Original Research

Providing an Intelligent Model to Detect Fraud in Financial Statements

Marzieh Poursaedi[a], Mahmood Hematfar [a, *], Seyed Enayatallah Alavi[b], Roya Nasirzadeh[c]

aDepartment of Accounting, Borujerd Branch, Islamic Azad University, Borujerd, Iran
bAssistant professor of computer Department, faculty of Engineering, Shahid Chamran university of Ahvaz. Ahvaz, Iran
cDepartment of Statistics, Faculty of Science, Fasa University, Fasa, Iran

ABSTRACT

Some companies manipulate financial statements to users and commit fraud. Therefore, effort to detect fraud is essential. Meanwhile, data mining techniques have increasingly popular. This article aimed to use an advanced model to detect fraudulent financial statements and compare it with the other methods. Crowd optimization algorithms have been considered to solve many optimization problems, but so far, they have not been used in fraud detection research to determine the optimal value of SVM parameters and optimize ANN architecture. In this research, for the first time, the PSO algorithm was used as one of the best innovative optimization algorithms for these optimizations due to its memory and high convergence speed, as well as having solutions for exiting from local optimal points and cooperation and information sharing between particles to detect fraud. For this purpose, the financial statements of companies admitted to the stock exchange from 2017 to 2023 were reviewed. The findings showed that the SVM-PSO method, with 89.86% accuracy, compared to the ANN-PSO method, with 80.43% accuracy, and the LR method, with 69.57% accuracy, performs better in identifying suspected fraudulent financial statements. Combining the PSO algorithm with the SVM method has proven superior to other methods due to SVM's high ability to reduce false negatives and PSO's ability to fine-tune its parameters. This combination can be used for high-accuracy financial statement fraud detection.

1 Introduction

Fraud in financial statements is undoubtedly a serious issue that can lead to significant losses for investors. In recent years, some companies around the world, such as Toshiba in 2015 [1], Luckin Coffee in 2019 [2], and Wirecard in 2020 [3], have faced failures and scandals that raise the role and responsibility of auditors concerning fraud and continuity. Activity in auditing financial statements is challenging. Such events have highlighted the gap between what auditors and financial statement audit users expect and the reality of auditing. The gap in these expectations reduces public trust and confidence in the financial reporting system [4]. As a result, auditors can be considered guardians whose

---

* Corresponding author. Tel.: +989123116609
E-mail address: *M.hematfar@iaub.ac.ir*

responsibilities in identifying risks of material misstatement due to fraud (fraud risks) and detecting material misstatements in financial statements due to fraud should not be underestimated. Changes in the macroeconomic and geopolitical environment in which companies operate may lead to new pressures, opportunities, or justifications for fraud, making auditors' role much more prominent [5]. Alternatively, the inability of auditors for detecting fraud in the face of a considerable amount of data is quite evident, which, in some cases, has led to fines for audit institutions. Today, as systems and activities become more complex and fraudulent information increases in modern business environments, up-to-date and innovative technologies are used to prevent and detect fraud. Therefore, developing and growing methods related to various data, including data mining, is a priority to detect financial fraud [6]. Therefore, since the methods of conducting audit proceedings in Iran are very traditional, focusing on improving the level of auditors' skills to identify fraud through vector machine learning can increase the confidence of stakeholders in audit reports. In other words, the evaluation of professional audit services to the stakeholders shows that the weakness in using advanced technologies in creating comprehensive assurance has caused this profession to suffer a kind of decline in the implementation of professional functions and this study tries to partially overcome the challenge in Iran's capital market by developing technological capacities in audit functions. Traditional methods are often limited due to updated standards and stricter regulatory guidelines on company performance.

To more accurately identify potential areas of fraud detection, newer aspects of data mining algorithms are needed. In reality, in these analytical processes, reaching an optimal point for placing companies based on fraud detection can have higher credibility in gaining the trust of financial decision-makers. Many swarm optimization algorithms have been introduced since the early 60s, all of which have shown their potential to solve many optimization problems. Still, they have not been used in fraud detection research to optimize the value of SVM parameters and neural network architecture. In fact, the lack of attention of prior researches in the development of crowd optimization algorithms has caused this study to help the researchers and standard setters of the audit profession by understanding the void created in the use of this algorithm to identify fraud procedures by auditors to have a more coherent understanding of the nature of fraud in capital market companies.

Particle congestion optimization algorithm is one of the best initiative optimization algorithms due to its memory, accuracy and speed of convergence, easy implementation and extension strategies from local optimization points, and cooperation and sharing of information in this study it has been used for the first time to optimize the value of SVM parameters and neural network architecture and present a new and different model of patterns presented by other researchers so far. In this regard, this study aims to detect suspected fraudulent financial statements based on the combination of particle swarm optimization algorithm with support vector machine and artificial neural network and by analyzing these methods and logistic regression (LR) method, evaluate their efficiency and answer the following questions:

1- Can combining a particle swarm optimization algorithm with a support vector machine (SVM-PSO) be suitable for detecting fraudulent financial statements?

2- Can combining a particle swarm optimization algorithm with an artificial neural network (ANN-PSO) be suitable for detecting fraudulent financial statements?

3- Which of LR methods and the ANN-PSO and the SVM-PSO is more suitable for detecting fraudulent financial statements?

2 Theoretical Fundamentals and Research Background

2.1 Fraud in Financial Statements

From both behavioral and legal standpoints, fraud represents a complex category of crime. As financial crimes become more prevalent, defining fraud remains a significant challenge. Currently, no universally accepted definition exists, complicating efforts to assess the nature and scope of offenses classified under financial fraud. In response to this issue, the Bureau of Justice Statistics, in collaboration with Stanford University's Financial Fraud Research Center and the Financial Industry Regulatory Authority's Investor Education Foundation, developed a fraud taxonomy to systematically categorize fraudulent activities. According to their framework, financial fraud involves deliberately deceiving a victim through misrepresentation, concealment, or omission of facts related to goods, services, or associated benefits and consequences. These deceptive actions pertain to entities that are either nonexistent, unnecessary, never intended for provision, or deliberately misrepresented, often with the intent of financial gain [7].The Association of Certified Fraud Examiners' report on employee embezzlement and fraud categorizes the risks of fraud into three main groups: fraudulent financial reporting, asset misappropriation, and corruption [8]. Fraudulent financial reporting is the intentional and dishonest presentation of facts to deprive a person of valuable assets. Conscious fraud committed by management harms investors and creditors through misleading financial statements [9]. Types of financial statement fraud include: misrepresentation of revenue, fraud related to inventory and cost of goods sold, overstatement of assets, misuse of off-balance sheet items, inadequate disclosure, and manipulation of liabilities [10]. According to the Association of Certified Fraud Examiners (ACFE), this type of fraud occurs with the awareness and approval of management, distinguishing it from earnings management, which involves permissible accounting practices. However, fraud in financial reporting often begins with earnings management and grows to lead to "manipulation of books or accounts."

Fraud in financial reporting is considered one of the significant obstacles to the economic development of the company, which reduces trust. Some studies have presented different methods for fraudulent activities [11]. For example, Achmad, Ghozali and Pamungkas defined fraud in financial reporting as including manipulation, falsification, or alteration of supporting documents and accounting records for the preparation of financial statements, omission, error, or deliberate obstruction of transactions, events, or information that led to the presentation financial statements, they know [12].There are various motivations behind financial reporting fraud, including the intention to enhance management's bonuses or rewards, the necessity of securing or retaining financial resources for the business, and the concealment of financial difficulties. Additionally, fraudulent reporting may be driven by efforts to obscure asset misuse for personal gain, prevent delisting from the stock market, address urgent personal financial needs, or respond to pressures from state-owned enterprises and family businesses. Other influencing factors include the desire to minimize tax obligations, external coercion, intense market competition, and psychological or social factors such as greed, ego, selfishness, social status comparison, revenge, managerial ideologies and beliefs, as well as cultural and normative influences within management [13].

