PalArch's Journal of Archaeology of Egypt / Egyptology

LITERATURE REVIEW: DETECTION OF FINANCIAL STATEMENT FRAUD

KiranMaka[1], Dr.S. Pazhanirajan[2], Dr. Sujata Mallapur[3]*

[1*]Research Scholar, Department of CSE, AnnamalaiUniveristy, Chidambaram.

[2]Assistant Professor, Department of CSE, AnnamalaiUniveristy, Chidambaram.

[3]Professor, Department of ISE, GodutaiEnggineering College for Women, Kalaburagi.

[1*]Kiran Maka

Kiran Maka, Dr.S. Pazhanirajan, Dr. Sujata Mallapur.Literature Review: Detection Of Financial Statement Fraud-- Palarch's Journal of Archaeology of Egypt/Egyptology 17(7), 4874-4886. ISSN 1567-214x

Keywords: Fraud, Financial Statement Fraud, Detection of Fraud, Financial Fraud Detection.

ABSTRACT:
In this work literature related to the detection of fraud in the financial statements are reviewed. Various typesof financial fraud are listed and challenges in the detection of fraud in financial statements are presented. Various techniques of data mining used in the fraud detection are explained. Importance of variables in the financial statement is discussed. Two important works in the fraud detection is explained with various features of the modeling. Also the approach adopted by these research works has been reviewed and best practices are presented. Also a method to remove some of the variables from the list of all variable available as part of the financial statement isexplained.

FINANCIAL FRAUD

Financial statement fraud is one of the components offinancial fraud that is very common in the industry. Before analyzing the methods used in the financial statement fraud, focus is laid on the literature connected with the financial fraud in general to understand the trends. There are many common methods adopted by people for both financial fraud and financial statement fraud. There are many sub categories in the financial fraud, namely, financial statement fraud, insurance auto fraud, automobile insurance fraud, corporate insurance fraud, health insurance fraud, accounting fraud, credit card fraud, money laundering, corporate financial fraud, occupational financial fraud, procurement transaction fraud, Tax fraud etc. The losses incurred by various industriesare:

- Insurance fraud: 70 BillionUSD

- Bank fraud: 1.5 BillionUSD
- Money laundering fraud: 50 BillionUSD
- Credit card fraud: 1.2 BillionUSD

Many organizations undergo losses due to financial fraud in various forms as listed above. Bank of America had agreed to pay around 16 Billion USD to get cleared off a case related to financial fraud [1]. Similarly, the Bixby energy systems [2] indulged in generating fraudulent financial statements related to salaries and commissions to the employees, plant operational capacity and initial public offerings in stock market. Hence any loss due to fraudulent statement is very huge and sometimes catastrophic to the institution in specific and to the economy in general. Therefore, this problem needs more research to prevent such frauds as more and more professional tools and methods are being employed by the fraudsters in this space.

Financial fraud can be detected by employing the tools or techniques available in data mining. The very first methodology adopted to determine the financial fraud is by using outlier detection technique [3]. Outlier detection is performed to determine the relationships, patterns and trends in a dataset [4]. There are many techniques in data mining [5] that can be used to detect the financial fraud suchas,

- Neuralnetworks
- Decisiontrees
- Support vectormachines
- Logisticregression
- Naïve Bayes.

The financial fraud detection was categorized into three major areas by Glancy and Yadav [6], and by Jans et al. [7] namely,

- Internal fraud
    - Financial statement fraud
    - Transaction fraud
- Creditfraud
- Insurancefraud

Of these, the financial statement fraud may be defined as the process of creating a misstatement intentionally to make the profitability appear attractive to stakeholders by modifying the financial values, whereas the transaction fraud involves snatching away some of the assets of the organization.

In the past, Ngai et al. has reviewed the literature related to detection of financial frauds in a systematic, well organized and comprehensive and in depth approach. The study [5] has focused in the literature belonging to a period between 1997 to 2008. MousaAlbashrawi has done a similar analysis on the literature of the period 2004 to 2015. Mousa has provided list of techniques that can be best used with certain contexts and the techniques that yield highest accuracy, classification frame works.

In this work, frauds related to the financial statements in specific and financial fraud in general are discussed and the literature of these research subjects is reviewed in detail. In Sec. II, the literature related to the financial statement fraud detection is explained. In Sec. III, various types of data mining techniques used in fraud detection are presented. In Sec. IV, two important works in the detection of fraud in financial statements are presented. Finally conclusions are presented in Sec. V.

