Bankruptcy Prediction Using Memetic Algorithm with Fuzzy Approach: Empirical Evidence from Iran

Gholamreza Karami

Assistant Professor of Accounting

Faculty of Management, University of Tehran, Tehran, Iran

P.O.Box: 14155-6311

Tel: 98-912-140-6910   E-mail: ghkarami@ut.ac.ir

Seyed Mostafa Seyed Hosseini (Corresponding author)

Faculty of Management, University of Tehran, Tehran, Iran

P.O.Box: 14155-6311

Tel: 98-935-473-7464   E-mail: m.hosseini.ut@gmail.com

Navid Attaran

Faculty of Management, University of Tehran, Tehran, Iran

P.O.Box: 14155-6311

Tel: 98-912-524-8093   E-mail: attarannavid@hotmail.com

Seyed Mojtaba Seyed Hosseini

Faculty of Engineering, Azad University, Branch of Mashhad, Mashhad, Iran

Tel: 98-915-307-5325   E-mail: mojtaba.hoseini@yahoo.com

Abstract

Several corporate failures have recently occurred, many have suffered from serious losses and most notably public confidence has deteriorated. In order to facilitate investor's decision making regarding potential investment opportunities, this paper seeks to demonstrate that it is model specific constraints that limits the usefulness of accounting information, not the nature of variables per se. Thus we develop a Hybrid model which is an adaptive Memetic Algorithm combined with fuzzy approach that generates and optimizes a set of if-then rules for bankruptcy prediction. Data are derived from Tehran Stock Exchange (TSE) data bank, adopting 18 variables all of which are accounting ones, between 2001 and 2009. Four out of five models used in this survey have either accomplished high degree of accuracy or low level of type I error; however experimental results show that in terms of both average accuracy in prediction and occurrence of type I and II errors, fuzzy memetic performs better than GA, MLP, C4.5 and LDA in comparison.

Keywords: Bankruptcy prediction, Memetic algorithm, Fuzzy learning, Accounting variables, Cross validation, Article 141 of commercial codes of Iran

1. Introduction and Literature Review

Numerous studies have been performed regarding bankruptcy prediction and its importance is becoming more vital due to recent failures, financial distress and most importantly its effect on various stakeholders such as stockholders, managers, creditors, potential investors and regulatory institutions. Warner(1977) indicates that the direct costs associated with bankruptcy (such as court costs, lawyer costs, and accountant fees) may be around 4% of the firm value, and that both direct and indirect costs (such as lost sales, lost profits, higher cost of credit, inability to issue new securities, and lost investment opportunities) may be around 28. Therefore, it is important to detect potential

insolvency at its early stages. (Premachandra et al, 2011). Aziz and Dar (2006) have cited the methodologies and findings of previous studies through a comprehensive literature review. Three main categories in predicting corporate failure include statistical models, artificial intelligence expert system models (AIES) and theoretical models (Aziz & Dar, Ibid). The initial models were parametric ones where subsequent research introduced non-parametric models as well as methods for improving their parameters (Davalos et al, 2009). Statistical bankruptcy classification methods include Univariate and Multivariate analysis (Altman, 1968). The insight of both practitioners and researchers, originating with the papers of Beaver and Altman, was that firms with certain financial structures have a greater probability of default and eventual bankruptcy than other types of firms (Martin et al, 2011). Primary multivariate methods include cluster analysis, factor analysis, MDA, multidimensional scaling, logit analysis (Ohlson, 1980), probit analysis (Zmijewski, 1984), Fischer's LDA (Fisher, 1936), Altman's Z score, and logit-probit (Zhang et al, 1999).