2.2 Research Background

Nahari Aghdam Qala Jough [14] found that the support vector machine model with radial kernel has the lowest RMSE and the highest accuracy criterion, and the support vector machine model with linear

kernel and Bayesian linear regression has the highest RMSE and the lowest accuracy criterion for modeling the financial fraud of companies in they were Tehran stock market. Also, the models of artificial neural network model, Bayesian linear regression and support vector machine model with linear kernel respectively had the lowest characteristic values and did not perform relatively well in detecting the existence of financial fraud in the companies present in the Tehran stock market. Schneider and Bruhl [15] investigated the ability of CEO characteristics to predict accounting fraud using machine learning techniques. Their findings demonstrated that nonlinear models, such as random forest and gradient boosting, outperformed linear models, suggesting a complex interplay between CEO attributes, financial data, and fraudulent activities.

Zhang et al. [16] developed a financial fraud detection model incorporating financial and non-financial variables through machine learning algorithms. Their findings indicated that the combined Stacking model performed significantly better in detecting financial fraud than individual models, including GBDT, RF, and AdaBoost, making it an effective approach for fraud identification. Zhao and Bai [17] proposed a method to predict and detect financial fraud. They investigated the LR, RF, XGBoost, SVM, and DT models separately, as well as three combined models. The results showed that a combined model of logistic regression with XGBoost for predicting and detecting the fraudulent activity of companies is better among all models and will also reduce the burden of doing it. Ashtiani and Raahemi [18] conducted a review of machine learning and data mining techniques, as well as various datasets utilized in financial fraud detection. Their findings suggest that future research should prioritize unsupervised and semi-supervised approaches, along with heuristic methods inspired by biological and evolutionary processes. Similarly, Mongwe et al. [19] explored the application of Bayesian logistic regression (BLR) for predicting financial statement audit outcomes based on financial ratios. Their study demonstrated that financial ratios could enhance the early detection of fraudulent activities and improve audit accuracy. Additionally, El-Bannany et al.

 [20] examined the use of machine learning (ML) techniques to predict potential financial statement fraud. Their results indicated that the SVM classifier outperforms other models, including logistic regression (LR), decision trees (DT), and neural networks (NN). Javadian Kootanaee et al. [21] proposed a model for predicting fraud in financial statements, integrating an improved ID3 decision tree with a support vector machine (SVM) as a hybrid approach to enhance performance and accuracy. Their model also incorporates a genetic algorithm and a multilayer perceptron neural network, with results indicating its effectiveness in fraud prediction. Similarly, Mohammadi et al. [22] developed a fraud detection model for financial reporting based on a genetic algorithm. Their findings demonstrated that the proposed model achieves high accuracy in identifying fraudulent companies. Furthermore, Rostamy-Malkhalifeh et al.

 [23] highlighted the exceptional predictive accuracy of the Adaptive Neuro-Fuzzy Inference System (ANFIS), suggesting it as a highly effective tool for detecting financial statement fraud. Craja et al. [24] explored a method for detecting fraud in financial statements by integrating financial ratio data with management opinions expressed in annual company reports. Their experimental findings indicated that the HAN technique provides promising classification outcomes. Similarly, Afruzianazar et al. [25] investigated fraud risk assessment factors, revealing that the involvement of management and employees in unethical practices, as well as the presence of written job descriptions and resources outlining personnel responsibilities, had the weakest correlation with organizational risk management performance in state-owned banks in East Azerbaijan. Additionally, Refahi Bakhsh et al. [26] analyzed the ranking of fraudulent financial reporting using data envelopment analysis. The findings indicate that investment

plays a crucial role in a country's development, and the DEA approach, by assessing company efficiency, can assist investors in making informed decisions. Omidi et al. [27] applied both supervised and unsupervised methods to detect, prevent, and correct financial statement fraud, utilizing techniques such as multi-layer feedforward neural networks (MFFNN), probabilistic neural networks (PNN), support vector machines (SVM), and log-linear models. Additionally, multinomial logit models (MLM) and discriminant analysis (DA) were employed. The experimental results demonstrated that MFFNN outperformed other models in accurately identifying fraudulent financial data. Similarly, Sadgali et al. [28] conducted an extensive analysis of financial fraud detection techniques using data mining approaches. They determined that artificial intelligence (AI) techniques outperform logistic regression in terms of both accuracy and effectiveness in fraud detection. Chen *et al.* [6] in research using Tree, ANNs, KNNs, and SVM techniques, developed approaches to detect fraud in financial statements of business groups and finally identified mechanisms that had high accuracy and ability to detect fraud. Among the techniques introduced, the support vector machine could detect fraud in financial statements. Lagusto [29] employed text analysis and machine learning methods to forecast fraudulent financial statements. The findings demonstrated that the prediction model implemented in the study successfully identified fraudulent financial statements. Similarly, Kopun [30] reviewed literature on data mining techniques for detecting fraud in financial reports. The results indicated that out of 110 financial and non-financial ratios examined in previous studies, eight were identified as the most significant in constructing a fraud detection model using data mining approaches. Hajek and Henriques [31] explored the development of an enhanced financial fraud detection system by integrating features derived from financial data and management comments in corporate annual reports.

Their findings revealed that hybrid models were more effective in correctly classifying fraudulent companies, whereas Bayesian Belief Networks (BBN) outperformed other methods in accurately classifying non-fraudulent firms. Omar et al. [32] examined fraud prediction in financial statements using artificial neural networks, concluding that this approach outperformed traditional statistical methods commonly used for detecting fraudulent financial reports.

Sorkun and Toraman [33] investigated data mining techniques for identifying fraud in electronic ledgers through financial statements. Their study demonstrated that data mining effectively detects fraudulent discrepancies between financial statements and electronic ledgers, with the Decision Stump Algorithm outperforming other methods. Previous studies in fraud detection have shown the use of different machine learning methods, however they have rarely addressed the optimization of these methods using metaheuristic algorithms. While these algorithms are highly flexible and applicable to various optimization problems, they can be combined with other methods and algorithms to increase efficiency. So, considering the necessity of developing methods for fraud detection, in this research, we have optimized ANN and SVM methods using the PSO algorithm.

Other findings suggest that cost stickiness has a positive impact on the relationship between institutional investors and passive institutional investors with conservatism [60]. The findings of some researchers showed that there is a significant relation between the stock market uncertainty changes in an economic boom and the investment risk in general, which is not significant in terms of the economic turndown. The Investment risk during both economic boom and recession is decreased by the unexpected increase in profit of each share and propagation of positive news. Although the risk is increased by the spread of negative forecasts in relation to shares [61]. Several researchers have examined the relationship between

ownership structure and stock liquidity of companies listed on the Tehran Stock Exchange. The effects of ownership structure were analyzed in two dimensions: ownership type and ownership concentration. The findings of their study indicated an inverse relationship between the level of institutional ownership, managerial ownership, and ownership concentration with liquidity. Moreover, a direct relationship was identified between the level of corporate ownership and liquidity[62]. With the aim of evaluating the effects of peer company performance, future competitive performance of companies, and other factors, researchers investigated the issue of manipulation of actual abnormal operations in companies listed on the Tehran Stock Exchange during 2013-2017. The results of their research showed that peer performance, future competitive performance, and correlation factors with peer companies affect the manipulation of actual abnormal operations [63].

3 Methodology

This research is applied in terms of purpose and the type of experimental research, and due to the use of historical information, it is in the category of quasi-experimental research. To collect data related to the theoretical foundations and background of the research, the library collection method was used, and the necessary data for the implementation of the techniques were extracted from the financial statement information obtained from the Codal Website and the database of the Stock Exchange Organization. The collected data were summarized using Excel software, then the desired variables were calculated. Finally, MATLAB software was used to measure the ability of the logistic regression technique, and Python software was used to measure the ability of the ANN-PSO and the SVM-PSO to detect financial statement fraud.

The research's statistical population consists of companies listed on the Tehran Stock Exchange from 2017 to 2023. This study employs a systematic exclusion sampling method, outlined as follows:

1- They do not belong to investment companies, banks, financial intermediation, holdings, leasing, or insurance companies.

2- Their financial year concludes at the end of March

3- Since March 20, 2017, I have been registered as a member of both the Tehran Stock Exchange and the over-the-counter stock exchange.

4- Their information can be prepared and available.

