Decision Support Systems 50 (2011) 491–500

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Detection of financial statement fraud and feature selection using data mining techniques

P. Ravisankar [a], V. Ravi [a,*], G. Raghava Rao [a], I. Bose [b]

[a] Institute for Development and Research in Banking Technology, Castle Hills Road #1, Masab Tank, Hyderabad 500 057, AP, India
[b] School of Business, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

Article history:
Received 20 November 2009
Received in revised form 14 June 2010
Accepted 3 November 2010
Available online 12 November 2010

Keywords:
Data mining
Financial fraud detection
Feature selection
t-statistic
Neural networks
SVM
GP

ABSTRACT

Recently, high profile cases of financial statement fraud have been dominating the news. This paper uses data mining techniques such as Multilayer Feed Forward Neural Network (MLFF), Support Vector Machines (SVM), Genetic Programming (GP), Group Method of Data Handling (GMDH), Logistic Regression (LR), and Probabilistic Neural Network (PNN) to identify companies that resort to financial statement fraud. Each of these techniques is tested on a dataset involving 202 Chinese companies and compared with and without feature selection. PNN outperformed all the techniques without feature selection, and GP and PNN outperformed others with feature selection and with marginally equal accuracies.

1. Introduction

Financial fraud is a serious problem worldwide and more so in fast growing countries like China. Traditionally, auditors are responsible for detecting financial statement fraud. With the appearance of an increasing number of companies that resort to these unfair practices, auditors have become overburdened with the task of detection of fraud. Hence, various techniques of data mining are being used to lessen the workload of the auditors. Enron and Worldcom are the two major scandals involving corporate accounting fraud, which arose from the disclosure of misdeeds conducted by trusted executives of large public corporations. Enron Corporation [17] was an American energy company based in Houston, Texas. Before its bankruptcy in late 2001, Enron was one of the world's leading electricity, natural gas, pulp and paper, and communications companies, with revenues amounting to nearly $101 billion in 2000. Long Distance Discount Services, Inc. (LDDS) began its operations in Hattiesburg, Mississippi in 1983. The company's name was changed to LDDS WorldCom [18] in 1995, and later it became WorldCom. On July 21, 2002, WorldCom filed for Chapter 11 bankruptcy protection in the largest such filing in US history at that time.

Financial statements are a company's basic documents to reflect its financial status [3]. A careful reading of the financial statements can indicate whether the company is running smoothly or is in crisis. If the company is in crisis, financial statements can indicate if the most critical thing faced by the company is cash or profit or something else. All the listed companies are required to publish their financial statements every year and every quarter. The stockholders can form a good idea about the companies' financial future through the financial statements, and can decide whether the companies' stocks are worth investing. The bank also needs the companies' financial statements in order to decide whether to grant loans to them. In a nutshell, the financial statements are the mirrors of the companies' financial status. Financial statements are records of financial flows of a business. Generally, they include balance sheets, income statements, cash flow statements, statements of retained earnings, and some other statements. A detailed description of the items listed in the various financial statements is given below:

- Balance sheet

A balance sheet is a statement of the book value of an organization at a particular date, usually at the end of the fiscal year. A balance sheet has three parts: assets, liabilities, and shareholders' equity. The

* Corresponding author. Tel.: +91 40 23534981x2042; fax: +91 40 23535157.
E-mail addresses: ravisankar_hcu@yahoo.co.in (P. Ravisankar), rav_padma@yahoo.com (V. Ravi), graghavarao.cse@gmail.com (G. Raghava Rao), bose@business.hku.hk (I. Bose).

0167-9236/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2010.11.006

difference between the assets and the liabilities is known as the 'net assets' or the 'net worth' of the company.

• Income statement

Income statements, also called Profit and Loss Statement for companies indicate how net revenue (money received from the sale of products and services before expenses are subtracted, also known as the 'top line') is transformed into net income (the result after all revenues and expenses have been accounted for, also known as the 'bottom line'). The purpose of the income statement is to show managers and investors whether the company made or lost money during the period under consideration.

• Cash flow statement

A cash flow statement is a financial statement that shows incoming and outgoing funds during a particular period. The statement shows how changes in balance sheet and income accounts affect cash and cash equivalents. As an analytical tool the statement of cash flows is useful in determining the short-term viability of a company, particularly its ability to pay bills.

• Statement of retained earnings

The statement of retained earnings, also known as 'statement of owners' equity' and 'statement of net assets' for non-profit organizations, explains the changes in company's retained earnings over the reporting period. It breaks down changes affecting the account, such as profits or losses from operations, dividends paid, and any other items charged or credited to retained earnings. Next, we will describe the key characteristics of financial fraud that can be observed through the financial ratios calculated on the basis of the financial statements published by companies.

1.1. Financial ratios

Financial ratios are a valuable and easy way to interpret the numbers found in financial statements. They can help to answer critical questions such as whether the business is carrying excess debt or inventory, whether customers are paying according to terms, whether the operating expenses are too high, and whether the company assets are being used properly to generate income.

• Liquidity

Liquidity measures a company's capacity to pay its liabilities in short term. There are two ratios for evaluating liquidity. They are:

1) Current ratio = Total current assets / Total current liabilities
2) Quick ratio = (Cash + Accounts receivable + Any other quick assets) / Current liabilities

The higher the ratios the stronger is the company's ability to pay its liabilities as they become due, and the lower is the risk of default.

• Safety

Safety indicates a company's vulnerability to risk of debt. There are three ratios for evaluating liquidity. They are:

1) Debt to equity = Total liabilities / Net worth
2) EBIT/Interest = Earnings before interest and taxes / Interest charges
3) Cash flow to current maturity of long-term debt = (Net profit + Non-cash expenses) / Current portion of long-term debt

• Profitability

Profitability ratios measure the company's ability to generate a return on its resources. There are four ratios to evaluate a company's profitability. They include:

1) Gross profit margin = Gross profit / Total sales
2) Net profit margin = Net profit / Total sales
3) Return on assets = Net profit before taxes / Total assets
4) Return on equity = Net profit before taxes / Net worth

• Efficiency

Efficiency evaluates how well the company manages its assets. There are four ratios to evaluate the efficiency of asset management:

1) Accounts receivable turnover = Total net sales / Accounts receivable
2) Accounts payable turnover = Cost of goods sold / Accounts payable
3) Inventory turnover = Cost of goods sold / Inventory
4) Sales to total assets = Total sales / Total assets

Financial statement fraud may be perpetrated to increase stock prices or to get loans from banks. It may be done to distribute lesser dividends to shareholders. Another probable reason may be to avoid payment of taxes. Nowadays an increasing number of companies are making use of fraudulent financial statements in order to cover up their true financial status and make selfish gains at the expense of stockholders. The fraud triangle is also known as Cressey's Triangle, or Cressey's Fraud Triangle. The fraud triangle seeks to explain what must be present for fraud to occur. The fraud triangle describes the probability of financial reporting fraud which depends on three factors: incentives/pressures, opportunities, and attitudes/rationalization of financial statement fraud [37,38]. The fraud triangle is depicted in Fig. 1, and it is discussed below.

