the space to optimally distinguish between different data samples. In fact, these hyperplanes differentiate between two groups so that they have the longest distance from the closest points from each group. The best hyperplane is the plane with the longest distance from both groups. This method classifies data by finding the best hyperplanes that separate a group's data from another group's data (Pradhan, 2012).

Research Background

Recent advancements in fraud detection have witnessed significant breakthroughs, employing cutting-edge technologies and methodologies. Ali et al. (2023) introduced a state-of-the-art fraud detection model utilizing the robust XGBoost algorithm. Their study focused on Middle Eastern and North African (MENA) companies and effectively addressed class imbalance issues using the Synthetic Minority Over-sampling Technique (SMOTE). Their findings demonstrated exceptional accuracy, with the XGBoost method achieving an impressive 96.05%.

Similarly, Lei et al. (2023) presented a comprehensive AI-driven methodology to mitigate corporate financial risk proactively. Their four-step approach encompasses data preprocessing, feature selection, categorization, and parameter optimization. Leveraging innovative techniques like the Chaotic Grasshopper Optimization Algorithm (CGOA) and Support Vector Machines (SVM), their model achieved an accuracy rate of 85.38%, showcasing superior predictive and decision-making capabilities.

Chen (2023) contributed to the field by evaluating machine learning models, including Random Forest, Gradient Boosting Decision Trees (GBDT), XGBoost, and LightGBM, to develop a robust financial statement fraud detection system for public businesses. This research also introduced an integrated feature selection technique and effectively tackled the issue of imbalanced data distribution, notably enhancing fraud identification. Among the models assessed, GBDT exhibited outstanding performance in terms of AUC and sensitivity.

Aftabi et al. (2023) proposed an innovative approach based on Generative Adversarial Networks (GANs) for fraud detection. Their method showcased remarkable effectiveness by compiling a new dataset from the annual financial statements of Iranian banks and extracting three types of features. Notably, their approach outperformed other classification techniques, achieving an accuracy of 99% with XGBoost and 100% with SVM in generating synthetic suspected fraud samples.

Kamrani and Abedini (2022) extracted two nonfinancial ratios and 19 financial ratios by conducting a literature review, using snowball sampling, and interviewing experts. They then used an artificial neural network and a support vector machine to predict and detect fraud. According to the results, the support vector machine outperformed the artificial neural network with a prediction power of 86%.

Rezaei, Nazemi Ardakani, and Naser Sadr Abadi (2020) used 41 financial and nonfinancial variables in a Bayesian network, a decision tree, a neural network, a support vector machine, and a combinatorial method for fraud risk detection. Their results indicated that the combinatorial method outperformed the other techniques with a prediction rate of 96.2% and a higher evaluation ability.

Cheng, Kao, and Lin (2021) employed data preprocessing techniques in addition to feature selection of missing values, management of unbalanced classes, merged features, and distance correlation for feature selection with four classifiers of neural network, decision tree, extra trees, and random forest. They reduced 72 financial ratios to 18 ratios. According to the results, the 18 ratios selected by features merged with the random forest classifier yielded an accuracy of 98.92%, higher than those of other methods.

Omidi, Min, Moradinaftchali, and Piri (2019) used five supervised methods, i.e., feedforward multilayer neural network, probabilistic neural network, support vector machine, polynomial linear, logarithmic model, and differential analysis, with 18 financial data for fraud risk prediction in financial statements. Their results indicated that the feedforward multilayer neural network outperformed the other methods in fraud risk detection with an accuracy above 90% in financial reports.

Tashdidi, Sepasi, Etemadi, and Azar (2019) selected 23 financial ratios with available information in Iran to propose a novel approach for fraud risk detection in financial statements by searching empirical evidence. Using the cross-entropy method, they then extracted 16 ratios as the best and most effective ratios. Moreover, they employed logistic regression, genetic algorithm, and artificial bee colony algorithm to classify companies as fraudulent and non-fraudulent categories. According to their results, the artificial bee colony algorithm outperformed the other methods in fraud risk prediction with an accuracy of 82.5%.

Jan (2018) used a neural network and a support vector machine to extract appropriate variables for fraud prediction from 22 financial and nonfinancial variables within 11 years. They obtained 10 and 3 variables from the neural network and the support vector machine, respectively. They also employed four decision tree techniques (i.e., CHAID, CART, C5.0, and QUEST) to analyze the accuracy of fraud risk detection in financial statements. According to their results, ten variables extracted by the artificial neural network and classification with the CART decision tree yielded the highest accuracy of fraud risk detection (90.21%) in financial statements.