Table 1: The statistical population of the research

Description	Company	Year-Company
Statistical population	799	
Systematic elimination	694	
Statistical sample containing noise	105	735
Noisy data		275
Statistical sample		460

3.1 Target variable

Fraud in financial statements is this research's output variable (goal). The selection of companies suspected of financial fraud was based on the methods of Mongwe, Mbuvha, and Marwala [19], as well as Mongwe and Malan [34]. This methodology assumes that firms receiving adjusted audit opinions

(i.e., rejected, non-commentary, or conditional reports) have a higher likelihood of engaging in financial fraud compared to those with unmodified (accepted) audit reports. Consequently, firms whose audit reports were initially rejected, non-commentary, or conditional were identified. Then, among the identified companies, those whose audit reports mention one of the following instances of fraud as the reason for their opinion type are identified as companies suspected of financial statement fraud upon verification of one of the instances:

1-Misrepresentation or improper recognition of revenue and realization of income

2-Overstatement of assets and year-end balances

3-Misclassification or failure to recognize costs and realized expenses

4-Understatement of liabilities and deceptive use of reserve accounts

5-Failure to prepare financial statements under the assumption of business continuity despite substantial doubt regarding ongoing operations, as indicated in audit reports

6-Non-compliance or improper application of accounting principles, standards, and estimates related to measurement, classification, recognition, presentation, or disclosure of key financial statement items

Fraud detection is inherently qualitative and assessed using a nominal measurement scale. In measuring this variable, financial statements suspected of fraud are assigned one, and healthy financial statements are assigned zero.

3.2 Input variables

The (independent) input variables used in this research are the ratios calculated from the items in the financial position statement, profit and loss statement, and cash flow statement of the sample companies. In this research, existing empirical evidence in this field has been used to select independent variables. According to previous studies [27,32,35,36,37] in the field of fraud detection in financial reporting, 24 financial ratios are used as follows:

The study evaluates various financial ratios, including the current ratio, current assets to total assets, fixed assets to total assets, working capital to total assets, total debt to total assets, retained earnings to total assets, sales to total assets, net profit to total assets, gross profit to total assets, total debt to equity, long-term debt to equity, cost of goods sold, gross profit to sales, net profit to sales, operating expenses to sales, operating profit to sales, net profit to equity, financial expenses to total debt, accounts receivable to sales, inventory to sales, cash balance to total assets, inventory to total liabilities, net profit to gross profit, and operating cash flow to total assets.

Among these features, according to the research by [38], the chi-square test was used to select the features with a more appropriate description of the objective under investigation (probability of fraud). This method is one of the most important supervised filter methods for feature selection and is used when at least one of the variables is qualitative [39]. Using this test, according to Fig.1, the selected features include gross profit ratios to total assets, accounts receivable to sales, and cash balance to total assets, which is consistent with previous research [10, 40, 41].
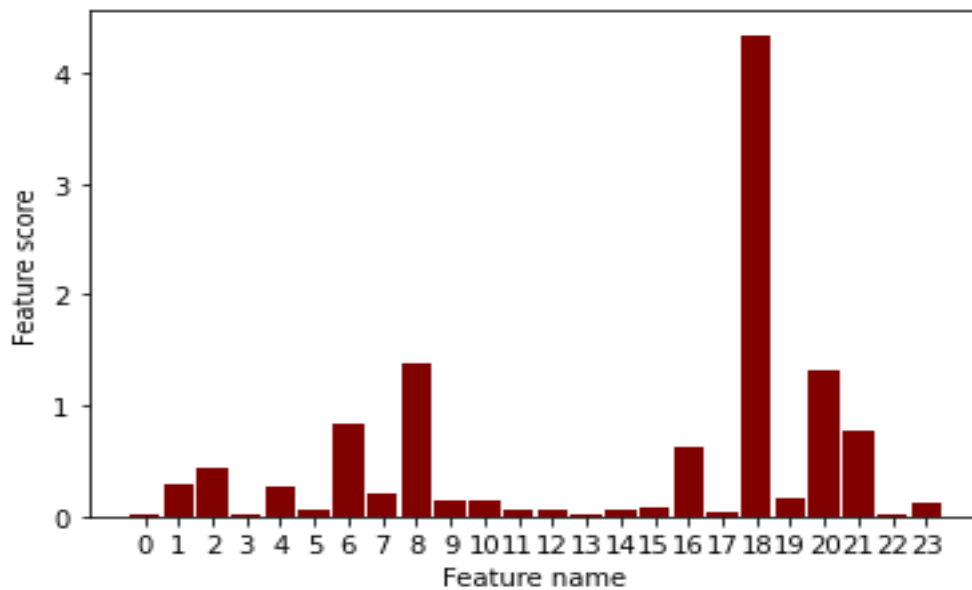
Fig. 1: Chi-Square Test

3.3 Research Procedure

In this study, out of 460 data points (measured according to the method described in the independent variable measurement section), 154 cases were identified as suspected fraud and 306 as non-fraudulent, represented as dummy variables (1 and 0). Subsequently, fraud detection in financial statements was conducted using LR, ANN-PSO and SVM-PSO methods. For data mining, machine learning techniques were applied to train the dataset within data mining software. Finally, newly obtained data were used as test data in the software system to assess the performance of the models.

3.4 Machine Learning

Machine learning is a process that enables a computer to learn without being programmed to do so. Machine learning applies statistics and algorithms to large amounts of data. One of the goals of machine learning is to produce artificial intelligence [42]. Machine learning algorithms include two types of learning with or without supervision [43]. Supervised learning methods require a labeled training dataset containing normal and abnormal samples to build predictive models [44]. Unsupervised learning techniques do not require training data [45]. Various sciences are related to machine learning. These techniques can be used in various fields, such as fraud detection, unauthorized access detection, medical error detection, data mining, and prevention [46]. With the progress of profound learning technology transformation, the value of machine prediction has also increased because machines can consider more factors and act more accurately, faster, and cost-effectively [47]. Although various machine learning methods have been investigated to detect fraud in financial statements, it is still not possible to say which method has the best performance because the performance of each method can be changed depending on various factors such as the number and limit of data set changes and other parameters. This research used support vector machine and artificial neural network methods optimized with particle swarm algorithms and logistic regression to detect fraud in financial statements.

3.5 Support Vector Machine

Support vector machine is a supervised machine learning algorithm and an effective method in solving classification and regression problems, which basically separates data samples that are points in space using a line. In data sets that are linearly separable, the discriminator H divides the samples, for example, into two groups such that the data in one group lie almost entirely on one side of H. Although there are infinitely many candidate discriminators, SVM chooses the discriminator that has the greatest distance from the nearest data points in that category, which we call marginal maximization [48]. However, it is possible that for many data in the real world such a separator does not exist. In this case, SVM uses a function called the kernel function (kernel) to map the data into a higher dimensional space so that this separation is possible. The most important kernel functions are sigmoid, linear, polynomial and RBF functions [49].

3.6 Artificial Neural Network (ANN)

The ANN is inspired by the structure of the human brain and is used as a mathematical structure to map between input and output numbers. Nerve cells (neurons) are the main components of artificial neural networks, including a set of inputs and outputs and many nodes or processor units called the transfer (activation) function. Fig. 2 shows the structure of a neuron and transfer functions in general:
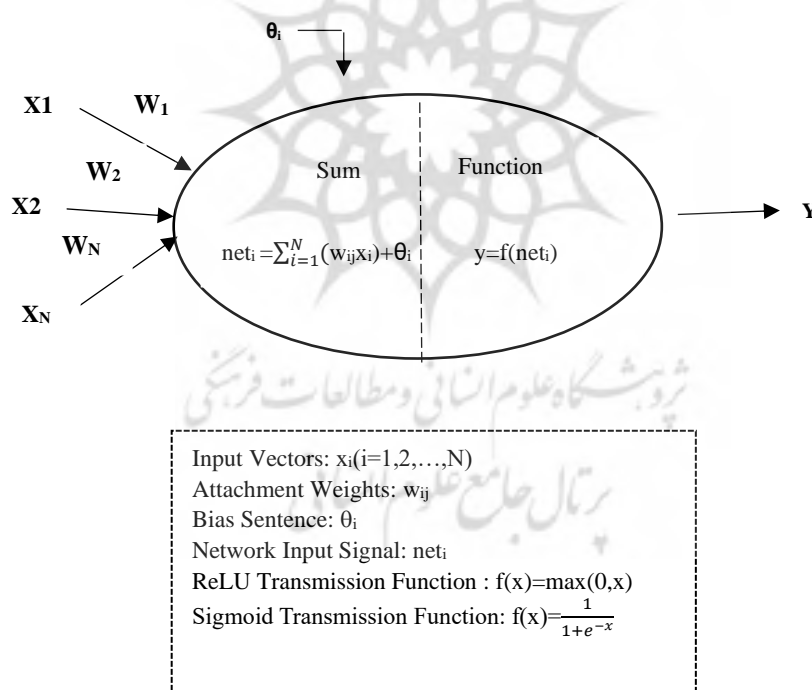

