Advances in Mathematical Finance & Applications, 6(2), (2021), 377-392
DOI: 10.22034/amfa.2019.1872783.1252

Published by IA University of Arak, Iran
Homepage: www.amfa.iau-arak.ac.ir

Presenting a Model for Financial Reporting Fraud Detection using Genetic Algorithm

Mahmood Mohammadi, Shohreh Yazdani*, Mohammadhamed Khanmohammadi

Department of Accounting, Damavand Branch, Islamic Azad University, Damavand, Iran

ARTICLE INFO

Article history:
Received 02 July 2019
Accepted 19 December 2019

Keywords:
financial reporting fraud
fraud detection
genetic algorithm
data mining

ABSTRACT

The main objective of this study was to present a model to detect financial reporting fraud by companies listed on Tehran Stock Exchange (TSE) using genetic algorithm. For this purpose, consistent with theoretical foundations, 21 variables were selected to predict fraud in financial reporting that finally, using statistical tests, 9 variables including SALE/EMP, RECT/SALE, LT/CEQ, INVT/SALE, SALE/TA, NI/CEQ, NI/SALE, LT/XINT, and AT/LT were selected as the potential financial reporting fraud indexes. Then, using genetic algorithm, the final model of fraud detection in financial reporting was presented. The statistical population of this study included 66 companies including 33 fraudulent and 33 non-fraudulent companies from 2011 to 2016. The results showed that the presented model with the accuracy of 91.5% can detect fraudulent companies.

1 Introduction

In the past decade, many cases of financial frauds by large companies in developed and developing countries are detected and reported. The frauds by Harris Scarf in Australia, Parmalat in Italy, Ahold in India, and Vivendi in France show that this problem has penetrated into all parts of the world and is not limited to famous companies such as Enron, World Com, Tyco, Lucent, etc. Organizations lose about 5% of their annual income in different frauds and 1.6% of their annual income in reporting fraud [6]. However, there have been many efforts to assess the accurate level of frauds, but obtaining accurate and acceptable statistics is not an easy task because a majority of frauds are not detected and even when discovered, they may be ignored. This is because the victim company does not want to have a distorted face in the public. Increased numbers of frauds in financial reporting and financial restatements that are almost related to the bankruptcy of large companies have caused concerns about the quality of financial reporting. For this reason, detection of important frauds in financial reporting has always been taken into consideration by investors, legislators, managers, and auditors. Financial reporting fraud can have detrimental effects on corporate reporting to the point where its nature is compromised. Despite economic consequences of fraud for companies and economy of the country, it is not seriously taken into consideration in capital market and research centers. Also, according to increased number of companies listed on Stock Exchange and the process of privatization and capital growth, the prediction of fraud

* Corresponding author. Tel.: +98-910-480-8957
E-mail address: *m_mohammadi@tci.ir*

© 2021. All rights reserved.
Hosting by IA University of Arak Press

reporting using methods other than conventional auditing methods according to the limitations is necessary. Most of studies have employed machine learning-based models such as logistic regression, linear detection analysis, and artificial neural networks. In this study, it has been attempted to present a model to detect fraud in financial reporting using genetic algorithm.

1.1 Genetic Algorithm

Genetic algorithms are general-purpose search and optimization procedure. They are inspired by the biological evolution principle of survival of the fittest. The genetic algorithm is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms, which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover in Fig 1. Genetic Algorithms does not deal directly with the parameters of the problem to be solved. They work with codes which represent the problem and produce codes which represent the solution. A typical genetic algorithm requires:

1. A genetic representation of the solution domain
2. A fitness function to evaluate the solution domain.
3. Genetic algorithms reported to have achieved great success because of their ability to exploit accumulated information about an initially unknown search space leading to moving subsequent searches into more useful subspaces. Genetic algorithms are also flexible tools and can be used in combination with other techniques including fuzzy logic and neural networks [35].

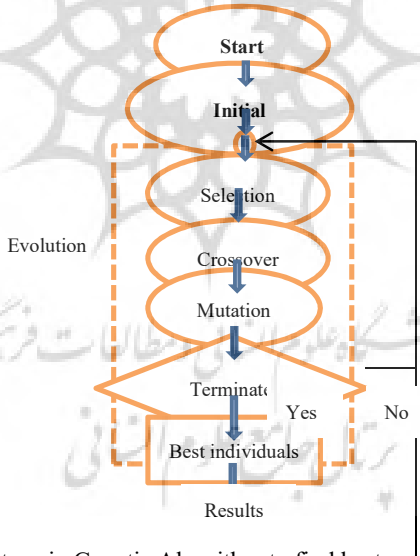

Fig. 1: Steps in Genetic Algorithm to find best optimal solution

To describe how the algorithm can be used for function optimization, consider a population of N individual χi, each represented by a chromosomal string of L allele values. Allele refers to the specific value of a gene position on the string. For example, consider the function f defined over a range of the real numbers [a,b]. a representation for each χi could be binary bit string, using standard binary encoding. Here each gene position, or allele value, takes an either 0, or 1 depending on the value of χi. the task is to maximize the output of f by searching the space [a,b]. In this case f is the fitness measure, or fitness function, and represents the environment in which candidate solutions are judged.The initial

population is generally chosen at random. Once the population is initialized the genetic algorithms evolutionary cycle can begin. The first stage of the genetic cycle is the evaluation of each of the population members. In the above this equates to evaluating each $f(\chi i)$ for all population members ($1 \leq i \leq N$). there then follows the repeated application of the biological operators. In the general case we have the following:

- Selection: selection is the process by which individuals survive from one generation to the next. A selection scheme is a means by which individuals are assigned a probability of survival based on their relative fitness to the population as a whole. Individuals with high fitness should have a high probability of surviving. Individuals with low fitness should have a low probability of surviving.
- Crossover: this is version of artificial mating. If two individuals have high fitness values, then the algorithm explores the possibility that a combination of their genes may produce an offspring with even higher fitness. Crossover represents a way of searching the space of possible solutions based on the information gained from the existing solutions.
- Mutation: if crossover is seen as a way of moving through the space based on past information, mutation represents innovation. Mutation is important, some forms of evolutionary algorithms rely on this operator as the only form of search (i.e., no crossover). In practice it is random adjustment in the individual's genetic structure.