When financial stability or profitability is threatened by economic, industry, or entity operating conditions, or excessive pressure exists for management to meet debt requirements, or personal net worth is materially threatened, the management will face the incentives or pressures to resort to fraudulent practice. Pressure can come in the form of peer pressure, living a lavish lifestyle, a drug addiction, and many other aspects that can influence someone to seek gains via financial fraud. When there are significant accounting estimates that are difficult to verify, or there is oversight over financial reporting, or high turnover or ineffective accounting internal audit, there are opportunities for fraud. For instance, a cashier can steal money out of the cash register because it is there. If the cashier is required to drop all cash into an underground safe for which he does not know the combination, opportunity will not exist. When inappropriate or inefficient communication and support of the entity's values is evident, or a history of violation of laws is known, or management has a practice of making overly aggressive or unrealistic forecasts, then there are risks of fraudulent reporting due to attitudes/rationalization. Rationalization is a grey area in the fraud triangle. Opportunities and incentives exist or they don't. Rationalization depends on the individuals and the circumstances they are facing [19,37]. Understanding the fraud triangle is essential to evaluating financial fraud. When someone is able to grasp the basic concept of the fraud triangle, they are able to better understand financial frauds, how they occur, why they occur, and what to do to stop them.

1.2. Variables related to financial statement fraud

Based on expert's knowledge, intuition, and previous research, it is important to identify some key financial items that are relevant for detection of financial statement fraud. These are listed below:

• Z-score: The Z-score is developed by Altman [2]. It is a formula for measurement of the financial health of a company and works as a tool to predict bankruptcy. It is used to detect financial statement

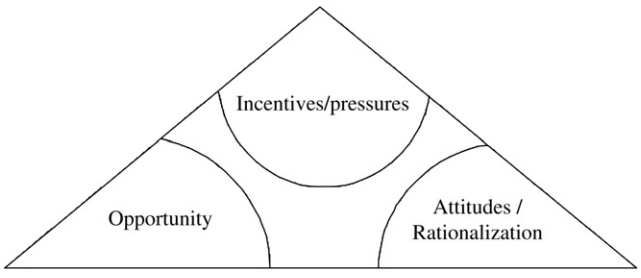

Fig. 1. Components of the fraud triangle.

fraud as well. The formula for Z-score for public companies is given by:

$$\text{Z-score} = (\text{Working capital}/\text{Total assets}^{*}1.2) + (\text{Retained earnings} \div \text{Total assets}^{*}1.4) + (\text{Earnings before income tax} \div \text{Total assets}^{*}3.3) + (\text{Market value of equity} \div \text{Book value of total}) + (\text{Liabilities}^{*}0.6 + \text{Sales} \div \text{Total assets}^{*}0.999)$$

- A high debt structure increases the likelihood of financial fraud as it shifts the risk from equity owner to the debt owner. So the financial ratios related to debt structure such as (i) Total debt/Total assets and (ii) Debt/Equity need to be carefully considered when searching for indications of fraud.
- An abnormal value reported as a measure of continuous growth such as sales to growth ratio is also a factor that may be indicative of fraudulent financial practice.
- Many items of the financial statements such as Accounts receivable, Inventories, Gross margin etc. can be estimated to some degree using subjective methods and different accounting methods can often lead to different values even for the same company.
- According to previous research, many other financial ratios can be considered for fraud detection, such as Net profit/Total assets, Working capital/Total assets, Net profit/Sales, Current assets/Current liabilities and so on.
- The tenure of CEO and CFO: According to the auditors' experience and previous research, the high turnover of CEO and CFO may indicate the existence of financial fraud in the company.
- Some qualitative variables such as previous auditor's qualifications can be considered to determine the likelihood of fraudulent book keeping.

Data mining has been applied in many aspects of financial analysis. Few areas where data mining techniques have already being used include: bankruptcy prediction, credit card approval, loan decision, money-laundering detection, stock analysis, etc. However, research related to the use of data mining for detection of financial statement fraud is limited. The main objective of this research is to predict the occurrence of financial statement fraud in companies as accurately as possible using intelligent techniques. Financial accounting fraud can be detected by a human expert by using his/her experiential/judgemental knowledge, provided he/she has sufficient expertise. However, in this case, human bias cannot be eliminated and the judgments tend to be subjective. Hence, we resort to data-driven approaches, which solely rely on the past data of fraudulent and healthy companies and their financial ratios. When data mining techniques (most of them barring a few statistical ones are artificial intelligence based) are employed to solve these problems, they work in an objective way by sifting through the records of fraudulent and healthy companies. In the process, they discover knowledge which can be used to predict whether a company at hand will perpetrate financial accounting fraud in future. Data mining techniques have another advantage in that they can handle a large number of records and financial ratios efficiently. According to Kirkos et al. [23], artificial intelligence methods have the theoretical advantage that they do not impose arbitrary assumptions on the input variables. An auxiliary aim of this research is to select the most important financial items that can explain the financial statement fraud. The results obtained using this research will be useful for auditors engaged in the prediction of financial statement fraud. In fact, emerging companies can carefully monitor those financial statements for getting long-term advantages in the competitive market. Further, it will be useful for investors who plan to invest in such companies.

The rest of the paper is organized as follows. Section 2 reviews the research done in the area of financial statement fraud detection. Section 3 provides an overview of the data mining techniques that are used in this paper. Section 4 describes the feature selection phase of data mining. Section 5 presents the results and discusses the implications of these results. Finally, Section 6 concludes the paper.

2. Literature review

There has been a limited use of data mining techniques for detection of financial statement fraud. The data mining techniques used include decision trees, neural networks (NN), Bayesian belief networks, case based reasoning, fuzzy rule-based reasoning, hybrid methods, logistic regression, and text mining. Extant research in this direction is reviewed in the following paragraphs.

According to Kirkos et al. [23], some estimates stated that fraud cost US business more than $400 billion annually. Spathis et al [42] compared multi-criteria decision aids with statistical techniques such as logit and discriminant analysis in detecting fraudulent financial statements. A novel financial kernel for the detection of management fraud is developed using support vector machines on financial data by Cecchini et al. [9]. Huang et al. [20] developed an innovative fraud detection mechanism on the basis of Zipf's Law. The purpose of this technique is to assist auditors in reviewing the overwhelming volumes of datasets and identifying any potential fraud records. Kirkos et al. [23] used the ID3 decision tree and Bayesian belief network to detect financial statement fraud successfully.

