DETECTION OF FINANCIAL INFORMATION MANIPULATION BY AN ENSEMBLE-BASED MECHANISM

Ching-Hui Shih[*], *Sin-Jin Lin*[†], *Ming-Fu Hsu*[‡]

Abstract: Complicated financial information manipulation, involving heightened offender knowledge of transactional procedures, can be damaging to the reputations of corporations and the auditors, as well as cause serious turbulence in financial markets. Unfortunately, most incidents of financial information manipulation involve higher level managers who are truly knowledgeable and comprehend the limitations of standard auditing procedures. Thus, there is an urgent need for additional detection mechanisms to prevent financial information manipulation. To address this problem, the author proposes an ensemble-based mechanism (EM) consisting of feature selection and extraction ensemble and extreme learning machine (ELM). The model not only counters the redundancy-removing problem, but also gives direction to auditors who need to allocate limited audit resources to abnormal client relationships during the auditing procedure and protect the CPA firms' reputation. The experimental results demonstrate that the model is a promising alternative for detecting financial information manipulation, and one that can ensure both the confidence of investors and the stability of financial markets.

Key words: *Feature selection and extraction ensemble, decision making, extreme learning machine, financial information manipulation*

Received: June 26, 2012 **DOI:** 10.14311/NNW.2014.24.028
Revised and accepted: October 12, 2014

1. Introduction

Since the boom of information technology and the invention of modern devices, there have been an extreme increase in financial information manipulation (deceptive cases) associated with all aspects of the real business world [64]. These deceptive cases are generally composed of credit card fraud, e-commerce transaction fraud, insurance fraud, telecommunication fraud, money laundering, computer

[*]Ching-Hui Shih, Department of Accounting and Information Systems, Asia University, 500, Lioufeng Rd., Wufeng, Taichung, Taiwan 41354, ROC, E-mail: `chshih168@gmail.com`

[†]Sin-Jin Lin – Corresponding Author, Department of Accounting, Chinese Culture University, 55, Hwa-Kang Rd., Yang-Ming-Shan, Taipei, Taiwan 11114, ROC, Tel.: +886 7-2241588; E-mail: `annman1204@gmail.com`

[‡]Ming-Fu Hsu, Department of Information Management, National Chi Nan University 1, University Rd., Puli, Nantou, Taiwan 545, ROC, E-mail: `hsumf0222@gmail.com`

©CTU FTS 2014

instruction fraud and financial statement fraud (FSF). In particular, FSF has had a dramatic adverse impact not only on the individual investor, but also on the overall stability of worldwide economies. According to a 2008 technical report from the Association of Certified Fraud Examiners [1], FSF imposes the highest per case cost and total cost to defrauded organizations, up to US$ 572 billion per year in funds. By undermining the reliability of financial statements and investors' confidence in the stock market, FSF also has numerous side effects on market participants, including higher risk premiums and a less efficient capital market [50].

The American Institute of Certified Public Accountants (AICPA) announced the SAS No. 82 to specifically acknowledge the responsibility of auditors in detecting FSF. While auditors are the last line of defense in cases of suspected fraud, many auditors lack the experience and expertise to tackle the related duties. This limitation is compounded by the sampling approach that auditors often use [40]. The problem with this auditing procedure is that it is contingent on the information gathered being representative of all items in the corporation. However, in an age of information explosion, auditors who execute limited auditing procedures find it complicated to identify valuable information from an overabundance of data. Unfortunately, most incidents of FSF involve top managers who are truly knowledgeable and have the necessary resources to outwit the system and fool many detection techniques [11]. Motivated by this challenge, numerous researches have attempted to construct models that will forecast the presence of FSF. Beasley [2] applied logistic regression to detect the FSF and indicated that the non-FSF corporations have a higher proportion of outside members than FSF corporations do. Fanning and Cogger [13] proposed a model in which they developed by an artificial neural network for detecting FSF, while Summers and Sweeney [58] established a cascaded logit model to evaluate the relationship between insider trading and FSF. Kirkos et al. [29] explored the effectiveness of detecting FSF with three different kinds of data mining techniques: support vector machine (SVM), decision tree (DT) and Bayesian network (BN). The results indicated that the BN performed a satisfactory job in terms of forecasting accuracy. Doumpos et al. [12] utilized a SVM to deal with the FSF problem. Many of the aforementioned studies emphasize the design of more sophisticated detecting models, but quite a few researchers [47] focused on dimensionality reduction, which is an inevitable preprocessing step in data mining. When there is an overabundance of irrelevant information, it is unlikely to discriminate and interpret the useful information very easily. Therefore, determining how to filter out redundant or irrelevant features from the original data is an essential issue to FSF.

Feature selection (FS) and feature extraction (FE) are two different methods for feature dimensionality reduction [34]; the former determines the set of input variables that actually affect the output from a pre-defined set of input candidates; these input variables reduce the original set of features into a linear or nonlinear transformed set [3]. The advantage of FS is that it not only simplifies the design and implementation of a detection model, but also increases the processing speed of calculation, leading to better response time [63]. Rough set theory (RST) has been used as such a dataset pre-processor with much success; however, it is reliant on a crisp dataset; essential information may be lost as a result of the quantization of the underlying numerical values [54]. Therefore, many researchers have greatly emphasized this problem and have introduced numerous methods to handle it [31]. Shen and Jensen [54] introduced an emerging approach, namely fuzzy rough set theory (FRST) which is a hybridized fuzzy set theory (FST), and rough set theory (RST). The basic concept of FRST is both interesting and systematic, since by considering the degree of belongingness of real-valued data to fuzzy sets, essential information loss may be recovered. A primary application of FRST is to eliminate irrelevant features without deterioration of the forecasting performance, reduce the storage capability and facilitate the speed of execution [22, 25].

The random forest (RF) belongs to the family of ensemble approaches which appeared in artificial intelligence at the end of nineties [16]. Each tree in RF casts a unit vote for the most popular class, and the final output of the classifier is determined by a majority vote of the tree [18, 57]. Since each tree in RF only utilizes a portion of the input variables, the computational complexity is considerably lighter than is the case with conventional bagging used in conjunction with a comparable tree-shaped classifier. Based on its mechanism of permutation evaluation, RF can measure the importance of each attribute or feature [5]. Although FRST and RF were widely employed in feature selection and attribute reduction [6, 10], few researchers examined the effectiveness of both approaches on the FSF issue.

In addition to FS, another way to eliminate the dimensionality reduction problem is by using FE to map high-dimensional data into a low-dimensional subspace [32, 61]. The data in this process enhances visualization by determining essential low-dimensional representations of high-dimensional data, as well as identifying the intrinsic data structure. Principal component analysis (PCA) and linear discriminant analysis (LDA), which aim to discover a linear subspace as a feature space, to preserve specific characteristics of observed data, are two well-known feature extraction techniques. In general, the linear subspace approach performs a satisfactory job on feature extraction for linear structure data; however, most real-world data are nonlinear. Recently, many approaches have been proposed to tackle the nonlinear structural data, such as isometric feature mapping (ISOMAP) [59] and local linear embedding (LLE) [53]. Unlike the aforementioned linear approach, both of these methods intend to calculate and preserve the geometric signatures of manifold; they also share the same fundamental phases: (1) the neighborhood is established in the input space; (2) a square matrix is calculated with as many rows as elements in the input data set; and (3) spectral embedding is computed by adopting the eigenvectors of this matrix [14].