Omar, Johari, and Smith (2017) predicted fraud risk in financial reports through logistic regression, support vector machine, multiple-criteria decision analysis, and artificial neural networks. They used ten financial ratios for fraud risk prediction. The results indicated that the artificial neural network outperformed the other methods in fraud risk prediction with an accuracy of 94.87%.

Ebrahimi and Khajavi (2017) used the correlation-based feature selection method to select the variables having the most significant effects on fraud risk detection in financial statements. For this purpose, they selected 40 financial and nonfinancial variables. The research results indicated the usefulness of the cash ratio, interest coverage ratio, accounts receivable to total assets, inventory to net sales ratio, the natural logarithm of sales, ratio of net income to sales, and ratio of current assets to total assets. They also adopted data mining methods (e.g., artificial neural network, Bayesian network, and random forest) for fraud risk prediction. According to their results, the random forest algorithm outperformed the other techniques in fraud risk prediction with an accuracy of 96.77%.

Zare Bahmanmiri and Malekian (2016) proposed a financial fraud risk prediction model. They conducted stepwise regression and elastic net tests in two steps in MATLAB. They selected seven financial ratios: the ratio of working capital to an asset, the ratio of accounts receivable to sales, the ratio of cash to current debt, the ratio of inventory to a current asset, the ratio of debt to equity, the ratio of gross income to asset, and absolute value of changes in current ratio. The logit test results indicated that 64.04% of the estimated model could be predicted.

These recent studies underscore the rapid evolution of fraud detection techniques, emphasizing the critical role of advanced algorithms and data-driven methodologies in effectively identifying fraudulent activities. They provide valuable insights that inform the current research and its contribution to this dynamic field.

Research Hypothesis

After the theoretical foundations and research background were reviewed, the research hypothesis was presented below:

The particle swarm optimization algorithm generated more significant financial ratios than the genetic and grey wolf optimization algorithms for fraud risk prediction.

The Particle Swarm Optimization (PSO) algorithm, due to its inherent characteristics of parallelism, population-based search, and local exploration, demonstrates superior capabilities and better performance in feature reduction and optimal variable extraction compared to genetic algorithms and Grey Wolf Optimization (GWO) algorithms. Therefore, based on this premise, the above hypothesis was formulated and investigated in general terms.

Research Methodology

From a results and outcomes perspective, this research is considered practical as it seeks to solve the practical problem of fraud detection. In terms of its objective can be described as descriptive-correlational because it aims to describe the relationships between elements in financial statements and fraud. The process is quantitative, relying on numerical data from companies' financial statements. Regarding the temporal dimension, it is retrospective because it utilizes historical financial statement data from past years.

Statistical Population and Research Sample

The statistical population included the companies listed on the Tehran Stock Exchange. The systematic conditional sampling method was employed to select the research sample. The fiscal years of companies were expected to end on March 20 (or March 21), and they were not supposed to be financial intermediaries such as investment companies, holdings, banks, and insurance companies. The necessary data of research variables should be available. Based on these conditions, 180 companies were selected.

Data Collection Tools

Both Iranian and foreign studies (e.g., books and papers) were reviewed through notes in a library method to collect the necessary data regarding theoretical foundations and research background. The necessary variables data were collected from financial statements and reports provided by independent auditors and authorized inspectors and published by the Tehran Stock Exchange. Rahavard Novin Software Suite and MS Excel were also employed for essential calculations. Moreover, metaheuristic and data mining methods were in MATLAB for data analysis and hypothesis testing. Metaheuristic and data mining methods were employed for data analysis, and these analyses were conducted using MATLAB R2020a (MATLAB 9.8).

Research Variables and Models

Dependent Variable: To define and detect fraud in financial statements as the dependent variable, Audit Standard 240, entitled Auditor's Responsibility, was reviewed along with the theoretical foundations of domestic and foreign studies regarding fraud to extract the most critical cases of fraud:

1) Overestimating and underestimating incomes and assets
2) Overestimating and underestimating costs and debts

In the paragraphs about the fundamentals of the auditor's opinion regarding the audit of financial statements, if there is a condition clause regarding the inventory being stagnant, where no allowance has been made for a reduction in value in the financial statements, discrepancies exist in the confirmations received from counterparties for receivables and payables (whether favorable or unfavorable), incorrect calculations related to accrued end-of-service benefits and taxes, or non-compliance with or incorrect application of the relevant standards for reevaluating fixed assets, investments, depreciation calculations, and so forth, it indicates either an understatement or overstatement of assets, liabilities, income, or expenses, depending on the nature of the event and the specific item.