Sohl and Venkatachalam [41] used back-propagation NN for the prediction of financial statement fraud. There are other researchers who used different NN algorithms to detect financial reporting fraud. Cerullo and Cerullo [10] explained the nature of fraud and financial statement fraud along with the characteristics of NN and their applications. They illustrated how NN packages could be utilized by various firms to predict the occurrence of fraud. Calderon and Cheh [8] examined the efficacy of NN as a potential enabler of business risk based auditing. They employed different methods using NN as a tool for research in the auditing and risk assessment domain. Further, they identified several opportunities for future research that include methodological issues related to NN modeling as well as specific issues related to the application of NN for business risk assessment. Koskivaara [25] investigated the impact of various preprocessing models on the forecast capability of NN when auditing financial accounts. Further, Koskivaara [26] proposed NN based support systems as a possible tool for use in auditing. He demonstrated that the main application areas of NN were detection of material errors, and management fraud. Busta and Weinberg [7] used NN to distinguish between 'normal' and 'manipulated' financial data. They examined the digit distribution of the numbers in the underlying financial information. The data analysis is based on Benford's law, which demonstrated that the digits of naturally occurring numbers are distributed on a predictable and specific pattern. They tested six NN designs to determine the most effective model. In each design, the inputs to the NN were the different subsets of the 34 variables. The results showed that NN were able to correctly classify 70.8% of the data on an average.

Feroz et al. [15] observed that the relative success of the NN models was due to their ability to 'learn' what were important. The perpetrators of financial reporting frauds had incentives to appear prosperous as evidenced by high profitability. In contrast to conventional statistical models replete with assumptions, the NN used adaptive learning processes to determine what were important in predicting targets. Thus, the NN approach was less likely to be affected by accounting manipulations. The NN approach was well suited to predicting the possible fraudsters because the NN 'learnt' the characteristics of reporting violators despite managers' intent to obfuscate misrepresentations. Brooks [6] also applied various NN models to detect financial statement fraud with great success. Fanning and Cogger [13] used NN (AutoNet) for detecting management fraud. The study offered an in-depth examination of important publicly available predictors of fraudulent financial statements. The study

reinforced the efficiency of AutoNet in providing empirical evidence regarding the merits of suggested red flags for fraudulent financial statements. Ramamoorti et al. [37] provided an overview of the multi-layer perceptron architecture and compared it with a Delphi study. They found that internal auditors could benefit from using NN for assessing risk. Zhang et al. [46] conducted a review of the published papers that reported the use of NN in forecasting during the time period 1988–98.

Aamodt and Plaza [1] and Kotsiantis et al. [27] used case based reasoning to identify the fraudulent companies. Further, Deshmukh and Talluru [12] demonstrated the construction of a rule-based fuzzy reasoning system to assess the risk of management fraud and proposed an early warning system by finding out 15 rules related to the probability of management fraud. Pacheco et al. [34] developed a hybrid intelligent system consisting of NN and a fuzzy expert system to diagnose financial problems. Further, Magnusson et al. [30] used text mining and demonstrated that the language of quarterly reports provided an indication of the change in the company's financial status. A rule-based system that consisted of too many if–then statements made it difficult for marketing researchers to understand key drivers of consumer behavior [22]. Variable selection was used in order to choose a subset of the original predictive variables by eliminating variables that were either redundant or possessed little predictive information.

3. Methodology

The dataset used in this research was obtained from 202 companies that were listed in various Chinese stock exchanges, of which 101 were fraudulent and 101 were non-fraudulent companies. The data also contained 35 financial items for each of these companies. Table 1 lists these financial items. Of these, 28 were financial ratios reflecting liquidity, safety, profitability, and efficiency of companies. We performed log transformation on the entire dataset to reduce its dimension. Then we normalized each of the independent variables of the original dataset during the data preprocessing stage. Furthermore, ten-fold cross-validation is performed to improve the reliability of the result. Then, we analyzed the dataset using six data mining techniques including MLFF, SVM, GP, GMDH, LR, and PNN. The block diagram in Fig. 2 depicts the data flow. We chose the six techniques because MLFF, GMDH and PNN fall under the NN category, SVM comes from statistical learning theory, GP is an evolutionary technique, and logistic regression is a traditional statistical technique for classification. Thus, these methods had varied background and different theories to support them. In this way, we ensured that the problem at hand is analyzed by disparate models, that had varying degree of difficulty in implementation and also exhibited varying degree of performance on different data mining problems. In other words, the problem is studied and analyzed comprehensively from all perspectives.

It is observed that some of the independent variables turned out to be much more important for the prediction purpose whereas some contributed negatively towards the classification accuracies of different classifiers. So, a simple statistical technique using the t-statistic is used to accomplish feature selection on the dataset by identifying the most significant financial items that could detect the presence of financial statement fraud. This is described in Section 4. The features having high t-statistic values were more significant than others. For feature selection, first we extracted the top 18 features (more than half of the total 35 financial items) from the original dataset. Then the dataset with the reduced feature set (only 18 financial items) was fed as input to the above mentioned classifiers, which resulted in new combinations such as t-statistic-MLFF, t-statistic-SVM, t-statistic-GP, t-statistic-GMDH, t-statistic-LR, and t-statistic-PNN. In order to conduct further analysis, we then extracted the top 10 features and the same process was repeated, i.e., the dataset with the reduced feature set (only 10 financial items) was fed as input to all of the above classifiers. A brief description of the different data mining techniques used in this research is provided below.

Table 1
Items from financial statements of companies that are used for detection of financial statement fraud.

No.	Financial items
1	Debt
2	Total assets
3	Gross profit
4	Net profit
5	Primary business income
6	Cash and deposits
7	Accounts receivable
8	Inventory/Primary business income
9	Inventory/Total assets
10	Gross profit/Total assets
11	Net profit/Total assets
12	Current assets/Total assets
13	Net profit/Primary business income
14	Accounts receivable/Primary business income
15	Primary business income/Total assets
16	Current assets/Current liabilities
17	Primary business income/Fixed assets
18	Cash/Total assets
19	Inventory/Current liabilities
20	Total debt/Total equity
21	Long term debt/Total assets
22	Net profit/Gross profit
23	Total debt/Total assets
24	Total assets/Capital and reserves
25	Long term debt/Total capital and reserves
26	Fixed assets/Total assets
27	Deposits and cash/Current assets
28	Capitals and reserves/Total debt
29	Accounts receivable/Total assets
30	Gross profit/Primary business profit
31	Undistributed profit/Net profit
32	Primary business profit/Primary business profit of last year
33	Primary business income/Last year's primary business income
34	Account receivable /Accounts receivable of last year
35	Total assets/Total assets of last year

3.1. Support vector machines (SVM)

SVM introduced by Vapnik [44] use a linear model to implement nonlinear class boundaries by mapping input vectors nonlinearly into a high-dimensional feature space. In the new space, an optimal separating hyperplane is constructed. The training examples that are closest to the maximum margin hyperplane are called support vectors. All other training examples are irrelevant for defining the binary class boundaries. SVM are simple enough to be analyzed mathematically. In this sense, SVM may serve as a promising alternative combining the strengths of conventional statistical methods that are more theory-driven and easy to analyze, and machine learning methods that are more data-driven, distribution-free and robust. Recently, SVM have been used in financial applications such as credit rating, time series prediction, and insurance claim frauds detection. These studies reported that the performance of SVM is comparable to and even better than other classifiers such as MLFF, case based reasoning, discriminant analysis, and logistic regression.

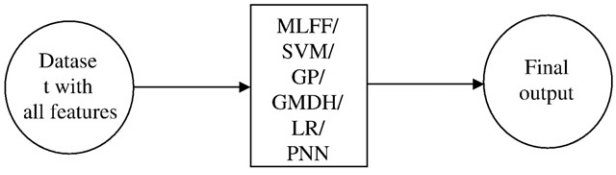

Fig. 2. Architecture of different classifiers.