However, the same classifier embedded in different FS and FE approaches will lead to quite different outcomes. This means there is no exact solution to the problem of what is the most expressive combination strategy (see Fig. 1). In addition, none of the related work considers the effectiveness of combining FS and FE in the same time. Determining how to yield a thorough evaluation of combination strategies and determine the optimal one is an emergent problem in FSF due to the high time-pressure audit environment. In fact, the problem of strategy (algorithm) selection is an active research domain in many fields, such as artificial intelligence, operation research and machine learning [55]. Rokach [51] suggested that the strategies selection can be translated into a multiple criteria decision making (MCDM) problem and that an MCDM algorithm can be utilized to systematically determine an adequate strategy. The ensemble-based mechanism (EM) with two-phase dimensionality reduction (FS and FE ensemble) can be used to alleviate the curse of dimensionality, facilitate data visualization, and decrease the storage requirement. Extreme learning machine (ELM) [9] shows as an effective learning approach to train single-hidden layer feed forward neural network (SLFNs) which have been extensively utilized in numerous research domains because of its feasibility of directly approximating nonlinear mappings by input data and generating models for a number of natural and artificial tasks that are complicated to tackle with by conditional parametric approaches [60, 65]. Thus, the ELM was taken as a base classifier. The empirical results derived from EM (FS and FE ensemble + ELM) can generate a suitable direction for auditors to make a reliable judgment, and effectively allocate limited audit resources to abnormal client relationships during the auditing period, as well as protect the CPA firms' reputation.

The rest of this paper is organized as follows. Section 2 presents the methodologies for combination strategy evaluation and introduces the hybrid intelligent model designed for detecting FSF. Section 3 describes the experimental setup, including the chosen dataset, the combination strategies of feature selection ensemble and the evaluation methods. Section 4 concludes the study.

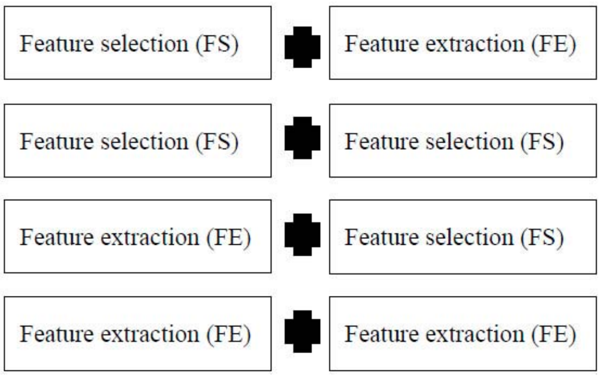

Fig. 1 *The structures of dissimilar combination strategies.*

2. Methodologies

2.1 Extreme learning machine: ELM

In this section, we present a brief overview of the extreme learning machine (ELM) technique [23-24]. For N arbitrary distinct instances $(\mathbf{x}_g, \mathbf{t}_g) \in \Re^n \times \Re^m$, the normal single-hidden-layer feedforward neural network (SLFN) with K hidden nodes and activation function $H(\cdot)$ is expressed as follows:

$$\sum_{i=1}^{K} \lambda_i H(\mathbf{x}_g; \gamma_i, \nu_i) = \mathbf{t}_g, \quad g = 1, \ldots, N \tag{1}$$

where $\gamma_i \in R$ and $\nu_i \in R^n$ denotes the randomly assigned bias of the i-th hidden node and the randomly assigned input weight vector connecting the i-th hidden and input nodes, respectively. The weight vector connecting the i-th hidden node to the output node was represented as λ_i. The output of the i-th hidden node with respect to the input instance $\mathbf{x}_g$ was expressed as $H(\mathbf{x}_g; \gamma_i, \nu_i)$. Sequentially, the Eq. (1) can be depicted as follows:

$$G\lambda = T \tag{2}$$

where

$$G = \begin{bmatrix} H(\mathbf{x}_1; \gamma_1, \nu_1) \cdots H(\mathbf{x}_1; \gamma_K, \nu_K) \\ \vdots \quad\quad \ddots \quad\quad \vdots \\ H(\mathbf{x}_N; \gamma_1, \nu_1) \ldots H(\mathbf{x}_N; \gamma_K, \nu_K) \end{bmatrix}_{N \times K} \tag{3}$$

$$\gamma = (\gamma_1^T \gamma_2^T \ldots \gamma_K^T)^T_{m \times K} \tag{4}$$

and the target output was represented in Eq. (5):

$$T = (\mathbf{t}_1^T \mathbf{t}_2^T \cdots \mathbf{t}_L^T)^T_{m \times N} \tag{5}$$

The output weights can be computed by determining the least-square solutions to the above linear structure is given as follows:

$$\hat{\lambda} = G^+ T$$

where G^+ denotes the Moore-Penrose generalized inverse of the hidden layer output matrix G. Computation of the output weights is done in a single step here. Thus this avoids any lengthy training process where the network parameters are modified iteratively with suitably chosen control parameters (such as learning rate and learning epochs, etc.) [52].

2.2 The MCDM approach: VIKOR

The VIKOR method was proposed by Opricovic [42] and Opricovic and Tzeng [43] for multi-criteria optimization of complicated problems. Opricovic [42] indicated that the VIKOR ranks alternatives in the occurrence of conflicting criteria by generating a multi-criteria ranking index ground on a specific evaluation of closeness to the ideal alternative. The VIKOR is expressed as follows:

Procedure 1: Calculate the best g_i^* and the worst g_i^- values of all the criteria functions, $i = 1, \ldots, n$. If the i-th function represents a benefit, then the following equations results:

$$\begin{aligned} g_i^* &= \left\{ \begin{array}{lll} \max_j & g_{ij}, & \text{for benefit criteria} \\ \min_j & g_{ij}, & \text{for cost criteria} \end{array} \right\}, j = 1, \ldots, J \\ g_i^- &= \left\{ \begin{array}{lll} \max_j & g_{ij}, & \text{for benefit criteria} \\ \min j & g_{ij}, & \text{for cost criteria} \end{array} \right\}, j = 1, \ldots, J \end{aligned} \tag{6}$$

where the number of alternatives is denoted as J, and the number of criteria is expressed as n.

Procedure 2: Calculate the values of X_j and Y_j, $j = 1, \ldots, J$ as

$$\begin{aligned} X_j &= \sum_{i=1}^{n} [w_i \, (g_i^* - g_{ij})/(g_i^* - g_i^-)] \\ Y_j &= \max_i [w_i \, (f_i^* - f_{ij}) \, / (f_i^* - f_i^-)] \end{aligned} \tag{7}$$

where the weight of the i-th criterion is expressed as w_i, and the ranking evaluation is measured by X_j and Y_j.

Procedure 3: Calculate the value of Z_j, for $j = 1, \ldots, J$ as

$$\begin{aligned} Z_j &= [v(X_j - X^*)/(X^- - X^*)] + [(1 - v)(Y_j - Y^*)/(Y^- - Y^*)] \\ X^* &= \min_j X_j, \, X^- = \max_j X_j \\ Y^* &= \min_j Y_j, \, Y^- = \max_j Y_j \end{aligned} \tag{8}$$

where X^* is the solution with the maximum group utility, Y^* is the solution with a minimum single regret of the opponent, and the weight of the strategy of the majority of criteria is represented as v. This compromise solution is stable within a decision making process, which could be: "voting by majority rule" (when $v > 0.5$ is need), or "by consensus" $v \approx 0.5$ or "with veto" $v < 0.5$ [41]. Followed by the prior researches [44], the value of v is set to 0.5

Procedure 4: Ranks the alternatives in decreasing order. There are three ranking lists: X, Y and Z.