3) Restated financial statements and significant yearly moderations
4) Tax differences from tax areas and insufficiency of savings for performance tax:

In the event of a tax dispute with tax authorities and insufficient provisions for tax liabilities in the financial statements, the independent auditor, based on the results of their examinations of the company's financial statements, will explicitly reference this issue in their audit report as an adjusted opinion (qualified opinion, disclaimer of opinion, or an adverse opinion) under one of the condition clauses in the fundamentals of the auditor's opinion.

5) Stagnant assets and items, such as inventory
6) The assumption of a company’s nonstop activity for several consecutive periods is doubted, and an auditor’s statement is conditional. However, the company is still supposed to present financial statements based on the continuity of its activities. For instance, consider a company where production was stopped two years ago with no sales.
7) Misuse of accounting standards for identification, measurement, classification, presentation, and disclosure

Suppose a company makes an error in applying a standard at any stage (identification, measurement, classification, presentation, or disclosure). In that case, the independent auditor will explicitly refer to these errors as condition clauses in the fundamentals of the auditor's opinion in the audit report on the company's financial statements.

Some Iranian studies (eg., Kamrani & Abedini (2022); Tashdidi et al. (2019), Ebrahimi and Khajavi (2017), Zare Bahmanmiri and Malekian (2016) & Kazemi (2016)) have confirmed the relationships between fraud cases and auditor statements. Hence, the paragraphs on condition and the other paragraphs of audit reports of companies with moderated statements (i.e., rejected statements, lack of statements, and conditional statements) were analyzed thoroughly. Of 1440 fiscal years (180 companies in 8 years), 532 fiscal years were identified as suspiciously fraudulent, whereas 908 were identified as non-fraudulent. The suspiciously fraudulent companies were represented by 1, whereas the non-fraudulent companies were represented by 0.

Independent Variable: Financial ratios were used as the independent or fraud predictor of financial statements in this study. After a review of the literature and theoretical foundations, financial ratios were extracted and classified into four categories: liquidity, leverage, efficiency, and profitability. Some of the similar and inverted ratios were excluded in the initial analysis. Finally, 96 financial ratios remained.

Results

In fraud detection, many data collection techniques are used for early detection, usually characterized by many distinct features and a small number of cases. Achieving a reliable result (e.g., classification of fraudulent and non-fraudulent reports) requires appropriate ratio of sample size to the number of features. Hence, feature selection is necessary for complicated problems such as fraud detection. In addition, feature selection can improve the classification accuracy due to data dimensionality reduction. Many feature selection methods have been developed so far, such as filter, wrapper, and embedded techniques, among which wrapper techniques can yield better accuracy and can be integrated with different optimization tools, such as metaheuristic search, for searching an optimal subset with high computational complexity. Metaheuristic methods were introduced to solve the feature selection problem because non-deterministic polynomial (NP)1 It is the best solution to many problems. Metaheuristic algorithms can select the optimal feature subset that classifies the results more accurately (Begum et al., 2019).

The genetic algorithm, grey wolf optimization algorithm, and particle swarm optimization algorithm were used as metaheuristic methods of nature-inspired evolutionary algorithms to select a subset of features (i.e., financial ratios). The results were saved as MATLAB tables.

[1] In the computational complexity theory, NP stands for non-deterministic polynomial and represents one of the most fundamental classes. It refers to the runtime. In many typical problems of computer science, especially the decision-making model, search and optimization problems are considered NP problems.

Selecting a Feature Subset Based on Metaheuristic Algorithms

According to the following figure, the fitness functions of GA, GWO, and PSO increased in the problem of selecting subsets of features by increasing iteration to an optimum with an error rate of zero.

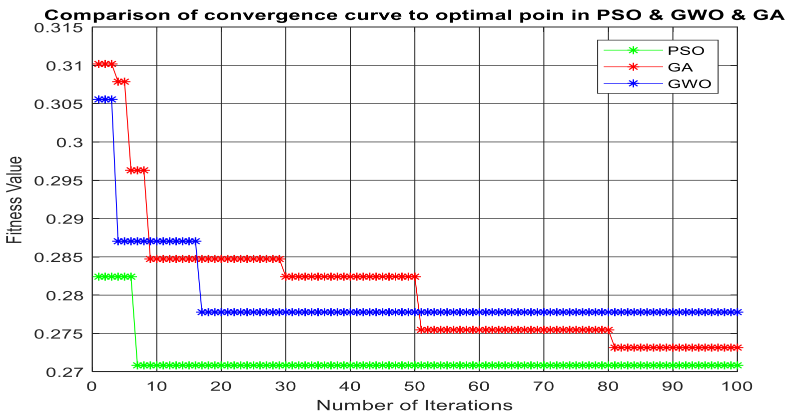

Figure 5. Convergence of fitness functions to an optimum in the proposed algorithms (Ref.: research findings)