3.2. Genetic programming (GP)

GP [28] is an extension of genetic algorithms (GA). It is a search methodology belonging to the family of evolutionary computation. GP randomly generates an initial population of solutions. Then, the initial population is manipulated using various genetic operators to produce new populations. These operators include reproduction, crossover, mutation, dropping condition, etc. The whole process of evolving from one population to the next population is called a generation. A high-level description of the GP algorithm can be divided into a number of sequential steps [14]:

- Create a random population of programs, or rules, using the symbolic expressions provided as the initial population.
- Evaluate each program or rule by assigning a fitness value according to a predefined fitness function that can measure the capability of the rule or program to solve the problem.
- Use the reproduction operator to copy existing programs into the new generation.
- Generate the new population with crossover, mutation, or other operators from a randomly chosen set of parents.
- Repeat the second to the fourth steps for the new population until a predefined termination criterion is satisfied, or a fixed number of generations is completed.
- The solution to the problem is the genetic program with the best fitness within all generations.

In GP, the crossover operation is achieved by reproduction of two parent trees. Two crossover points are then randomly selected in the two offspring trees. Exchanging sub-trees, which are selected according to the crossover point in the parent trees, generates the final offspring trees. The offspring trees are usually different from their parents in size and shape. Then, mutation operation is also considered in GP. A single parental tree is first reproduced. Then a mutation point is randomly selected from the reproduction, which can be either a leaf node or a sub-tree. Finally, the leaf node or the sub-tree is replaced by a new leaf node or a randomly generated sub-tree. Fitness functions ensure that the evolution goes toward optimization by calculating the fitness value for each individual in the population. The fitness value evaluates the performance of each individual in the population.

GP is guided by the fitness function to search for the most efficient computer program that can solve a given problem. A simple measure of fitness [14] is adopted for the binary classification problem and is given as follows:

$$\text{Fitness} = \frac{\text{No. of samples classified correctly}}{\text{No. of samples used for training during evaluation}}$$

The major considerations in applying GP to pattern classification are:

- GP based techniques are free of the distribution of the data, and so no *a priori* knowledge is needed about the statistical distribution of the data.
- GP can directly operate on the data in its original form.
- GP can detect the underlying but unknown relationship that exists among data items and express it as a mathematical expression.
- GP can discover the most important discriminating features of a class during the training phase.

3.3. Multi-layer feedforward neural network (MLFF)

MLFF is one of the most common NN structures, as they are simple and effective, and have found home in a wide assortment of machine learning applications. MLFF starts as a network of nodes arranged in three layers—the input, hidden, and output layers. The input and output layers serve as nodes to buffer input and output for the model, respectively, and the hidden layer serves to provide a means for input relations to be represented in the output. Before any data is passed to the network, the weights for the nodes are random, which has the effect of making the network much like a newborn's brain—developed but without knowledge. MLFF are feed-forward NN trained with the standard back-propagation algorithm. They are supervised networks so they require a desired response to be trained. They learn how to transform input data into a desired response. So they are widely used for pattern classification and prediction. A multi-layer perceptron is made up of several layers of neurons. Each layer is fully connected to the next one. With one or two hidden layers, they can approximate virtually any input–output map. They have been shown to yield accurate predictions in difficult problems [39].

3.4. Group method of data handling (GMDH)

GMDH was introduced by Ivakhnenko [21] in 1966 as an inductive learning algorithm for modeling complex systems. It is a self-organizing approach that tests increasingly complicated models and evaluates them using some external criterion on separate parts of the data sample. GMDH is partly inspired by research in perceptrons and learning filters. GMDH has influenced the development of several techniques for synthesizing (or 'self-organizing') networks of polynomial nodes. GMDH attempts a hierarchic solution by trying out many simple models, retaining the best of these models, and building on them iteratively to obtain a composition (or feed-forward network) of functions as the model. The building blocks of GMDH, or polynomial nodes, usually have the quadratic form:

$$z = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2 + w_5x_1x_2$$

where x_1 and x_2 are inputs, w_i is the coefficient (or weight) vector w, and z is the node output. The coefficients are determined by solving the linear regression equation with $z = y$, where y represents the response vector. The GMDH develops on a data set. The data set including independent variables (x_1, x_2,..., x_n) and one dependent variable y is split into a training and testing set. During the process of learning a forward multilayer NN is developed by observing the following steps:

- In the input layer of the network n units with an elementary transfer function $y = x_i$ are constructed. These are used to provide values of independent variables from the learning set to the successive layers of the network.
- When constructing a hidden layer an initial population of units is generated. Each unit corresponds to the Ivakhnenko polynomial form:

$$y = a + bx_1 + cx_2 + dx_1^2 + ex_1x_2 + fx_2^2 \text{ or}$$
$$y = a + bx_1 + cx_2 + dx_1x_2$$

where y is an output variable; x_1, x_2 are two input variables; and $a, b, \ldots f$ are parameters.

- Parameters of all units in the layer are estimated using the learning set.
- The mean square error between the dependent variable y and the response of each unit is computed for the testing set.
- Units are sorted in terms of the mean square error and just a few units with minimal error survive. The rest of the units are deleted. This step guarantees that only units with a good ability for approximation are chosen.
- Next the hidden layers are constructed so that the mean square error of the best unit decreases.
- Output of the network is considered as the response of the best unit in the layer with the minimal error.

Table 2
Top 18 items selected from financial statements of companies by t-statistic based feature selection.

No.	Financial items
1	Net profit
2	Gross profit
3	Primary business income
4	Primary business income/Total assets
5	Gross profit/Total assets
6	Net profit/Total assets
7	Inventory/Total assets
8	Inventory/Current liabilities
9	Net profit/Primary business income
10	Primary business income/Fixed assets
11	Primary business profit/Primary business profit of last year
12	Primary business income/Last year's primary business income
13	Fixed assets/Total assets
14	Current assets/Current liabilities
15	Capitals and reserves/Total debt
16	Long term debt/Total capital and reserves
17	Cash and deposits
18	Inventory/Primary business income

The GMDH network learns in an inductive manner and builds a function (called a polynomial model), that results in the minimum error between the predicted value and expected output. The majority of GMDH networks use regression analysis for solving the problem. The first step is to decide the type of polynomial that the regression will find. The initial layer is simply the input layer. The first layer is created by computing regressions of the input variables and then choosing the best ones. The second layer is created by computing regressions of the values in the first layer along with the input variables. This means that the algorithm essentially builds polynomials of polynomials. Again, only the best are chosen by the algorithm. These are called survivors. This process continues until a pre-specified selection criterion is met.