Procedure 5: Generate the alternative b', which is measured by Z and ranks the best, as a compromise solution if the following two conditions are satisfied [48]:

(a) $Z(b'') - Z(b') \geq 1 - (J - 1)$
(b) Alternative b' is ranked the best by X and/or Y.

If only condition (b) is violated, the alternatives b' and b'' are taken as compromise solutions, where b'' is measured by Z and is ranked second. If condition (a) is violated, alternatives $b', \ldots, b^M$ are viewed as compromise solutions, where b^M is evaluated by Z and is ranked the M-th according to the relation $Z(b^M) - Z(b') < 1(J - 1)$ for maximum M.

3. The ensemble-based mechanism: EM

Fig. 2 illustrates the architecture of EM for detecting FSF, with financial and corporate governance data collected from public websites. Not all of the data features were informative or essential for an FSF task, as some of the collected data were redundant or useless. Without a data cleaning process, raw data causes confusion in the mining procedure and leads to improper judgments. It is acknowledged that high-dimensional data pose numerous challenges, such as complicated computational ability, a huge storage requirement, and inferior outcome performance. Thus, this research utilizes FS to identify the optimal subset, which is essential and sufficient for solving the task, and employs FE to discover meaningful low-dimensional representations of high-dimensional data and simultaneously find the underlying pattern structure. As no previous study has examined the usefulness

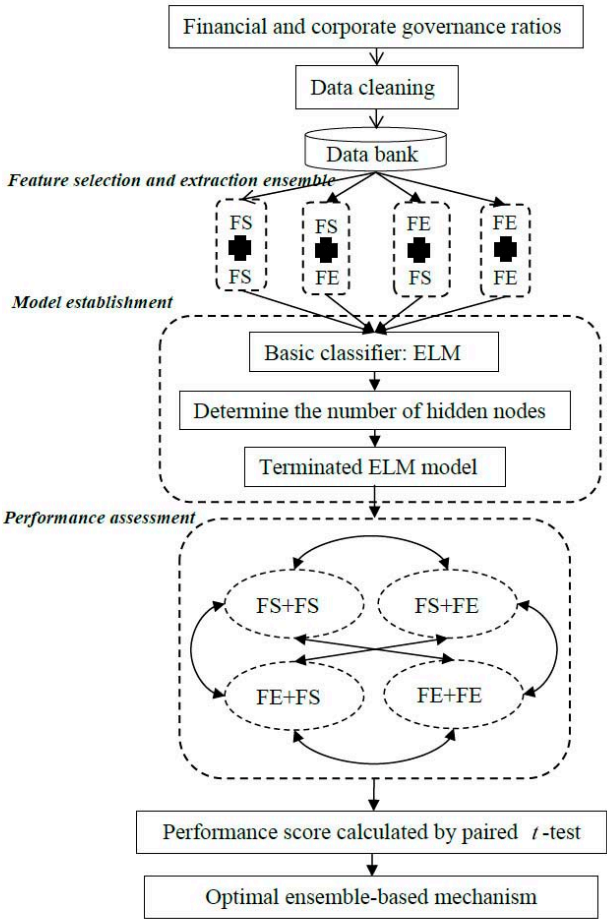

Fig. 2 *The ensemble-based mechanism (EM).*

of combining FS and FE at the same time, this present one fills the gap in the literature to investigate the usefulness of the combined strategy that is grounded on ensemble learning. Its fundamental idea helps tackle any error made by a singular technique, decreasing the biased forecasting output and increasing the forecasting quality.

To examine the effectiveness of combination strategies (that is, the combination of FS and FE), this study sets up four dissimilar types: (1) FS+FS, (2) FS+FE, (3) FE+FS, and (4) FE+FE. The original dataset generated four dissimilar selected outcomes that were injected into ELM (that is, a basic classifier) to construct the forecasting mechanism. To build this mechanism, we divide the selected data into two: a training sample (e.g., 112 corporations) and a holdout sample (e.g., 28 corporations). A five-fold cross-validation is then performed only within the training sample (that is, pick up 90 corporations for training and 22 for validating), with the parameters (numbers of hidden neurons) selected on the basis of the results in the validation sub-sample (cross-validation in the 112 corporations). The number of hidden neurons is generated randomly, and the iteration procedure is manually set to 500. Once we have selected the parameters with the best forecasting performance, we re-run the model using the whole training sample of 112 corporations and evaluate it using the holdout sample. This thereby establishes the well-defined ELM.

How to select an optimal ensemble strategy in combination with ELM can be translated into an MCDM task by calculating the performance score. The score is computed by performing paired -tests for each classifier at the 5% significance level in five-fold cross-validation under three assessing measurements: overall accuracy, sensitivity, and specificity. Since there is no "best" combination strategy for FSF detection under dissimilar criteria and different environments, the purpose of the paired t-tests is to assess whether the superior or inferior performance of one classifier over another is statistically significant. Finally, we execute an MCDM algorithm to determine the appropriate strategy.

The resulting EM has a wide range of applicability. It can be used to assist auditors in optimally allocating their limited amount of audit resources, eliminate audit failure and litigation risk, and protect the reputation of CPA firms.

4. A numerical example and experiment results

4.1 Dataset

This study takes publicly listed electronics corporations from 2003 to 2010 as the research sample. The reason for this choice of data is that the government has invested considerable effort and resources into their industry and the bulk of overall investors have financed this prosperous industry. The data were collected from Taiwan Economic Journal (TEJ) and the Gre-Tai Security Market (GTSM).

4.2 Selected features

Most related studies in the literature merely consider the usefulness of financial characteristics. This study further considers the effectiveness of corporate governance variables so as to suitably illustrate the whole aspect of corporate operating status. We present the financial and corporate governance characteristics as follows.

4.2.1 Financial characteristics

The variables determined as candidates for participation in the input vector are based on previous studies linked to the research domain of FSF [13, 29, 56]. Persons [49] indicated that a corporation with a high debt ratio may increase the possibility of FSF, because a higher debt ratio shifts the risk from equity owners and managers to debt owners. Higher level controllers may be window dressing the financial statements due to their needs to satisfy debt covenants, implying that a higher level of debt may increase the possibility of financial statement fraud.

This study evaluates the possibility of fraud via total debt to total assets (A9: TD/TA). The requirement for continuing progress (that is, the corporate's sustainability) is another motivation for manipulating financial statements. Corporations which cannot meet past performances may be eager to use artificial sales or profits to maintain previous trends. As a growth measurement, we utilize earnings before interest and tax to total equity (A1: EBIT/TE), sales to total assets (A2: S/TA), and net free cash flow to total assets (A10: NFCF/TA). Financial accounts permitting subjective estimations are more complicated to detect and thus are prone to dishonest financial activity. In this category are accounts receivable, inventory, and sales, which we respectively evaluate by accounts receivable to sales (A4: AR/S), inventory to total assets (A8: I/TA), and accounts receivable to total assets (A11: AR/TA). In the logistic regression research on FSF, the detection executed by Spathis et al. [56] indicates that the ratios of net profit to total assets (A5: NP/TA), working capital to total assets (A6: WC/TA), and net profit to sales (A3: NP/S) exhibit significant coefficients. Furthermore, Kirkos et al. [29] suggested that the ratio of gross profit to total assets (A7: GP/TA) also demonstrates significance.