FINANCIAL STATEMENT FRAUD

Financial statement may be defined as a basic document of a company that reflects the financial transactional summary of a quarter or a year and current status [29]. The financial statement may be used to understand if the company isrunning smoothly without any hassles or in a crisis. All companies report the financial statement every quarter and year to the regulatory agencies. The stakeholders review the financial statements and take important decisions to further invest into or withdraw investments from the company. Also, companies use these financial statements to avail loans from financial institutions like banks. The financial statements include income statements, balance sheets, statements of retained earnings, cash flow statements, and other statements as the companies determine asappropriate.

It is the responsibility of the accountant of the organization to detect the financial statement fraud. In majority of the cases, the financial statement frauds are caused by the accountants themselves. In such cases, auditing by external agencies is responsible to detect the frauds. Given the large number of companies that undergo the auditing by the regulatory or auditing agencies, it becomes too big a task to complete the auditing and detect the true fraudulent statements. It is very time consuming if the manual auditing is conducted. Hence there is a need to develop an automatic fraud detection system to verify if a given statement is indeed a fraudulent statement or genuine statement. Two major frauds happened in the years 2001 and 2002. The two companies involved in such major fraudulent financial statement were Enron [30] and Worldcom [31]. Enron was based in Houston, Texas and was involved in the business of pulp and paper, natural gas, electricity and communications. The revenue of Enron was around 100 Billion USD. The Worldcom filed for bankruptcy in 2002, which was the largest in the history ofUSA.

Varioustechniquesusedinthedetectionoffinancialstatement fraud are neural networks, decision trees, logistic regression, support vector machines, naïve bayes, fuzzy rule based, hybrid techniques etc. Kirkos et al. [8] estimated in 2007 that losses incurred due to fradulenet statement in USA were amounting to a total of 400 billion USD. Spathis et al. used various statistical methods like discriminant analysis and logit regression to determine the fraudulency in the financial statements [32]. Support vector machines were used by Cecchini et al. to verify the veracity of the financial statements [18]. They developed a novel kernel for this purpose. Huangetal. adopted Zipf’s law to determine the exaggerated volumesin the financial statements [33]. This method was developed to assist the auditors during the audit process. It also helped auditors to detect potential financial frauds from the statements. Kirkos et al. used advanced methods like neural networks based on Bayesian belief network and decision trees to determine the frauds in the financial statements[8].

Neural networks have been in use due to its ability to establish the non linear relationships between the input and outputs. It has found many of its application to solve prediction problems in the areas of computer vision, natural language processing and general prediction problems. It has also been extensively used in the detection of fraud in the financial statements. Sohl et al. [34] have used the neural network to determine the fraud in the financial statement. They used back propagation algorithm in the neural network. Cerullo et al. [35] explained various problems in the financial statements in various typesof industries and they also provided a list of methods based on neural networks that suits the type of financial statement in a particular industry. They illustrated the methods to use different neural network packages to determine specificfrauds. Calderon et al. [36] studies various methods based on neural networks to determine the risks associated with the audit process. The neural networks were employed as a tool tocheck risk associated with in the financial statement. It was also proposed to identify many opportunities in the detection of fraud in the financial statements. It has been established that the business risk can be ascertained with the help of neural networks very effectively. Koskivaara [37] analyzed various methods of preprocessing and its implication of detecting the fraud in financial statements. The effectiveness of the preprocessing methods was analyzed with the help of neural networks. Koskivaara [38] proposed that neural networks can be used as a tool in the audit process to aid the auditor to determine the veracity of different components in the financial statement. It has been demonstrated that the neural networks can be used to detect material errors and intentionally created flaws in the statement. Busta et al. [39] have used the neural networks to differentiate between the good, bad or manipulated data. Authors used very novel methods of determining the distribution of the digits occurring in the statement. This is known as Benford’s law. The naturally occurring digits in a number in the statement follow certain distribution and pattern. Six varieties of neural networks were used by the authors to determine the effective model. The models have been provided with 34 variables. Neuralnetworks were able to predict to an accuracy of 71% on an average. Feroz et al. [40] highlighted that neural networks and edge over the other statistical methods since neural networks learn only what is needed and important. The neural networks are based on adaptive learning methods to find if it is an important feature to predict the fraud. The neural networks are very robust and are not affected by the manipulations in the accounting process. The neural networks are capable of learning features related to the characteristics of the fraudsters since these characteristics are evident in the form of the fraudulent entries in the financial statements. Brooks [41] also tried various neural network models to determine thepresenceof fraud in the financial statements. Similarly. Fanning and Cogger [42] developed neural networks to determine the management fraud. Authors examined the publicly available variables that can be used in the modeling activity. The model was based on Autonet and it provided an empirical evidence to derive the important features in the detection of fraud in financial statements. Ramamoorti et al. [43] have developed a neural network model based on the multilayer perceptron and it has been demonstrated that these models could be used to assess the risk associated with the presence of fraud in the financial statements. Zhang et al. [44] reviewed the literature published between 1988 to 1998 where the neural networks were used. Aamodt et al.[45] and Kotsiantis et al. [46] adopted a different approach of using case by case approach to determine the fraudulent balance sheets. Probability of management fraud was computed by Deshmukh et al. [47] by developing a methodology with 15 rules. Also an early warning system was developed by the authors to alert the presence of a fraud. Pacheco et al. [48] proposed a hybrid model in which a neural network and a fuzzy expert system were integrated to analyze the financial statements.Magnusson et al. [49] used natural language processing to analyze the statements made in the quarterly financial reports to assess the status.