Having applied the biological operators, the process is repeated until either the population converges (all members are the same) or some fixed control parameter is violated (such as a set number of generations). Fig. 2 depicts a typical form of crossover and mutation defined for a binary encoding of the search space. Holland crossover, depicted above, picks a position m, $1 \leq m \leq L$ (where L is the string length) at random and builds two offspring from two parents by swapping all bits in the positions $m \leq j \leq L$ on both strings. It should be noted that bit positions in this form of crossover are preserved. This allows the algorithm to exploit any linearity within the fitness function mutation for binary encodings is generally defined as a small probability that a bit value changes [3].

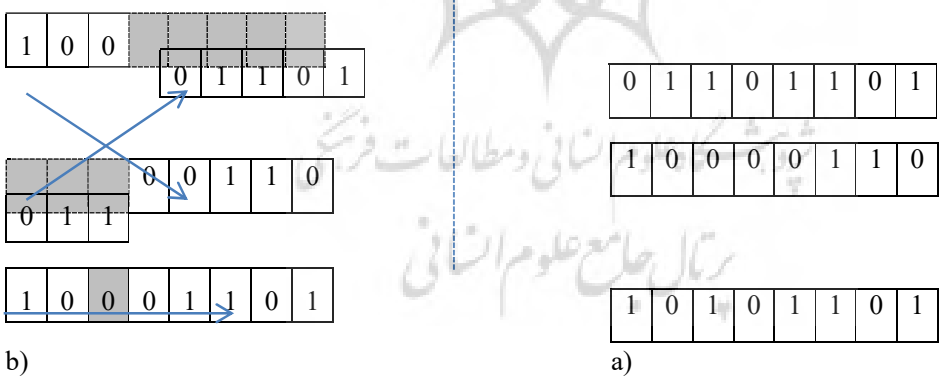

Fig. 2: Genetic operators a) Crossover b) Mutations

To complete the terminology, a set of chromosomes of an individual is referred to as its genotype, which defines a phenotype with certain fitness. A genetic algorithm is a parallel search algorithm with centralized control. The centralization comes from the selection regime. The fact that it is a parallel search relates to the fact that there is a population of candidate solutions, which sample the search space

simultaneously [3]. Genetic algorithms have many strengths well suited to the problem of fraud detection. They provide the capability to learn class boundaries that are non-linear functions of multiple variables, allowing solutions that cannot be achieved by linear methods. Furthermore, genetic algorithms perform explicit feature selection during learning. This is significant benefit in financial reporting fraud detection, as there is little consensus among experts about variables that consistently indicate fraudulent behavior. The flexibility in design of genetic algorithms allows the incorporation of logic to handle missing data and the definition of output formats using problem-domain language to enhance user understanding of the results. While prior research enhances our understanding of fraud indicators and prediction methods, this research rarely uses genetic algorithm. Our paper makes at least two important contributions. First, by introducing new method and showing that method improve prediction performance of fraud. the methodology and procedure in this paper can easily be extended to other domains of financial crimes. Second, the introduction and evaluation of this method makes an important contribution to practice. Better prediction models can, for example, help the capital market and external auditors improve their identification of potentially fraudulent accounting practices [59].

2 Theoretical Foundations and Research Background

Study on fraud detection in financial reporting necessitates identification of potential indexes of fraud detection. Studies on fraud detection in financial reporting have focused on financial and non-financial indexes affecting fraud detection. According to the primary studies on frauds in financial reporting by Feroz, Pastena, and Park [30], fraudulent companies almost have distorted inventories and amounts receivable. Later, Beneish [12] analyzed the differences between fraudulent and non-fraudulent companies and introduced amounts receivable and debt collection period as separate variables of two groups. Beneish [13], developing his suggested model identified debt collection prediction, gross profit margin, asset growth index, sales growth index, and accruals (capital change in non-working capital plus dispreciation) as potential fraud detection indexes. Ettredge, Sun, Lee, and Anandarajan [27] found that deferred taxes, auditor change, market value to book value ratio, and income growth are the evidences for reporting fraud. Also, according to Brazel, Jones, and Zimbelman [14], unusual growth in financial and non-financial variables can influence fraud prediction.

Some of these studies have tested their hypotheses using available literature about profit management and corporate governance (Beasley [9], Dechow, Sloan and Sweeney [22], Summers and Sweeney [58], Beneish [13], Sharma [57], Erickson, Hanlon and Maydew [25], Lennox and Pittman [42], Feng, Luo and Shevlin [29], Perols and Lougee [51], Armstrong, Larcker and Ormazabal [5], and Markelevich and Rosner [47]). Also, in these studies, other non-financial variables based on auditing standards have been used (Loebbecke, Eining and Willingham [44], Beneish [12], Lee, Ingram and Howard [40], Apostolou, Hassell and Webber [4], Kaminski, Wetzel and Guan [38], Etttedge, Sun, Lee and Anandarajan [27], Jones, Krishnan and Melendrez [37], Brazel, Jones and Zimbelman [14], and Dechow, Ge, Larsen and Sloan [22]). Persons [53], Summers and Sweeney [58], and Dechow et al. [23] were among the researchers who used logistic regression by financial ratios and presented a model to detect fraudulent financial reporting. Recent studies that have been conducted by logistic regression show that rapid growth of assets is accompanied by increased need for cash and external financing and this has a significant relationship with fraud. Also, increased percentage of shares owned by executives and board of directors has a significant relationship with fraud likelihood in financial statements. Green and Choi [34], Fanning and Cogger [28], and Feroz et al. [30] used multilayer neural networks to propose a model for fraud detection in financial reporting. Kaminski et al. [38], using linear detection analysis, concluded that financial ratios have limited power to detect or predict fraudulent reporting. Mc Kee [49], using a

