Expert Systems with Applications 37 (2010) 2292–2306

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Performance evaluation for classification methods: A comparative simulation study

Yong Soo Kim *

Division of Industrial and Management Engineering, Sungkyul University, 400-10, Anyang 8-dong, Manan-gu, Anyang-City, Gyeonggi-do 430-742, Republic of Korea

ARTICLE INFO

Keywords:
Classification
Data mining
Statistical method
Artificial neural network
Decision tree
Logistic regression

ABSTRACT

In this article, the performance of classification methods was empirically compared while varying the number of classes of dependent variables, the number of independent variables, the types of independent variables, the number of classes of the independent variables, and the sample size. Our study employed 324 simulated examples, with artificial neural networks and decision trees as the data mining techniques, and logistic regression as the statistical method. In the performance study, we use the misclassification errors as the metric and come up with some additional findings: (i) for continuous independent variables, a statistical technique (i.e., logistic regression) was superior to data mining techniques (i.e., artificial neural network and decision tree) when dependent variable has binary values, while the artificial neural network was best when the number of classes of dependent variable was three or more; (ii) for continuous and categorical independent variables, logistic regression performs better than artificial neural network and decision tree in the case of small number of independent variables and small sample size, while artificial neural network was best in other cases; and (iii) the artificial neural network performance improved faster than that of other methods as the number of independent variables and the number of classes of dependent variables increases.

1. Introduction

The difficulties posed by prediction and classification problems have resulted in a variety of problem-solving techniques. For example, data mining methods comprise artificial neural networks and decision trees, and statistical techniques include linear regression and logistic regression. However, it is difficult to compare the performance of the techniques and determine the best one since their performance is data-dependent (Kim, 2008). Prediction techniques deal with dependent variable with continuous values while classification techniques handle the dependent variable that has categorical values such as ‘A’, ‘B’, and ‘C’.

Quite a lot of research has been done to compare the performance of data mining and statistical approaches to solving prediction and classification problems. In prediction problems, comparison studies between statistical and data mining techniques have been carried out as follows. Gorr, Nagin, and Szczypula (1994) compared linear regression, stepwise polynomial regression, and neural networks in the context of prediction student GPAs. Although they found that linear regression performed overall, none of the methods performed significantly better than the ordering index used by the investigator. Shuhui, Wunsch, Hair, and Giesselmann (2001) reported that neural networks performed better than linear regression for wind farm data, while Hardgrave, Wilson, and Walstrom (1994) experimentally showed that neural networks did not significantly out perform statistical techniques in predicting the academic success of students entering the MBA program. Cao, Leggio, and Schniederjans (2005) utilized artificial neural networks and linear regression to predict stock price movement and compared the predictive power of linear models from financial forecasting literature to that of univariate and multivariate neural network models. Their results showed that neural networks outperformed the linear regression compared. Kim, An, and Kang (2004) examined the performance of three cost estimation models based on linear regression, artificial neural network and case-based reasoning of the data of 530 historical costs and found that the best artificial neural network model gave more accurate estimation results than either the linear regression and case-based reasoning models. Raaymakers and Weijters (2003) used artificial neural network and linear regression methods to predict the makespan of job sets in batch process industries and compared their performances. It was shown that both methods were robust for changes in the number of jobs, the averages processing time, a more unbalanced workload and for different resource configuration, but the prediction accuracy of artificial neural networks appeared significantly better than the linear regression models. Wang and Elhag (2007) utilized artificial neural network, linear regression and the evidential reasoning in order to model bridge risks. It was found that the artificial neural networks outperform the evidential reasoning and linear regression for the considered case study. Subbanarasimha, Arinze, and Anadarajan (2000) demonstrated linear

* Tel.: +82 31 467 8311; fax: +82 31 467 8067.
E-mail addresses: yskim@sungkyul.ac.kr, yskim95@kaist.ac.kr

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.07.043

regression performed better than neural networks when the distribution of the dependent variable was skewed, and Kumar (2005) expanded on Subbanarasimha et al. (2000) result, developing a hybrid method that improved the prediction accuracy. These comparison studies had mainly considered a specific data set or the distribution of the dependent variable. Other unexplored criteria, however, affect the performance of decision problem techniques such as sample size, characteristics of the independent and dependent variables. Therefore, Kim (2008) empirically compared the performance of data mining and statistical techniques while varying the number of independent variables, the number of classes of the independent variables, and the sample size.

In addition, a few comparison studies of classification methods have been performed as follows. Tam (1991) and Tam and Kiang (1992) compare the neural networks, decision tree analysis, discriminant analysis, logistic regression and k Nearest Neighbor method for the data on bank failures. It was shown that artificial neural networks have better predictive accuracy than the other methods. Salchenberger, Cinar, and Lash (1992) evaluate the ability of a neural network to predict thrift institution failures by comparing it with the best logit model for the data. It was found that the artificial neural networks could achieve better predictive accuracy than the logit model. Desai, Crook, and Overstreet (1996) explored the ability of neural networks and traditional techniques such as linear discriminant analysis and logistic regression in building credit scoring models in the credit union environment and compared their performances with customized credit scoring model. It was indicated that customized neural networks offered a very promising avenue if the measure of performance was percentage of bad loans correctly classified. However, if the measure of performance was percentage of good and bad loans correctly classified, logistic regression models were comparable to the neural networks approach. The issue of sample size is investigated by Patuwo, Hu, and Hung (1993) in classification problems. With an increase in sample size, it was found that prediction accuracy of artificial neural networks had better than the other methods.

As seen above, the conventional comparison studies of classification methods had considered a specific data set or the sample size. That is, a variety of data types were not handled in classification problems while various data types were considered in prediction problems (Kim, 2008).

In this article, the performance of data mining and statistical methods for classification problem was empirically compared based on the number of independent variables, the types of independent variables, the number of classes of the independent variables, the number of classes of the dependent variables and the sample size. That is, in this article, Kim's (2008) approach has been applied to compare the performance of classification methods. Our study employed 324 simulated examples, with artificial neural networks and decision trees as the data mining techniques and logistic regress as the statistical method.

In addition to these general comparison result, we used the misclassification error as the metric and determined the following: for continuous independent variables, a statistical technique (i.e., logistic regression) was superior to data mining techniques (i.e., artificial neural network and decision tree) when dependent variable has binary values, while the artificial neural network was best when the number of classes of dependent variable was three or more; for continuous and categorical independent variables, logistic regression performs better than artificial neural network and decision tree in the case of small number of independent variables and small sample size, while artificial neural network was best in other cases; and the artificial neural network performance improved faster than that of other methods as the number of independent variables and the number of classes of dependent variables increases.