3.5. Logistic regression (LR)

According to Panik [35], when dealing with logistic regression, the response variable is taken to be dichotomous or binary (it takes on only two possible values), i.e., $y_i = 0$ or 1 for all $i = 1,...,n$. For instance, we can have a situation in which the outcome of some process of observation is either a success (we record a 1) or failure (we record a 0), or we observe the presence (1) or absence (0) of some characteristic or phenomenon. In addition, dichotomous variables are useful for making predictions, e.g., we may ask the following: Will an individual make a purchase of a particular item in the near future?

$$\text{Here } y_i = \begin{cases} 1, yes; \\ 0, No. \end{cases}$$

According to Williams et al. [45], LR is a commonly used approach for performing binary classification. It learns a set of parameters, $\{w_0, w\}$, that maximizes the likelihood of the class labels for a given set of training data. Let $x_i \in R^d$ denote a (column) vector of d features representing the ith data point, and $y_i \in \{0,1\}$ denote its corresponding class label (e.g., clutter or mine). For a labeled (training) data point, y_i is known; for an unlabeled (testing) data point, y_i is unknown. Under the LR model, the probability of label $y_i = 1$ given x_i is given by Eq. (1):

$$\psi_i \equiv p(y_i = 1 | x_i) = \frac{exp\left(w_0 + w^T x_i\right)}{1 + exp(w_0 + w^T x_i)} \quad (1)$$

where $w_0 \in R$ and $w \in R^d$ are the LR intercept and coefficients respectively. For a set of N independent labeled data points, $\{x_i, y_i\}_{i=1}^N$, the log-likelihood of the class labels can be written as Eq. (2):

$$l(w_0, w) = \sum_{i=1}^{n} \left[(1-y_i)\, log(1-\psi_i) + y_i\, log\psi_i\right] \quad (2)$$

To maximize the log-likelihood in Eq. (2), a standard optimization approach can be employed, since the gradient (and Hessian) of Eq. (2) with respect to $\{w_0, w\}$ can be readily calculated. Once the LR parameters $\{w_0, w\}$ have been learned, the probability that an unlabeled testing data point x_i belongs to each class can be obtained using Eq. (1).

3.6. Probabilistic neural network (PNN)

PNN is a feed-forward NN involving a one pass training algorithm used for classification and mapping of data. PNN was introduced by Specht [43] in 1990. It is a pattern classification network based on the classical Bayes classifier, which is statistically an optimal classifier that seeks to minimize the risk of misclassification. Any pattern classifier places each observed data vector $x = [x_1, x_2, x_3 \ldots x_N]^T$ in one of the predefined classes c_i, $i = 1, 2,..., m$ where m is the number of possible classes. The effectiveness of any classifier is limited by the number of data elements that the vector x can have and the number of possible classes m. The classical Bayes pattern classifier [40] implements the Bayes conditional probability rule that the probability $P(c_i|x)$ of x being in class c_i is given by:

$$P(c_i|x) = \frac{P(x|c_i)P(c_i)}{\sum_{j=1}^{m} P\left(x|c_j\right)P\left(c_j\right)} \quad (3)$$

where $P(x|c_i)$ is the conditioned probability density function of x given set c_i, $P(c_j)$ is the probability of drawing data from class c_j. Vector x is said to belong to a particular class c_i if $P(c_i|x) > P(c_j|x)$, $\forall j = 1, 2,..., m$ and $j \neq i$. This input x is fed into each of the patterns in the pattern layer. The summation layer computes the probability $P(c_i|x)$ that the given input x is included in each of the classes c_i that is represented by the patterns in the pattern layer. The output layer selects the class for which the highest probability is obtained in the summation layer. The input is then made to belong to this class. The effectiveness of the network in classifying input vectors depends on the value of the smoothing parameter.

4. Feature selection

Feature selection is critical to data mining and knowledge based authentication. The problem of feature selection has been well studied in areas where datasets with a large number of features are available, including machine learning, pattern recognition, and statistics. Piramuthu [36] observed that about 80% of the resources in a majority of data mining applications are spent on cleaning and preprocessing the data, and developed a new feature selection method based on Hausdorff distance for analyzing web traffic data. Feature selection is of paramount importance for any learning algorithm which when poorly done (i.e., a poor set of features is selected) may lead to problems related to incomplete information, noisy or irrelevant features, not the best set/mix of features, among others [45]. Mladenic and Grobelnik [31] reviewed various feature selection methods in the context of web mining. Chen and Liginlal [11] developed a maximum entropy based feature selection technique for knowledge based authentication.

In this study, we employed a feature selection phase by using the simple t-statistic technique. t-statistic is one of the efficient feature

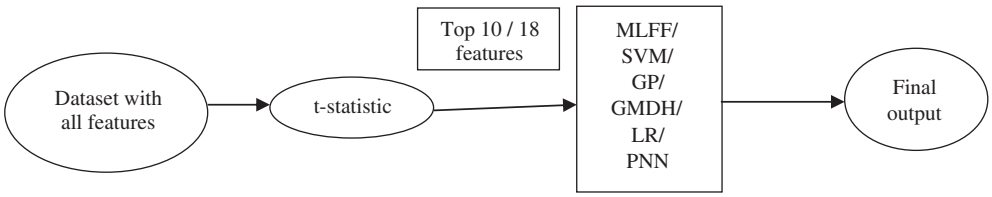

Fig. 3. Architecture of different classifiers after feature selection.

selection techniques. The features are ranked according to the formula shown below [16,29]. In fact, Liu et al. [29] were the first to propose t-statistic for the purpose of feature selection in the field of bioinformatics.

$$\text{t}-\text{statistic} = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (4)$$

where μ_1 and μ_2 represent the means of the samples of fraudulent companies and non-fraudulent companies for a given feature respectively, σ_1 and σ_2 represent the standard deviation of the samples of fraudulent companies and non-fraudulent companies for a given feature respectively. n_1 and n_2 represent the number of samples of fraudulent companies and non-fraudulent companies for a given feature. The t-statistic values are computed for each feature and the top 18 features with the highest t-statistic values are considered in the first case and the top 10 features are considered in the second case. A high t-statistic value indicates that the feature can highly discriminate between the samples of fraudulent and non-fraudulent companies. The top 18 financial features that are selected by the t-statistic based feature selection are shown in Table 2. The feature subset formed with the top 18 features is fed as input to MLFF/SVM/GP/GMDH/LR/PNN for classification purpose in the first case. Similarly, the feature subset formed with the top 10 features is fed as input to MLFF/SVM/GP/GMDH/LR/PNN for classification purpose in the second case. The block diagram for all these combinations is shown in Fig. 3. Ten-fold cross-validation is used to ensure better validity of the experiments. It should be noted that the t-statistic is employed for feature selection for each fold separately. It is observed that the same set of features did not turn out to be best in each fold. Hence, we followed a frequency based approach, whereby, the frequency of occurrence of each of the features in top slots is computed and the features are then sorted in the descending order of the frequency of occurrences. In this manner, we selected the top 10 and top 18 features and reported them in Table 2.

5. Results and discussion

The dataset analyzed in this paper comprised 35 financial items for 202 companies, of which 101 were fraudulent and 101 were non-fraudulent. Since the financial items had a wide range, we first performed natural logarithmic transformation, and then normalization during the data preprocessing phase. We employed the GP as implemented in the tool Discipulus (available at www.rmltech.com and downloaded on 20th August, 2008). For MLFF, GMDH, and PNN, we employed Neuroshell 2.0 [33] and for SVM and LR we used KNIME 2.0.0 [24].