Input Vectors: $x_i (i=1,2,\ldots,N)$
Attachment Weights: w_{ij}
Bias Sentence: θ_i
Network Input Signal: net_i
ReLU Transmission Function : $f(x)=\max(0,x)$
Sigmoid Transmission Function: $f(x)=\frac{1}{1+e^{-x}}$

Fig. 2: The Structure of a Neuron of an ANN

A layer in an artificial neural network consists of a collection of parallel neurons. A neural network can comprise multiple layers, including hidden and output layers, which are sequentially connected to

generate the final output. Each input within the network is processed through an activation (transfer) function, influenced by its associated weight, to determine the inputs for the subsequent layers [50].

3.7 Logistic Regression (LR)

Logistic regression is a widely recognized machine learning and data science algorithm classified under supervised learning techniques [51]. It is applicable to both classification and regression tasks [52]. As a statistical model, logistic regression assesses the influence of quantitative or qualitative variables on a binary-dependent variable. Unlike linear regression, it does not assume a linear relationship between dependent and independent variables, despite its classification as a regression method [53].

3.8 Particle Swarm Optimization Algorithm (PSO)

The Particle Swarm Optimization (PSO) technique, inspired by the collective movement of birds, was introduced by Kennedy and Eberhart in the mid-1990s [54]. In this approach, an initial set of randomly selected candidate solutions (particles) explores the search space through iterative updates, aiming to find the optimal solution. Each particle benefits from shared information within the search space, facilitating collective optimization.

PSO simulates the behavior of birds, fish, and other animals as they forage for food or evade predators, making it an evolutionary optimization method [55]. PSO was specifically designed to address nonlinear optimization problems involving continuous variables. Unlike other evolutionary approaches, such as genetic algorithms, PSO can be implemented with minimal programming effort. It has gained significant recognition due to its efficiency in solving complex, computationally intensive, and sometimes infeasible optimization problems. A foundational aspect of PSO was introduced by Boyd and Richerson [56], who explored decision-making and the concept of learning. Their study identified two primary sources of information influencing decision-making: personal experience, which reflects an individual's past choices and their outcomes, and social experience, which involves observing and evaluating the choices made by others.

Individuals assess both their own successes and the collective experiences of their peers to inform their decisions. Building on this conceptual foundation, Kennedy and Eberhart formulated PSO by simulating bird flocking behavior. In this framework, each particle aims to optimize an objective function by considering two key aspects: its personal best-known position (pbest) and the best-known position within its neighborhood (gbest). The particle updates its position based on these values and its velocity. The velocity update process follows the equation:

$$v_i^{k+1} = wv_i^k + c_1 \text{ rand}_1 \times (pbest_i - s_i^k) + c_2 rand_2 \times (gbest - s_i^k) \tag{1}$$

where, v_i^k is the velocity of particle i in the kth iteration, w is the weighting function and c_j is a weighting coefficient. rand is a random number between 0 and 1, s_i^k is the current position of particle i in the kth iteration, $pbest_i$ is the pbest of point i, and gbest is the gbest of the group. In this equation, the first term maintains the particle's previous velocity, ensuring continuity in movement. The second and third terms introduce velocity adjustments, enabling the particle to explore new regions of the search space.

Without these adjustments, a particle would continue in its previous direction until reaching a boundary. The first term contributes to the diversity of the search process, while the latter terms balance indi-

vidual learning with collective learning. Consequently, particles iteratively adjust their positions to converge toward either their personal best positions (pbest) or the globally best position (gbest), enhancing the efficiency of the optimization process.
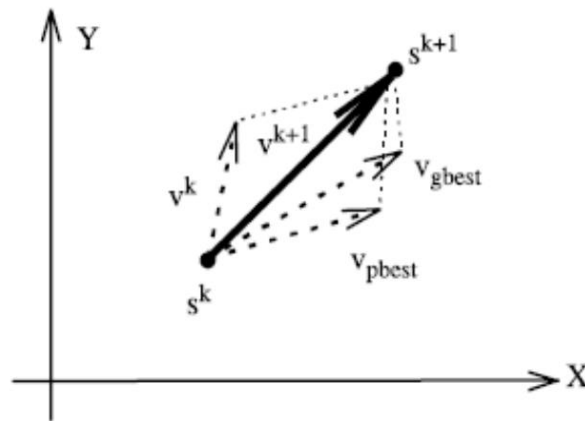

Fig. 3: The Concept of Modifying the Search Point with PSO [57]

The position of the search point in the solution (response) space can be adjusted using equation (2):

$$s_i^{k+1}=s_i^k+v_i^{k+1} \tag{2}$$

Each particle modifies its current position using the combination of vectors shown in Fig. 3. In fact, PSO uses several search points and the search points gradually approach the optimal point using pbests and gbests.
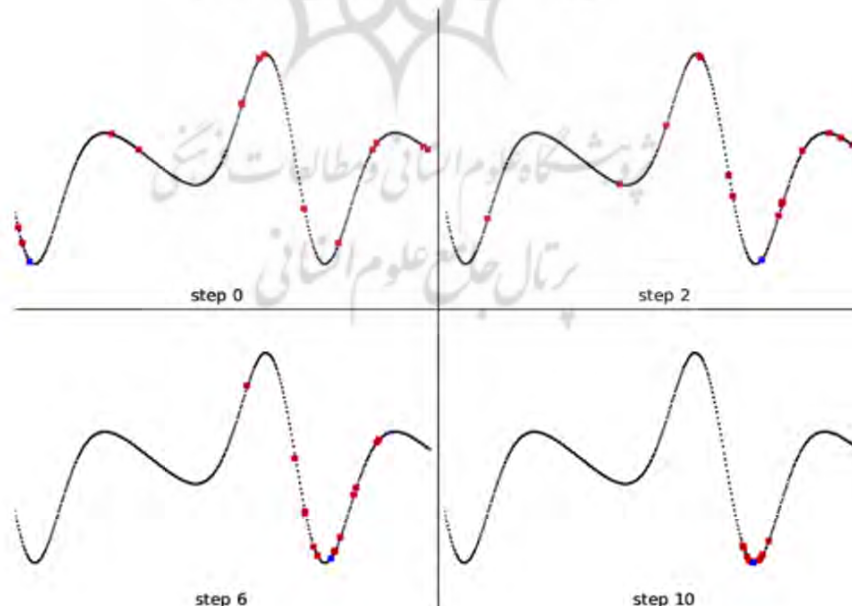

Fig. 4: The Concept of Searching with Particles in the Solution Space (Answer) PSO [57]

4 Findings

Descriptive statistics of the research variables are shown in the table below:

Table 2: Descriptive statistics of central tendency and dispersion indices of variables