mixture of logistic regression techniques, neural network, and decision tree designed fraud prediction model. Also, Ravisankar [57] compared five categorization techniques (i.e. artificial intelligence-based techniques and genetic programming) and presented a model to detect fraudulent companies. Perols [51] compared categorization algorithms to detect fraud in financial reporting. The results of his study showed that logistic regression enjoys from better efficiency compared with other algorithms. Alden et al. [3] investigated the profitability of categorizers based on fuzzy rules in detecting fraudulent financial reporting models. In total, 32 financial variables were selected as potential predictors of financial reporting. The results of fuzzy rules-based categorizers with genetic algorithm and Marco's estimation of distribution algorithm indicate the profitability of evolutional models to detect fraudulent financial reporting. Varian [59] signified the importance of data analysis. He recommended researchers to follow the recent advances instead of using conventional models. Lin et al. [43] in a study found that 18 fraud factors have been used. The results of logistic regression data mining, decision tree, and artificial neural networks indicated the accuracy of decision tree and artificial neural networks relative to logistic regression. In the end, in order to improve the achievements of the study, a comparison was made between the judgments of experts and data mining techniques. Finally, Perols et al. [52] concluded that to detect different frauds, only one general model cannot be taken into consideration. Therefore, to detect different types of fraud, suitable solutions for each type of fraud are needed. We summarize three types of methodologies used in literature in Table 1. Some of these researches always used of one or several methods, and compared their results with previous researches.

Table 1: Methodologies used in prior financial reporting fraud detection studie

methodology	Method(s)	References
Statistical model	Discriminant Analysis	Kaminski et al. [38], Fanning and Cogger [28]
	Logistic Regression	Bell[10], Dechow et al.[23], Persons[53], Summers and Sweenwy[59], Fanning and Cogger [28], Feroz et al. [30], Lin et al. [43], Ravisankar [57], Jahanshad and Sardarizadeh [36], Etemadi and Zolghi [26], Dastjerdi et al[21], Etemadi and Zolghi [26], Maham and Torabi [45]
	Probit	Benish[13]
Data mining	Neural network	Fanning and Cogger [28], Green and Choi [34], Feroz et al. [30], Lin et al. [43], Ravisankar [57]
	Decision Tree	Gupta et al. [35]
	Support Vector Machine	Ravisankar [57], Cecchini et al. [17]
	Genetic Programming	Ravisankar [57], Gupta et al. [35]
Text mining	Linguistic cues based method	Larcker and Zakolyukina [40]
	A computational fraud detection model	Glancy and Yadav [32]

Some recent studies have used advanced categorization techniques such as support vector machine, decision tree, genetic algorithm, and adaptive learning models (e.g. studies by Gupta and Gill [35] and Whiting et al. [63] and also data mining models such as studies by Glancy and Yadav [32], Goel and Gangolly [33], and Larcker and Zakolyukina [40]). A list of prior studies using evolutionary techniques for financial reporting fraud detection is summarized in Table 2. Additional methodological details are also provided, such as the size, methods employed and overall accuracy, when available.

In Iran, various studies are conducted on fraud detection and some of them are as follows. Maham and Torabi [45] determined the risk of fraud in financial reporting using some of financial and non-financial indexes by logistic regression. The results indicated that the presented model is able to determine fraud in financial reporting at an acceptable confidence level.

Table 2: Sample size, fraud cases and accuracy of prior financial reporting fraud detection studies

study	Sample size	Fraud cases	Method(s)	Overall accuracy (%)
Persons [53]	206	103	Logistic Regression	n/a
Fanning and Cogger [28]	204	102	Logistic Regression Discriminant Analysis Neural Networks	50 52 63
Green and Choi [34]	172	86	Neural Networks	n/a
Feroz et al. [30]	132	42	Neural Networks Logistic Regression	81 70
Lin et al. [43]	200	40	Neural Networks Logistic Regression	76 79
Kaminski et al. [38]	158	79	Discriminant Analysis	n/a
Ravisankar [57]	202	101	Support Vector Machine Genetic Programming Logistic Regression Neural Networks	72 89 91 91
Gupta et al. [35]	114	29	Decision Tree Genetic Programming	95 88
Cecchini et al. [17]	2312	107	Support Vector Machine	88
Dechow et al. [23]	1301	57	Logistic Regression	76
Beneish [13]	3538	149	Probit	49
Jahanshad and Sardarizadeh [36]	80	40	Logistic Regression	n/a
Etemadi and Zolghi [26]	68	34	Logistic Regression	84

Etemadi and Zolghi [26] investigated the use of logistic regression in detecting fraudulent financial reporting in companies listed on Stock Exchange. In this study, using 9 financial ratios and information of 34 companies with fraud signs and 34 companies without fraud signs and also logistic regression, it was attempted to develop a suitable model with the accuracy level of 83.8%. Jahanshad and Sardarizadeh [36] investigated the relationship between financial criterion (income growth difference) and non-financial criterion (staff number growth) and fraudulent financial reporting in companies listed on Stock Exchange. The results indicated a significant negative relationship between income growth, staff number growth, and fraudulent financial reporting. Mashayekhi and Hoseinpour [48] investigated the relationship between real earning management and accrual-based earning management in companies suspected of fraud in TSE. These companies were selected according to a series of factors related to false accounting information in financial reporting. The results of this study indicated that in suspected companies, real earning management has a significant negative effect on accrual-based earning management. Rahimian and Hajiheidari [56] aimed to detect frauds using the adjusted Benish model and recognize financial ratios sensitive to fraud. the results of this study shows that sales to total assets and equity to total assets ratios are two financial ratios sensitive to frauds. The model has an accuracy rate of 69/1 percent in classifying the total sample. Dastjerdi et al. [21] analyzed the text of boards reports

and used two methods including the convex optimization (CVX) method and least absolute shrinkage and selection operator (LASSO) regression method. The results indicated that both methods can detect the managers high fraud risk index with a precision between 82/55 and 91/25 percent. The LASSO method was significantly more precise than the CVX method.