The sensitivity is the measure of the proportion of the number of fraudulent companies predicted correctly as fraudulent by a particular model to the total number of actual fraudulent companies. The specificity is the measure of the proportion of the number of non-fraudulent companies predicted as non-fraudulent by a model to the total number of actual non-fraudulent companies. In all cases, we presented the average accuracies, sensitivities, specificities, and area under the Receiver Operating Characteristic curve (AUC) for the test data, averaged over 10-folds. We ranked the classifiers based on AUC. First, the results of the 10-fold cross-validation method for the stand-alone techniques viz. MLFF, SVM, GP, GMDH, LR, and PNN without feature selection are presented in Table 3. From Table 3 we observe that PNN with 98.09% accuracy and 98.09% sensitivity outperformed all other classifiers (as indicated by bold faced numerals in the Table 3). GP yielded the next best result with 94.14% accuracy and 95.09% sensitivity. We also observe that PNN is the best classifier among all others in terms of AUC as well. The best results obtained by Bose and Wang [5], who employed canonical discriminant analysis (CDA), classification and regression tree (C&RT) and exhaustive pruning NN on the same dataset are also presented in Table 3 for ease of comparison. From Table 3 we can observe that the results obtained in this study are always superior to the results obtained by them for all cases, except SVM and LR.

As the next step, we used t-statistic for feature selection and extracted the most important features. First, we considered the top 18 features for constructing the reduced feature subset. Later, this feature subset is fed to all the above classifiers for the purpose of classification. The average results of all the classifiers over all folds with 18 features are presented in Table 4. From Table 4 we observe that GP outperformed other classifiers with 92.68% accuracy and 90.55% sensitivity, whereas PNN came close behind with 95.64% accuracy and 91.27% sensitivity (as indicated by bold faced numerals in Table 4). Furthermore, results based on AUC indicated that GP yielded highest accuracy followed by PNN, which yielded marginally less accuracy. This makes us infer that the selected feature subsets

Table 3
Average results of dataset with all features using 10-fold cross-validation.

Classifier	Accuracy	Sensitivity	Specificity	AUC
MLFF	78.36	80.21	76.35	7827.90
SVM	70.41	55.43	84.13	6978.00
GP	94.14	95.09	93.05	9407.10
GMDH	93.00	91.46	95.18	9331.85
LR	66.86	63.32	70.66	6699.10
PNN	**98.09**	**98.09**	**98.09**	**9809.00**
CDA [5]	71.37	61.96	80.77	7136.5
C&RT [5]	72.38	72.40	72.36	7238
Exhaustive pruning NN [5]	77.14	80.83	73.45	7714

Table 4
Average results of dataset with reduced features (top 18 features selected by t-statistic) and using 10-fold cross-validation.

Classifier	Accuracy	Sensitivity	Specificity	AUC
MLFF	78.77	76.98	81.28	7912.80
SVM	73.41	72.07	75.04	7355.55
GP	**92.68**	**90.55**	**95.27**	**9290.95**
GMDH	90.68	93.46	88.34	9089.95
LR	70.36	62.91	78.88	7089.50
PNN	95.64	91.27	94.16	9271.75

Table 5
Average results of dataset with reduced features (top 10 features selected by t-statistic) and using 10-fold cross-validation.

Classifier	Accuracy	Sensitivity	Specificity	AUC
MLFF	75.32	67.24	82.79	7501.65
SVM	72.36	73.60	69.68	7164.35
GP	89.27	85.64	93.16	8939.95
GMDH	88.14	87.44	89.25	8834.40
LR	70.86	65.23	76.46	7084.45
PNN	**90.77**	**87.53**	**94.07**	**9079.85**

Table 7
t-statistic values of average AUCs of GP compared to that of other classifiers with (top 18 features) feature selection.

Classifier compared	t-statistic at 10% level of significance
MLFF	5.13*
SVM	5.28*
GMDH	0.69
LR	6.40*
PNN	0.08

The * indicates that the result is statistically significant.

have a high discriminatory power and the 'left-over' features have very little to contribute to the success of financial fraud detection. Furthermore, in order to conduct an exhaustive study over this dataset, in the second set of experiment we considered only the top 10 features (based on the values of the t-statistics) for constructing the reduced feature subset. The top 10 features can be seen in the first ten rows of Table 2. We repeated the experiments as in the first case. The average results for all the classifiers over all folds with 10 features are presented in Table 5. From Table 5 we observe that PNN outperformed other classifiers with 90.77% accuracy and 87.53% sensitivity (as indicated by bold faced numerals in Table 5), whereas GP came second with 89.27% accuracy and 85.64% sensitivity. Moreover, results based on the AUC indicated that PNN yielded the highest accuracy followed by GP, which yielded only marginally less accuracy.

In order to find out whether the difference in average AUCs is statistically significant or not, we conducted a t-test between the top performer and the remaining classifiers (i) without feature selection, (ii) with feature selection including top 18 features, and (iii) with feature selection including top 10 features. In case of the dataset without feature selection, the t-statistic values between the average AUCs obtained by PNN and that of other classifiers are presented in Table 6. From Table 6 we observe that t-statistic values are more than the critical value of the test statistic, which is 1.73 at the 10% level of significance. Thus, we infer that PNN significantly outperformed other classifiers without feature selection. In case of the dataset with feature selection and considering only the top 18 features, the t-statistic values between the average AUCs obtained by GP and that of other classifiers are presented in Table 7. From this table we can observe that t-statistic values are more than 1.73 in case of MLFF, SVM and LR, whereas those values are less than 1.73 in case of PNN and GMDH. From these results we can say that the GP significantly outperformed all classifiers except GMDH and PNN. Considering only the top 10 features, the t-statistic values between the average AUCs obtained by PNN and that of other classifiers are presented in Table 8. From this table we can observe that t-statistic values are more than 1.73 in case of MLFF, SVM and LR, whereas those values are less than 1.73 in case of GP and GMDH. From these results we can say that PNN outperformed all classifiers except GP and GMDH.

When we take a close look at the top 10 and top 18 features shown in Table 2, we observe that most of these features are associated with the firm's ability to generate profit or income. Among the top 10 features, eight features are associated with the profitability of the firm. A closer looks reveals that among the top 10 features, four are associated with primary business income, and five are associated with either gross or net profit earned by the firm. This indicated that a fraudulent firm usually tried to inflate the profit or the income figures in order to create an impressive financial statement. Any unusual income or profit figures should be a reason for suspicion and further investigation by an auditor.