Table 1 Summary of Literature of Financial Statement Fraud Detection

Type of Fraud	Data mining techniques	Reference
Financial statement fraud	Text mining	6
Financial statement fraud	Discriminant analysis, Decision trees, Neural networks and Bayesian belief networks	8
Financial statement fraud	Decision trees, Neural networks, Support vector machines and kNN, Bayesian network	9
Financial statement fraud	Genetic algorithm	10, 11,12
Financial statement fraud	Growing hierarchical self- organizing map	12
Financial statement fraud	Logit model and fuzzy logic	13
Financial statement fraud	Neural networks, decision tree and logistic regression	14
Financial statement fraud	CART	15
Financial statement	Discriminant analysis,, kNN, Neural networks, Supportvector machines,multi-	16

fraud	grouphierarchical discrimination	

Financial statementfraud	Self-organizing maps and unsupervised clustering	17
Financial statement fraud	Support vector machines with custom financial kernel	18
Financial statement fraud	Three-phase cutting plane algorithm	19
Financial statement fraud	Neural networks	20
Financial statement fraud	Logistic regression	21
Financial statement fraud	Response surface method	22
Financial statementfraud	Support vector machines, genetic programming, multi- layer feedforward, group method of data handling,logistic regression, and Neural networks	23
Financial statementfraud	Logistic regression, bagging, Support vector machines, Neural networks, C4.5decision tree and stacking	24
Financial statementfraud	Logistic regression, C 4.5 decision tree, Naïve Bayes, locally weighted learning (LWL), and Support vectormachines	25
Financial statementfraud	Probit regression, logistic regression, random forests, stochastic gradient boosting, rule ensemble, and partiallyadaptive estimators	26
Financial statement fraud	Neural networks, decision tree, and logistic model	27
Financial statement fraud	Disriminant analysis and Logit regression	28

DATA MINING ALGORITHMS

Popular algorithms in the prediction of fraud in the financial statements are support vector machines, genetic algorithms, neural networks, logistic regression, discriminant analysis, naive bayes etc. The support vector machines algorithm was first introduced by Vanpik [50]. The methodology adopted in the support vector machines is using a linear model to separate non linear class boundaries. It can be achieved by finding a corresponding function in the feature space. The feature space is usually a high dimensional space. In order to separate the boundaries, a hyper-plane is created. Of all the trainingsamples, the records that are very close to the hyper-planes are known as support vectors, since these are the points on which the current position of the hyper plane is based on. If these points are modified, then the position of the hyper planes also gets modified. The performance of the support vector machines have been found be reasonably comparable with that of other methods like genetic algorithms, neural networks, logistic regression, discriminant analysis and naive bayes in financial applications like credit rating, fraud detectionetc.