3 Proposed Methodology

The main objective of the present study is to present a model to detect fraud in financial reporting of Iranian companies using genetic algorithm. Therefore, the present study, in terms of approach, is a developmental-applied, because it designs a model to detect fraud in financial reporting and uses this model to detect fraudulent and non-fraudulent companies. In terms of method, this study is a causal-correlational study. Also, in terms of relationship with the environment, this study is a quasi-experimental study. Since after data collection and estimations, the results are generalized, this study is among inductive studies. Data were collected through documentation, audited financial statements, documents, and reports published by Stock Exchange and databases. The population of this study included companies listed on TSE over the five-year period from 2011 to 2016. For sample selection, the following conditions were considered: 1. The company should be a manufacturing company; therefore, investment and holding companies and financial and financial intermediation institutions were excluded due to their different nature; 2. The fiscal year of these companies should end in March 20; 3. Their information should be available for the time period of 2011 to 2016; 4. Companies that are covered by Article 141 of the Commercial Code are excluded from the sample. Therefore, for sample selection, a list of companies that had the above conditions during 2011-2016 and were among fraudulent and non-fraudulent companies were selected. According to the above mentioned, the sample includes 66 companies that were active during this time period and their financial statements are available. Thus, since this period has been used to build a model, 330 observations are made during five years where 165 companies were non-fraudulent and 165 companies were fraudulent.

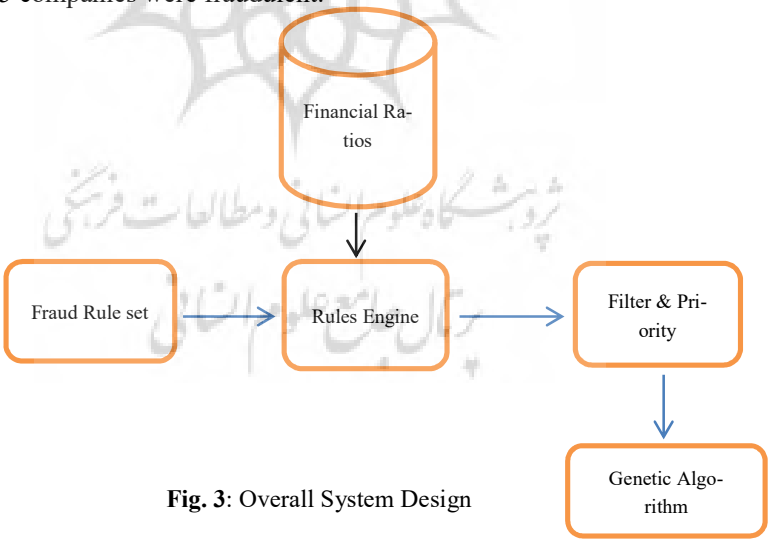

Fig. 3: Overall System Design

3.1 Architectural Design

During architectural design, some recent studies are identified. The following modules are identified in the proposed system (Fig. 3). The overall system architecture describes work structure of the system

Table 3: Fraud Predictors

Predictor	Definition	Reference
Change in Inventory	(INVT _ INVTt_1)/Average total assets	Dechow et al. [23]
Change in Recaivables	(RECT _ RECTt_1)/Average total assets	Dechow et al. [23]
Deferred Tax Expense	TXDI/ATt_1	Dechow et al. [23]
Sales to Employees	SALE/EMP	Dechow et al. [23], Perols and loge [51]
Accounts Receivable to Sale	RECT/SALE	Green and Choi [34], Feroz at al. [30], Lin et al. [43], Kaminski et al. [38], Perols et al. [51]
Accounts Receivable to Total Assets	RECT/AT	Dechow et al. [23], Green and Choi [34], Benish [12], Benish [13], Lee et al. [41], Lin et al. [43], Perols et al. [51]
Current minus prior year Inventory to Sales	INVT/SALE _ INVTt_1/SALEt_1	Summers and Sweeney [58], Perols et al. [51]
Days in Receivables index	(RECT/SALE)/(RECTt_1/SALEt_1)	Benish [13], Chen and Storey [20], Perols et al. [51]
Debt-to-Equity	LT/CEQ	Fanning and Cogger [28]
Gross Margin	(SALE _ COGS)/SALE	Green and Choi [34], Lin et al. [43], Chen and Storey [34], Perols et al. [51]
Inventory to Sales	INVT/SALE	Kaminski et al. [38], Perols et al. [51]
Net Sales	SALE	Green and Choi [34], Lin et al. [43], Cechini et al. [17], Perols et al. [51]
Sales to Assets	SALE/TA	Fanning and Cogger [28], Kaminski et al. [38], Chen and Storey [20], Cechini et al. [17], Perols et al. [51]
The number of Auditor Turnovers	if if AU ,. AUt_1, then 1, else 0 if AUt_1 ,. AUt_2, then 1, else 0 AUt_2 ,. AUt_3 then 1, ELSE 0	Feroz at al. [30], Perols et al. [51]
% change in Sales	(SALE _ SALEt_1)/SALEt_1	Cechini et al. [17]
% change in Sales to Assets	(SALE/AT _ SALEt_1/ATt_1)/(SALEt_1/ATt_1)	Cechini et al. [17]
Return On Assets	NI/AT	Cechini et al. [17]
Return On Equity	NI/CEQ	Cechini et al. [17]
Return On Sales	NI/SALE	Cechini et al. [17]
Liabilities to Interest Expenses	LT/XINT	Cechini et al. [17]
Assets to Liabilities	AT/LT	Cechini et al. [17]

in the following way:

1) The financial ratios in the data warehouse is subjected to the rules engine which consists of the financial reporting fraud rule set; and

2) The filter and priority module sets the priority for the data and then sends it to the genetic algorithm which performs its functions and generates output.

3.2 Variables and Research Model

The dependent variable of this study was fraud in financial reporting that has qualitative nature and nominal scale. In order to measure this variable, fraudulent companies get 1 and non-fraudulent companies get 0. Since there is no defined theoretical framework in financial reporting to detect and categorize economic entities as fraudulent and non-fraudulent, the employed criteria to categorize companies as fraudulent companies, like many available studies, are as follows:

- Unacceptable auditing statement
- Tax disputes with the tax area according to the statement of income tax saving and tax filing and audit report clause
- Significant severance package adjustments and financial restatements.