Non-parametric statistical models are based on machine learning techniques such as Artificial Neural Networks (ANN); thereafter rule induction techniques have been introduced to improve the limitations and performance of the earlier models. However, they are not consistent in outperforming the "classical statistical" models (Davalos et al, Ibid). Until 1980's discriminant analysis was the dominant method in failure prediction (Back et al, 1996). Since the early 1990's, neural networks, and Multi-Layer Perceptron neural networks in particular, have been widely used to design bankruptcy prediction models. These neural networks make it possible to get around the statistical constraints of discriminant analysis, the main technique used to design such models since Altman (du Jardin, 2010). Neural networks perform classification tasks in a way intended to emulate brain processes (Aziz & Dar, Ibid). There are generally three layers in neural network, the first layer—input layer, the second—hidden layer and the third layer—output layer (Martin et al, Ibid). Later on Genetic algorithms (GA) developed which have played an important role in conducting bankruptcy models, Based on the idea of genetic inheritance and Darwinian Theory of natural evolution (survival of the fittest genes), GA works as a stochastic search technique to find an optimal solution to a given problem from a large number of solutions .Varetto was the first person who introduced a bankruptcy classification model based on a genetic algorithm. Using GA we can choose the financial ratio set dynamically by considering features such suitability of bankruptcy model for an industry, Number of ratios used for prediction, suitability of classification model, all of which can increase the accuracy of the prediction. Neural network is preferred for non-linear data with learning capacity, while its lack of explanatory power due to rational of the decision being made is a black box (Martin et al, Ibid). Fuzzy Neural Network adds rules to Neural Network which could overcome the black box issue although learning capacity is reduced. It is anticipated that Genetic Fuzzy Neural Network improves the learning efficiency. These multi-classifier models can enhance bankruptcy classification by combining the strengths of the different models.

There are two types of multi-classifier models, one, typically called hybrid, involves an optimized model focused on manipulating the parameters for a classifier model that generates a classification. A second type of multi-classifier model combines the output of several classifiers into a single classifier; known as Ensemble, which performs better than single classifiers but is more time consuming to develop since the contribution of each classifier needs to be determined and in some cases, different combinations need to be examined. Bagging is a commonly used method for aggregating classifiers into an ensemble (Davalos et al, Ibid).Inductive learning on the other hand can improve the performance of bankruptcy models (Tsai, 2008). Rule induction generates a model in human terms in the form of if-then rules. Decision tree algorithms and Genetic algorithms have also been successfully applied to rule induction logic. Of course, in terms of increasing accuracy in prediction, several factors are involved, such as the nature of variables, their number and relevance.

The reliability of bankruptcy classification models developed using only financial ratios is in question since there is doubt about the validity and reliability of the accounting information used for the ratios (Agarwal, Taffler, 2008). The main difficulty in choosing a set of predictors is the large number of variables that could be used. Analyzing nearly 200 scientific papers dealing with corporate failure prediction published in the last 50 years, Philippe du Jardin found that more than 500 different ratios were used in the final models (du Jardin, Ibid). In addition, the relevance of particular ratios depends on changes in the environment (Tsai, Ibid). Thus, classifier models have included other information such as firm population characteristics, macro-economic factors, and market variables (Shumway, 2001). Ohlson (Ibid) suggested trying all the possible and viable combinations of variables and models to find an optimal set of variables, albeit this approach is computationally not feasible (Varetto, 1999).

2. Research Hypotheses

In order to evaluate the accuracy power and occurrence of type I error regarding bankruptcy predictions, independent sample t-test is used between FM and GA, MPL, C4.5 and LDA predictions, thus four hypotheses are proposed:

H1: Fuzzy Memetic achieves higher average accuracy and lower type I error comparing to Genetic Algorithm in bankruptcy prediction.

H2: Fuzzy Memetic achieves higher average accuracy and lower type I error comparing to MLP in bankruptcy prediction.

H3: Fuzzy Memetic achieves higher average accuracy and lower type I error comparing to C4.5in bankruptcy prediction.

H4: Fuzzy Memetic achieves higher average accuracy and lower type I error comparing to LDA in bankruptcy prediction.

3. Research Methodology

Memetic algorithms just like genetic algorithms are rule-base models in bankruptcy prediction field. The model presented in this paper is a hybrid one, developed by Memetic algorithm with fuzzy approach. Regarding accuracy and occurrence of type I error this paper seeks model comparison among Genetic Algorithm(GA), Multi-Layer Perceptron(MLP), Linear Discriminant Analysis (LDA), C4.5 (Decision Tree) along with the presented model of this paper, FM (Fuzzy Memetic).