After 100 iterations, the fitness function ($F = Max\ f(x) = \sum_{i=1}^{M} s_j x_j - p \times m\ \ x_j \in \{0,1\}$) was employed to obtain the information regarding the best financial ratios selected by each algorithm. The following table presents the information, including the value of the fitness function, accuracy, and iteration for extracting the best financial ratios:

Table 1. Results of extracting financial statements in each algorithm

Algorithm	Fitness Function	Accuracy	Iteration
GA	0.2917	70.83%	89
GWO	0.2940	70.60%	31
PSO	0.2708	72.92%	19

(Ref.: research findings)

In light of the obtained results, it becomes evident that the primary hypothesis, focusing on selecting optimal financial ratios, has been rigorously validated. Specifically, the particle swarm optimization algorithm has outperformed competing algorithms such as grey wolf optimization and genetic algorithms. This achievement is significant as it has facilitated the identification of more meaningful financial ratios for predicting the likelihood of financial fraud. Tables 2 to 4 report the financial ratios extracted by each of the algorithms after 100 iterations:

Table 2. Financial ratios selected by GA

Financial Ratio	Frequency of Feature in 30 Iterations	Financial Ratio	Frequency of Feature in 30 Iterations
Total debts to total assets	22	Total debts to equity	17
Working capital to total assets	22	Retained earnings and loss to equity	29
Accounts receivable to sales	30	Inventory turnover	13
Accounts receivable to total assets	30		

(Ref.: research findings)

Table 3. Financial ratios selected by GWO

Financial Ratio	Frequency of Feature in 30 Iterations	Financial Ratio	Frequency of Feature in 30 Iterations
Total debts to total assets	18	Gross income to total assets	22
Net income to total assets	17	Cash balance to total assets	17
Working capital to total assets	21	Net income to gross income	18
Accounts receivable to sales	23	Retained earnings and loss to equity	22
Instant asset to current debt	15		

(Ref.: research findings)

Considering the values obtained for accuracy, sensitivity, specificity, and the F1-score in the table above, the second hypothesis, which pertains to the support vector machine classifier compared to other classifiers such as k-nearest neighbors and Bayesian networks, has been substantiated. It is now evident that the support vector machine classifier is more effective in predicting the likelihood of financial fraud, as supported by the superior performance indicated in the evaluation metrics.

Table 4. Financial ratios selected by PSO

Financial Ratio	Frequency of Feature in 30 Iterations	Financial Ratio	Frequency of Feature in 30 Iterations
Total debts to total assets	26	Net income to gross income	17
Working capital to total assets	22	Current assets to current debt	15
Inventory to current asset	23	Cash balance to current debt	16
Accounts receivable to sales	27	Retained earnings and loss to equity	23
Accounts receivable to total asset	25	Long-term debt to equity	18
Gross income to total assets	29		

(Ref.: research findings)

Learning techniques were first trained to analyze and evaluate the financial ratios extracted from the proposed algorithms. For this purpose, 70% of data (i.e., 1008 data including 376 data from suspiciously fraudulent companies and 632 data from non-fraudulent companies) were given as training data to the proposed MATLAB algorithms to determine each model's training percentage. Finally, 30% of the data (i.e., 432 data including 156 data of suspiciously fraudulent companies and 276 non-fraudulent companies) were used as test data in MATLAB to evaluate the proposed algorithms and analyze fraud risk prediction in financial statements.

Results of Evaluating Financial Ratios in Fraud Risk Prediction

The following criteria were employed to evaluate the financial ratios extracted from each algorithm in fraud risk prediction:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{TP+TN+FP+FN} \quad (8)$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (9)$$

$$Recall = \frac{\mathrm{TP}}{TP+FN} \quad (10)$$

$$F - measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (11)$$

Tables 5 to 7 present the brief results of financial ratios extracted by the proposed algorithms and the SVM classifier for fraud risk prediction in financial statements through the abovementioned evaluation criteria in 30 iterations with testing data:

Table 5. Brief results of GA with SVM classifier

Statistics	Accuracy	Precision	Recall	F-Measure
Mean	0.7647	0.6934	0.6248	0.6572
Max	0.7755	0.7092	0.6410	0.6734
Min	0.7430	0.6666	0.5769	0.6185
SD	0.0079	0.0115	0.0173	0.0133