The article is organized as follows: Section 2 illustrates the generation of the data sets and analysis methods for the empirical study. The experimental results are described in Section 3, and the conclusions and future research directions are presented in Section 4.

2. Data analysis

2.1. Simulated data generation

In this section, we describe the 324 simulated classification problems that we generated to evaluate the performance of the decision tree, neural network, and logistic regression techniques. First, Table 1 shows 36 simulated examples with continuous independent variables. These 36 examples were obtained from the linear model, where x_i was randomly selected in the range [0, 1], and ε was normally distributed with mean 0 and standard deviation 1. The number of independent variables was set to three, five or seven and the sample size was set to 100, 500, 1000, or 10,000. In addition, continuous values of a dependent variable(y) were converted to categorical classes such as 'A', 'B' or 'C'. The number of classes of the dependent variable was set to two, three or four.

In the case of two categorical classes, a value of a dependent variable was converted to category 'A' when it is less than 50% and as 'B' when greater than 50%. For three categorical classes, the dependent variable was categorized as 'A' when it was less than 25%, as 'B' when greater than 75%, and as 'C' otherwise. Finally, for four categorical classes, the dependent variable was categorized as 'A' when it was less than 25%, 'B' as when it was greater than 25% and less than 50%, 'C' as when it was greater than 50% and less than 75%, and as 'D' otherwise.

Some continuous independent variables in Table 1 were converted to categorical variables in Tables 2, in which three independent variables were considered. In the case of two categorical classes, a value of a continuous independent variable was converted to category 'A' when it is less than 50% and as 'B' when greater than 50%. For three categorical classes, each continuous variable was categorized as 'A' when it was less than 25%, as 'B' when greater than 75%, and as 'C' otherwise.

Table 2 describes 48 simulated examples with three independent variables including categorical variables. These 48 examples were obtained from Table 1. As mentioned above, a continuous variable x_i was converted to a categorical variable c_i (see to Table 2). In this manner, 96 and 144 simulated examples with five and seven independent variables including categorical variables were generated, respectively.

2.2. Data analysis methods

In this section, the artificial neural network (ANN), decision tree analysis (DT), and logistic regression (LR) techniques are applied to the 324 simulated examples to evaluate their prediction accuracy. Each example was randomly divided into two sets, a training set and a test set. The training set consisted of 70% of the data while the remainder was assigned to the test set. For simplicity, our performance comparisons only considered the misclassification error of the test set. The analyses were performed using "SAS Enterprise Miner".

The ANN employed in this study was a multilayer feed-forward network trained by a backpropagation algorithm. The number of hidden layers was set to either one or two. For each hidden layer, the number of hidden neurons varied between 3 and 15 to identify the best ANN structure. The learning rate and momentum were set to 0.1 and 0.9, respectively. A low learning rate ensures a continuous descent on the error surface, and a high momentum is able to

Table 1
Simulated examples with continuous independent variables.

ID	No. of classes of dependent variable	No. of independent variables	Sample size	Relationship	Conversion of a dependent variable
S1	2	3	100	$y = 1 + 3x_1 + 2x_2 + 2x_3 + \varepsilon$	y is converted to two categorical values. y has a value of 'A' or 'B'
S2			500		
S3			1000		
S4			10,000		
S5		5	100	$y = 1 + 3x_1 + 2x_2 + 2x_3 + x_4 + x_5 + \varepsilon$	
S6			500		
S7			1000		
S8			10,000		
S9		7	100	$y = + 3x_1 + 2x_2 + 2x_3 + x_4 + x_5 + 0.5x_6 + 0.5x_7 + \varepsilon$	
S10			500		
S11			1000		
S12			10,000		
S13	3	3	100	$y = 1 + 3x_1 + 2x_2 + 2x_3 + \varepsilon$	y is converted to three categorical values. y has a value of 'A', 'B' or 'C'
S14			500		
S15			1000		
S16			10,000		
S17		5	100	$y = 1 + 3x_1 + 2x_2 + 2x_3 + x_4 + x_5 + \varepsilon$	
S18			500		
S19			1000		
S20			10,000		
S21		7	100	$y = 1 + 3x_1 + 2x_2 + 2x_3 + x_4 + x_5 + 0.5x_6 + 0.5x_7 + \varepsilon$	
S22			500		
S23			1000		
S24			10,000		
S25	4	3	100	$y = 1 + 3x_1 + 2x_2 + 2x_3 + \varepsilon$	y is converted to four categorical values. y has a value of 'A', 'B', 'C', or 'D'
S26			500		
S27			1000		
S28			10,000		
S29		5	100	$y = 1 + 3x_1 + 2x_2 + 2x_3 + x_4 + x_5 + \varepsilon$	
S30			500		
S31			1000		
S32			10,000		
S33		7	100	$y = 1 + 3x_1 + 2x_2 + 2x_3 + x_4 + x_5 + 0.5x_6 + 0.5x_7 + \varepsilon$	
S34			500		
S35			1000		
S36			10,000		

speed up the training process (Sarle, 1994; Yeh, Hamey, & Westcott, 1998). These values are typically used for ANN training (Ting, Yunus, & Salleh, 2002).

For DT, we varied the splitting criterion and used two parameters for pre-pruning: 'minimum number of observations in a leaf' and 'observations required for a split search.' The splitting criterion was set to '*F*-test at 10% significance level', 'Entropy reduction' or 'Gini reduction.' The 'minimum number of observations in a leaf' and 'observations required for a split search' parameters were set to either 5% and 10% or 10% and 20% of sample size, respectively. Thus, six decision trees were generated.

Finally, LR used the 'Logit' and 'Probit' as the link function and all independent variables were considered.

3. Experimental evaluation

Computational results (i.e., Misclassification errors) are summarized in Tables 5–8 for ANN, DT and LR classification methods (*M*), sample size (*S*), number of independent variables (*V*), number of categorical variables (*CA*), number of classes of categorical dependent variables (*CL_Y*), and number of classes of independent categorical variables (*CL_X*).