The data used in this survey are derived from Tehran stock exchange data bank for the period between 2001 and 2009. For classifying bankrupt from non-bankrupt firms, we used Article 141 of commercial codes of Iran which states: "had the cumulative losses of a firm reaches at least halve of the firms legal capital, the board is responsible for holding an extraordinary general meeting of shareholders, deciding between liquidation and survival". Data from prior year (one year to actual bankruptcy) is used in bankruptcy prediction. Table 1 presents the number of total companies, consisting of bankrupts and non-bankrupts in the mentioned period. We randomly selected 45 bankrupted and 45 non-bankrupted firms from various industries; the next step was to divide the samples into training and test groups with the same proportion.

3.1 Descriptive Statistics

Table 2 presents 18 variables selected for this analysis along with central and dispersion parameters, their standard deviation, and the result of independent sample's t-test in order to compare the means of the sample groups. The result of t-test shows that there are significant differences in two groups and those ratios such as liquidity (CA/CL), leverage (TL/TA) and profitability (ROA, NI/NR, OR/TA) are appropriate measures for classifying bankrupt from non-bankrupt.

3.2 Learning of Memetic Algorithm with Fuzzy Approach

Memetic Algorithms (MAs) are a class of stochastic global search heuristics in which Evolutionary Algorithms-based approaches are combined with problem-specific solvers. The hybridization is meant to either accelerate the discovery of good solutions, for which evolution alone would take too long to discover, or to reach solutions that would otherwise be unreachable by evolution or a local method alone. It is assumed that the evolutionary search provides for a wide exploration of the search space while the local search can somehow zoom-in on the basin of attraction of promising solutions (Krasnogor, 2006). When integrating local search with evolutionary search we are faced with the dilemma of what to do with the improved solution that is produced by the local search. To wit, suppose that individual i belongs to the population p in generation t and that the fitness of i is f(i). Furthermore, suppose that the local search produces a new individual i' with f(i') < f(i) for a minimization problem. The designer of the algorithm must now choose between two alternative options. Either (option 1) he/she replaces i with i', in which case P = P −{i}+{i'} and the genetic information in i is lost and replaced with that of i', or (option 2) the genetic information of i is kept but its fitness should alter: f(i)= f(i'). The first option is commonly known as Lamarckian learning while the second option is referred to as Baldwinian Learning (Baldwin, 1896). The fuzzy memetic (FM) model proposed in this research is based on Lamarckian approach with modest changes. While in Lamarckian learning algorithms all the genetic information of I is replaced with i', we substitute those attributes of I with that of i' according to improvements made by these changes in the final fitness function. The logic behind this is due to fuzzy perspective executed in this article. The superiority level of each individual in population comparing to another one is uncertain. In Lamarckian approach the level of superiority is neglected and those with lower advantages even by modest percentages are wholly substituted, which reduces the leaning capacity of the memetic algorithm.

Molga (2005) has used several benchmark functions for optimization needs; hence we chose three most popular of them, namely Ackley, Griewank and Rastrigin functions to compare our model with Genetic Algorithm (GA) and Standard Memetic Algorithm (MA). For assessing the efficiency of proposed model, three statistical measures were used. First, the Mean which is achieved via continuously running the algorithm 30 times, second the variance of results used in computing the Mean and finally Best Results of each algorithm. Each operation was performed in

similar condition and each on 210 seconds. Most notably is the Variance of the model which shows great stability, and its low risk toward bankruptcy prediction. The results are presented in tables 3, 4 and 5. As it is shown in each function, FM algorithm achieves the lowest amount; however, the best results are randomly obtained, thus to reliably compare the algorithms, it's appropriate to use the Means of each algorithm. On the other hand variance which represents Stability of results in continuous operations and its related risk is a vital element in bankruptcy prediction models. Regarding the results of three benchmark functions for optimization, we expect that proposed algorithm would achieve the lowest variance and highest accuracy.

Despite of usage of any model in bankruptcy prediction, crisp logic which divides firms to bankrupt and non-bankrupt along with ignoring deviation (distance) from bankruptcy frontier reduces accuracy level in prediction. On the other hand, fuzzy logic can define a grade quality and that enables us to differentiate between bulks of variables. With having article 141 of commercial codes of Iran in mind, we used a fuzzy approach toward the coverage ratio of accumulated losses to legal capital. However due to non-existence of such trade law in other countries, researchers can perform their surveys based on other assumptions.