When the present dataset of 35 dimensions (financial items) is visualized using the tool Neucom [32] in the principal component dimensions by plotting the first principal component on x-axis and the second principal component on y-axis, we noticed three predominant clusters and nine outliers. This provided a possible reason for the spectacular performance of PNN because PNN is tolerant to outliers [4]. While comparing the dataset with and without feature selection, it is noticed that even after reducing the number of features to almost one third of the original number, the change in accuracies is at most 5% in all the cases except PNN, where the accuracies are reduced by 8%. From this we can infer that the t-statistic is a simple and efficient feature selection technique for picking up very significant features that ensured better accuracies. Based on our experiments, we conclude that PNN without feature selection outperformed methods such as MLFF, SVM, GP, GMDH, and LR. After feature selection, GP performed well compared to all other techniques, and PNN yielded marginally less accuracies when top 18 features are selected. Similarly, PNN outperformed all other techniques when top 10 features are selected. Also, we concluded that our results are much superior to an earlier study on the same dataset.

It should be noted that while all the techniques have equal cost, the technique that is preferred and recommended is totally dictated by the dataset at hand. Since accuracy is a major concern for financial analysts, we should select that technique which yields less misclassifications and consumes less time. This is because the performance of all of these techniques depends on the dataset on which they are used. Having said that, everything else (i.e., accuracies, sensitivity, specificity, etc.) being equal, we should select that technique which is less cumbersome, easy to understand, and easy to implement.

Table 6
t-statistic values of average AUCs of PNN compared to that of other classifiers without feature selection.

Classifier compared	t-statistic at 10% level of significance
MLFF	7.84
SVM	15.66
GP	2.11
GMDH	2.49
LR	11.58

Table 8
t-statistic values of average AUCs of PNN compared to that of other classifiers with (top 10 features) feature selection.

Classifier compared	t-statistic at 10% level of significance
MLFF	5.36*
SVM	5.69*
GP	0.41
GMDH	0.83
LR	6.35*

The * indicates that the result is statistically significant.

6. Conclusion and future research directions

This paper presents the application of intelligent techniques to predict financial statement fraud in companies. The dataset consisting of 202 Chinese companies is analyzed using the stand-alone techniques like MLFF, SVM, GMDH, GP, LR, and PNN. Then, t-statistic is used for feature subset selection and top 18 features are selected in the first case and top 10 features are selected in the second case. With the reduced feature subset the classifiers MLFF, SVM, GMDH, GP, LR, and PNN are invoked again. Results based on AUC indicated that the PNN was the top performer followed by GP which yielded marginally less accuracies in most of the cases. Also, the results obtained in this study are better than those obtained in an earlier study on the same dataset. Ten-fold cross-validation is performed throughout the study. Prediction of financial fraud is extremely important as it can save huge amounts of money from being embezzled. Our study is an important step in that direction that highlights the use of data mining for solving this serious problem.

With regards to the future research directions, we can extend this work by extracting 'if–then' rules from different classifiers. These rules can be helpful for easy understanding of the prediction process for the end user because they make the knowledge learnt by these techniques transparent. This type of knowledge elicitation can help in providing early warning. In addition to the data mining techniques used in this research, hybrid data mining techniques that combine two or more classifiers can be used on the same dataset. Also, text mining algorithms for sentiment analysis of the textual description of the financial statements can be used together with data mining algorithms for assessing the financial items in the financial statements to provide better prediction of financial statement fraud.

Acknowledgments

We are very thankful to Mr. Frank Francone for giving us permission to use the Discipulus tool (demo version) for conducting various numerical experiments reported in this paper. We want to thank the three anonymous reviewers for their insightful comments which helped to improve the quality of this paper.

References

[1] A. Aamodt, E. Plaza, Case-based reasoning: foundational issues, methodological variations, and system approaches, Artificial Intelligence Communications 7 (1) (1994) 39–59.
[2] E.I. Altman, Financial ratios, discriminant analysis and prediction of corporate bankruptcy, The Journal of Finance 23 (4) (1968) 589–609.
[3] W.H. Beaver, Financial ratios as predictors of failure, Journal of Accounting Research 4 (1966) 71–111.
[4] D.P. Berrar, C.S. Downes, W. Dubitzky, Multiclass cancer classification using gene expression profiling and probabilistic neural networks, Proceedings of the Pacific Symposium on Biocomputing, vol. 8, 2003, pp. 5–16.
[5] I. Bose, J. Wang, Data mining for detection of financial statement fraud in Chinese companies, Working Paper, The University of Hong Kong, 2008.
[6] R.C. Brooks, Neural networks: a new technology, The CPA Journal Online, http://www.nysscpa.org/cpajournal/old/15328449.htm1994.
[7] B. Busta, R. Weinberg, Using Benford's law and neural networks as a review procedure, Managerial Auditing Journal 13 (6) (1998) 356–366.
[8] T.G. Calderon, J.J. Cheh, A roadmap for future neural networks research in auditing and risk assessment, International Journal of Accounting Information Systems 3 (4) (2002) 203–236.
[9] M. Cecchini, H. Aytug, G.J. Koehler, and P. Pathak. Detecting Management Fraud in Public Companies. http://warrington.ufl.edu/isom/docs/papers/DetectingManagementFraudInPublicCompanies.pdf
[10] M.J. Cerullo, V. Cerullo, Using neural networks to predict financial reporting fraud: Part 1, Computer Fraud & Security 5 (1999) 14–17.
[11] Y. Chen, D. Liginlal, A maximum entropy approach to feature selection in knowledge-based authentication, Decision Support Systems 46 (1) (2008) 388–398.
[12] A. Deshmukh, L. Talluru, A rule-based fuzzy reasoning system for assessing the risk of management fraud, International Journal of Intelligent Systems in Accounting, Finance & Management 7 (4) (1998) 223–241.
[13] K.M. Fanning, K.O. Cogger, Neural network detection of management fraud using published financial data, International Journal of Intelligent Systems in Accounting, Finance, and Management 7 (1) (1998) 21–41.
[14] 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.
[15] 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.
[16] X. Fu, F. Tan, H. Wang, Y.Q. Zhang, R. Harrison, Feature similarity based redundancy reduction for gene selection, Proceedings of the International Conference on Data Mining (Las Vegas, NV, USA), June 26–29, 2006.
[17] http://en.wikipedia.org/wiki/Enron.
[18] http://en.wikipedia.org/wiki/MCI_Inc.
[19] http://www.examiner.com/x-17547-Financial-Fraud-Examiner~y2009m7d17-Financial-Fraud-101-Understanding-the-Fraud-Triangle.
[20] 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.
[21] 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.
[22] Y. Kim, Toward a successful CRM: variable selection, sampling, and ensemble, Decision Support Systems 41 (2) (2006) 542–553.
[23] E. Kirkos, C. Spathis, Y. Manolopoulos, Data mining techniques for the detection of fraudulent financial statement, Expert Systems with Applications 32 (2007) 995–1003.
[24] KNIME 2.0.0. http://www.knime.org
[25] E. Koskivaara, Different pre-processing models for financial accounts when using neural networks for auditing, Proceedings of the 8th European Conference on Information Systems, vol. 1, 2000, pp. 326–332,, Vienna, Austria.
[26] E. Koskivaara, Artificial neural networks in auditing: state of the art, The ICFAI Journal of Audit Practice 1 (4) (2004) 12–33.
[27] S. Kotsiantis, E. Koumanakos, D. Tzelepis, V. Tampakas, Forecasting fraudulent financial statements using data mining, International Journal of Computational Intelligence 3 (2) (2006) 104–110.
[28] J.R. Koza, Genetic programming: on the programming of computers by means of natural selection, MIT press, Cambridge, MA, 1992.
[29] H. Liu, J. Li, L. Wong, A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns, Genome Informatics 13 (2002) 51–60.
[30] 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.
[31] D. Mladenic, M. Grobelnik, Feature selection on hierarchy of web documents, Decision Support Systems 35 (1) (2003) 45–87.
[32] Neucom, http://www.aut.ac.nz/research/research-institutes/kedri/research-centres/centre-for-data-mining-and-decision-support-systems/neucom-project-home-page#download.
[33] Neuroshell 2.0, Ward Systems Inc. http://www.wardsystems.com.
[34] 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, vol. 2, 1996, pp. 1007–1012,, New Orleans, LA, USA.
[35] M. Panik, Regression Modeling Methods, Theory, and Computation with SAS, CRC Press, 2009.
[36] S. Piramuthu, On learning to predict web traffic, Decision Support Systems 35 (2) (2003) 213–229.
[37] S. Ramamoorti, A.D. Bailey Jr., R.O. Traver, Risk assessment in internal auditing: a neural network approach, International Journal of Intelligent Systems in Accounting, Finance & Management 8 (3) (1999) 159–180.
[38] M. Ramos, Auditor's responsibility for fraud detection, Journal of Accountancy 195 (1) (2003) 28–35.
[39] G.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning Internal Representations by Error Propagation, MIT Press, Cambridge, MA, 1986.
[40] M.F. Selekwa, V. Kwigizile, R.N. Mussa, Setting up a probabilistic neural network for classification of highway vehicles, International Journal of Computational Intelligence and Applications 5 (4) (2005) 411–423.
[41] J.E. Sohl, A.R. Venkatachalam, A neural network approach to forecasting model selection, Information & Management 29 (6) (1995) 297–303.
[42] C. Spathis, M. Doumpos, C. Zopounidis, Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques, European Accounting Review 11 (3) (2002) 509–535.
[43] D.F. Specht, Probabilistic neural networks, Neural Networks 3 (1990) 110–118.
[44] V. Vapnik, Adaptive and learning systems for signal processing, in: HaykinS. (Ed.), Statistical Learning Theory, John Wiley and Sons, 1998.
[45] D.P. Williams, V. Myers, M.S. Silvious, Mine classification with imbalanced data, IEEE Geoscience and Remote Sensing Letters 6 (3) (2009) 528–532.
[46] 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.