Table 3 shows the misclassification errors for ANN, DT and LR when the independent variables were continuous. The classification methods of ANN and LR consistently performed better than DT. Furthermore, when *CL_Y* was set to 2, LR was superior or equal to ANN in almost all cases, although the differences between the misclassification errors decreased as *S* increased. However, when *CL_Y* was set to 3 or 4, ANN outperformed LR except for some cases.

Table 4 gives the experimental results for the case when three independent variables included one or two categorical independent variables. It is not shown that any method outperforms the other. However, we have known that the performance of DT decreased as *CL_Y* increased.

Table 5 shows that for five independent variables that included one, two, three or four categorical variables, ANN outperformed the other methods in many cases except for *CL_Y* = 2. Note that when *CL_Y* was set two, the performances of three methods were identical. However, as *CL_Y* increased, ANN performed better than LR and DT in many cases. Especially, when *CL_X* was set two, the misclassification errors for ANN remained smaller than for LR and DT as *CA* increased.

Table 6 illustrates the experimental results for the case when seven categorical independent variables. The results in Table 6 are similar to those in Table 5. That is, it is not concluded that any method performed better than the other methods in the case when *CL_Y* was set to two. However, in the case of *CL_Y* = 3 or *CL_Y* = 4, ANN outperformed DT and LR, and the differences between the misclassification errors increased as *CA* increased.

To access the effects of various parameters on misclassification errors in a more succinct manner, we applied analysis of variance (ANOVA) to the experimental data given in each table. The experimental setting for each approach can be regarded as a full factorial design (Montgomery, 2000). For example, the factors for the proposed approach included the prediction method (denoted by *M* with three levels of LR, ANN, and DT), sample size (denoted by *S*

Table 2
Simulated examples with three independent variables, including categorical variables.

ID	Original ID before conversion	No. of classes of dependent variable	No. of categorical variables	No. of classes of independent categorical variables	Sample size	Relationship	Description
S37	S1	2	1	2	100	$y = 1 + 3x_1 + 2x_2 + 2c_1 + \varepsilon$	x_3 is converted to categorical variable c_1. c_1 has a value of 'A' or 'B'. y has a value of 'A' or 'B'
S38	S2				500		
S39	S3				1000		
S40	S4				10,000		
S41	S1		1	3	100	$y = 1 + 3x_1 + 2x_2 + 2c_1 + \varepsilon$	x_3 is converted to categorical variable c_1. c_1 has a value of 'A', 'B' or 'C'. y has a value of 'A' or 'B'
S42	S2				500		
S43	S3				1000		
S44	S4				10,000		
S45	S1		2	2	100	$y = 1 + 3x_1 + 2c_1 + 2c_2 + \varepsilon$	x_2 and x_3 are converted to categorical variables c_1 and c_2, respectively. c_1 and c_2 have a value of 'A' or 'B'. y has a value of 'A' or 'B'
S46	S2				500		
S47	S3				1000		
S48	S4				10,000		
S49	S1		2	3	100	$y = 1 + 3x_1 + 2c_1 + 2c_2 + \varepsilon$	x_2 and x_3 are converted to categorical variables c_1 and c_2, respectively. c_1 and c_2 have a value of 'A', 'B' or 'C'. y has a value of 'A' or 'B'
S50	S2				500		
S51	S3				1000		
S52	S4				10,000		
S53	S13	3	1	2	100	$y = 1 + 3x_1 + 2x_2 + 2c_1 + \varepsilon$	x_3 is converted to categorical variable c_1. c_1 has a value of 'A' or 'B'. y has a value of 'A', 'B' or 'C'
S54	S14				500		
S55	S15				1000		
S56	S16				10,000		
S57	S13		1	3	100	$y = 1 + 3x_1 + 2x_2 + 2c_1 + \varepsilon$	x_3 is converted to categorical variable c_1. c_1 has a value of 'A', 'B' or 'C'. y has a value of 'A', 'B' or 'C'
S58	S14				500		
S59	S15				1000		
S60	S16				10,000		
S61	S13		2	2	100	$y = 1 + 3x_1 + 2c_1 + 2c_2 + \varepsilon$	x_2 and x_3 are converted to categorical variables c_1 and c_2, respectively. c_1 and c_2 have a value of 'A', 'B'. y has a value of 'A', 'B', or 'C'
S62	S14				500		
S63	S15				1000		
S64	S16				10,000		
S65	S13		2	3	100	$y = 1 + 3x_1 + 2c_1 + 2c_2 + \varepsilon$	x_2 and x_3 are converted to categorical variables c_1 and c_2, respectively. c_1 and c_2 have a value of 'A', 'B' or 'C'. y has a value of 'A', 'B' or 'C'
S66	S14				500		
S67	S15				1000		
S68	S16				10,000		
S69	S25	4	1	2	100	$y = 1 + 3x_1 + 2x_2 + 2c_1 + \varepsilon$	x_3 is converted to categorical variable c_1. c_1 has a value of 'A' or 'B'. y has a value of 'A', 'B', 'C' or 'D'
S70	S26				500		
S71	S27				1000		
S72	S28				10,000		
S73	S25		1	3	100	$y = 1+ 3x_1 + 2x_2 + 2c_1 + \varepsilon$	x_3 is converted to categorical variable c_1. c_1 has a value of 'A', 'B' or 'C'. y has a value of 'A', 'B', 'C' or 'D'
S74	S26				500		
S75	S27				1000		
S76	S28				10,000		
S77	S25		2	2	100	$y = 1 + 3x_1 + 2c_1 + 2c_2 + \varepsilon$	x_2 and x_3 are converted to categorical variables c_1 and c_2, respectively. c_1 and c_2 have a value of 'A' or 'B'. y has a value of 'A', 'B', 'C' or 'D'
S78	S26				500		
S79	S27				1000		
S80	S28				10,000		
S81	S25		2	3	100	$y = 1 + 3x_1 + 2c_1 + 2c_2 + \varepsilon$	x_2 and x_3 are converted to categorical variables c_1 and c_2, respectively. c_1 and c_2 have a value of 'A', 'B' or 'C'. y has a value of 'A', 'B', 'C' or 'D'
S82	S26				500		
S83	S27				1000		
S84	S28				10,000		

Table 3
Experimental results: misclassification errors for ANN, DT and LR in the case when independent variables consist of continuous variables.