3.3 Model Presentation

The model presented in this paper provides an if-then rule; each rule is associated with a chromosome consisting of N genes. Each gene has four fields:

Ci (gene1(X1, Le, V1, Q1) , gene2(X2 ,Le2 ,V2 , Q2),……, geneN(Xn, Len, Vn, Qn))

Xi: variable, Lei: logical equation, Vi: value, Qi: quality

Logical equation is simply "<" or ">" related to cut-off point, value is the cut-off point determined by the rule and Qi is the quality of the variable in the final fitness. To this end, we used this element for improvement of the next generation in population. If a variable deteriorates the final accuracy of the final rule, its changes should be greater than those variables which are acting accordingly due to final accuracy. Qi is derived from the total misclassification resulted from mentioned variable to the total predictions (in this paper total prediction equals 40). On one side mutation changes inappropriate variables with a random rate, while this rate can be lower, equal or greater than Qi. On the other hand, we defined a rule in the model which performs a comparison between the random rate and Qi, if Qi>random rate, then the Qi is the preferred rate for due changes in variables in next generation. While one of the aims of this paper is to reduce the type I error, we shall shift our focus from type II error to type I error. Thus, we multiplied Qi of the variables which caused type II error by 0.5.

For any prediction extracted by rules, there can be four different outcomes,

- *True positive (tp) - the rule predicts that the firm is non-bankrupt and it is not.*
- *False positive (fp) – the rule predicts that the firm is non-bankrupt but it is. (Type I error)*
- *True negative (tn) - the rule predicts that the firm is bankrupt and it is;*
- *False negative (fn) - the rule predicts that that the firm is bankrupt but it is not. (Type II error)*

Due to any prediction generated by the rules, rewards and penalties was prescribed. For determining the amount of these rewards and penalties, we assigned scores to the firms. This assignment is based on the distance of firms from bankruptcy frontier. Due to article 141 of commercial codes of Iran which states, had the cumulative losses of a firm reaches at least halve of the firms legal capital, the firm is a valid case of bankruptcy. Thus by computing coverage ratio (retained earnings/legal capital) of firms, a valid measure for score assignment is made.

$$Si=|Max \ coverage \ ratio| - |coverage \ ratio \ i| \qquad (1)$$

In above equation, Si is the score assigned to firm i, Max coverage ratio is the highest ratio in each two categories (bankrupt and non-bankrupt). The pitfall of most models in bankruptcy prediction is usually caused by those samples which are close to bankruptcy frontier, resulting in misclassification of firms, however those far above or below from frontier do not impose any errors. The logic behind this equation is that by identifying firms which are close to bankruptcy frontier (their absolute value of coverage ratio are small), the assigned scores are higher and for those with higher coverage ratio, assigned scores are lower, clearly because their identification are a lot easier. The score coefficients of any prediction are as follow:

- *true positive (tp) = +Si*
- *false positive (fp) (Type I error) =-1.5 Si*
- *true negative (tn) = + Si*
- *false negative (fn) (Type II error) = -Si*

Altman (1968) states that the cost of type I error can be 30 to 60 times as much as the cost of type II error, thus to account for this difference, we assigned 1.5*Si penalty to rules which caused type I error. In this way, those rules which have the lowest type I error are considered the optimized ones. Davalos (2009) has reached the lowest type I error, however along with lowest accuracy whereas using fuzzy approach and assigning scores through mentioned logic, enhanced accuracy is expected.

3.4 Fitness Function

Carvalho (2002) used the function below to compute fitness,

$$\text{Accuracy} = (tp+tn) / N(\text{size of population}) \qquad (2)$$

In this paper, we rewrite this equation using scores;

$$\text{Accuracy} = [\ [\textstyle\sum si(tp) + \sum si(tn) + \sum si(fn) + \sum si(fp)] / \sum si(\text{when all prediction would be true})] \qquad (3)$$

Our goal in this function is to increase accuracy, which can be reached via maximizing the numerator. Thus,

$$\text{Fitness} = \textstyle\sum Si_s \qquad (4)$$

Clearly occurrence of errors would result in lower accuracy, however there is a difference between type I and II error and their impact on accuracy.