(Ref.: research findings)

Table 6. Brief results of GWO with SVM classifier

Statistics	Accuracy	Precision	Recall	F-Measure
Mean	0.7583	0.6680	0.6588	0.6631
Max	0.7755	0.6928	0.6795	0.6860
Min	0.7338	0.6242	0.6218	0.6346
SD	0.0112	0.0190	0.0184	0.0143

(Ref.: research findings)

Table 7. Brief results of PSO with SVM classifier

Statistics	Accuracy	Precision	Recall	F-Measure
Mean	0.8060	0.7120	0.7773	0.7432
Max	0.8148	0.7236	0.7884	0.7546
Min	0.7824	0.6782	0.7500	0.7151
SD	0.0076	0.0110	0.0130	0.0098

(Ref.: research findings)

Upon examining the values obtained for evaluation metrics such as accuracy, sensitivity, specificity, and the F1-score in the tables above, in comparison to numerical analyses, the research hypotheses outlined in the study have been further corroborated. It is now evident that using a combined classification method can effectively reduce misclassifications in financial ratios compared to other classification techniques. This approach has demonstrated its efficacy and superiority over alternative classification methods in reducing false positives and improving prediction accuracy, thus confirming our research hypotheses.

Results of Analysing Models Based on Confusion Matrix

The number of classes in a confusion matrix depends on the number of rows and columns. There are two classes (i.e., suspiciously fraudulent companies and non-fraudulent companies) in this study; hence, the confusion matrix consists of the following elements:

True Positive (TP): The suspiciously fraudulent financial statements identified correctly

False Positive (FP): The suspiciously fraudulent financial statements identified wrongly as non-fraudulent

True Negative (TN): The non-fraudulent financial statements identified correctly

False Negative (FN): The non-fraudulent financial statements identified wrongly as suspiciously fraudulent

Table 8 presents the confusion matrix results obtained from the proposed algorithms and the SVM classifier through test data:

Table 8. Test data Confusion matrix results proposed models with SVM classifier

Algorithm	GA	GWO	PSO
TP Mean	100	106	123
TN Mean	235	229	229
TP Mean + TN Mean	335	335	352
FP Mean	41	47	47
FN Mean	56	50	33
FP Mean + FN Mean	97	97	80

(Ref.: research findings)

Results of Analyzing Precision and Error in Predicting Financial Ratios

The following table reports the results of analyzing precision and error in predicting financial ratios extracted by GA, GWO, and PSO with an SVM classifier through test data.

Table 9. Results of precision and error in proposed models with SVM classifier

Algorithm	Prediction Precision			Prediction Error		
	GA	GWO	PSO	GA	GWO	PSO
SVM	0.6934	0.6680	0.7120	0.3066	0.3320	0.2880

(Ref.: research findings)

The area under the Curve of ROC

The receiver operating characteristic (ROC) curve demonstrates the 2D presentation of results yielded by the proposed methods. The x-axis and the y-axis represent the values of TP and FP, respectively. In this method, a common criterion is to calculate the area under the curve of the ROC.

The efficiency of each algorithm was determined through the SVM classifier based on the values of accuracy, recall, precision, TP, and FP in the ROC. The areas under the ROC curve were 74,70%, 80,30%%, and 83,03% for GA–SVM, GWO–SVM, and PSO–SVM, respectively. The results indicate that the proposed models predicted suspiciously fraudulent and non-fraudulent companies.

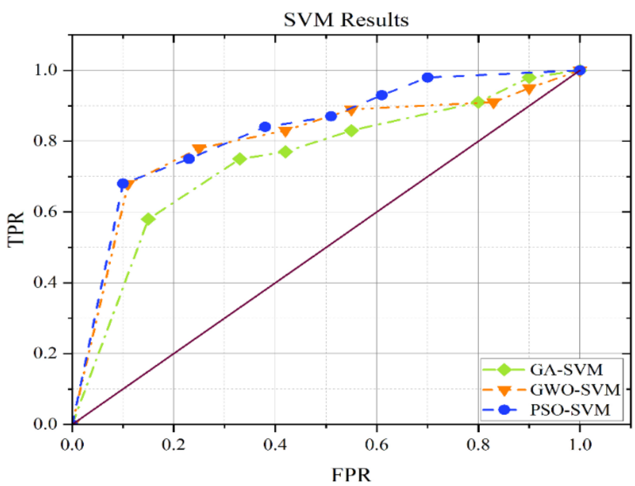

Figure 6. Efficiency evaluation of algorithms (Ref.: research results)