Variable	Mean	Median	Std. Deviation	Skewness	Kurtosis	Minimum	Maximum
CA/CL	0.0606	0.0446	0.06789	6.501	64.897	0	1
CA/TA	0.6653	0.6883	0.21015	-0.564	-0.342	0	1
FA/TA	0.2577	0.2078	0.19583	1.016	0.625	0	1
WC/TA	0.6101	0.6166	0.14385	-0.398	0.983	0	1
TD/TA	0.2769	0.2729	0.12674	0.617	1.751	0	1
RE/TA	0.6357	0.6391	0.12083	-0.715	2.430	0	1
SALE/TA	0.1307	0.1032	0.10775	3.000	13.788	0	1
NI/TA	0.5413	0.5261	0.12418	0.326	0.711	0	1
GP/TA	0.5876	0.5920	0.12897	-0,687	2.008	0	1
TD/TE	0.3516	0.3499	0.03167	9.663	269.334	0	1
LTD/TE	0.8188	0.8185	0.03543	-18.074	407.787	0	1
CGS/SALE	0.3043	0.3096	0.09522	0.278	4.046	0	1
GP/SALE	0.6959	0.6904	0.09553	-0.265	3.997	0	1
NI/SALE	0.2937	0.2879	0.04168	7.469	121.570	0	1
OE/SALE	0.6071	0.6101	0.04558	-2.380	51.180	0	1
OI/SALE	0.3929	0.3899	0.04558	2.380	51.180	0	1
NI/TE	0.9406	0.9429	0.03841	-21.243	501.298	0	1
IE/TD	0.1790	0.1296	0.17060	1.067	0.848	0	1
REC/SALE	0.0346	0.0216	0.05295	9.705	155.645	0	1
INV/SALE	0.0157	0.0093	0.05642	13.718	203.863	0	1
CASH/TA	0.0913	0.0624	0.10601	3.344	16.988	0	1
INV/TD	0.1172	0.0941	0.10691	3.218	17.713	0	1
NP/GP	0.3656	0.3643	0.03716	5.274	150.529	0	1
OCF/TA	0.7086	0.6970	0.07866	-0.391	9.114	0	1
Y	0.3967	0	0.48955	0.423	-1.826	0	1

The most important central tendency measure is the mean, which represents the equilibrium point and center of gravity of the distribution and is a good indicator of the centrality of the data. For example, the mean of the current asset to total asset ratio variable is 0.6653, indicating that most of the data related to this variable is centered around this point. The median is another central tendency measure; for the current asset to total asset ratio, it is 0.6883, meaning half of the data is greater than this value and half is less.Dispersion parameters also indicate the extent to which data is scattered from each other or their spread relative to the mean. Standard deviation is one of the most important dispersion parameters.The higher the standard deviation, the lower the accuracy of the variable. For example, among the input variables, current assets to total assets has the highest standard deviation at 0.21015, indicating lower accuracy for this variable compared to others.

4.1 Logistic Legression

Table 3 shows the result of estimating and predicting the probability of fraud using the linear logistic regression method:

Table 3: Confusion Matrix of Linear Logistic Legression

		0	64	5	%92.75
True Label	1	42	27	%39.13
			%60.38	%84.38	%65.94
			0	1	
	Predicted Label				

As the table shows, 64 healthy financial statements and 27 suspected fraud financial statements are correctly predicted in their group. The cost function for this model was 0.5024.In many cases, the linear logistic regression is unsuitable for the data; in other words, the nonlinear logistic regression gives a better answer. For this purpose, we fitted a rank two regularized nonlinear logistic regression to these data. In this constant regression model, we enter all features $X\_i, i=1,...,24$ and their binary product, $X\_i X\_j, I, j=1,...,24$, as featured in the model. This way, 325 features have been considered for this model. We have used the appropriate cost and gradient function for this type of regression.Table 4 shows the result of estimating and predicting the probability of fraud using the nonlinear logistic regression method:

Table 4: Confusion Matrix of Non-Linear Logistic Regression

		0	59	10	%85.50
True Label		1	32	37	%53.62
			%64.84	%78.72	%69.57
			0	1	
		Predicted Label			

As the table shows, 59 healthy financial statements and 37 suspected fraud financial statements are correctly predicted in their group. The cost function for this model was 0.1956.

4.2 ANN-PSO Method

Feedforward multilayer networks are among real-world applications' most important and common neural networks.

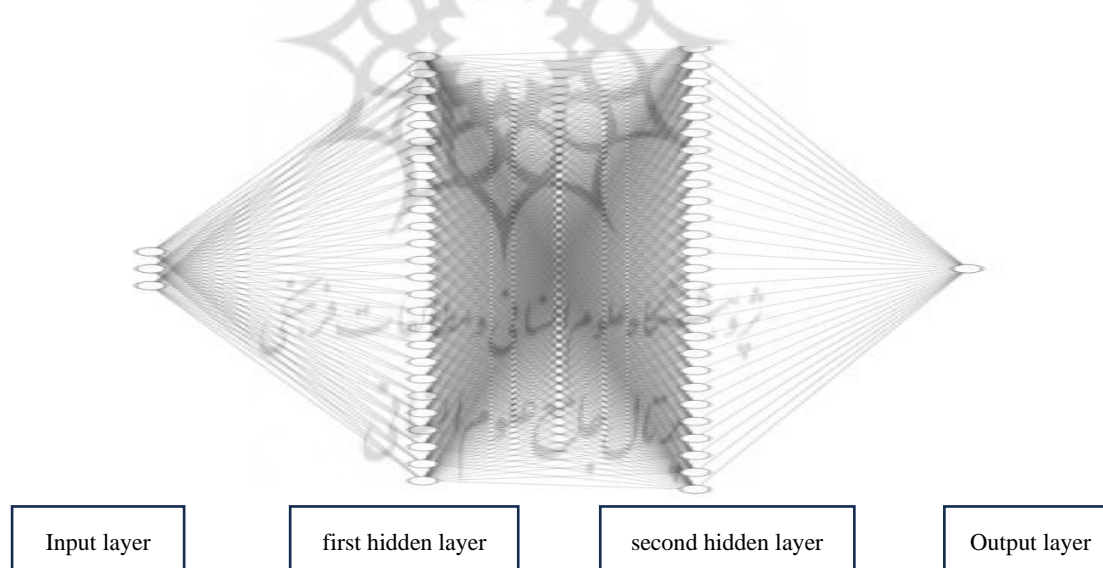| Input layer | first hidden layer | second hidden layer | Output layer |

Fig. 5: An Example of ANN with Two Hidden Layers

This research uses a feedforward multi-layer neural network to solve binary classification problems. In the training of feedforward neural networks, the flow of information is forward. In other words, in feedforward propagation, data enters the network through the input layer, and by applying calculations on them using activation functions in hidden layers, the output of each layer is transferred to its next layer until, finally, the output of the network in Determine the last layer. In the feedforward propagation

stage, the activation function of neural networks is considered a "Gate" that sends each layer's inputs to the next layer [58].The artificial neural networks are expected to reach an optimal trained state by minimizing the discrepancy between predicted and actual values. When presented with diverse input data, the network is anticipated to generate predictions with minimal deviation from the corresponding real values. Before implementing the artificial neural network, the input and output data were normalized between 0 and 1, and this network's training and testing process was done using the normalized values. The artificial neural network toolbox of Python software was used to implement the artificial neural network.

This toolbox randomly assigns initial weight and bias values to the network every time it is executed. Allocation of weights and biases is one of the most influential factors in the performance of artificial neural network training, even if the architecture and other neural network parameters are fixed. This research used the particle swarm optimization algorithm to determine the optimal neural network architecture, including the number of hidden layers, neurons in each layer, repetition, and batch size.70% and 30% of the total data set were used to train and test the network. The number of input and output layer neurons is 3 and 1, respectively. The evaluation showed that the optimal network is as described in the following Table:

Table 5: Optimum Network Architecture

The number of neurons in the first hidden layer	107
The number of neurons in the second hidden layer	119
No. of repetition	12
Size of bunches	95

In this architecture, the transfer function of the hidden layers was considered RELU, and the output layer's transfer function was considered sigmoid. Finally, the ADAM algorithm was used to optimize network weights and biases. The considered parameters for the particle swarm optimization algorithm are shown in Table 6.

Table 6: Parameters of Particle Swarm Optimization Algorithm for Neural Network Optimization

No. of particles	10
No. of repetitions	5
Recognition acceleration (C1)	0.5
Cumulative acceleration (C2)	0.3
Inertia weight (W)	0.3

The particle swarm algorithm provides the highest accuracy in this case. Table 6 shows the result of estimating and predicting the probability of fraud using the ANN-PSO:

Table 7: Confusion Matrix of ANN-PSO

True Label	0	54	15	%78.26
	1	12	57	%82.61
		%81.82	%79.17	%80.43
		0	1	
Predicted Label				

As the table shows, 54 healthy financial statements and 57 suspected fraud financial statements are correctly predicted in their group.