Another methodology used in the determination of fraud in financial statements is with Genetic algorithms [51]. Genetic algorithm is a search algorithm to find the optimal route to reach the global minima or near global minima. Genetic algorithm is also known as a heuristic algorithm or evolutionary computation algorithm. The Genetic algorithm is based on reproduction, crossover and mutation of generations. The evolution of next population from the current population is known as generation. The common steps [52] involved in the Genetic algorithmare:

- Create the initial set ofpopulations
- Evaluate the fitness value for each thegenerations created.
- Select two or four generations that has the bestfitness values.
- Generate new population using cross overand mutation
- Repeat the above three steps until desired fitnessvalue isachieved.
- Stop the iteration once the desired fitness value is achieved and choose the generation with thedesired fitness value as thesolution.

Ivakhnenko [53] has introduced an induction learning methodology. The inductive learning algorithm is suitable when the systems are complex. It is based on testing complex systems and evaluating with external criteria. The motivation behind the inductive learning algorithm is perceptrons and learning filters. In this approach, simple models are built hierarchically, and then it retains only those models that are best performing. The simple models are utilized to form a complex model. The common steps involved in the inductive learning algorithm are:

- Learning set is used to estimate the weights of the perceptrons in thelayers.
- The error is estimated between the actual output and predicted output. The error is computed in the form of mean squareerror.
- Only those units that have lower mean square error are selected.
- Add the hidden layers to reduce the mean square error further down.
- Of all the units available, the one with lowest mean square error is thesolution.

Logistic regression [54, 55] is used when output is the classification of the records. The classification is binary in majority of the cases. For example, a binary classification can be in the form of good or bad, positive or negative, fraudulent or genuine. In majority of the cases the outcome required is toguess if a certain process is going to be a success or failure. If it is a success, it is coded as 1 and for failure, it is coded as 0. The records that are used to estimate the parameters of the logistic regression model are labeled with 1 or 0 based on the historical observations. A mathematical model, usually alinear model is defined. The number of parameters of the model is one more than the number of features selected if it is a linear model. For example, if eighteen features are there in the data, then the total number of parameters is nineteen. Thenineteenth parameter is the bias term. Some of the features [23] used in the detection of fraud in the financial statementare:

- Gross profit of theorganization
- net profit of theorganization
- Primary businessincome
- Ratio of primary business income to Totalassets
- Ratio of inventory to Primary businessincome
- Ratio of inventory to Totalassets
- Ratio of inventory to Currentliabilities
- Ratio of gross profit to Totalassets
- Ratio of net profit to Totalassets
- Ratio of net profit to Primary businessincome
- Ratio of primary business income to Fixedassets
- Ratio of primary business profit to Primarybusiness profit of previousyear
- Ratio of primary business income to previousyear's primary businessincome
- Ratio of fixed assets to totalassets
- Ratio of current assets to currentliabilities
- Ratio ofcapitals and reserves to totaldebt
- Ratio of long term debt to total capital andreserves
- Cash andDeposits

RESULTS AND DISCUSSION

Inthissection,twoimportantworksinthedetectionoffraudin financial statements are reviewed. Ravisankar et al. [23] derived 18 variables that are very important in the detection of financial statement fraud. The terms represent fraud triangle. These features are derived to represent the profitability, liquidity, efficiency and safety of the banking system. The dataset contained almost 2660 financial statements of different companies. These statements belonged to those companiesthat were listed in a China’s stock market. All these variables were first normalized and then three clusters were formed. Clusters were named as fraudulent, non-fraudulent and suspicious. Of the 2660 financial statements, each one was categorized into any of these three clusters. The suspicious cluster was subjected to further investigation as the algorithm was not able to clearly distinguish between the fraudulent and non-fraudulentcategories.

In another research, Ravisankar et al. [23] 202 Chinese companies were considered in the analysis. The data set is a balanced dataset in which there is equal number of fraudulent and non-fraudulent cases. Since each variable has its own range and some of the numbers are huge, a logarithmic transformation was carried out. Multi-layer feed forward neural network, Group method of datahandling andProbabilistic neural network [32] were used in this work. The logistic regression and support vector machines were also used from KNIME

[56]. Authors proved that Probabilistic neural network [32] yielded best accuracy of 98.09% among all the methods tried.