3.5 Model Validation

This paper has used 5-fold cross validation in order to gauge the generalizability of proposed algorithm, and in the meantime it enables us to compare the performance of two or more different algorithms and to find out the best algorithm for the available data. Cross-Validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model. In typical cross-validation, the training and validation sets must cross-over in successive rounds in a way that each data point has a chance of being validated against. The basic form of cross-validation is k-fold. Other forms of cross-validation are special cases of k-fold or involve repeated rounds of k-fold cross-validation. Kohavi (1995) also obtained desirable results for 10-fold cross-validation with empirical decision trees (C4.5). Values of K small as 5 or even 2 may work even better if you analyze several different random k-way splits of the data to reduce the variability of the cross-validation estimate.

4. Empirical Result

The cross-validated prediction results are presented for each model separately. To study the consequences of different model selection approaches we have applied corresponding statistical method to test the predictive ability of constructed models. In this study, we used 5 fold cross validation method for all models, in our algorithm, each fold generates 6 rules, resulting total 30 rules and then we applied these rules to all available data which are derived from one year prior to bankruptcy .

FM generates the highest accuracy (94.1%) with relatively small standard deviation (0.7%) which implies its low risk regarding bankruptcy prediction. GA ranks second from both accuracy (92.5%) and standard deviation (1.80%) perspective. MLP has lower accuracy (86.4%) comparing to two mentioned methods, but its standard deviation of accuracy is close to LDA (8.83%). Decision Tree (C4.5) has relatively lower accuracy (87.7%) along with higher standard deviation (8.2%). LDA has reached sounded predictive ability (89.6%) despite the fact that its standard deviation of accuracy is higher than FM and GA (8.84%). Type I error is 7%, 10.1% and 10.4% for FM, MLP and GA respectively. However FM's remarking low standard deviation of type I error (0.13%) demonstrates its strong stability in numerous operations, as was expected in benchmark function results. MLP and LDA have much higher deviation regarding type I error, 8.92% and 10.1% respectively. The result from independent sample t-test which is depicted in table 7 shows there is significant discrepancy between average accuracy and type I error of FM and other four models predictions, hence all research hypotheses are accepted.

Hwang (2007) has considered the sum of both error rates (type I and type II) to be important, since type II error can also deteriorate model reliability, and it can further impose opportunity cost to investors, here, type II error for FM is 4.7%, a relatively low rate, however this rate for MLP and LDA is as high as 17.03%, 10.7% respectively. We examined five different classifiers using cross validation. The model presented in this paper is able to classify bankrupt firms from non-bankrupts better than mentioned models, jointed with high degree of stability, higher accuracy and relatively low type I error and II.

5. Conclusion

The failure prediction research has suffered from lack of any unified theory since the 1930's when first empirical studies on this subject were published. (Back et al, Ibid). In this paper we proposed a Fuzzy memetic algorithm which

belongs to Hybrid model family. The primary goal of this model was to improve bankruptcy prediction accuracy rate while attempting to reduce occurrence of type I error, not ignoring the consequences of type II error since it can impose opportunity cost to investors. Grice and Ingram (2001) reported that Altman's Z-score model declined when applied to various industries. The samples of this paper are a mixture of different industries and variables used are obtained from firm's financial statements. To achieve the optimal prediction rule, 5 fold cross validation was used while extracting 6 rules from each fold. The empirical results show that the anticipated high degree of accuracy is a valid case. However this research has some constraints in practice since determining a benchmark in order to segregate bankrupt firms form non-bankrupt ones is related to Article 141 of commercial codes of Iran. Thus, applicability of the hypothesized coverage ratio is somehow related to each bankruptcy code in each country. However, it is demonstrated that different model results are somehow caused by its specific constraints, not the very nature of the variables.

References

Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance, 32*, 1541–1551. http://dx.doi.org/10.1016/j.jbankfin.2007.07.014

Altman, E. (1968). Financial ratios, discriminant analysis and the classification of corporate bankruptcy. *The Journal of Finance, 23*(4). 589-609. http://dx.doi.org/10.2307/2978933

Aziz, M.A., & Dar, H.A. (2006). Predicting corporate bankruptcy: where we stand. *Corporate Governance (Bradford), 6*(1). 18-33. http://dx.doi.org/10.1108/14720700610649436

Back, B. Laitinen, T. Sere, K., & Van Wezel, M. (1996). Choosing bankruptcy predictors using discriminant analysis, logit analysis, and genetic algorithms. *Turku Centre for Computer Science, 40*.