CL_Y														
2					3					4				
V	S	Misclassification errors			V	S	Misclassification errors			V	S	Misclassification errors		
		ANN	DT	LR			ANN	DT	LR			ANN	DT	LR
3	100	0.2267	0.2333	*0.2000*	3	100	0.2600	0.3867	*0.2267*	3	100	*0.3467*	0.4467	0.3733
	500	0.2667	0.3000	*0.2000*		500	0.3667	0.4333	*0.3333*		500	*0.4000*	0.4667	0.4333
	1000	*0.2067*	0.2167	0.2167		1000	*0.3000*	0.3998	0.3267		1000	*0.4567*	0.4999	0.4700
	10,000	*0.2170*	0.2357	*0.2170*		10,000	*0.3287*	0.3933	0.3323		10,000	*0.4803*	0.5280	0.4899
5	100	0.2667	0.2777	*0.2000*	5	100	*0.2000*	0.4000	0.2333	5	100	*0.4000*	0.5000	*0.4000*
	500	0.2367	0.2333	*0.1800*		500	*0.3733*	0.4467	0.3800		500	*0.4867*	0.5600	0.5067
	1000	0.2467	0.2633	*0.2267*		1000	*0.3100*	0.3987	0.3433		1000	0.5133	0.5355	*0.4867*
	10,000	*0.2110*	0.2487	0.2117		10,000	0.3253	0.3910	*0.3220*		10,000	*0.4434*	0.5527	0.4657
7	100	0.2666	0.2555	*0.2333*	7	100	*0.1667*	0.2667	0.2000	7	100	*0.4333*	0.5000	*0.4333*
	500	0.2700	0.2300	*0.1667*		500	*0.3933*	0.4933	*0.3933*		500	0.5267	0.5867	*0.5200*
	1000	0.2067	0.3200	*0.2000*		1000	*0.3100*	0.3544	0.3300		1000	*0.4767*	0.5103	*0.4767*
	10,000	*0.2060*	0.2540	*0.2060*		10,000	*0.3373*	0.4493	0.3410		10,000	*0.4633*	0.5597	0.4900

Table 4
Experimental results: misclassification errors for ANN, DT and LR with three independent variables, including categorical variables ($V = 3$).

CL_Y																	
2						3						4					
CA	CL_X	S	Misclassification errors			CA	CL_X	S	Misclassification errors			CA	CL_X	S	Misclassification errors		
			ANN	DT	LR				ANN	DT	LR				ANN	DT	LR
1	2	100	*0.4000*	*0.4000*	0.4333	1	2	100	*0.5000*	0.7000	*0.5000*	1	2	100	0.6667	0.8000	*0.6000*
		500	0.2667	*0.2200*	0.2600			500	0.4200	*0.4000*	0.4667			500	0.5467	*0.5333*	0.6200
		1000	0.3767	*0.3533*	0.3567			1000	0.4400	*0.4300*	0.4400			1000	*0.5633*	0.5800	0.5933
		10,000	0.3423	0.3770	*0.3310*			10,000	0.4503	0.4513	*0.4487*			10,000	0.6137	0.6317	*0.6133*
	3	100	*0.4000*	0.4667	0.4333		3	100	0.5000	*0.4333*	0.4667		3	100	0.7000	0.7333	*0.6667*
		500	0.2733	*0.2467*	0.2600			500	0.5333	0.5333	*0.4800*			500	*0.5400*	0.6000	0.5667
		1000	*0.3500*	0.3767	0.3733			1000	*0.4067*	0.4900	0.4367			1000	0.5833	*0.5733*	0.5900
		10,000	0.3333	0.3383	*0.3303*			10,000	*0.4467*	0.4537	0.4500			10,000	*0.6117*	0.6277	0.6147
2	2	100	*0.4000*	0.5000	*0.4000*	2	2	100	0.5333	0.7000	*0.4667*	2	2	100	0.7333	0.8000	*0.6333*
		500	0.3400	*0.2867*	0.3267			500	0.4400	*0.4267*	0.4800			500	0.6200	*0.6000*	0.6400
		1000	*0.3567*	0.3767	0.3767			1000	0.4500	*0.4367*	0.4400			1000	0.5967	*0.5900*	0.6133
		10,000	*0.3543*	0.3937	0.3553			10,000	*0.4573*	0.5023	0.4590			10,000	*0.6247*	0.6360	0.6260
	3	100	*0.3663*	0.4667	0.4000		3	100	0.5667	*0.4000*	0.5000		3	100	0.7333	0.7333	*0.6667*
		500	*0.2000*	0.2533	0.2200			500	0.4600	0.5333	*0.4333*			500	*0.5333*	0.5800	0.5800
		1000	0.3600	*0.3333*	0.3500			1000	*0.4200*	0.4933	0.4467			1000	*0.5667*	0.5700	0.5800
		10,000	0.3443	*0.3436*	0.3443			10,000	*0.4553*	0.4607	0.4603			10,000	*0.6176*	0.6183	0.6223

Table 5
Experimental results: misclassification errors for ANN, DT and LR with five independent variables, including categorical variables ($V = 5$).