These criteria were selected because about the first criterion, significant fraud can lead to unacceptable statement and about the second criterion, tax dispute is mainly resulted from false interpretation of tax regulations and false employment of the clauses and in some case, delay in tax detection and liquidity maintenance are not acceptable. About the third criterion, manipulation of accruals, especially profit and loss in severance package, causes representation of financial statements and fraud in financial statements. The independent variables in this study are experimental evidences in this context. For this purpose, previous studies on fraud detection in financial reporting were studied accurately. Studies such as Green and Choi [34], Beneish [13], Summers and Sweeney [59], Fanning and Cogger [28], Lee et al. [41], Feroz et al. [30], Bell and Carcello [10], Lin et al. [43], Kaminski et al. [38], Chen and Storey [18], Perols and Loungee [51], Cecchini et al. [17], Dechow et al. [23], Perols [51], and Perols et al. [52] presented indexes of fraud in fraudulent financial reporting. The first investigation provided us with 22 financial and non-financial variables that information related to their estimation in Iran was not available and these ratios are indicated in Table 3.

Table 4: Descriptive Statistics for Fraud Predictor Variables

Sig		Non-Fraudulent Firms		Fraudulent Firms		variables	Symbol
p-value	t-test	Standard	Mean	Standard	Mean		
0.000	-9/686	0.128	0.241	0.165	-0.187	SALE/EMP	X1
0.000	-9.625	0.174	0.152	0.925	-1.163	RECT/SALE	X2
0.000	-7.443	0.223	0.049	0.661	-1.970	LT/SEQ	X3
0.000	-3.830	0.470	1.205	0.255	-0.354	INVT/SALE	X4
0.000	-7.471	0.298	0.437	0.405	-0.425	SALE/TA	X5
0.000	-4.935	0.468	1.155	0.374	0.615	NI/CEQ	X6
0.000	-7.316	0.345	0.448	0.433	-0.501	NI/SALE	X7
0.000	8.924	0.215	0.602	0.273	1.313	LT/XINT	X8
0.000	-8.040	1.375	13.525	1.316	10.678	AT/LT	X9

From the 21 initial variables, using t-test, variables that their means in two groups were significant at an error level of 1% were selected as independent variables. These variables included SALE/EMP, RECT/SALE, LT/CEQ, INVT/SALE, SALE/TA, NI/CEQ, NI/SALE, LT/XINT, and AT/LT. In Table 4, mean and standard deviation of the selected variables and significance test are shown for two research groups. As can be seen, all variables are significant at a level of 99%. Therefore, it can be said that a significant difference exists between two groups of the study.

3.2 Building Fraud Prediction Model using Genetic Algorithm

Using MATLAB version 2016, at the beginning of model building and in determining the initial parameters, 9 variables, 5 rules, 100 chromosomes for the initial population of the study, and 330 independent variables were determined. In this study, the designed algorithm coding was string algorithm and the potential answers are expressed as string and non-binary. In the next stage, to create the initial population, the sample including 300 years/companies was added to the software as the input. After creating the initial population, appropriateness of the answers with the competency function was assessed. For this purpose, the more appropriate answer, the higher competency. Therefore, the chromosome that is more competent participates in the next generation with a higher likelihood and more sequences are created and inappropriate chromosomes are removed. After creating the initial population, the competency function is estimated for the population. Then, the median is estimated. The chromosomes that their competency functions are less than the median will be removed as inappropriate answers and from the removed population, other answers are created by genetic operators and the previous stages are repeated. In this stage, chromosomes are selected for combination and creating the next generation. In this model, roulette wheel has been used to select chromosomes. After selecting the chromosomes, to create the next generations, a random number between 1 and 5 (number of rules) is created. By the resulted number, the location of two new chromosomes is created. Since only one point has been considered for integration, the integration type in this model is single-point or single-location. The final integration rate of this model is 0.7. The next stage is mutation operator. This operator causes movement in search space and storage of lost information. The mutation rate in this model is 0.07. The last stage before running the algorithm is characterization of algorithm stopping condition. In the designed model in this study, algorithm stopping occurs with one of two following events: 1. The best chromosomes are not transformed after running algorithm for 25 times; 2. Algorithms are repeated for 500 times.

4 Analysis and Findings

4.1 Running the Genetic Algorithm Model

After running the algorithm for 135 times, the obtained chromosome by the model was not transformed. In other words, the first stopping condition occurred. The obtained chromosome in this stage is as follows:

X2 < 1/12, X4 < 0/983, X5 < 0/135, X7 < 0/485, X9 < 12/189

The resulted answer in this stage is fraud prediction model in financial reporting. Therefore, if we have X2<1/12, X4<0/983, X5<0/135, X7<0/485, X9<12/189, the company is fraudulent; otherwise, the company is non-fraudulent.

4.1 Testing the Discriminating Power of Genetic Algorithm Model

The results of testing the model using data related to model estimation are presented in Table 5. As can be seen, of 165 fraudulent companies, this model has categorized 147 companies in this group

correctly and only 18 companies are categorized into non-fraudulent companies group. Also, of 165 non-fraudulent companies, 155 companies are categorized correctly and only 10 companies are categorized into fraudulent companies group. The results state that the model has categorized 89% of the fraudulent group and 94% of the non-fraudulent group correctly. Therefore, the first-type error (error in failure to detect fraudulent companies) was 11% and the second-type error (error in identifying non-fraudulent companies) was 6%. The general results of the model state that almost 91.5% of the whole sample is categorized correctly.

Table 5: Detection accuracy of classification model

	Fraudulent Firms	Non-Fraudulent Firms	Overall accuracy (%)
Fraudulent Firms	147	18	89
Non-Fraudulent Firms	10	155	94
Total accuracy			91.5

The results indicate that the model has high discriminating power to differentiate fraudulent companies from non-fraudulent companies. Therefore, model presentation can be used.