Baldwin, J. (1896). A new factor in evolution. *American Naturalist, 30*, 697-710.

Carvalho, D. R., & Freitas, A.A. (2004). A Hybrid Decision Tree/Genetic Algorithm Method for Data Mining. *Information Sciences, 163*(1-3). 13-35. http://dx.doi.org/10.1016/j.ins.2003.03.013

Davalos, S. Leng, F. Feroz, E.H., & Cao, Z. (2009). Bankruptcy Classification of Firms Investigated by the US Securities and Exchange Commission: An Evolutionary Ensemble Computing Model Approach. *International Journal of Applied Decision Sciences, 2*(4). 360 - 388. http://dx.doi.org/10.1504/IJADS.2009.031180

Du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing, 73*, 2047–2060. http://dx.doi.org/10.1016/j.neucom.2009.11.034

Fisher, R. A. (2010). The use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics. 7*, 179-188. http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x

Grice, J.S., & Ingram, R.W. (2001). Tests of the generalizability of Altman's bankruptcy classification model *Journal of Business Research, 54*(1), 53-61. http://dx.doi.org/10.1016/S0148-2963(00)00126-0

Hwang, R.C. Cheng, k.F., & Lee, H.C. (2007). A Semi parametric Method for Predicting Bankruptcy. *Journal of Forecasting, 26*. 317–342. http://dx.doi.org/10.1002/for.1027

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *14th International Joint Conference on Artificial Intelligence, (IJCAI-95)*. 1137–43.

Krasnogor, N. Aragon, A., & Pacheco, J. (2006). *Metaheuristic Procedures for Training Neural Networks, (*1st ed.). Berlin: Springer.

Martin, A. Saranya,V. G. Gayathri, P., & Venkatesan, P. (2011). Hybrid Model for bankruptcy prediction using genetic algorithm, fuzzy c-means and mars. *International journal of soft computing, 2*(1), 13-24.

Molga, M., & Smutnicki, C. (2005). Test functions for optimization needs.[online] Available: http://www.zsd.ict.pwr.wroc.pl/files/docs/functions.pdf,.

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*, 109–131. http://dx.doi.org/10.2307/2490395

Premachandra, I.M Chen , Y. Watson, J. (2011). DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment. *Omega: the international Journal of Management Science. 39*. 620-626.

Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard model. *Journal of Business*, *74*, 101–124. http://dx.doi.org/10.1086/209665

Tsai, C.F. (2009). Feature Selection in Bankruptcy Classification. *Knowledge-Based Systems, 22*(2), 120-127. http://dx.doi.org/10.1016/j.knosys.2008.08.002

Varetto, F. (1998). Genetic Algorithms Applications in the Analysis of Insolvency Risk. *Journal of Banking and Finance, 22*, 1421-1439. http://dx.doi.org/10.1016/S0378-4266(98)00059-4

Warner, J.B. (1977). Bankruptcy costs: some evidence. *The journal of Finance*, *16-18*, 337-347. http://dx.doi.org/10.2307/2326766

Zhang, G. Hu, M.Y. Patuwo, B.E., & Indro, D. C. (1999). Artificial Neural Networks in Bankruptcy Classification General Framework and Cross-Validation Analysis. *European Journal of Operational Research*, 116, 16-32. http://dx.doi.org/10.1016/S0377-2217(98)00051-4

Zmijewski, M. (1984). Methodological Issues Related to the Estimation of Financial Distress prediction Models. *Journal of Accounting Research, 20*, 59-82. http://dx.doi.org/10.2307/2490859

Table 1. Number of bankrupt and non-bankrupt firms

Year	Firms	Bankrupt Firms	Non-bankrupt firms
2001	313	12	301
2002	335	10	325
2003	393	5	388
2004	406	14	392
2005	416	18	398
2006	424	7	417
2007	426	6	420
2008	428	9	419
2009	429	13	416
Firm/Year	3570	94	3476