In a study [57] 31 financial indices have been considered in the analysis. Of the 31, seven indices were not considered as these seven variables have very correlation with the remaining variables. These variables are excluded to avoid the multi- collinearity issue. The seven indices that have high correlation with the remaining variablesare:

- Liquidity
- Profitabilityratios
- Activityratios
- Solvencyratios
- Marketrations
- Accrued income
- Cashflow

The data that was collected as part of this work was normalized. The normalization is carried out in the form of standardization. Each value in the variable is subtracted by the mean of the variable and divided by the standard deviation. The missing value imputations were performed. The missing values were estimated using the statistical distribution. Standard normal distribution was used to predict the missing values in the variables. A multivariate normal distribution was used in the prediction of fraud [57-62]. The companies have been rated as A, B and U. All the financial statement has been categorized as A, B or U-Rated. The A-rated companies have constituted 68.89%, B-rated as 7.6% and U rated companies have constituted 23.51%. It has been concluded that 68.89% of the financial statements have very few anomalies and 23.51% of the companies have significant anomalies.

CONCLUSIONS

In this literature review, literature related to the detection of financial fraud in general and detection of financial statement fraud in specific is reviewed and important observations are presented. Two important works are discussed in detail in the results section. It has been identified that 18 variables are considered as the important features in finding the fraud in the financial statements. The variables that occur in the financial statements have a huge range of values and hence it needed to be transformed, normalized and standardized. The variables needed to be imputed with statistical distributions orprediction models if there are any values missing in the data. It is also highlighted that finding these variables that have high correlation with other variables in the data is very important and hence it needed to be removed from the list of variables as it will result in multi collinearity. In most of the literature, the financial statements are classified into fraudulent, non-fraudulent and suspicious categories. The fraudulent and non- fraudulent category has very few anomalies in the detection. In case suspicious category, there are some confusing anomalies which makes it difficult to say in certain that it belong to a fraudulent or non-fraudulent category. Hence more research is required to understand these kinds of financial statements. Of all the models tried by various researchers, theprobabilisticneural network models yielded very good results in detecting the fraud in the financial statements.

REFERENCES

Fox News. Bank of America pays $16.5 bn to settle financial fraud case. Fox News Latino: http://latino.foxnews.com/latino/news/2014/08/21/bank-america-pays-165-bn-to-settle-financial-fraud-case/,2014.

IRS. Examples of Corporate Fraud Investigations. IRS: http://www.irs.gov/uac/Examples-of-- Corporate-Fraud-Investigations-Fiscal-Year-2014, 2014.

Jayakumar, G.D.S., & Thomas, B.J. A New Procedure of Clustering based on Multivariate Outlier Detection. *Journal of Data Science* 2013; 11:69-84.

Hassani H, Gheitanchi S, Yeganegi MR. Onthe Application of Data Mining to Official Data. *Journal of Data Science* 2010; 8:75-89.

Ngai E, Hu Y, Wong Y, Chen Y, Sun X. The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature. *Decision Support Systems* 2011; 50:559–569.

Glancy FH, Yadav SB. A Computational Model for Financial Reporting Fraud Detection. *Decision Support Systems* 2011; 50:595–601.

Jans M, Werf JM, Lybaert N, Vanhoof K. A Business Process Mining Application for Internal Transaction Fraud Mitigation. *Expert Systems with Applications* 2011; 38: 13351–13359.

Kirkos E, Spathis C, Manolopoulos Y. Data Mining Techniques for the Detection of Fraudulent Financial Statements. *Expert Systems with Applications* 2007; 32: 995– 1003.

Kotsiantis S, Koumanakos E, Tzelepis D, Tampakas V. Forecasting Fraudulent Financial Statements Using Data Mining. *International Journal of Computational Intelligence* 2006; 3: 104–110.

Chai W, Hoogs BK, Verschueren BT. Fuzzy Ranking of Financial Statements for Fraud Detection. *In proceeding of IEEE International Conference on Fuzzy System* 2006; 152– 158.

Hoogs B, Kiehl T, Lacomb C, Senturk, D. A Genetic Algorithm Approach to Detecting Temporal Patterns Indicative of Financial Statement Fraud. Intelligent Systems in Accounting, *Finance and Management* 2007; 15:41-56.

Huang SY, Tsaih RH, Yu F. Topological Pattern Discovery and Feature Extraction for Fraudulent Financial Reporting. *Expert Systems with Applications* 2014; 41: 4360– 4372.

Lenard MJ, Watkins AL, Alam P. Effective Use of Integrated Decision Making: An Advanced Technology Model for Evaluating Fraud in Service-Based Computer and Technology Firms. *The Journal of Emerging Technologies in Accounting* 2007; 4:123-137.