4.3 SVM-PSO Method

In this research, the RBF kernel function has been used to map input vectors (feature vectors) to a space with higher dimensions, and in this new space, the classification of vectors is done more efficiently and appropriately.
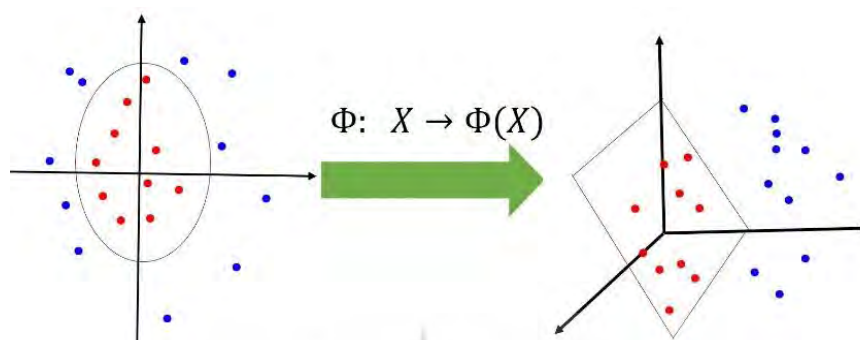

Fig. 6: Mapping Samples to a Space with Higher Dimensions in the Model Generation Process in the SVM Method [59]

The shape of the RBF kernel is as follows:

$$K(x_i,x_j)=e^{\gamma} \|x_i-x_j\|^2 \tag{3}$$

where x_i and x_j are vectors in the input space.

In order to properly generalize the support vector machine method, the values of parameters γ and C must be chosen correctly. The γ parameter determines how influential each training data is, and the C parameter controls the margin. In this research, in order to determine the optimal value for these parameters, the particle swarm algorithm has been used.

Before implementing the support vector machine, the input and output data were normalized between 0 and 1, and the training and testing process of this machine was done using the normalized values. Python software was used to implement support vector machine. In order to train and test the machine, 70% and 30% of the total data set were used, respectively. An equal number of test data have been selected from both groups, that is, 69 suspected fraud data and 69 healthy data have been tested in order to investigate the applicability of that model in the future. The evaluation showed that the optimal value of the parameters is as described in the following Table:

Table 8: The Optimal Value of the Support Vector Machine Parameters

Parameter γ	0.95
Parameter C	2.03

The considered parameters for particle swarm optimization algorithm are shown in Table 9.

Table 9: The Parameters of the SVM-PSO

No. of particles	10
No. of repetitions	5
Recognition acceleration (C1)	0.5
Cumulative acceleration (C2)	0.3
Inertia weight (W)	0.3

The particle swarm algorithm provides the highest accuracy in this case.

Table 10 shows the result of estimating and predicting the probability of fraud using the SVM-PSO:

Table 10: Confusion Matrix of SVM-PSO

	0	60	9	%86.96
True Label	1	5	64	%92.75
		%92.31	%87.67	%89.86
		0	1	
Predicted Label				

As the table shows, 60 healthy financial statements and 64 suspected fraud financial statements are correctly predicted in their group.According to the abovementioned, the models' performance Table is as follows.

Table 11: Performance of the Models

Method	Precision	Property	Sensitivity	Error Rate	Accuracy	Type 1 Error (False Positive)	Type 2 Error (False Negative)
Linear Logistic Regression	%60.38	%39.13	%92.75	%34.05	%65.94	%60.87	%7.25
Nonlinear Logistic Regression	%64.84	%53.62	%85.50	%30.43	%69.75	%46.38	%14.50
ANN-PSO	%81.82	%82.61	%78.26	%19.57	%80.43	%17.39	%21.74
SVM-PSO	%92.31	%92.75	%86.96	%10.14	%89.86	%7.25	%13.04

The accuracy for the linear logistic regression model is 65.94%. Simply put, this model correctly predicted 65.94% of the data. The accuracy for the nonlinear logistic regression model is 69.57%. In simple words, this model correctly predicted 69.75% of the data. The accuracy for the ANN-PSO model is 80.43%. In simple words, this model correctly predicted 80.43% of the data. . The accuracy for the SVM-PSO model is 89.86%. In simple words, this model correctly predicted 89.86% of the data.

5 Discussion and Conclusions

Detecting financial fraud is a challenging endeavor. Continuous compliance and the complex nature of fraudulent activities make it necessary to use the latest technologies to combat fraud. Empowering an intelligent fraud detection system with more research in this field will lead to more active design of an efficient and effective detection program. This paper explores the potential of an advanced intelligent model to add to the development of advanced fraud detection methods in financial statements.

This research aimed to detect fraudulent financial statements using logistic regression methods and an

artificial neural network optimized with a particle swarm algorithm and support vector machine optimized with a particle swarm algorithm. For this purpose, the financial statements of 105 companies (including 735 year-companies) were collected, and the information from these financial statements was used to calculate the input variables. After removing the noise data, 460 data were entered into the algorithm. Noisy data (275 out of 735) were removed by: 1-Eliminating top/bottom 20% extreme values (outliers) based on empirical optimization. 2-Normalizing remaining data to [0, 1] range for model stability .Final cleaned dataset: 460 samples. This approach aligns with common practices in financial ML [6]. After the implementation of the methods, it was found that the support vector machine method optimized with the particle swarm algorithm performs better than the logistic regression method and artificial neural network optimized with a particle swarm algorithm for identifying financial statements suspected of fraud, and this model can be suggested for detecting fraud in financial statements. Also, the results show that the selection of financial ratios has been done correctly. Compare to other methods, SVM has a high ability to reduce false negative errors, which is very important in fraud detection. With proper parameter tuning, this algorithm can prevent overfitting and provide a more generalizable model. Due to its ability to find optimal decision boundaries, SVM has high accuracy in data classification. By fine-tuning SVM parameters through PSO, the model can even achieve higher accuracy in classifications. PSO, due to its high flexibility, can adapt well to various SVM kernels and different problems. Given its simple implementation, it can be easily run alongside SVM. Furthermore, due to its nature, PSO usually achieves the optimal solution faster than some other optimization algorithms. In addition, ANNs are powerful tools for fraud detection and can identify complex and unusual patterns in data that may not be detectable by traditional methods. They can analyze large volumes of data quickly and accurately, which is highly beneficial in detecting financial fraud. Using this technology can reduce operational costs associated with fraud detection and increase efficiency. The PSO algorithm can automatically design the architecture of a neural network. This algorithm, using the collective behavior of particles, accelerates the search process for the optimal architecture and reduces the time required to design the network. The optimal architecture designed by PSO can improve the performance of the neural network and increase prediction accuracy. The results obtained from this research are consistent with the results of the studies by [18], [19], [20], [6]. Studies [19, 20, 6] used machine learning methods such as ANN, SVM, and BLR to detect fraud. However, the difference between these studies and this one is the lack of optimization of these methods with metaheuristic algorithms. Study [18] prioritized unsupervised and semi-supervised approaches along with evolutionary algorithms for fraud detection, while this research uses a combination of supervised methods with metaheuristic algorithms.From a managerial point of view, this study advises the policy makers to take into account the necessary measures to spread as much as possible vector machine learning skills in auditing functions. For this purpose, the higher authorities and supervisors of the auditing profession can motivate the auditors to identify the fraud and financial distortions of the owners by holding training courses in order to ensure the absence of possible frauds by the owners before submitting the audit reports.

Suggestions for future research:

1-In this research, feature selection method is used to reduce dimensions. It is suggested to use feature extraction method in future researches for this purpose.

2- Other meta-heuristic methods for optimization, such as mixed frog mutation algorithms, colonial competition, invasive weeds, fireflies, etc., are suggested.

3- According to the research results, using the SVM-PSO is suggested to predict such things as the movement trend of stock prices, companies' financial helplessness, and the Iranian market's exchange rate.

4- In this study, investment companies, banks, financial intermediation, holdings, leasing, and insurance companies were excluded from the sample selection; it is suggested that these companies should also be investigated in future studies.