4.2.2 Corporate governance characteristics

A public corporation represents a legal entity with limited liability, delegated management under a board structure, tradable stock, and investor ownership [21]. These characteristics make a corporation the most attractive form of business organization [38], but they also cause some potential agency problems. A corporation with a dispersed ownership structure frequently leads to a conflict of interest between management and shareholders. Managers may forgo the objective to maximize shareholders' wealth and undertake actions that enlarge their personal interests, but not the value of the firm. A well-established corporate governance structure can help shareholders effectively monitor managerial action and prevent managers from misusing corporate wealth [8]. Due to a lack of coordination among small shareholders, it is difficult for them to monitor management. Thus, they have to rely on an external monitoring organization [26], which this study evaluates according to institutional ownership (A14: IO). Less dispersed than in the U.S. or U.K., East Asia's ownership structure is another critical defect of corporate governance [33].

The ultimate controllers of a firm usually strengthen their controlling power by means of pyramid structures and crossholdings, tend to pledge their shareholdings as loan collateral, and manipulate financial reports to prevent a decrease in stock prices. This study evaluates this phenomenon via director ownership (A12: DO), pledged shares of directors (A13: PSD), pyramid structure (A15: PS), and crossholdings (A16: CS). The definition of a pyramid structure is that the shareholder exercises control through at least one publicly traded company [33]. Transparency and information disclosure are integral to corporate governance [46]. A firm with greater transparency and better information disclosure can eliminate any risk due to information asymmetry between management and shareholders. One of the sources of the East Asia financial crisis of 1997-1998 was weak corporate governance [39]. Therefore, most countries have introduced laws and regulations to ensure the clarity of corporate operations. Taiwan regulators generated a ranking system for publicly traded firms to enhance investor confidence. This study further considers the effectiveness of the transparency indicator proposed by the Securities and Futures Institute (SFI). The indicator is assessed by 88 criteria, divided into five main categories: (a) timeliness of financial reporting; (b) compliance with regulated disclosure; (c) completeness of annual reports; (d) suitability of financial forecasts; and (e) integrity of information displayed on official websites. The indicator has five ranks, from superior to inferior (A^+, A, B, C, and C^-). This study uses a numerical number from one to five to represent the different rankings.

The ratios derived from financial reporting may be contaminated by some degree of error (e.g., outlier, extreme value). Thus, to construct a more accurate FSF detection model, we eliminate abnormal cases at the top 1% and the bottom 1% of each ratio [27]. The sample includes 35 FSF corporations and 2100 non-FSF corporations. We employ a matched pair experimental design (paired by 1: industry; 2: products, 3: capitalization, and 4: values of assets) on 140 sample corporations, including 35 corporations identified as being FSF, that have been cited as: (a) violating the Securities and Exchange Act; (b) misreporting financial statements on purpose; and (c) having manipulated financial earnings. The identifications are based on records from the Securities and Futures Investor Protection Center of Taiwan (SFIPCT) and the Financial Supervisory Commission of Taiwan (FSCT). Tab. I lists the features used herein. Tab. II presents the descriptive statistics of

Financial attributes	Symbols	Description
A1	EBIT/TE	Earnings before interest and tax to total equity
A2	S/TA	Sales to total assets
A3	NP/S	Net profit to sales
A4	AR/S	Account receivable to sales
A5	NP/TA	Net profit to total assets
A6	WC/TA	Working capital to total assets
A7	GP/TA	Gross profit to total assets
A8	I/TA	Inventory to total assets
A9	TD/TA	Total debt to total assets
A10	NFCF/TA	Net free cash flow to total assets
A11	AR/TA	Account receivable to total assets
Corporate governance		
A12	DO	Director ownership
A13	PSD	Pledge shares of directors
A14	IO	Institutional ownership
A15	PS	Pyramid structure
A16	CS	Cross shareholding
A17	TI	Transparency indicator

Tab. I *Research attributes description.*

Financial attributes	FSF		Non-FSF		Hypothesis
	Mean	S.D.	Mean	S.D.	$H_0 : \mu_{FSF} = \mu_{Non-FSF}$ $H_1 : \mu_{FSF} \neq \mu_{Non-FSF}$
A1: EBIT/TE	3.576	38.987	0.112	0.155	P-value=0.001
A2: S/TA	0.672	0.441	0.960	0.559	P-value=0.598
A3: NP/S	-0.586	0.807	0.088	0.147	P-value=0.000
A4: AR/S	0.144	0.125	0.202	0.117	P-value=0.99
A5: NP/TA	-0.249	0.395	0.074	0.100	P-value=0.000
A6: WC/TA	-0.155	0.406	0.290	0.191	P-value=0.000
A7: GP/TA	0.057	0.090	0.189	0.119	P-value=0.062
A8: I/TA	0.126	0.180	0.116	0.086	P-value=0.078
A9: TD/TA	0.785	0.331	0.342	0.155	P-value=0.000
A10: NFCF/TA	0.242	0.204	0.224	0.171	P-value=0.235
A11: AR/TA	0.243	0.234	0.228	0.098	P-value=0.000
A12: DO	15.171	10.331	25.314	12.789	P-value=0.1
A13: PSD	17.649	26.782	3.233	10.301	P-value=0.000
A14: IO	20.985	16.170	30.901	21.058	P-value=0.121
A15: PS	1.943	0.236	1.781	0.416	P-value=0.000
A16: CS	1.914	0.284	1.476	0.501	P-value=0.000
A17: TI	4.629	0.490	2.457	0.680	P-value=0.006

Tab. II *The descriptive statistics of each feature.*

each feature and the statistical significance of the mean differences between the two groups (FSF and Non-FSF).

4.3 Feature selection and feature extraction ensemble

We perform FS and FE to overcome the curse of the dimensionality problem and accelerate the calculation performance. During the process, the study considers the individual technique (FS or FE) as well as the FS and FE ensemble grounded on ensemble learning. The combination strategy is able to counteract any errors made by a singular approach and determines the most representative features that are 'agreed upon' by all FS and FE approaches. We take the ELM with outstanding generalization ability as a basic classifier.

This study performs overall forecasting accuracy, sensitivity, and specificity analyses to derive the performance score, with the mathematical formulation expressed in Equation (9). Sensitivity analysis evaluates how well a classifier can distinguish abnormal records. In the case of FSF detection, abnormal cases are fraud, default, or error accounts. A classifier with satisfactory sensitivity can help auditors better alleviate fraud losses than a classifier with unsatisfactory sensitivity. Specificity analysis looks at how well a classifier can distinguish normal records. However, abnormal records (that is, the firm with financial manipulation) are usually more essential than normal records. Thus, the cost of sensitivity is higher than the cost of specificity.

$$\begin{aligned}
&Overall \quad accuracy = (TN + TP) \,/\, (TP + FP + FN + TN)\\
&Sensitivity = (TP) \,/\, (TP + FP)\\
&Specificity = (TN) \,/\, (TN + FP)\\
&TP : True_Positive \;\; ; \quad TN : True_Negative\\
&FP : False_Positive \;\; ; \quad FN : False_Negative
\end{aligned} \tag{9}$$

4.4 Forecasting results

Tab. III presents the effectiveness of ensemble learning. The experimental results indicate that an ensemble technique can provide a preferable forecasting outcome than just using only one method, which agrees with the success of ensemble learning that is grounded on the diversity of individual methods [28]. Executing FS and FE ensemble learning results in some potential benefits, such as facilitating data visualization, defying the curse of dimensionality, and improving forecasting performance. The presented EM designed for detecting FSF is a two-stage dimensionality reduction mechanism (combining FS and FE). The combination strategy selection can be translated into an MCDM task by deriving the performance score, and the MCDM algorithm can tackle the selection task.