Liou FM. Fraudulent Financial Reporting Detection and Business Failure Prediction Models: A Comparison. *Managerial Auditing Journal* 2008; 23:650-662.

Bai B, Yen J, Yang X. False Financial Statements: Characteristics of China's Listed Companies and CART Detecting Approach. *International*

Journal of Information Technology & Decision Making 2008; 7:339–359.

Gaganis C. Classification Techniques for the Identification of Falsified Financial Statements: A Comparative Analysis. *International Journal of Intelligent Systems in Accounting and Finance Management* 2009; 16: 207-229.

Deng Q, Mei G. Combining Self-Organizing Map and K- Means Clustering for Detecting Fraudulent Financial Statements. *In IEEE International Conference on Granular Computing* 2009;126-131.

Cecchini M, Aytug H, Koehler G, Pathak P. Detecting Management Fraud in Public Companies. *Management Science* 2010; 56: 1146-1160.

Dikmen B, Küçükkocaoğlu G. The Detection of Earnings Manipulation: The Three-Phase Cutting Plane Algorithm Using Mathematical Programming. *Journal of Forecasting* 2010; 29: 442-466.

Kapardis MK, Christodoulou C, Agathocleous M. Neural Networks: The Panacea in Fraud Detection? *Managerial Auditing Journal* 2010; 25:659-678.

Dechow P, Ge W, Larson C, Sloan R. Predicting Material Accounting Misstatements. *Contemporary Accounting Research* 2011; 28: 1-16.

Zhou W, Kapoor G. Detecting Evolutionary Financial Statement Fraud. *Decision Support Systems* 2011; 50: 570– 575.

Ravisankar P, Ravi V, Rao GR, Bose I. Detection of Financial Statement Fraud and Feature Selection Using Data Mining Techniques. *Decision Support Systems* 2011; 50: 491– 500.

Perols J. Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. *Auditing: A Journal of Practice & Theory* 2011; 30:19-50.

Humpherys SL, Moffitt KC, Burns MB, Burgoon JK, Felix WF. Identification of Fraudulent Financial Statements Using Linguistic Credibility Analysis. *Decision Support Systems* 2011; 50:585–594.

Whiting DG, Hansen JV, Mcdonald JB, Albrecht C, Albrecht WS. Machine Learning Methods for Detecting Patterns of Management Fraud. *Computational Intelligence* 2012; 28: 505-527.

Lin CC, Chiu AA, Huang SY, Yen DC. Detecting the Financial Statement Fraud: The Analysis of the Differences between Data Mining Techniques and Experts' Judgments. *Knowledge-Based Systems* 2015; 89:459-470.

Li SH, Yen DC, Lu WH, Wang C. Identifying the Signs of Fraudulent Accounts Using Data Mining Techniques. *Computers in Human Behavior* 2012; 28:1002–1013.

W.H. Beaver, Financial ratios as predictors of failure, *Journal of Accounting Research* 4 (1966)71–111.

http://en.wikipedia.org/wiki/Enron.

http://en.wikipedia.org/wiki/MCI_Inc.

Neuroshell 2.0, Ward Systems Inc.http://www.wardsystems.com.

S.-M. Huang, D.C. Yen, L.-W. Yang, J.-S. Hua, An investigation of Zipf's Law for fraud detection, *Decision Support Systems 46* (1) (2008)70–83.

J.E. Sohl, A.R. Venkatachalam, A neural network approach to forecasting model selection, *Information & Management 29* (6) (1995)297–303.

M.J. Cerullo, V. Cerullo, Using neural networks to predict financial reporting fraud: Part 1, *Computer Fraud & Security 5* (1999)14–17.

T.G. Calderon, J.J. Cheh, A roadmap for future neural networks research in auditing and riskassessment,*InternationalJournalofAccountingInformationSystems3* (4) (2002) 203–236.

E. Koskivaara, Different pre-processing models for financial accounts when using neural networks for auditing, *Proceedings of the 8th European ConferenceonInformation Systems,* vol. 1, 2000, pp. 326–3328, Vienna, Austria.

E. Koskivaara, Artificial neural networks in auditing: state of the art, *The ICFAI Journal of Audit Practice 1* (4) (2004) 12–33.

B. Busta, R. Weinberg, Using Benford's law and neural networks as a review procedure, *Managerial Auditing Journal 13* (6) (1998)356–366.