CL_Y																	
2						3						3					
CA	CL_X	S	Misclassification errors			CA	CL_X	S	Misclassification errors			CA	CL_X	S	Misclassification errors		
			ANN	DT	LR				ANN	DT	LR				ANN	DT	LR
1	2	100	0.3333	0.2333	*0.2000*	1	2	100	*0.2667*	0.4000	0.3667	1	2	100	0.5333	*0.4000*	0.5667
		500	0.3067	0.2800	*0.2733*			500	*0.4533*	0.5267	0.4933			500	*0.5267*	0.5733	0.5933
		1000	*0.2633*	0.3200	0.2867			1000	0.4567	*0.4467*	0.4933			1000	*0.5467*	0.6433	0.5833
		10,000	0.3240	0.3483	*0.3193*			10,000	*0.4356*	0.4580	0.4377			10,000	*0.5987*	0.6070	0.6370
	3	100	0.2667	*0.2000*	0.3000		3	100	0.4667	*0.3333*	*0.3333*		3	100	*0.3667*	0.4333	0.4667
		500	0.3333	*0.2800*	0.3067			500	*0.3933*	0.4467	0.4067			500	*0.5553*	0.6133	0.5733
		1000	0.3400	0.3367	*0.2833*			1000	*0.3871*	0.4194	0.4194			1000	*0.5677*	0.6000	0.5742
		10,000	*0.3217*	0.3413	0.3233			10,000	0.4679	0.5000	*0.4615*			10,000	0.6090	0.7115	*0.5833*
2	2	100	0.4000	0.4333	*0.3667*	2	2	100	*0.4000*	0.4333	*0.4000*	2	2	100	*0.5333*	0.6000	0.6667
		500	*0.3267*	0.3800	*0.3267*			500	0.5533	*0.5400*	0.5533			500	*0.6667*	0.6933	0.7067
		1000	*0.3167*	0.3300	0.3200			1000	*0.4933*	0.5200	0.5200			1000	*0.6167*	0.6400	0.6467
		10,000	*0.3527*	0.3753	*0.3527*			10,000	*0.4613*	0.4680	0.4643			10,000	0.6377	0.6467	*0.6343*
	3	100	*0.3000*	0.3333	0.3333		3	100	0.4000	0.4667	*0.3667*		3	100	*0.5333*	0.7000	0.6000
		500	*0.3467*	0.4000	0.3667			500	0.4733	0.4933	*0.4600*			500	0.6600	0.6867	*0.6467*
		1000	0.3400	0.4000	*0.3300*			1000	0.4452	0.4581	*0.4322*			1000	*0.5677*	0.6129	0.6065
		10,000	*0.3522*	0.3667	0.3560			10,000	0.4745	*0.4423*	0.4871			10,000	0.6667	0.7115	*0.6538*
3	2	100	0.3667	0.4333	*0.3000*	3	2	100	*0.3333*	0.5000	0.5000	3	2	100	*0.5667*	0.6333	0.6333
		500	*0.1400*	0.2914	0.3657			500	0.5733	*0.5400*	0.5533			500	*0.7000*	0.7067	0.7267
		1000	0.3333	0.3233	*0.3200*			1000	*0.4833*	0.5133	0.4967			1000	*0.6100*	0.6233	0.6133
		10,000	0.3627	0.3850	*0.3550*			10,000	*0.4667*	0.4883	0.4867			10,000	0.6350	0.6703	*0.6287*
	3	100	0.4333	*0.3333*	*0.3333*		3	100	*0.4000*	0.4667	0.4667		3	100	*0.4667*	0.7333	0.6667
		500	*0.3733*	0.4000	0.4333			500	*0.4800*	0.4933	0.4933			500	*0.6667*	0.7067	0.7000
		1000	0.3767	0.3933	*0.3600*			1000	*0.4194*	0.4645	0.4581			1000	*0.6000*	0.6258	0.6516
		10,000	0.3680	0.3753	*0.3627*			10,000	0.5128	*0.4679*	0.5000			10,000	*0.6603*	0.6667	0.6795
4	2	100	0.4667	*0.3000*	0.4333	4	2	100	*0.3667*	0.5667	0.4667	4	2	100	*0.5000*	0.5333	0.6000
		500	0.3267	0.3200	*0.3133*			500	*0.4867*	0.5400	0.5000			500	*0.5467*	0.7000	0.6133
		1000	0.3100	*0.2933*	0.3000			1000	*0.4667*	0.4933	0.4776			1000	*0.5663*	0.6433	0.5733
		10,000	0.3393	0.3543	*0.3353*			10,000	*0.4383*	0.4743	0.4423			10,000	0.6210	0.6417	*0.6183*
	3	100	0.3333	*0.2667*	0.3000		3	100	*0.4000*	*0.4000*	0.4333		3	100	0.5333	*0.4667*	0.6000
		500	0.3267	0.3133	*0.2867*			500	0.3867	*0.3600*	0.3933			500	*0.5600*	0.6133	0.5800
		1000	*0.3100*	0.3567	0.3167			1000	0.4194	*0.4129*	*0.4129*			1000	*0.5806*	0.6581	0.6065
		10,000	0.3473	0.3827	*0.3433*			10,000	0.4744	*0.4615*	0.4872			10,000	0.6410	*0.6282*	0.6603

with four levels of 100, 500, 1000, and 10,000), number of independent variables (denoted by V with three levels of one, three, and five), number of categorical variables (denoted by CA with two, four or six levels), number of classes of categorical dependent variables (denoted by CL_Y with two levels of two and three) and number of classes of categorical independent variables (denoted by CL_X with two levels of two and three).

Table 7 shows the ANOVA table for the continuous independent variables. The three-way interaction effect (i.e., $S \times V \times M$) was assumed to be negligible. Note that the p-values (or the least significant probabilities) for the main effects CL_Y, S, V, and M as well as for interaction effects $CL\_Y \times S$, $CL\_Y \times V$ and $CL\_Y \times M$ were 'small,' and therefore considered statistically significant (Montgomery, 2000). This is also illustrated in Figs. 1–4, which show

Table 6
Experimental results: misclassification errors for ANN, DT and LR with seven independent variables, including categorical variables ($V = 7$).