We calculate performance scores by a paired -test with a significant level of 5% for all scheme pairs and compute the performance score for each scheme as follows [48]. (1) For the FSF database, evaluate 5-fold cross-validation outcomes of the individual performance measure (paired t-test with a significance level of 5%) for two schemes. The null hypothesis can be expressed as the two schemes being the same. If the statistical test is significant, it indicates that one scheme is better than the other scheme, and the performance scores of the superior and inferior schemes

Accuracy : **C1**: (86.00) > **C2**: (80.80) > **C3**: (80.00) > **C4**: (79.60) > **C5**: (76.80) > **C6**: (76.00) > **C7**: (75.20) > **C8**: (74.00) > **C9**: (67.60)
Sensitivity : **C1**: (80.65) > **C3**: (68.17) > **C2**: (60.01) > **C4**: (53.34) > **C5**: (52.18) > **C6**: (48.01) > **C7**: (43.84) > **C8**: (43.18) > **C9**: (18.16)
Specificity : **C5**: (87.51) > **C2**: (87.11) > **C1**: (87.08) > **C7**: (84.13) = **C6**: (84.13) > **C3**: (83.77) > **C4**: (83.72) > **C8**: (82.68) > **C9**: (81.58)
The definition of each combination strategy: **C1:**FS+FS (FRS+RF); **C2:** FS+FE (FRS+ISOMAP); **C3:** FE+FE (ISOMAP+LLE); **C4:**FE+FS (LLE+FRS); **C5:** FE (ISOMAP); **C6:** FE (LLE); **C7:** FS (RF); **C8:** FS (FRS); **C9:** None

Tab. III *The results of each combination strategy.*

equal 1 and -1, respectively. If the statistical test is not significant, it indicates that the null hypothesis cannot be rejected and then the performance scores for the two schemes equal 0. (2) Repeat the same procedures and we get the performance scores of all schemes for the specific measurements (overall accuracy, sensitivity, and specificity). The higher the score is, the better the classifier performs in the evaluation procedure.

Tab. IV illustrates the performance score of the three assessment criteria. Outstanding performance is highlighted in boldface and italicized. Similar to the evaluation results in Tab. IV, no scheme (specific combination strategy) exhibits optimal performance for all measurements. Moreover, a scheme with optimal scores on some evaluations may perform poorly on other evaluations. For example, the combination strategy C1 (FS+FS) achieves the highest score on accuracy and sensitivity, while its performance on specificity evaluation is in the middle. Therefore, the MCDM algorithm can deal with the above problem and yield an adequate ranking by a reasonable procedure.

Performance score			
Scheme	Accuracy	sensitivity	specificity
C1	***6***	***8***	0
C2	1	3	0
C3	0	6	0
C4	1	1	***1***
C5	0	1	0
C6	0	1	0
C7	0	-4	0
C8	0	-4	0
C9	-8	-8	-1

Tab. IV *The performance score of each combination strategy.*

4.5 Robust test

Due to the priorities being considerably dependent on the subjective judgment of the users, the stability of the terminal ranking under varying determinant weights should be tested [7]. It is thus necessary to execute a robust test under different conditions that reflect divergent aspects of the relative importance of the determinants. Following Paelinck [45], we perform the robust test by the extreme weight method. After modifying the weights of the determinant under three cases (case1: 1-0-0; case2: 0.5-0.5-0; case3: 0.33-0.33-0.33), Fig. 3 presents the results of the test. Another robust test is performed by adjusting the inherent parameter (v) of VIKOR, with the results expressed in Fig. 4. According to Figs. 3 and 4, the ensemble-based mechanisms (C1, C2, C3, and C4) outperform other singular mechanisms. This finding is in accordance with the success of ensemble learning that is grounded on the diversity of individual methods. In comparison to the other eight combination strategies, the feature selection ensemble (C1) posts the outstanding performance.

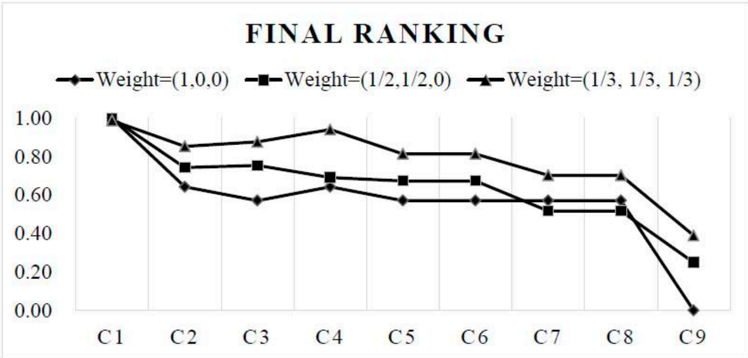

Fig. 3 *The result of VIKOR approach (weight).*

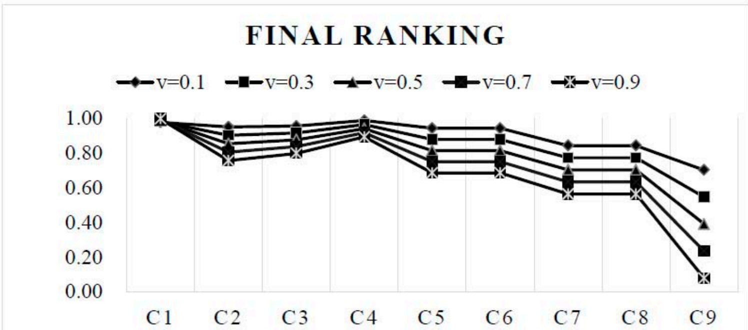

Fig. 4 *The outcome of VIKOR approach (parameter).*

The C1 combination strategy achieves optimal forecasting quality and identifies the most essential attributes that are 'agreed upon' by the two-phase procedure (see Tab. VI). The result indicates that profitability and debt structure are essential financial elements to FSF detection, because management could manipulate financial reports to hit analyst expectations or have a certain debt structure. Taiwan's corporate governance structure has numerous weaknesses in contrast to Europe and the U.S. – that is, its ownership structure is less dispersed, and the ultimate decision makers often strengthen their controlling power by means of a pyramid structure and crossholdings and pledge shareholdings as collateral for loans. Thus, they have a higher incentive to prevent their company's stock price from falling by manipulating financial figures. The public sector in Taiwan announced Corporate Governance Best Practice Principles in 2002 and proposed the Information Transparency and Disclosure ranking system in order to strengthen corporate governance. An interesting finding in this study is that the transparency indicator (A17: TI) has been selected under the C1 combination strategy, meaning that corporates with higher transparency will reduce information asymmetry between management and shareholders and eliminate the agency problem, thus decreasing the possibility of manipulated financial reports.