References

[1] Banks, J. E., Toshiba accounting scandal a case study in corporate governance failure, *18th International Conference on Human Rights, E-Commerce,Marketing, and Management (HERMM),* Dubai, UAE, Jan 1-3, 2018, Doi: 10.17758/EIRAI.DIR0118105

[2] He, J., The Analysis of Luckin Coffee's Accounting Scandal, *Highlights in Business, Economics and Management,* 2023; 24(2024): 2572-2576.

[3] Teichmann, F., Boticiu, S. R., Sergi, B., Wirecard scandal. a Commentary on the Biggest Accounting Fraud in Germany's Post-war History, *Journal of Financial Crime*, 2023; 2(3): 37-56. Doi: 10.1108/JFC-12-2022-0301.

[4] Donnelly, A., Hartman, M., Building Public Confidence in Audit: Fraud, Going Concern, Perception, *International Federation of Accountants*, NewYork, September 25, 2020.

[5] Dorminey, J., Fleming, A. S., Kranacher, M. J., and Riley Jr, R. A., The Evolution of Fraud Theory, *Issues in Accounting Education*, 2012; 27(2): 555-579. Doi; 10.2308/iace-50131

[6] Chen, Y. J., Liou, W. C., Chen, Y. M., and Wu, J. H., Fraud Detection for Financial Statements of Business Groups, *International Journal of Accounting Information Systems*, 2019; 32(1): 1-23. Doi: 10.1016/j.accinf.2018.11.004

[7] Golladay, K. A., Snyder, J. A., Financial Fraud Victimization: An Examination of Distress and Financial Complications, *Journal of Financial Crime*, 2023; 30(6): 1606-1628. Doi: 10.1108/JFC-08-2022-0207

[8] Khamainy, A.H., Ali, M. and Setiawan, M.A., Detecting financial statement fraud through new fraud diamond model: the case of Indonesia, *Journal of Financial Crime*, 2022; 29(3): 925-941. https://doi.org/10.1108/JFC-06-2021-0118

[9] Setayesh, M. H., and Monfared, R., Fraudulent Financial Reporting from the Perspective of the Fraud Pentagon Theory, *Journal of Applied Research in Financial Reporting*, 2023; 12(22): 267-300. (in Persian).

[10] Rahimian, N., and Haji Heydari, R., Fraudulent Financial Statement Detection Using: Adjusted-M-score-Beneish Models and Financial Ratios, *Empirical Research In Accounting*, 2019; 9(1): 47-70. Doi:10.22051/jera.2018.15993.1713. (in Persian).

[11] Sukmadilaga, C., Winarningsih, S., Handayani, T., Herianti, E., and Ghani, E. K., Fraudulent Financial Reporting in Ministerial and Governmental Institutions in Indonesia: an Analysis Using Hexagon Theory, *Economies*, 2022; 10(4): 1-14. Doi: 10.3390/economies10040086

[12] Achmad, T., Ghozali, I., Pamungkas, I. D., Hexagon Fraud: Detection of Fraudulent Financial Reporting in State-owned Enterprises Indonesia, *Economies*, 2022; 10(1): 1-16. Doi:10.3390/economies10010013

[13] Sallal, F., Bagherpour Velashani, M. A., Saei, M. J., Fraudulent Financial Reporting Motivations in Emerging Markets, *Journal of Financial Crime*, 2021; 28(3): 892-905. Doi: 10.1108/JFC-09-2020-0188

[14] Nahari Aghdam Qala Jough, J., Rezaei,N., Aghdam Mazrae, Y., and Abdi, R., Comparing The Performance of Machine Learning Techniques in Detecting Financial Frauds, *Advances in Mathematical Finance & Applications*, 2024; 9(3): 1006-1023. Doi:10.71716/amfa.2024.22101813

[15] Schneider, M., Brühl, R., Disentangling the Black Box Around CEO and Financial Information-Based Accounting Fraud Detection: Machine Learning-Based Evidence from Publicly Listed U.S. Firms, *Journal of Business Economics*, 2023; 93(1): 1591–1628. Doi: 10.1007/s11573-023-01136-w

[16] Zhang, L., Wang, D., Xie, C., Liu, S., Chi, L., Ma, X., and Ren, F. F., The Effects of Tai Chi on the Executive Functions and Physical Fitness in Middle-aged Adults with Depression: a Randomized Controlled Trial, *Evidence-Based Complementary and Alternative Medicine*, 2022; 2022: 1-16. Doi: 10.1155/2022/1589106

[17] Zhao, Z., Bai, T., Financial Fraud Detection and Prediction in Listed Companies Using SMOTE and Machine Learning Algorithms, *Entropy*, 2022; 24(8): 1-17. Doi: 10.3390/e24081157

[18] Ashtiani, M. N., Raahemi, B., Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review, *IEEE Access*, 2022; 10(1): 72504-72525. Doi: 10.1109/ACCESS.2021.3096799

[19] Mongwe, W. T., Mbuvha, R., Marwala, T., Bayesian Inference of Local Government Audit Outcomes, *Plos one*, 2021; 16(12): 1-19. Doi: 10.1371/journal.pone.0261245

[20] El-Bannany, M., Dehghan, A. H., Khedr, A. M., Prediction of financial statement fraud using machine learning techniques in UAE, *18th International Multi-Conference on Systems, Signals & Devices (SSD),* Monastir, Tunisia, March 22-25, 2021, Doi: 10.1109/SSD52085.2021.9429297

[21] Javadian Kootanaee, A., Poor Aghajan, A. A., Hosseini Shirvani, M., A Hybrid Model Based on Machine Learning and Genetic Algorithm for Detecting Fraud in Financial Statements, *Journal of Optimization in Industrial Engineering*, 2021; 14(2): 169-186. Doi: 10.22094/JOIE.2020.1 877455.1685

[22] Mohammadi, M., Yazdani, Sh., Khanmohammadi, M., Presenting a Model for Financial Reporting Fraud Detection using Genetic Algorithm, *Advances in Mathematical Finance & Applications*, 2021; 6(2): 377-392. Doi: 10.22034/amfa.2019.1872783.1252

[23] Rostamy-Malkhalifeh, M., Amiri, M., Mehrkam, M., Predicting Financial Statement Fraud Using Fuzzy Neural Networks, *Advances in Mathematical Finance & Applications*, 2021; 6(1): 137-145. Doi: 10.22034/amfa.2020.1892431.1370

[24] Craja, P., Kim, A., Lessmann, S., Deep Learning for Detecting Financial Statement Fraud, *Decision Support Systems*, 2020; 139(2): 47-71. Doi: 10.1016/j.dss.2020.113421

[25] Afruzianazar, A., Rezaei, N., Hajiha, Z., and Pakmaram, A., Optimal Banking Performance Model based on ERM, *Advances in Mathematical Finance & Applications*, 2023; 8(1): 273-285. Doi: 10.22034/AMFA.2020.1900625.1435

[26] Refahi Bakhsh, S., Banimahd, B., Kheradyar, S., and Ooshaksaraei, M., The Ranking of Fraudulent Financial Reporting By Using Data Envelopment Analysis: Case of Pharmaceutical Listed Companies, *Advances in Mathematical Finance & Applications*, 2020; 5(1): 69-80. Doi: 10.22034/amfa.2019.1863571.1193

[27] Omidi, M., Min, Q., Moradinaftchali, V., and Piri, M., The Efficacy of Predictive Methods in Financial Statement Fraud, *Discrete Dynamics in Nature and Society*, 2019; 2019(4): 1-12. Doi: 10.1155/2019/4989140

[28] Sadgali, I., Sael, N., Benabbou, F., Performance of Machine Learning Techniques in the Detection of Financial Frauds, *Procedia computer science*, 2019; 148(C): 45-54. Doi: 10.1016/j.procs.2019.01.007

[29] Lagusto, D., Predicting fraudulent financial statement using textual analysis and machine-learning techniques, *M.A. thesis, University of Ritsumeikan Asia Pacific, Beppu, Ōita, Japan*, 2018.

[30] Kopun, D., A Review of the Research on Data Mining Techniques in the Detection of Fraud in Financial Statements, *Journal of Accounting and Management*, 2018; 8(1): 1-18. https://api.semanticscholar.org/CorpusID:202358955.