Discussion and Conclusion

Despite the importance of financial fraud and its detrimental effects on the economy, the fraud risk always exists and never leaves organizations alone. Given recent IT advances, methods of fraud have become more complicated and faster. Moreover, it is impossible to determine the independent variables affecting fraud risk prediction in financial statements. Since statistical methods can better predict continuous linear data than discrete nonlinear data

(Ranganathan et al., 2017), they cannot predict fraud risk correctly. Hence, it is essential to develop methods of fraud risk detection in financial statements. Metaheuristic methods were used in this study to extract appropriate financial ratios for fraud prediction in financial statements. Table 10 presents an overview of results regarding performance evaluation criteria in the proposed algorithms with the SVM classifier.

Table 10. Results of performance evaluation criteria for financial ratios extracted by the proposed methods with the SVM classifier for fraud prediction

Algorithm	GA	GWO	PSO
Accuracy	76.47%	75.83%	80.60%
Precision	69.34%	66.80%	71.20%
Recall	62.48%	65.88%	77.73%
F-measure	65.72%	66.31%	74.32%
Prediction error	30.66%	33.20%	28.80%
Confusion matrix results (TP mean + TN mean)	335	335	352
ROC	74.70%	80.30%	83.03%

(Ref.: research findings)

The research hypothesis was confirmed: the financial ratios extracted by the PSO algorithm outperformed the other financial ratios extracted by the other two algorithms in all performance evaluation criteria.

Table 11 indicates the consistency between the financial ratios extracted by the PSO algorithm and the financial ratios used in some of the previous studies:

Table 11. Results of analyzing the extracted financial ratios in comparison with the previous ratios

Financial Ratio	Previous Studies
Total debts to total assets	Chang *et al.* (2021); Jan (2018); Omar *et al.* (2017); Kamrani & Abedini (2022); Rezaei *et al.* (2020); Kazemi (2016)
Working capital to total assets	Omar *et al.* (2017); Tashdidi *et al.* (2019); Bahmanmiri & Malekian (2016)
Inventory to current asset	Bahmanmiri & Malekian (2016)
Accounts receivable to sales	Omar *et al.* (2017); Kamrani & Abedini (2022); Tashdidi *et al.* (2019)
Accounts receivable to total assets	Chang *et al.* (2012); Kamrani & Abedini (2022); Rezaei *et al.* (2020)
Gross income to total assets	Omar *et al.* (2017); Bahmanmiri & Malekian (2016)
Net income to gross income	Tashdidi *et al.* (2019)
Current asset to current debt	Omidi *et al.* (2019); Jan (2018); Rezaei *et al.* (2020); Tashdidi *et al.* (2019); Kazemi (2016)
Retained earnings and loss to equity	Kazemi (2016)

Long-term debt to equity	Omidi *et al.* (2019); Tashdidi *et al.* (2019)

(Ref.: research findings)

The research results are consistent with the findings of previous studies indicating the informational content and usefulness of financial ratios in fraud risk prediction.

According to the research results, the following suggestions are made:

1) It is difficult, specialized, and time-consuming to identify the companies committing fraud in their financial statements, and many users of financial statements lack the necessary and sufficient expertise in this regard. Hence, an organization or institution should consider addressing fraudulent financial reporting more seriously than ever. Moreover, a specialized council should be formed to identify fraudulent companies and to publicize their information.
2) Since the 11 introduced financial ratios have the highest capability of predicting fraud, the users of financial statements and auditors are advised to benefit from those ratios in their analyses.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest concerning the research, authorship and, or publication of this article.

Funding

The authors received no financial support for the research, authorship and, or publication of this article.

References

Aftabi, S. Z., Ahmadi, A., & Farzi, S. (2023). Fraud detection in financial statements using data mining and GAN models. Expert Systems with Applications, 227, 120144.

Ahmadpour, Ahmad, and Yaghizadeh Far, Mahmoud. (2013). Financial Management, Volume One, 9th Edition, Babolsar: Mazandaran University Publications.

Alden, M. E., Bryan, D. M., Lessley, B. J., & Tripathy, A. (2012). Detection of financial statement fraud using evolutionary algorithms. Journal ofEmerging Technologies in Accounting. 9(1): 71-94.

Alem Tabriz, A., Zandieh, M., & Mohammad Rahimi, A. (2008). *Metaheuristic Algorithms in Composite Optimization*, *SAMT Publications*, Tehran, Iran.

Ali, A. A., Khedr, A. M., El-Bannany, M., & Kanakkayil, S. (2023). A Powerful Predicting Model for Financial Statement Fraud Based on Optimized XGBoost Ensemble Learning Technique. Applied Sciences, 13(4), 2272.