Model	Ensemble-based mechanism (EM)+ Combination strategy (C1)	
Scenario	S1: With A17: TI	S2: Without A17: TI
AVG. Accuracy	86.00	81.14
AVG. Sensitivity	80.65	77.14
AVG. Specificity	87.08	82.48

Tab. V *The research outcomes (%).*

C1 combination strategy: FS+FS (FRS+RF)	
Feature	Description
A3: NP/S	Net profit to sales
A5: NP/TA	Net profit to total assets
A9: TD/TA	Total debt to total assets
A13: PSD	Pledge shares of directors
A17: TI	Transparency indicator

Tab. VI *The essential features selected by C1 combination strategy.*

The study further examines the effectiveness of the transparency indicator (A17: TI) under two dissimilar scenarios: (1) Scenario 1 (S1): model with TI attribute; and (2) Scenario 2 (S2): model without TI attribute. According to the outcome expressed in Tab. V, the TI attribute impacts not only on forecasting accuracy, but also the other two assessing criteria (sensitivity and specificity), which follows the current global trend of strengthening corporate governance. To examine the feasibility of the proposed EM (C1+ELM), this study further compares the model with three other classifiers illustrated in Tab. VII: SVM, DT, and BN. Overall, EM still outperforms the other three classifiers.

Yang and Wu [62] stated that the class imbalance problem is one of the current challenges in data mining, appearing when the cases in one or several classes, known as majority classes, outnumber the cases of the other classes, called minority classes [15, 37, 41]. In a class-imbalance problem, the minority classes are usually the more essential classes. Due to the matched pair design, the relative distribution of each class is not skewed. However, most real-life datasets are extremely unbalanced and most developed forecasting models are designed to minimize forecasting errors rather than consider the relative distribution of each class. Thus, the forecasted results by the aforementioned models might be problematic in empirical applications. Methods that over-sample the minority class to match the size of the majority class (over-sampling) and methods that under-sample the majority class to match the size of the minority class (under-sampling) are rather effective in overcoming the class imbalance problem [20]. Regardless of matched pair design, the original dataset is quite skewed (FSF: Non-FSF= 1:60).

The study further examines the effectiveness of EM under two dissimilar conditions: (a) under-sample the prevalent class by random sampling; and (b) over-sample the rare class by synthetic minority over-sampling technique (SMOTE). Tabs. 8 and 9 illustrate the results. According to the research finding, the presented EM still outperforms the other classifiers under these two dissimilar conditions.

Assessing criteria
Accuracy: EM (86.00) > SVM (80.00) > BN (77.60) > DT (75.60)
Sensitivity: EM (80.65) > SVM (64.17) > BN (55.01) > DT (53.03)
Specificity: EM (87.08) > SVM (84.61) > BN (83.90) > DT (82.23)

( ): Average accuracy

Tab. VII *Performance in terms of accuracy, sensitivity and specificity.*

Condition 1: Under-sample by random sampling
Assessing criteria
Accuracy: EM (84.00) > SVM (71.00) > DT (66.00) > BN (64.00)
Sensitivity: EM (76.32) > SVM (63.68) > DT (51.98) > BN (51.02)
Specificity: EM (89.48) > SVM (78.45) > DT (75.43) > BN (74.46)

( ): Average accuracy

Tab. VIII *The comparison results for the synthetic dataset.*

Condition 2: Over-sample by SMOTE
Assessing criteria
Accuracy: EM (96.21) > SVM (94.60) > DT (86.38) > BN (70.48)
Sensitivity: EM (95.67) > SVM (94.52) > DT (85.05) > BN (68.43)
Specificity: EM (96.76) > SVM (94.67) > DT (87.71) > BN (72.52)

( ): Average accuracy

Tab. IX *The comparison results for the synthetic dataset.*

5. Conclusion

The detection of FSF is an essential and challenging issue that has gained widespread attention in recent years as the number of fatally deceptive cases has increased.

Unfortunately, most incidents of FSF have been aroused by top managers who are truly knowledgeable and have the necessary resources to outwit the system and fool any detection techniques [11]. Corporations may also be reluctant to admit that they have a fraud or security problem within their operations. Managers may not wish to open their corporation to enquiry or analysis by outside groups, including academic researchers, lest it affect their reputation in the market [17]. It is rare for external auditors to be granted access to raw, unsanitized data. Additionally, practical analysis of fraud incidents is made harder because the data can be incomplete, over-abundant or poorly organized; therefore, performing straightforward data mining techniques to detect FSF has many shortcomings. One of the critical challenges is finding how to identify representative information from over-abundant data.

FS and FE are dissimilar methods used to alleviate the curse of dimensionality problem [17]. Both methods can simplify the detection model as well as increase the processing speed of calculation, leading to better response time. However, prior studies considering the dimensionality reduction task (FS and FE) were usually based on only one decided method, i.e., the decided dimensionality reduction method is supposed to determine informative features for FSF. Nevertheless, applying dissimilar dimensionality reduction methods is likely to generate dissimilar outcomes; therefore, if it were possible to utilize a variety of dissimilar dimensionality reduction methods and sequentially combine the selection outcomes, it would be possible to not only realize the most essential features, but also the improved, advanced detection power it has over utilizing one singular dimensionality reduction method. According to our empirical results, the forecasting model with FS or FE can yield better performance than forecasting model with no FS or no FE technique. This finding is corresponded to prior works [19, 30, 35-36]. The benefits of dimensionality reduction (FS or FE) include decreasing the computational complexity, saving the storage space, enhance the forecasting quality and interpreting complicated dependencies among attributes.

While auditors are the last line of defense in detecting FSF, numerous auditors lack the expertise and experience to tackle the related risk. Thus, this study presented an ensemble-based mechanism (EM) incorporating dimensionality reduction ensemble and ELM to detect FSF. The experimental results indicated that dimensionality reduction ensemble can yield superior performances to utilizing singular dimensionality reduction method. This finding shows that combining a set of accurate and diverse techniques will lead to a powerful forecasting quality. One of the interesting finding is that the forecasting model with FS + FS ensemble strategy poses outstanding performance among whole ensemble strategies. FS ensemble strategy not only can sound the forecasting performance, eliminate the forecasting error, but also can identify the most essential features which are “agree upon” by FS ensemble approaches. Transparency index (A17: TI) is identified by FS ensemble strategy and it has been regarded as an extremely important part in the quality of corporate governance. That is, the corporate with sounded corporate governance structure will make better informative disclosures. The corporate with higher transparency are more valuable, less risky, less volatile and pay out more dividends [4]. The study further examines the effectiveness of Transparency indicator (A17: TI) under two dissimilar scenarios. According to our research outcome, the

Transparency indicator (A17: TI) not only can increase the forecasting accuracy, but also enhance the other two assessing criteria (sensitivity and specificity).

Prior studies have indicated that the classifier performance will vary under dissimilar measurement criteria and different situations. The MCDM algorithm can be utilized to tackle this and related issues, as well as to determine an optimal scheme. The proposed optimal FSF detection model incorporating FS ensemble with ELM can assist auditors in identifying essential information from an overabundant dataset which expands quickly, not only in rows (objects), but also in columns (attributes) in this era of information explosion. By utilizing the proposed model, auditors can simultaneously screen a large amount of corporations and direct their attention to those having a higher potentiality for manipulating financial statements, thus assisting auditors in allocating their limited auditing resources to abnormal client relationships, and thus decreasing the possibility of audit failure, as well as protecting the CPA firms' reputation. Moreover, auditors can utilize the proposed model in peer reviews when evaluating potential clients to assess what decisions other auditors would make in similar circumstances, thereby controlling the quality within corporations and providing defense in lawsuits. Through numerous, sophisticated examinations, the model is a promising alternative for FSF detection that can ensure both the confidence of investors and the stability of stock markets.

Feature research is needed to establish an active detection module that is both effective and efficient. Another possible suggestion is to utilize the EM for other databases in order to further examine the practicability of the proposed mechanism.