CL_Y																	
2						3						4					
CA	CL_X	S	Misclassification errors			CA	CL_X	S	Misclassification errors			CA	CL_X	S	Misclassification errors		
			ANN	DT	LR				ANN	DT	LR				DT	ANN	LR
1	2	100	*0.2667*	*0.2667*	*0.2667*	1	2	100	*0.3667*	0.4333	0.4333	1	2	100	0.5333	*0.5000*	0.5333
		500	0.2600	0.2667	*0.1787*			500	*0.3933*	0.4067	0.4267			500	*0.5133*	0.5667	0.5667
		1000	0.2600	0.3100	*0.2533*			1000	0.3700	0.4233	*0.3667*			1000	*0.5133*	0.5800	0.5333
		10,000	0.2243	0.2520	*0.2023*			10,000	*0.3537*	0.3893	0.3570			10,000	0.5103	0.5400	*0.4990*
	3	100	*0.2000*	0.2667	0.2667		3	100	0.3333	0.3000	*0.2667*		3	100	*0.4333*	0.6000	0.5000
		500	0.2333	0.2333	*0.1667*			500	*0.3867*	0.4400	0.4200			500	*0.5133*	0.5663	0.5663
		1000	0.2300	0.2800	*0.1723*			1000	*0.3600*	0.4133	0.3878			1000	0.5067	0.5567	*0.4967*
		10,000	*0.2150*	0.2503	0.2197			10,000	*0.3377*	0.4060	0.3400			10,000	*0.4720*	0.5433	0.4952
2	2	100	0.2333	0.3333	*0.2000*	2	2	100	*0.3667*	0.4333	0.4223	2	2	100	*0.4667*	0.5000	*0.4667*
		500	0.2667	*0.2467*	0.2667			500	*0.3567*	0.4200	0.4333			500	*0.5300*	0.5867	0.5667
		1000	0.2600	0.3067	*0.2500*			1000	*0.3667*	0.4167	0.4067			1000	*0.5167*	0.5500	0.5300
		10,000	*0.2330*	0.2447	0.2343			10,000	*0.3703*	0.3953	0.3940			10,000	0.5240	0.5500	*0.5227*
	3	100	*0.1667*	0.2667	0.1800		3	100	0.3333	*0.3000*	*0.3000*		3	100	*0.3333*	0.5000	0.4333
		500	0.1959	0.2800	*0.1767*			500	*0.3773*	0.4733	0.4400			500	*0.5067*	0.5400	0.5267
		1000	0.2200	0.2567	*0.2033*			1000	*0.3600*	0.4067	0.3867			1000	*0.4933*	0.5500	0.5500
		10,000	*0.2220*	0.2620	0.2223			10,000	*0.3410*	0.3440	0.3423			10,000	*0.4883*	0.5377	0.4913
3	2	100	*0.2333*	*0.2333*	0.2667	3	2	100	*0.3000*	0.4333	0.3667	3	2	100	*0.4333*	0.5667	0.4667
		500	0.2767	*0.2400*	0.2733			500	*0.3900*	0.4200	0.4333			500	*0.5267*	0.5800	0.5733
		1000	*0.2767*	0.2967	0.2833			1000	*0.3933*	0.4167	0.4267			1000	*0.5267*	0.5433	0.5600
		10,000	*0.2427*	0.2533	0.2450			10,000	*0.3600*	0.3880	0.3823			10,000	*0.5277*	0.5550	0.5317
	3	100	*0.1667*	0.2667	0.1800		3	100	0.2667	0.3667	*0.2333*		3	100	*0.4333*	0.5000	0.4667
		500	0.2467	0.3267	*0.1923*			500	*0.3933*	0.4467	0.4333			500	*0.5222*	0.5333	0.5533
		1000	0.2267	0.2467	*0.2033*			1000	*0.3633*	0.3933	0.3933			1000	*0.4967*	0.5367	0.5123
		10,000	*0.2250*	0.2587	0.2280			10,000	*0.3453*	0.4153	*0.3453*			10,000	*0.4722*	0.5430	0.5033
4	2	100	*0.2000*	0.2333	0.3000	4	2	100	*0.3667*	0.4333	0.4000	4	2	100	*0.4000*	0.5667	0.4667
		500	0.2600	*0.2400*	*0.2400*			500	*0.3600*	0.4200	0.4223			500	*0.5533*	0.5800	0.5733
		1000	*0.2767*	0.2967	0.2867			1000	*0.3768*	0.4232	0.4067			1000	0.5500	*0.5433*	0.5600
		10,000	0.2540	0.2741	*0.2490*			10,000	0.3803	0.3867	*0.3800*			10,000	*0.5340*	0.5487	0.5357
	3	100	*0.1667*	0.2667	0.2000		3	100	*0.2667*	0.3667	0.3000		3	100	*0.4000*	0.5000	0.5000
		500	0.2600	0.3200	*0.2133*			500	*0.3867*	0.4333	0.3933			500	*0.5067*	0.5333	0.5333
		1000	0.2700	0.2500	*0.2023*			1000	*0.3667*	0.4367	0.3833			1000	*0.4900*	0.5400	0.5300
		10,000	0.2313	0.2530	*0.2253*			10,000	0.3517	0.4133	*0.3503*			10,000	*0.4822*	0.5430	0.5010
5	2	100	*0.2000*	0.2667	0.3000	5	2	100	*0.3667*	0.4000	0.4000	5	2	100	*0.3667*	0.5333	0.5000
		500	0.2600	*0.2400*	0.2600			500	*0.3933*	0.4200	0.4200			500	0.5773	*0.5667*	0.5867
		1000	0.3100	*0.2767*	0.3033			1000	*0.3967*	0.4100	0.4100			1000	0.5500	*0.5433*	0.5633
		10,000	*0.2487*	0.2517	0.2540			10,000	*0.3817*	0.3887	0.3850			10,000	*0.5350*	0.5463	0.5357
	3	100	*0.2333*	0.2667	*0.2333*		3	100	*0.2667*	0.3667	0.3667		3	100	*0.3667*	0.5333	0.4667
		500	0.2333	0.3267	*0.2023*			500	*0.3733*	0.4267	0.4200			500	*0.5133*	0.5333	0.5667
		1000	0.2533	0.2500	*0.2133*			1000	*0.3800*	0.4467	0.4332			1000	*0.4833*	0.5400	0.5100
		10,000	0.2380	0.2690	*0.2303*			10,000	*0.3467*	0.4133	0.3533			10,000	*0.4885*	0.5427	0.5010
6	2	100	*0.1667*	0.3333	0.2333	6	2	100	*0.2333*	0.4000	0.4000	6	2	100	*0.3667*	0.5333	0.4333
		500	0.2933	*0.2400*	0.2667			500	*0.4067*	0.4200	0.4333			500	*0.5200*	0.5333	0.5867
		1000	0.2900	*0.2800*	0.2867			1000	*0.3867*	0.4100	0.4200			1000	*0.5200*	0.5433	0.5500
		10,000	*0.2480*	0.2627	0.2490			10,000	*0.3833*	0.3867	0.3856			10,000	0.5487	0.5477	*0.5360*
	3	100	*0.2333*	0.2667	0.2667		3	100	*0.2333*	0.3667	0.3894		3	100	*0.2667*	0.5333	0.5333
		500	0.3267	0.2333	*0.1933*			500	*0.3933*	0.4267	0.4567			500	*0.5000*	0.5333	0.5200
		1000	0.2670	0.2500	*0.2023*			1000	*0.3400*	0.4467	0.3733			1000	*0.4900*	0.5400	0.5233
		10,000	*0.2323*	0.2690	0.2347			10,000	0.3560	0.4120	*0.3523*			10,000	*0.4799*	0.5427	0.5053

Table 7
ANOVA table for continuous independent variables.