5 Discussion and Conclusion

Fraud prediction in financial reporting is one of the important discussions in financial and audit fields, because with a correct understanding and prediction of fraud likelihood and employment of necessary measures, the heavy costs can be prevented. The main objective of this study was to present a reliable model to detect fraud in financial reporting in companies listed on TSE. For this purpose, 21 variables were studied as the predictors of fraudulent financial reporting. This study uses 50/50 distribution of fraudulent and non-fraudulent firms may not reflect the real world distribution of fraudulent firms. This one-to-one matched dataset of firms can exert both positive and negative influences on the classification performance and results. However, with a one-to-one matched sample, the genetic algorithm can learn to generalize pattern of fraud from balanced mix of examples, without bias toward any one particular class, such as non-fraudulent firms. This condition may improve the ability of the model to evolve fuzzy rule based classifiers that accurately discriminate the fraudulent firms from the non-fraudulent firms. After comparing the mean values of these variables, 9 variables including SALE/EMP, RECT/SALE, LT/CEQ, INVT/SALE, SALE/TA, NI/CEQ, NI/SALE, LT/XINT, and AT/LT were selected as potential indexes of fraudulent financial reporting. After running the genetic algorithm, 5 variables of RECT/SALE, INVT/SALE, SALE/TA, NI/SALE, and AT/LT were selected as the final indexes of fraudulent financial reporting.

Generally, it can be said that these 5 variables can discover fraudulent companies, but since fraudulent financial reporting is influenced by many factors other than financial ratios, by adding other factors, it is possible to increase percentage and ability of prediction. Also, the deceptive appearance of fraudulent financial reporting makes the simulation of fraudulent data complicated and causes many problems in creating fraud detection models in financial reporting. The results indicate that the presented model has a prediction ability of 91.5%. We conclude that genetic algorithm is a successful technique for detecting discriminatory patterns in challenging domains characterized by high dimensionality and pervasive missing value. The pattern generated by this study is easily translated to domain-appropriate

language and, therefore, easily understood by users. Furthermore, the pattern is capable of identifying potentially fraudulent behavior despite occasional missing value, and provide low false-positive rates, making them practical for use by any groups. Regulatory bodies and external auditors can consider the proposed method in this study for assessing the fraud risk for a firm or other legal party. We believe that future studies should take into account the fact that fraudsters will change their tactics to hide fraud. Future researches can be extended by improving the methodology to incorporate more non-financial indicators that is germane to the domain of fraud detection. Furthermore, the methodology can be combined with extant study from other domains to predict different phenomena in real world, such as bankruptcy, abnormal returns, credit card fraud, etc. According to the findings of the study, the suggestions are as follows:

1. Investors are recommended to use this model to assess Iranian companies and decide on their stock trading.

2. Creditors, banks, and other financing institutions are recommended to use this model to assess risks as one of the risk assessment indexes to grant facilities.

3. The Stock Exchange should use this model to list companies and assess them and provide the capital market activists with the results.

4. The audits are recommended to use the presented model before accepting the work and commenting on the company's financial statements.

5. Universities and research institutes and researchers, according to the findings of this study that has focused on fraud detection in financial reporting, can rely on the presented model in future studies and development of scientific theories.

References

[1] Abbasi, A., Albrecht, C., Vance, A., and Hansen, J., *Meta Fraud: A meta-learning framework for detecting financial fraud*, MIS Quarterly, 2012, **36**(4), P.1293-1327. Doi:10.2307/41703508.

[2] Agarwal, R., Dhar, V., *Editorial-Big Data, data science, and analytics: The opportunity and challenge for IS research,* Information Systems Research, 2014, **25** (3), P. 443-448. Doi:10.1287/isre.2014.0546.

[3] Alden, M.E., Bryan, D.M., Lessley, B.J., Tripathy, A., *Detection of Financial statement fraud using Evolutionary Algoriyhms*, Journal of Emerging Technologies in Accounting, 2012, **9**(1), P.71-94. Doi:10-2308/jeta-50390.

[4] Apostolou, B., Hassell, J., and Webber, S., *Forensic expert classification of management fraud risk factors*, Journal of Forensic Accounting, 2000, **1**(2), P.181-192. Doi:10.2308/bria.2001.13.1.1.

[5] Armstrong, C.S., Larcker, D.F., Ormazabal, G., Taylor, D, J., *The relation between equity incentives and misreporting: The role of risk-taking incentives*, Journal of Financial Economics, 2013, **109**(2), P. 327-350. Doi: 10.1016/j.jfineco.2013.02.019.

[6] Association of Certified Fraud Examiners (ACFE), *Report to the Nation on Occupational Fraud and Abuse*. Austin, TX: ACFE, 2014.

[7] Athey, S., and Imbens, G., *A measure of robustness to misspecification*, American Economic Review, 2015, **105**(5), P. 476-480. Doi:10.1257/aer.p20151020.

[8] Bayley, L., Taylor, S., *Identifying earnings management: A financial statement analysis (red flag) approach,* In Proceedings of the American Accounting Association Annual Meeting, Sarasota, FL: AAA, 2007.

[9] Beasley, M., *An empirical analysis of the relation between the board of director composition and financial statement fraud,* The Accounting Review, 1996, **71**(4), P. 443-465.

[10] Bell, T., Carcello, J., *A decision aid for assessing the likelihood of fraudulent financial reporting*, Auditing: A Journal of Practice & Theory, 2000, **19**(1), P. 169-184. Doi:10.2308/aud.2000.19.1.169.

[11] Bellman, R., *Adaptive Control Processes: A Guided Tour*, Princeton, NJ: Princeton University Press, 1961.

[12] Beneish, M., *Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance*, Journal of Accounting and Public Policy, 1997, **16**(3), P.271-309. Doi:10.1016/50278-4254(97)00023-9.

[13] Beneish, M., *Incentives and penalties related to earnings overstatements that violate GAAP*, The Accounting Review, 1999, **74**(4), P. 425-457. Doi:10.2308/accr.1999.74.4.425.