Acknowledgement

We thank Editor-in-Chief Dr. Z. Votruba and anonymous reviewers for helpful comment and suggestions that improved the paper. We thank Dr. Zeng Jhih-Hong and Dr. Chun-Chie Huang for technical supports. The author would like to thanks Ministry of Science and Technology of the Republic of China, Taiwan for financially supporting this work under Contract No. 103-2410-H-034 -029.

References

[1] ASSOCIATION OF CERTIFIED FRAUD EXAMINERS. *2008 Report to the Nation on Occupational Fraud & Abuse* [online]. USA: Association of Certified Fraud Examiners, 2008 [viewed 2014-10-08]. Available from: http://www.acfe.com/uploadedFiles/ACFE_Website/Content/documents/2008-rttn.pdf.

[2] BEASLEY M. An empirical analysis of the relation between board of director composition and financial statement fraud. *Accounting Review.* 1996, 71(4), pp. 443–466.

[3] BHATT R.B., GOPAL M. On fuzzy-rough sets approach to feature selection. *Pattern Recognition Letters.* 2005, 26(7), pp. 965–975, doi: 10.1016/j.patrec.2004.09.044

[4] BROWN L.D., CAYLOR M.L. Corporate governance and ?rm performance. *Journal of Accounting and Public Policy.* 2006, 25(4), pp.409–434, doi: 10.1016/j.jaccpubpol.2006.05.005

[5] BREIMAN L. Random Forests. *Machine Learning.* 2001, 45(1), pp. 5–32, doi: 10.1023/A:1010933404324

[6] CAO D.S., XU Q.S., LIANG Y.Z., CHEN X., Li H.D. Automatic feature subset selection for decision tree-based ensemble methods in the prediction of bioactivity. *Chemometrics and Intelligent Laboratory Systems.* 2010, 103(2), pp. 129–136, doi: 10.1016/j.chemolab.2010.06.008

[7] CHANG C.W., WU C.R., LIN C.T., CHEN H.C. An application of AHP and sensitivity analysis for selecting the best slicing machine. *Computers and Industrial Engineering.* 2007, 52(2), pp. 296–307, doi: 10.1016/j.cie.2006.11.006

[8] CRESPI R., RENNEBOOG L. Is (institutional) shareholder activism new? Evidence from UK shareholder coalitions in the pre-Cadbury era. *Corporate Governance: An International Review.* 2010, 18(4), pp. 274–295, doi: 10.1111/j.1467-8683.2010.00795.x

[9] DENG W., CHEN L. Color image watermarking using regularized extreme learning machine. *Neural Network World.* 2010, 20(3), pp. 317–330.

[10] DERRAC J., CORNELIS C., GARCIA S., HERRERA F. Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection. *Information Sciences.* 2012, 186(1), pp. 73–92, doi: 10.1016/j.ins.2011.09.027

[11] DELOITTE FORENSIC CENTER *Ten things about financial statement fraud: A review of SEC enforcement releases 2000-2006* [online]. United States: Deloitte Forensic Center, 2007 [viewed 2014-10-08]. Technical report. Available from: http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/us_forensic_tenthings_fraud01072008.pdf

[12] DOUMPOS M., GAGANIS C., PASIOURAS F. Explaining qualifications in audit reports using a support vector machine methodology. *International Journal of Intelligent Systems in Accounting, Finance and Management.* 2005, 13(4), pp. 197–215, doi: 10.1002/isaf.268

[13] FANNING M.K., COGGER K.O. Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting, Finance and Management.* 1998, 7(1), pp. 21–41, doi: 10.1002/(SICI)1099-1174

[14] GAO X., LIANG J. The dynamical neighborhood selection based on the sampling density and manifold curvature for isometric data embedding. *Pattern Recognition Letters.* 2011, 32(2), pp. 202–209, doi: 10.1016/j.patrec.2010.08.005

[15] GAO M., HONG X., CHEN S., HARRIS C.J. A combined SMOTE and PSO based RBF classifier for two-class imbalanced problems. *Neurocomputing.* 2011, 74(17), pp. 3456–3466, doi: 10.1016/j.neucom.2011.06.010

[16] GENUER R., POGGI J.M., CHRISTINE T.M. Variable selection using random forests. *Pattern Recognition Letters.* 2010, 31(14), pp. 2225–2236, doi: 10.1016/j.patrec.2010.03.014

[17] GOODE S., LACEY D. Detecting complex account fraud in the enterprise: The role of technical and non-technical controls. *Decision Support Systems.* 2011, 50(4), pp.702–714, doi: 10.1016/j.dss.2010.08.018

[18] GISLASON P.O., BENEDIKTSSON J.A., SVEINSSON J.R. Random Forests for land cover classification. *Pattern Recognition Letters.* 2006, 27(4), pp. 294–300, doi: 10.1016/j.patrec.2005.08.011

[19] GUYON I., ELISSEEFF A. An introduction to variable and feature selection. *Journal of Machine Learning Research.* 2003, 20(3), 1157–1182, doi: 10.1162/153244303322753616

[20] GUO H., VIKTOR H.L. Learning from imbalanced data sets with boosting and data generation: The data boost-IM approach. *ACM SIGKDD Explorations* 2004, 6(1), pp. 30–39, doi: 10.1145/1007730.1007736

[21] HANSMANN H., KRAAKMAN R. Organizational law as asset partitioning. *European Economic Review.* 2000, 44(4-6), pp. 807–817, doi: 10.1016/S0014-2921(00)00046-5

[22] HU Q., YU D., GUO M. Fuzzy preference based rough sets. *Information Sciences.* 2010, 180, pp. 2003–2022, doi: 10.1016/j.ins.2010.01.015

[23] HUANG G.B., ZHU Q.Y., SIEW C.K. Extreme learning machine: Theory and applications. *Neurocomputing.* 2006, 70 (1-3), pp. 489–501, doi: 10.1016/j.neucom.2005.12.126

[24] HUANG G.B., CHEN L. Enhanced random search based incremental extreme learning machine. *Neurocomputing.* 2008, 71(16-18), pp. 3460–3468, doi: 10.1016/j.neucom.2007.10.008

[25] HE Q., WU C., CHEN D., ZHAO S. Fuzzy rough set based attribute reduction for information systems with fuzzy decisions. *Knowledge-Based Systems.* 2011, 24(5), pp. 689–696, doi: 10.1016/j.knosys.2011.02.009

[26] JENSEN M. Takeovers: their causes and consequences. *Journal of Economic Perspectives.* 1988, 2(1), pp. 21–48, doi: 10.1257/jep.2.1.21

[27] KIM H.S., SOHN S.Y. Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research.* 2010, 201(3), pp. 838–846, doi: 10.1016/j.ejor.2009.03.036

[28] KITTLER J., HATEF M., DUIN R.P.W., MATAS J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1998, 20(3), pp. 226–239, doi: 10.1109/34.667881

[29] KIRKOS E., SPATHIS C., MANOLOPOULOS Y. Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications.* 2007, 32(4), pp. 995–1003, doi: 10.1016/j.eswa.2006.02.016

[30] KOHAVI R., JOHN G.H. Wrappers for feature subset selection. *Artificial Intelligence.* 1997, 97(1-2), pp. 273–324, doi: 10.1016/S0004-3702(97)00043-X

[31] KRUSINSKA E., SLOWINSKI R., STEFANOWSKI J. Discriminant versus rough sets approach to vague data analysis. *Applied Stochastic Models and Data Analysis.* 1992, 8(1), pp. 43–56, doi: 10.1002/asm.3150080107