Pediredla Ravisankar is working as a Software Engineer in Capgemini, Hyderabad, since February, 2010. He obtained his M.Tech (Information Technology) with specialization in Banking Technology and Information Security from UoH and IDRBT, Hyderabad (2009) and M.Sc. (Physics) from UoH, Hyderabad (2007). He has published papers in Knowledge-Based Systems, Information Sciences, International Journal of Data Mining, Modeling and Management and an IEEE conference paper. He is nominated for Marquis Who's Who in the world for 2011. His research interests include data mining, soft computing, evolutionary algorithms, neural networks and their applications.

Vadlamani Ravi is an Associate Professor in the Institute for Development and Research in Banking Technology (IDRBT), Hyderabad, since April 2010. He obtained his Ph.D. in Soft Computing from Osmania University, Hyderabad and RWTH Aachen, Germany (2001); MS (Science and Technology) from BITS, Pilani (1991) and M.Sc. (Statistics & Operations Research) from IIT, Bombay (1987). Prior to joining IDRBT, he worked as a Faculty at the Institute of Systems Science (ISS), National University of Singapore for three years. Earlier, he worked as Assistant Director at the Indian Institute of Chemical Technology (IICT), Hyderabad. He was deputed to RWTH Aachen (Aachen University of Technology) Germany under the DAAD Long Term Fellowship to carry out advanced research during 1997–1999. In a career spanning 22 years, Dr. Ravi has worked in the applications of Fuzzy Computing, Neuro Computing, Soft Computing, Data Mining, Global/Multi-Criteria/Combinatorial Optimization and Multivariate Statistics in Financial Engineering, Software Engineering, Reliability Engineering, Chemical Engineering, Environmental Engineering, Chemistry, Medical Entomology, Bioinformatics and Geotechnical Engineering. He published 93 papers in refereed International / National Journals / Conferences and invited chapters in edited volumes. He edited a Book on "Advances in Banking Technology and Management: Impact of ICT and CRM", published by IGI Global, USA, 2007. Further, he is a referee for 25 International Journals of repute in Computer Science, Operations Research, Computational Statistics, Economics and Finance. Moreover, he is an Editorial board member of International Journal of Information Systems in the Service Sector (IJISSS), IGI Global, USA, International Journal of Data Analysis Techniques and Strategies (IJDATS), Inderscience Publications, Switzerland, International Journal of Information and Decision Sciences (IJIDS), Inderscience Publications, Switzerland, International Journal of Information Technology Project Management (IJITPM), IGI Global, USA. His current research interests include Bankruptcy Prediction, CRM, Churn Prediction, FOREX rate prediction, Risk Modeling and Asset Liability Management through Optimization, Software reliability prediction, Software development cost estimation. He is listed in Marquis Who's Who in the World 2009, 2010: Marquis Who's Who in Science and Engineering in 2011. Also, he is an Invited Member of the 2000 Outstanding Intellectuals of the 21st Century 2009/2010 and 100 Top Educators in 2009 both published by International Biographical Center, UK.

Gundumalla Raghava Rao is working as Research Associate for IDRBT since May 2009. He holds an M.Tech (Computer Science & Engineering) from National Institute of Technology, Rourkela in 2008. He holds a B.Tech (Computer Science & Engineering) from M.I.T.S, Rayagada under Biju Patnaik University of Technology, Orissa. His research interests include data mining.

Indranil Bose is an associate professor of Information Systems at the School of Business, The University of Hong Kong. Prior to that, he was a faculty member at the University of Texas at Arlington and at the University of Florida. He holds a B.Tech. from the Indian Institute of Technology, MS from the University of Iowa, MS and Ph.D. from Purdue University. His research interests are in telecommunications, information security, data mining, and supply chain management. His publications have appeared in *Communications of the ACM*, *Communications of AIS*, *Computers and Operations Research*, *Decision Support Systems*, *Electronic Commerce Research & Applications, Ergonomics*, *European Journal of Operational Research*, *Information & Management, Information Systems and e-Business Management, Journal of the American Society for Information Science and Technology, Journal of Organizational Computing and Electronic Commerce*, *Operations Research Letters,* among others. His research is supported by several grants from academia and industry. He serves as Associate Editor/Editorial Review Board Member of *Communications of AIS, Information & Management, Journal of Global Information Technology and Management*, *Information Resources management Journal*, *International Journal of Information Systems and Supply Chain Management, Journal of Database Management,* etc. He has also served as guest editor for *Communications of AIS, Decision Support Systems, European Journal of Information Systems,* and *Journal of Organizational Computing and Electronic Commerce.*