Beck, L., & Ajzen, I. (1991). Predicting dishonest actions using the theory of planned behavior. Journal of research in personality, 25(3), 285-301.

Begum, S., Ansari, A. A., Sultan, S., & Dam, R. (2019). A hybrid model for optimum gene selection of microarray datasets. In *Recent Developments in Machine Learning and Data Analytics* (pp. 423-430). Springer.

Bianchi, L., Dorigo, M., Gambardella, L. M., & Gutjahr, W. J. (2009). A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing*, *8*(2), 239-287.

Brazel, J. F., Jones, K. L., & Zimbelman,M. F. (2009). Using nonfinancial measures to assess fraud risk. Journal of Accounting Research. 47 (5): 1135-1166.

Committee of Audit Standards (2020). *Standards and Principles of Accounting and Auditing: Audit Standards*. *Audit Organisation Publications*: Tehran, Iran.

Chen, Y. (2023). Financial Statement Fraud Detection based on Integrated Feature Selection and Imbalance Learning. Frontiers in Business, *Economics and Management*, 8(3), 46-48.

Cheng, C.-H., Kao, Y.-F., & Lin, H.-P. (2021). A financial statement fraud model based on synthesized attribute selection and a dataset with missing values and imbalanced classes. *Applied Soft Computing*, *108*, 107-487.

Christie, A. A. (1990). Aggregation of test statistics: An evaluation of the evidence on contracting and size hypotheses. Journal of accounting and economics, 12(1-3), 15-36.

Charalambos T. Spathis, (2002), Detecting false financial statements using published data:some Evidence from Greece, Managerial Auditing Journal, Vol. 17, Iss 4, pp. 179 -191.

Dalnial, H., Kamaluddin, A., Sanusi, Z. M., & Khairuddin, K. S. (2014). Accountability in financial reporting: detecting fraudulent firms. *Procedia-Social and Behavioral Sciences*, *145*, 61-69.

Etemadi, H., & Zolfi, H. (2013). *Use of Logistic Regression in Identification of Fraudulent Financial Reporting*. *Audit Knowledge*, (51): pp. 145–163.

Eberhart, R. C., Y. Shi & J. Kennedy. (2001). Swarm intelligence. Amsterdam: Elsevier.

El-Shorbagy, M. A., & Hassanien, A. E. (2018). Particle swarm optimization from theory to applications. *International Journal of Rough Sets and Data Analysis (IJRSDA)*, *5*(2), 1-24.

Goldberg, D.E. (1989). Genetic Algorithms in Search, Optimization and Machine Learning.

Goldman, P; & Hoffman, H. (2009). *Practical Guidelines for Fraud Risks and Anti-Fraud Controls*, translated by Amir Pourianasab & Mohsen Ghazi, *Dehkhoda Publications*: Tehran, Iran.

Hidayattullah, S., Surjandari, I., & Laoh, E. (2020). Financial Statement Fraud Detection in Indonesia Listed Companies using Machine Learning based on Meta-Heuristic Optimization. 2020 International Workshop on Big Data and Information Security (IWBIS).

Jan, C.-l. (2018). An effective financial statements fraud detection model for the sustainable development of financial markets: Evidence from Taiwan. *Sustainability*, *10*(2), 513.

Khajavi, S., & Ebrahimi, M. (2017). *Modeling Variables Affecting Fraud Detection in Financial Statements through Data Mining Techniques*. *Financial Accounting*, (33): pp. 41–62.

Khani, A. (2007). *Fraudulent Activities and Future of Certified Accounting*. *Auditor* (38): p. 44.

Kazemi, T. (2016). *Identifying Cases of Fraud Risk in Financial Statements of Iran and Evaluating Fraud Risk Detection Methods*. Doctoral Dissertation in Accounting, Faculty of Economy and Social Sciences, Shahid Chamran University of Ahvaz, Ahvaz, Khuzestan, Iran.

Kamrani, H., & Abedini, B. (2022). *Developing a Model for Fraud Risk Detection in Financial Statements through Artificial Neural Network and Support Vector Machine at TSE-Listed Companies*. *Accounting Knowledge and Management Auditing*, (41): pp. 285–314.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2009). Governance matters VIII: aggregate and individual governance indicators, 1996-2008. World bank policy research working paper, (4978).

Kim, Y. J., B. Baik & S. Cho. (2016). Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning. Expert Systems with Applications 62: 32-43.