Source	Degrees of freedom	Sum of squares ($\times 10^2$)	Mean squares ($\times 10^2$)	F	p-Value
CL_Y	2	100.9676	54.8381	653.48	0.000
S	3	6.9303	2.3101	27.53	0.000
V	2	0.6409	0.3205	3.82	0.027
M	2	9.6747	4.8374	57.64	0.000
$CL\_Y \times S$	6	5.9622	0.9937	11.84	0.000
$CL\_Y \times V$	4	1.0030	0.2507	2.99	0.025
$CL\_Y \times M$	4	1.8765	0.4691	5.59	0.001
$S \times V$	6	0.7515	0.1253	1.49	0.194
$S \times M$	6	0.5836	0.0973	1.16	0.339
$V \times M$	4	0.0424	0.0106	0.13	0.972
Error	68	5.7064			
Total	107	142.8475			

the four main effects and interaction effects $CL\_Y \times S$, $CL\_Y \times V$ and $CL\_Y \times M$. Since all main effects were involved in the interaction effects, their effects on misclassification errors could not be assessed independently. That is, the effects of *CL_Y*, *S*, *V*, and *M* on misclassification errors must be assessed using their main and interaction plots together.

Figs. 1 and 2 show that performance of LR was better than of the other techniques when $CL\_Y = 2$, while ANN outperformed LR and DT when $CL\_Y = 3$ and $CL\_Y = 4$. In other words, a smaller *CL_Y* facilitated better performance by LR, while a lager *CL_Y* improved the performance of ANN (see Fig. 2).

Note in Figs. 3 and 4 that the performances for different *CL_Y* show different patterns depending on *V* and *S*. Nevertheless, the performance of misclassification errors consistently decreased as the number of classes of dependent variables increased.

The ANOVA results for three independent variables, including categorical variables, are summarized in Table 8. Here, the *p*-values indicate that three main effects *CL_Y*, *S* and *M* and interaction effects $M \times S$, $CL\_Y \times S$ and $CL\_X \times CA$ were statistically significant (see also Fig. 5 for the main effects and Figs. 6–8 for the interaction effects). Fig. 6 shows that LR performed best when $S = 100$, while ANN was superior when $S > 100$. In Fig. 7, it is shown that the performances for different *S* have different patterns depending on *CL_Y*. Nevertheless, the performances of misclassification errors increased as *CL_Y* decreased. In addition, the performances for different *CA* showed different patterns depending on *CL_X*; i.e., when $CL\_X = 3$, the performances of $CA = 2$ were better than of $CA = 1$, while, in the case of $CL\_X = 2$, the performances of $CA = 1$ were superior to those of $CA = 2$.

In Table 9, ANOVA results are shown for five independent variables, including categorical variables. The main effects *CL_Y*, *S*, *CA*, and *M* and the interaction effects $CL\_Y \times M$, $CL\_Y \times CL\_X$, $CL\_Y \times M$, $S \times CL\_X$ and $S \times CA$ were statistically significant. In Figs. 9 and 10, that performance of LR was better than of the other techniques, while ANN outperformed LR and DT when $CL\_Y = 3$ and $CL\_Y = 4$. The results are similar to that of experiments for continuous independent variables (see Fig. 2). Fig. 11 shows that the performances for different *CL_Y* have different patterns depending on *CL_X*. Nev-

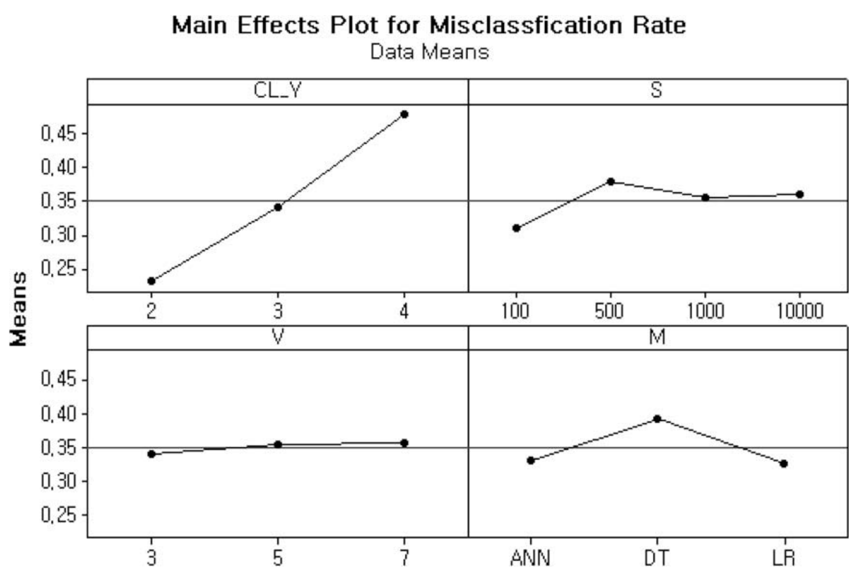

Fig. 1. Main effect plots for the continuous independent variables.

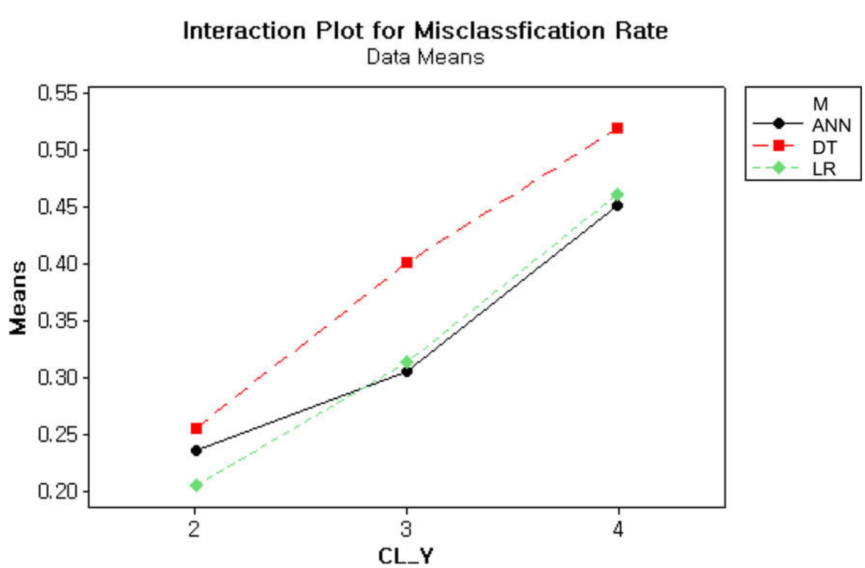

Fig. 2. Interaction plot of *CL_Y* and *M* for the continuous independent variables.