Table 2. Descriptive statistics and t-test data

Row	Variable	Definition	Min	Max	Mean	Std.dev	T-test
1	C/NS	Cash over net sale	0.003	0.528	0.053	0.070	0.257
2	C/TA	Cash over total assets	0.001	0.889	0.044	0.096	0.851
3	C/CL	Cash over current liabilities	0.002	1.097	0.071	0.135	0.288
4	CA/CL	Current assets over current liabilities	0.007	2.3	1.061	0.381	0.028
5	C/TL	Cash over total liability	0.002	1.011	0.094	0.208	0.042
6	TL/TA	Total liabilities over total assets	0.029	0.945	0.709	0.194	0.00
7	NI/TL	Net income over total liabilities	-0.168	2.472	0.212	0.370	0.00
8	ROS	Net income over net sale	-0.604	1.116	0.138	0.264	0.00
9	ROE	Return on equity	-0.65	3.81	0.331	0.643	0.00
10	ROA	Return on assets	-0.134	0.581	0.091	0.138	0.00
11	NS/TA	Net sale over total assets	0.099	2.613	0.841	0.419	0.074
12	NI/FA	Net income over fix assets	-0.222	59.688	1.948	8.350	0.05
13	OI/TA	Operating income over total assets	-0.141	0.654	0.112	0.141	0.00
14	RE/TA	Retained earnings over total assets	-0.139	0.428	0.070	0.124	0.00
15	EBIT/TA	Earnings before interest and taxes over total assets	-0.054	0.663	0.136	0.139	0.00
16	NS/OE	Net sale over owners' equity	0.368	30.12	4.559	4.109	0.145
17	CA/NS	Current assets over net sale	0.005	7.952	1.101	1.153	0.854
18	Growth	Sale growth in the past 3 years	-0.833	4.498	0.506	0.777	0.177

Table 3. Minimizing 100-dimensional Ackley function

algorithm	Mean	variance	Best result
GA	2.8790e+000	1.9370e-002	2.6665e+000
MA	5.0886e-001	5.9523e-003	3.8527e-001
FM	1.7914e-003	2.2273e-006	4.3375e-005

Table 4. Minimizing 100-dimensional Griewank function

Algorithm	Mean	variance	Best result
GA	2.4330e+000	9.3362e-002	1.9475e+000
MA	9.9118e-001	3.6275e-003	8.2449e-001
FM	6.8825e-004	2.1434e-006	5.9892e-008

Table 5. Minimizing 100-dimensional Rastrigin function

algorithm	Mean	variance	Best result
GA	2.9685e+001	1.3718e+001	2.3188e+001
MA	1.5130e+000	1.0415e-001	8.8490e-001
FM	9.3813e-005	2.0153e-008	1.0623e-010

GA: Genetic Algorithm, MA: Memetic Algorithm, FM: fuzzy Memetic

N: number of dimensions (variables)

Table 6. Cross validation results

Model	Ave accuracy	Type I error	Type II Error	St.dev Accuracy	St.dev Type I
FM	94.1%	7%	4.7%	0.7%	0.13%
GA	92.5%	10.4%	4.80%	1.80%	4.90%
MLP	86.4%	10.1%	17.03%	8.83%	8.92%
C4.5	87.7%	15.5%	8.8%	4.3%	8.2%
LDA	89.6%	11.1%	10.7%	8.84%	10.1%

Table 7. Independent-Sample T-test Results

H1	Ave accuracy	Error Type I	H2	Ave accuracy	Error Type I
FM	94.10%	7%	FM	94.10%	7%
GA	92.50%	10.40%	MLP	86.40%	10.10%
T-test statistics	3.986	3.456	T-test statistics	4.774	2.101
Sig.(2-tailed)	0.000	0.001	Sig.(2-tailed)	0.000	0.048
H3	Ave accuracy	Error Type I	H4	Ave accuracy	Error Type I
FM	94.10%	7%	FM	94.10%	7%
C4.5	87.70%	15.50%	LDA	89.60%	11.10%
T-test statistics	4.01	0.31	T-test statistics	2.744	2.23
Sig.(2-tailed)	0.000	0.025	Sig.(2-tailed)	0.001	0.033