Kirkos, S., C. Spathis & Y. Manolopoulos. (2007). Data Mining Techniques for the Detection of Fraudulent Financial Statements, Expert Systems with Applications 32: 995-1003.

Lazzem, S., & Jilani, F. (2017). The impact of leverage on accrual-based earnings management: The case of listed French firms, Research in International Business and Finance, In Press, 12 July.

Lei, Y., Qiaoming, H., & Tong, Z. (2023). Research on Supply Chain Financial Risk Prevention Based on Machine Learning. Computational Intelligence and Neuroscience, 2023.

Mirjalili, S., & Lewis, A. (2013). S-shaped versus V-shaped transfer functions for binary particle swarm optimization. Swarm and Evolutionary Computation, 9, 1-14.

Nath, S.S., G. Mishra., J. Kar., S. Chakraborty & N. Dey. (2014). A survey of image classification methods and techniques. In 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT) IEEE: 554-557.

Occupational Fraud. (2022).A Report To Nations, https://acfepublic.s3.us-west-mazonaws.com.

Omar, N., Z. A. Johari & M. Smith. (2017). Predicting fraudulent financial reporting using artificial neural network. Journal of Financial Crime. 24(2):362-387.

Omidi, M., Q. Min., V. Moradinaftchali & M. Piri. (2019). The efficacy of predictive methods in financial statement fraud. Discrete Dynamics in Nature and Society 4989140: 1-12.

Persons, Obeua S., (2011). Using Financial Statement Data to Identify Factors Associated with Fraudulent Financial Reporting, Journal of Applied Business

Research 3: 38-46.

Pradhan, A. (2012). Support vector machine-a survey. International Journal of Emerging Technology and Advanced Engineering 2(8): 82-85.

Raeisi Nafchi, A., & Dastgir, M. (2018). *Proposing a Model for Identification of Risk Factors Affecting Fraud in Banks*. *Financial Accounting*, (41): pp. 23–45.

Rezaee, Z., & Riley, R. (2011). Financial Statement Fraud Defined (Vol. 196). John Wiley & Sons.

Rezaei, M., Nazemi Ardakani, M., & Naser Sadr Abadi, A. (2020). *Fraud Detection in Financial Ratios through Financial Reporting*. *Management Accounting*, (45): pp. 141–153.

Ross, S., Randolph, W., & Jordan, B. (2013). *Modern Financial Management (Vol. 1)*, translated by Ali Jahankhani & Mojtaba Shouri, *SAMT Publications*, Tehran, Iran.

Sorunke, O. A. (2016). Personal ethics and fraudster motivation: The missing link in fraud triangle and fraud diamond theories. International Journal of Academic Research in Business and Social Sciences, 6(2), 159-165.

Sedighi Kamal, L. (2013). *Fraud in Financial Statements Based on Reports by Association of Certified Fraud Examiners*. *Auditor*, (120): pp. 64–116.

Taghavi, M., & Nazak, N. (2006). *Use of Evolutionary Algorithms in Data Mining*. *First International Conference on Research Methods in Engineering*, pp. 15–23

Tashdidi, E., Sepasi, S., Etemadi, H., & Azar, A. (2019). *A Novel Approach to Prediction and Detection of Fraud in Financial Statements through Bee Colony Optimization Algorithm*. *Accounting Knowledge*, (12): pp. 139–167.

Tehrani, Reza. (2011). Financial Management, 9th Edition, Tehran: Negah Danesh Publications.

Vakilifard, H., Jabarzadeh Kangarlouei, S., Pourreza, A., & Ahmadi, S. (2009). *Analysis of Fraud Characteristics in Financial Statements*. *Accountant*, (21): pp. 36–41.

Wolfe, D. T., & Hermanson, D. R. (2004). The fraud diamond: Considering the four elements of fraud.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. IEEE Transactions on neural networks, 16(3), 645-678.

Yazdani Fazalabadi, Fereshte. (2016). Identifying and Ranking the effective factors of fraud in financial statement based on fraud triangle using AHP method, master's thesis in accounting, Qom University.

Zareh Bahmanmiri, M, & Malekian Kaleh Basi, E. (2015). *Fraud Prediction in Financial Statements through Financial Ratios*. *Financial Management Outlook*, (12): pp. 65–80.

Bibliographic information of this paper for citing:

Nemati, Zahra; Mohammadi, Ali; Bayat, Ali & Mirzaei, Abbas (2024). Fraud Risk Prediction in Financial Statements through Comparative Analysis of Genetic Algorithm, Grey Wolf Optimization, and Particle Swarm Optimization. *Iranian Journal of Finance*, 8(1), 98-130.