[14] Brazel, J. F., Jones, K, L., Zimbelman, M.F., *Using nonfinancial measures to assess fraud risk,* Journal of Accounting Research, 2009, **47**(5), P. 1135-1166. Doi:10.1111/j.1475-679x.2009.00349x.

[15] Brown, B., Chui, M., Manyika, J., *Are you ready for the era of ''Big Data''?* McKinsey Quarterly, 2011, **4**, P.24–35.

[16] Caskey, J., Hanlon, M., *Dividend policy at firms accused of accounting fraud*, Contemporary Accounting Research, 2013, **30**(2), P. 818–850. Doi:10.111/j.1911-3846.2012.01173x.

[17] Cecchini, M., Koehler, G., Aytug, H., Pathak, P., *Detecting management fraud in public companies*, Management Science, 2010, **56**(7), P.1146-1160. Doi:10.1287/mnsc.1100.1174.

[18] Chan, P., Stolfo, S., *Toward scalable learning with non-uniform class and cost distributions: A case study in credit card fraud detection*, In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. New York, NY, 1998.

[19] Chawla, N.V., Bowyer, K, W., Hall, L.O., Kegelmeyer, W, P., *SMOTE: Synthetic minority oversampling technique*, Journal of Artificial Intelligence Research, 2002, **16**, P. 321-357. Doi:10.1613/jair.953.

[20] Chen, H., Chiang, R. H., Storey, V.C., *Business intelligence and analytics: From Big Data to big impact*, MIS Quarterly, 2012, **36**(4), P.1165-1188. Doi:10.2307/41703503.

[21] Dastjerdi, A.R., Foroghi, D., Kiani, G.H., *Detecting managers fraud risk using test analysis: evidence from Iran*, Journal of Applied Accounting Research, 2019, **20**(2), P. 154-171. Doi:10.1108/JAAR-01-2018-0016.

[22] Dechow, P. M., Sloan, R.G., Sweeney, A.P., *Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC*, Contemporary Accounting Research, 1996, **13**(1), P.1-36. Doi:10.1111/j.1911-3846.1996.tb00489.x.

[23] Dechow, P.M., Ge, W., Larson, C.R., Sloan, R.G., *Predicting material accounting misstatements*, Contemporary Accounting Research, 2011, **28**(1), P.17-82. Doi:10.1111/j.1911-3846.2010.01041x.

[24] Duin, P.W.R., Tax, M.J.D., *Experiments with classifier combining rules*, In Proceedings of the International Workshop on Multiple Classifier Systems, 2000.

[25] Erickson, M., Hanlon, M., and Maydew, E.L., *Is there a link between executive equity incentives and accounting fraud?* Journal of Accounting Research, 2006, **44**(1), P.113-143. Doi:10.1111/j.1475-679x.2006.00194x.

[26] Etemadi, H., Zolqhy, H., *Using Logistic Regression to Identify Fraudulent Financial Reporting*, Journal of Audit Science, 2013, **13**(51), P.163-145, (In Persian)

[27] Ettredge, M. L., Sun, L., Lee, P., and Anandarajan, A.A., *Is earnings fraud associated with high deferred tax and/or book minus tax levels?* A Journal of Practice & Theory, 2008, **27**(1), P.1-33. Doi:10.2308/aud.2008.27.1.1.

[28] Fanning, K., Cogger, K., *Neural network detection of management fraud using published financial data,* International Journal of Intelligent Systems in Accounting, Finance and Management, 1998, **7**(1), P. 21-41. Doi:10.2308/aud.2008.27.1.1.

[29] Feng, M., Ge, W., Luo, S., and Shevlin, T., *Why do CFOs become involved in material accounting manipulations?* Journal of Accounting and Economics, 2011, **51**(1), P. 21-36. Doi: 10.1016/j.jacceco.2010.09.005.

[30] Feroz, E., Kwon, T., Pastena, V., and Park, K., *The efficacy of red-flags in predicting the SEC's targets: An artificial neural networks approach*, International Journal of Intelligent Systems in Accounting, Finance and Management, 2000, **9**(3), P. 145-157. Doi:10.1002/1099-1174.

[31] Galar, M., Ferna´ndez, A., Barrenechea, E., Bustince, H., and Herrera, F., *A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches*, IEEE Transactions on Systems, Man and Cybernetics, Part C: Application sand Reviews, 2012, **42**(4), P.463-484. Doi:10.1109/TSMCC.2011.2161285.

[32] Glancy, F. H., Yadav, S.B., *A computational model for financial reporting fraud detection*, Decision Support Systems, 2011, **50**(3), P. 595-601. Doi: 10.1016/j.dss.2010.08.010.

[33] Goel, S., Gangolly, J., *Beyond the numbers: Mining the annual reports for hidden cues indicative of financial statement fraud*, Intelligent Systems in Accounting, Finance and Management, 2012, **19**(2), P.75–89. Doi:10.1002/isaf.1326.

[34] Green, B. P., Choi, J.H., *Assessing the risk of management fraud through neural network technology*, Auditing: A Journal of Practice & Theory, 1997, **16**(1), P. 14–28.

[35] Gupta, R., Gill, N.S., *A solution for preventing fraudulent financial reporting using descriptive data mining techniques*, International Journal of Computer Applications, 2012, **58**(1), P. 22–28.

[36] Jahanshad, A., Sardarizadeh, S., *Relation between difference of financial measure (Revenue growth) and nonfinancial measure (Employee growth) with fraudulent financial reporting*, Journal of Accounting Research, 2014, **4**(13), P.181-198, (In Persian).

[37] Jones, K. L., Krishnan, G, V., Melendrez, K.D., *Do models of discretionary accruals detect actual cases of fraudulent and restated earnings? An empirical analysis*, Contemporary Accounting Research, 2008, **25**(2). P.499-531. Doi:10.1506/car.25.2.80

[38] Kaminski, K., Wetzel, S., Guan, L., *Can financial ratios detect fraudulent financial reporting?* Managerial Auditing Journal, 2004, **19**(1). P.15-28. Doi:10.1108/02686900410509802.