[32] KUDELKA M., HORAK Z., VOZENILEK V., SNASEL V. Orthophoto feature extraction and clustering. *Neural Network World.* 2012, 22(2), pp. 103–121, doi: 10.14311/NNW.2012.22.007

[33] LA PORTA R., LOPEZ-de-SILANCES F., SHLEIFER A. Corporate ownership around the world. *Journal of Finance.* 1999, 54(2), pp. 471–517, doi: 10.3386/w6625

[34] LIU Y. Dimensionality reduction and main component extraction of mass spectrometry cancer data. *Knowledge-Based Systems.* 2012, 26(1), pp. 207–215, doi: 10.1016/j.knosys.2011.08.006

[35] LIN S.J., CHANG C., HSU M.F. Multiple extreme learning machines for a two-class imbalance corporate life cycle prediction. *Knowledge-Based Systems.* 2013, 39(1), pp. 214–223, doi: 10.1016/j.knosys.2012.11.003

[36] LIN S.J., HSU M.F. Enhanced risk management by an emerging multi-agent architecture. *Connection Science.* 2014, 26(3), pp.245–259, doi: 10.1080/09540091.2014.908821

[37] LIN S.C., CHANG Y.I., YANG W.N. Meta-learning for imbalanced data and classification ensemble in binary classification. *Neurocomputing.* 2009, 73(1-3), pp. 484–494, doi: 10.1016/j.neucom.2009.06.015

[38] MARTYNOVA M., RENNEBOOG L. Evidence on the international evolution and convergence of corporate governance regulations. *Journal of Corporate Finance.* 2011, 17(5), pp. 1531–1557, doi: 10.1016/j.jcorpfin.2011.08.006

[39] MITTON T. A cross-firm analysis of the impact of corporate governance on the East Asian financial crisis. *Journal of Financial Economics.* 2002, 64(2), pp. 215–241, doi: 10.1016/S0304-405X(02)00076-4

[40] MORTON S. Strategic auditing for fraud. *Accounting Review.* 1993, 68(4), pp. 825–839.

[41] OH S.H. Error back-propagation algorithm for classification of imbalanced data. *Neurocomputing.* 2011, 74(6), pp. 1058–1061, doi: 10.1016/j.neucom.2010.11.024

[42] OPRICOVIC S. *Multicriteria optimization of civil engineering systems* Belgrade, Serbia, 1998 PhD thesis, University of Belgrade [viewed 2014-10-12]. Available from: http://scholar.google.es/scholar?hl=es&q=Opricovic,+S.+(1998).+&btnG=&lr=#3.

[43] OPRICOVIC S., TZENG G.H. Multicriteria planning of post-earthquake sustainable reconstruction. *Computer-Aided Civil and Infrastructure Engineering.* 2002, 17(3), pp. 211–220, doi: 10.1111/1467-8667.00269

[44] OPRICOVIC S., TZENG G.H. Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research.* 2004, 156(2), pp. 445–455, doi: 10.1016/S0377-2217(03)00020-1

[45] PAELINCK J.H.P. Qualitative multiple criteria analysis, environment protection and multiregional development. *Papers in Regional Science.* 1976, 36(1), pp. 56–59, doi: 10.1007/BF01944375

[46] PATEL S.A., BALIC A., BWAKIRA L. Measuring transparency and disclosure at firm-level in emerging markets. *Emerging Markets Review.* 2002, 3(4), pp. 325–337, doi: 10.1016/S1566-0141(02)00040-7

[47] PASIOURAS F., GAGANIS C., ZOPOUNIDIS C. Multicriteria decision support methodologies for auditing decisions: The case of qualified audit reports in the UK. *European Journal of Operational Research.* 2007, 180(3), pp. 1317–1330, doi: 10.1016/j.ejor.2006.04.039

[48] PENG Y., WANG G., KOU G., SHI Y. An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing.* 2011, 11(2), pp. 2906–2915, doi: 10.1016/j.asoc.2010.11.028

[49] PERSONS O. Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research.* 1995, 11(3), pp. 38–46.

[50] PEROLS J.L., LOUGEE B.A. The relation between earnings management and financial statement fraud. *Advances in Accounting.* 2011, 27(1), pp. 39–53, doi: 10.1016/j.adiac.2010.10.004

[51] ROKACH L. Ensemble-based classifiers. *Artificial Intelligence Review.* 2010, 33(1-2), pp. 1–39, doi: 10.1007/s10462-009-9124-7

[52] RONG H.J., ONG Y.S., TAN A.H., ZHU Z. A fast pruned-extreme learning machine for classification problem. *Neurocomputing* 2008, 72(1-3), pp. 359–366, doi: 10.1016/j.neucom.2008.01.005

[53] ROWEIS S.T., LAWRANCE K.S. Nonlinear dimensionality reduction by locally linear embedding. *Science.* 2000, 290(5500), pp. 2323–2326, doi: 10.1126/science.290.5500.2323

[54] SHEN Q., JENSEN R. Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern Recognition.* 2004, 37(7), pp. 1351–1363, doi: 10.1016/j.patcog.2003.10.016

[55] SMITH-MILES K.A. Cross-disciplinary perspectives on meta-learning for algorithm selection. *ACM Computing Surveys.* 2008, 41(1), pp. 1–25, doi: 10.1145/1456650.1456656

[56] SPATHIS C., DOUMPOS M., ZOPOUNIDIS C. Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques. *European Accounting Review.* 2002, 11(3), pp. 509–535, doi: 10.1080/0963818022000000966

[57] SULAIMAN S., SHAMSUDDIN S.M., ABRAHAM A., SULAIMAN S. Intelligent web caching using machine learning methods. *Neural Network World.* 2011, 21(5), pp. 429–452.

[58] SUMMERS S.L., SWEENEY J.T. Fraudulently misstated financial statements and insider trading: an empirical analysis. *Accounting Review.* 1998, 73(1), pp. 131–146.

[59] TENENBAUM J., SILVA V., LANGFORD J.C. A global geometric framework for nonlinear dimensionality reduction. *Science.* 2000, 290(5500), pp. 2319–2323, doi: 10.1126/science.290.5500.2319

[60] WANG Y., CAO F., YUAN Y. A study on effectiveness of extreme learning machine. *Neurocomputing.* 2011, 74(16), pp. 2483–2490, doi: 10.1016/j.neucom.2010.11.030

[61] WU S., SUN M., YANG J. Stochastic neighbor projection on manifold for feature extraction. *Neurocomputing.* 2011, 74(17), pp. 2780–2789, doi: 10.1016/j.neucom.2011.03.036

[62] YANG Q., WU X. 10 challenging problems in data mining research. *International Journal of Information Technology and Decision Making.* 2006, 5(4), pp. 597–604, doi: 10.1142/S0219622006002258

[63] YU B., XU Z. A comparative study for content-based dynamic spam classification using four machine learning algorithms. *Knowledge-Based Systems.* 2008, 21(4), pp. 355–362, doi: 10.1016/j.knosys.2008.01.001

[64] ZHOU W., KAPOOR G. Detecting evolutionary financial statement fraud. *Decision Support Systems.* 2011, 50(3), pp.570–575, doi: 10.1016/j.dss.2010.08.007

[65] ZONG W., HUANG G.B. Face recognition based on extreme learning machine. *Neurocomputing.* 2011, 74(16), pp. 2541–2551, doi: 10.1016/j.neucom.2010.12.041