[31] Hajek, P., Henriques, R., Mining corporate annual reports for intelligent detection of financial statement fraud–a comparative study of machine learning methods, *Knowledge-Based Systems*, 2017; 128(1): 139-152. Doi: 10.1016/j.knosys.2017.05.001

[32] Omar, N., Johari, Z. A., Smith, M., Predicting Fraudulent Financial Reporting Using Artificial Neural Network, *Journal of Financial Crime*, 2017; 24(2): 362-387. Doi: 10.1108/JFC-11-2015-0061

[33] Sorkun, M. C., Toraman, T., Fraud Detection on Financial Statements Using Data Mining echniques, *Intelligent Systems and Applications in Engineering*, 2017; 5(3): 132-134. Doi: 10.18201/ijisae.2017531428

[34] Mongwe, W. T., Malan, K. M., A Survey of Automated Financial Statement Fraud Detection with Relevance to the South African Context, *South African Computer Journal*, 2020; 32(1): 74-112. Doi: 10.18489/sacj.v32i1.777

[35] Chen, S., Detection of Fraudulent Financial Statements Using the Hybrid Data Mining Approach, *SpringerPlus*, 2016; 5(1): 1-16. Doi: 10.1186/s40064-016-1707-6

[36] Tangod, K., Kulkarni, G., Detection of Financial Statement Fraud Using Data Mining Technique and Performance Analysis, *International Journal of Advanced Research in Computer and Communication Engineering*, 2015; 4(7): 549-555. Doi: 10.17148/IJARCCE.2015.47124

[37] Kotsiantis, S., Method of Financing, *Australian Accounting Review*, 2006; 16(38): 538-542.

[38] Abbasi, E., and Fahimi, M., Fraud Detection Model in Financial Statements by Using Financial Equity Instruments, *Accounting & Auditing Studies*, 2021; 36(9): 99-122. Doi: 10.22034/iaas.2020.128138. (in Persian).

[39] Momeni, M., and Faal Ghayoumi, A., Statistical Analysis with spss, *Tehran, Ketab e no*, 2024. (in Persian).

[40] Kanapickiene, R., and Grundiene, Z., The Model of Fraud Detection in Financial Statements by Means of Financial Ratios, *Procedia - Social and Behavioral Sciences*, 2015; 213(2015): 321-327. Doi: 10.1016/j.sbspro.2015.11.545

[41] Tashdidi, E., Sepasi, S., Etemadi, H., and Azar, A., New Approach to Predicting and Detecting Financial

Statement Fraud, Using the Bee Colony, *Journal of Accounting Knowledge*, 2019; 10(3): 139-167. Doi: 10.22103/Jak. 2019. 13616.2927. (in persian).

[42] Clarke, S. L., Parmesar, K., Saleem, M. A., and Ramanan, A. V., Future of Machine Learning in Paediatrics, *Archives of Disease in Childhood*, 2022; 107(3): 223-228. Doi: 10.1136/archdischild-2020-321023

[43] Sohail, A., Arif, F., Supervised and Unsupervised Algorithms for Bioinformatics and Data Science, *Progress in biophysics and molecular biology*, 2020; 151(2): 14-22. Doi: 10.1016/j.pbiomolbio.2019.11.012

[44] Han, K., Liu, L., Song, Y., Liu, Y., Qiu, C., Tang, Y., Teng, Q., and Liu, Z., An Effective Semi-Supervised Approach for Liver CT Image Segmentation, *IEEE Journal of Biomedical and Health Informatics*, 2022; 26(8): 3999-4007. Doi: 10.1109/JBHI.2022.3167384

[45] Teixeira, M., Pereira, T., Silva, F., Cunha, A., and Oliveira, H. P., Unsupervised approach for malignancy assessment of lung nodules in computed tomography scans using radiomic features, *44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Glasgow, Scotland, United Kingdom, July 11-15, 2022, Doi: 10.1109/EMBC48229.2022.9871704

[46] Domingues, D., Filippone, M., Michiardi, P., Probabilitic modeling for novelty detection with applications to fraud identification, *Ph.D. thesis, University of Sorbonne. Paris, France*, 2019.

[47] Achakzai, M. A.,Peng, J., Detecting financial statement fraud using dynamic ensemble machine learning, *International Review of Financial Analysis*, 2023; 89(2): 1-19. Doi: 10.1016/j.irfa.2023.102827

[48] Ding, C., Bao, T. Y., Huang, H. L., Quantum-Inspired Support Vector Machine, *IEEE transactions on neural networks and learning systems*, 2022; 33(12): 7210–7222. Doi: 10.1109/TNNLS.2021.3084467

[49] Gabere, M. N., Hussein, M. A., Aziz, M. A., Filtered Selection Coupled with Support Vector Machines Generate a Functionally Relevant Prediction Model for Colorectal Cancer, *OncoTargets and therapy*, 2016; 9(2016): 3313–3325. Doi: 10.2147/OTT.S98910

[50] Dey, P., Artificial Neural Network in Diagnostic Cytology, *CytoJournal*, 2022; 19(27): 1-22. Doi: 10.25259/Cytojournal_33_2021

[51] Tsang, K. C., Pinnock, H., Wilson, A. M., and Shah, S. A., Application of Machine Learning Algorithms for Asthma Management with mHealth: a clinical review, *Journal of Asthma and Allergy*, 2022; 15(7): 855-873. Doi: 10.2147/JAA.S285742

[52] Diao, Y., Chen, Q., Liu, Y., He, L., Sun, Y., Li, X., Chen, Y., Li, G. and Zhao, G., A Fuzzy Granular Logistic Regression Algorithm for sEMG-based Cross-Individual Prosthetic Hand Gesture Classification, *Journal of Neural Engineering*, 2023; 20(2): 1-12. Doi: 10.1088/17412552/acc42a

[53] Grant, S. W., Hickey, G. L., Head, S. J., Statistical Primer: Multivariable Regression Considerations and Pitfalls, *European Journal of Cardio-Thoracic Surgery*, 2019; 55(2): 179-185. Doi: 10.1093/ejcts/ezy403

[54] Eberhart, R. C., Kennedy, J., A new optimizer using particle swarm theory, *6th International Conference on Micro Machine and Human Science*, New York, October 4, 1995.

[55] Xu, L., Muhammad, A., Pu, Y., Zhou, J., and Zhang, Y., Fractional-Order Quantum Particle Swarm Optimization, *Plos one*, 2019; 14(6): 1-16. Doi: 10.1371/journal.pone.0218285

[56] Richerson, P. J., Boyd, R., The Evolution of Subjective Commitment to Groups: A Tribal Instincts Hypothesis. In R. M. Nesse (Ed.), *Evolution and the capacity for commitment*, 2001; 184–220.

[57] Waring, B., Practical Optimization of Petrolium Production Systems, *United States, CreateSpace Independent Publishing Platform,* 2015.

[58] Yang, Z., Zhang, H., Sudjianto, A., and Zhang, A., An Effective SteinGLM Initialization Scheme for Training Multi-Layer Feedforward Sigmoidal Neural Networks, *Neural Networks*, 2021; 139(6): 149-157. Doi: 10.1016/j.neunet.2021.02.014

[59] Soleiman Habib, M., Improving scalability of support vector machines for biomedical nameentity recognition - Scientific figure on researchgate, *Ph.D. thesis, University of Colorado, Boulder, United States,* 2015.

[60] Mohamadi,M., Zanjirdar,M., On the Relationship between different types of institutional owners and accounting conservatism with cost stickiness, Journal of Management Accounting and Auditing Knowledge, 2018;7(28): 201-214

[61] Zanjirdar, M., Moslehi Araghi, M., The impact of changes in uncertainty, unexpected earning of each share and positive or negative forecast of profit per share in different economic condition, Quarterly Journal of Fiscal and Economic Policies,2016;4(13): 55-76.

[62] Nekounam,J., Zanjirdar, M., Davoodi Nasr,M. Study of relationship between ownership structure liquidity of stocks of companies accepted in Tehran Stock Exchange, Indian Journal of Science and Technology,2012;5(6): 2840-2845

[63] Rahmani, A., Zanjirdar, M., Ghiabi H., Effect of Peer Performance, Future Competitive Performance, and Factors of Correlation with Peer Companies on Manipulation of Abnormal Real Operations, Advances in Mathematical Finance and Applications, 2021;6(1):57-70