E.H. Feroz, T.M. Kwon, V. Pastena, K.J. Park, 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 9* (3) (2000)145–157.

R.C. Brooks, Neural networks: a new technology, The CPA Journal Online,http://www.nysscpa.org/cpajournal/old/15328449.htm1994.

K.M. Fanning, K.O. Cogger, Neural network detection of management fraud using published financial data,*International Journal of Intelligent Systems inAccounting,Finance, and Management 7* (1) (1998) 21–41.

S. Ramamoorti, A.D. Bailey Jr., R.O. Traver, Risk assessment in internal auditing: a neural network approach, *International Journal of Intelligent SystemsinAccounting, Finance & Management 8* (3) (1999) 159–180.

G. Zhang, B.E. Patuwo, M.Y. Hu, Forecasting with artificial neural networks: the state of the art, *International Journal of Forecasting 14* (1) (1998)35–62.

A. Aamodt, E. Plaza, Case-based reasoning: foundational issues, methodological variations, and system approaches, *Artificial Intelligence Communications 7*(1)(1994) 39–59.

S. Kotsiantis, E. Koumanakos, D. Tzelepis, V. Tampakas, Forecasting fraudulent financial statements using data mining, *International Journal ofComputationalIntelligence 3* (2) (2006) 104–110.

A. Deshmukh, L. Talluru, A rule-based fuzzy reasoning system for assessing the risk of management fraud, *International Journal of Intelligent SystemsinAccounting, Finance & Management 7* (4) (1998) 223–241.

R. Pacheco, A. Martins, R.M. Barcia, S. Khator, A hybrid intelligent system applied to financial statement analysis, *Proceedings of the 5th IEEE conference onFuzzySystems, 2*, 1996, pp. 1007–10128, New Orleans, LA, USA.

C. Magnusson, A. Arppe, T. Eklund, B. Back, H. Vanharanta, A. Visa, The language of quarterly reports as an indicator of change in the company's financial status, *Information & Management 42* (4) (2005)561–574.

V. Vapnik, *Adaptive and learning systems for signal processing, in: HaykinS. (Ed.), Statistical Learning Theory,* John Wiley and Sons,1998.

J.R. Koza, *Genetic programming: on the programming of computers by means* *of natural selection,* MIT press, Cambridge, MA, 1992.

K.M. Faraoun, A. Boukelif, Genetic programming approach for multi-category pattern classification applied to network intrusion detection, *International Journal of Computational Intelligence and Applications 6* (1) (2006) 77– 99.

A.G. Ivakhnenko, The group method of data handling—a rival of the method of stochastic approximation, *Soviet Automatic Control 13* (3) (1966)43–55.

R. Pacheco, A. Martins, R.M. Barcia, S. Khator, A hybrid intelligent system applied to financial statement analysis, *Proceedings of the 5th IEEE conference on Fuzzy Systems, 2,* 1996, pp. 1007–10128, New Orleans, LA,USA.

D.P. Williams, V. Myers, M.S. Silvious, Mine classification with imbalanced data, *IEEE Geoscience and Remote Sensing Letters 6* (3) (2009)528–532.

KNIME 2.0.0. http://www.knime.org

Tran, V., Lokanan, M., &Hoai Nam, V. (Accepted). Detecting Anomalies in Financial Statements Using Machine Learning Algorithm: The Case of Vietnamese Listed Firms. *Asian Journal of AccountingResearch.*

Lokanan, M.E. (2017), “Theorizing Financial Crimes as Moral Actions”, *European Accounting Review*, pp. 1–38.

Lokanan, M.E. and Sharma, S. (2018), “A Fraud Triangle Analysis of the Libor Fraud”, *Journal of Forensic and Investigative Accounting, 10*(2), 187–212.

Li, Z. (2016). "Anomaly detection and predictive analytics for financial risk management." https://rucore.libraries.rutgers.edu/rutgers-lib/49363/ (accessed 21 September2018).

Harymawan, I. & Nurillah, D. (2017) "Do Reputable Companies Produce a High Quality of Financial Statements?", *Asian Journal of Accounting Research, 2* Issue: 2, pp.1- 7, https://doi.org/10.1108/AJAR-2017-02-02-B001

Hajek, P. and Henriques, R. (2017). "Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods", *Knowledge-Based Systems*, *125*(15), 139-152.