[39] Kittler, J., Hatef, M., Duin, R.P.W., Matas, J., *On combining classifiers*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, **20**(3), P. 226-239. Doi:10.1109/34.667881.

[40] Larcker, D. F., Zakolyukina, A. A., *Detecting deceptive discussions in conference calls*, Journal of Accounting Research, 2012, **50**(2), P. 495-540. Doi:10.1111/j.1475-679x.2012.00450x.

[41] Lee, T. A., Ingram, R, W., Howard, T.P., *The difference between earnings and operating cash flow as an indicator of financial reporting fraud*, Contemporary Accounting Research, 1999, **16**(4), P.749-786. Doi:10.1111/j.1911-3846.1999tb00603x.

[42] Lennox, C., Pittman, J.A., *Big five audits and accounting fraud*, Contemporary Accounting Research, 2010, **27**(1), P.209-247. Doi:10.1111/j.1911-3846.2010.01007x.

[43] Lin, J., Hwang, M., Becker, J., *A fuzzy neural network for assessing the risk of fraudulent financial reporting*, Managerial Auditing Journal, 2003, **18**(8), P.657-665. Doi:10.1108/02686900310495151.

[44] Loebbecke, J.K., Eining, M.M., Willingham, J.J., *Auditors' experience with material irregularities: Frequency, nature, and detectability*, Auditing: A Journal of Practice & Theory, 1989, **9**(1), P.1–28.

[45] Maham, K., Torabi, A., *Providing Financial Reporting Fraud Risk Rating Model*, National Conference on Economic Jihad (With Emphasis On National Production, Supporting Iranian Labor and Capital), University of Mazandaran, 2012, (In Persian)

[46] Maloof, M., *Learning when data sets are imbalanced and when costs are unequal and unknown*, In Proceedings of the 20th International Conference on Machine Learning. Washington, 2003.

[47] Markelevich, A., Rosner, R.L., *Auditor fees and fraud firms*, Contemporary Accounting Research, 2013, **30**(4), P. 1590–1625. Doi:10.1111/1911-3846.12013.

[48] Mashayekhi, B., Hosseinpour, A.H., *The relationship between real earnings management and accrual earnings management in companies suspected of fraud listed in tehran stock exchange*, Empirical Studies in Financial Accounting Quarterly, 2016, **12**(49), P. 29-52. (In Persian).

[49] McKee, T.E., *A meta-learning approach to prediction financial statement fraud*, Journal of Emerging Technologies in Accounting, 2009, **6**(1), P. 5-26. Doi:10.2308/jeta.2009.6.1.5.

[50] Perols, J.L., *Financial statement fraud detection: An analysis of statistical and machine learning algorithms*, Auditing: A Journal of Practice & Theory, 2011, **30**(2), P. 19-50. Doi:10.2308/ajpt-50009.

[51] Perols, J.L., Lougee, B, A., *The relation between earnings management and financial statement fraud*, Advances in Accounting, 2011, **27**(1), P. 39-53. Doi: 10.1016/j.adiac.2010.10.004.

[52] Perols, J.L., Bowen, R, M., and Zimmermann, C., *Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Prediction*, The Accounting Review, 2017, **92**(2), P. 221-245. Doi:10.2308/accr-51562.

[53] Persons, O., *Using financial statement data to identify factors associated with fraudulent financial reporting*, Journal of Applied Business Research, 1995, **11**, P. 38-46. Doi:10.19030/jabr.v11i3.5858.

[54] Phua, C., Alahakoon, D., Lee, V., *Minority report in fraud detection: Classification of skewed data*, SIGKDD Explorations, 2004, **6**(1), P.50-59. Doi:10.1145/1007730.1007738.

[55] Price, R.A., Sharp, N.Y., and Wood, D.A., *Detecting and predicting accounting irregularities: A comparison of commercial and academic risk measures*, Accounting Horizons, 2011, **25**(4), P.755-780. Doi:10.2308/acch-50064.

[56] Rahimian, N., Hajiheidari, R., *Fraudulent Financial Statement detection Using: Adjusted-M-score-Beneish models and financial ratios,* Empirical Research in Accounting, 2019, **8**(3), P.47-70. Doi: 10.22051/jera.2018.15993.1713.

[57] Ravisankar, P., Ravi, V., Rao, G.R., Bose, I., *Detection of financial statement fraud and feature selection using data mining techniques*, Decision Support System, 2011, **50**, P. 491-500. Doi: 10.1016/j.dss.2010.11.006.

[58] Sharma, V., *Board of director characteristics, institutional ownership, and fraud: Evidence from Australia*, Auditing: A Journal of Practice & Theory, 2004, **23**(2), P.105-117. Doi:10.2308/aud.2004.23.2.105.

[59] Summers, S. L., and Sweeney, J.T., *Fraudulently misstated financial statements and insider trading: An empirical analysis*, The Accounting Review, 1998, **73**(1), P. 131-146.

[60] Varian, H.R., *Big Data: New tricks for econometrics*, Journal of Economic Perspectives, 2014, **28**(2), P.3-27. Doi:10.1257/jep.28.2.3.

[61] Walter, E., *Harnessing tomorrow's technology for today's investors and markets*, Speech Presented at American University School of Law, Washington, DC, 2013.

[62] Weiss, G., *Mining with rarity: A unifying framework*, ACM SIGKDD Explorations Newsletter, 2004, **6**(1), P.7-19. Doi:10.1145/1007730.1007734.

[63] Whiting, D. G., Hansen, J.V., McDonald, J, B., Albrecht, C., and Albrecht, W. S., *Machine learning methods for detecting patterns of management fraud*, Computational Intelligence, 2012, **28**(4), P.505-527. Doi:10.1111/j.1476-8640.2012.00425.x.

[64] Witten, I. H., Frank, E., *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition. San Francisco, CA: Morgan Kaufmann Publishers, 2005.

[65] Yang, Q., Wu, X., *10 challenging problems in data mining research*, International Journal of Information Technology and Decision Making, 2006, **5**(4), P.597-604. Doi:10.1142/S0219622006002258.