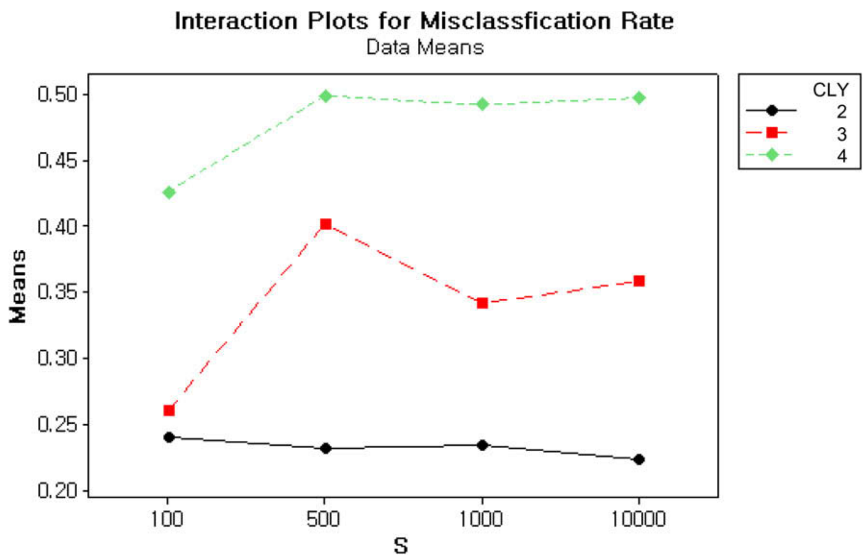

Fig. 3. Interaction plot of *CL_Y* and *S* for the continuous independent variables.

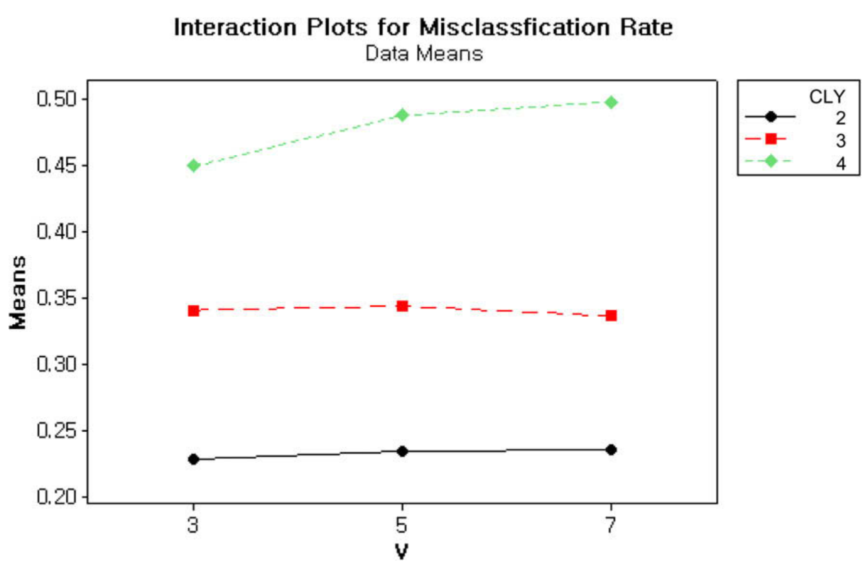

Fig. 4. Interaction plot of *CL_Y* and *V* for the continuous independent variables.

Table 8
ANOVA table for three independent variables, including categorical variables.

Source	Degrees of freedom	Sum of squares ($\times 10^2$)	Mean squares ($\times 10^2$)	F	p-Value
CL_Y	2	180.2919	90.1460	573.70	0.000
CL_X	1	0.5202	0.5202	3.31	0.072
S	3	25.4504	8.4825	53.99	0.000
CA	1	0.2838	0.2838	1.81	0.182
M	2	1.4185	0.7092	4.51	0.013
$CL\_Y \times CL\_X$	2	5.9622	0.9937	11.84	0.000
$CL\_Y \times S$	6	6.5938	1.0990	6.99	0.000
$CL\_Y \times CA$	2	0.0547	0.0274	0.17	0.840
$CL\_Y \times M$	4	0.1408	0.0352	0.22	0.924
$CL\_X \times S$	3	0.4022	0.1341	0.85	0.468
$CL\_X \times CA$	1	0.7699	0.7699	4.90	0.029
$CL\_X \times M$	2	0.1002	0.0501	0.32	0.728
$S \times CA$	3	0.0509	0.0170	0.11	0.955
$S \times M$	6	2.9897	0.4983	3.17	0.007
$CA \times M$	2	0.0483	0.0242	0.15	0.858
Error	103	16.1844			
Total	143	235.3619			

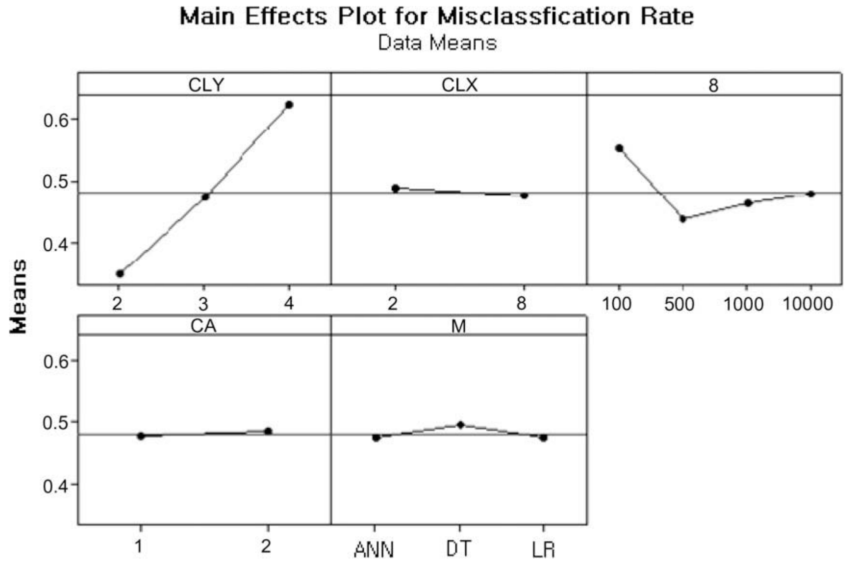

Fig. 5. Main effect plots for the three independent variables, including categorical variables.

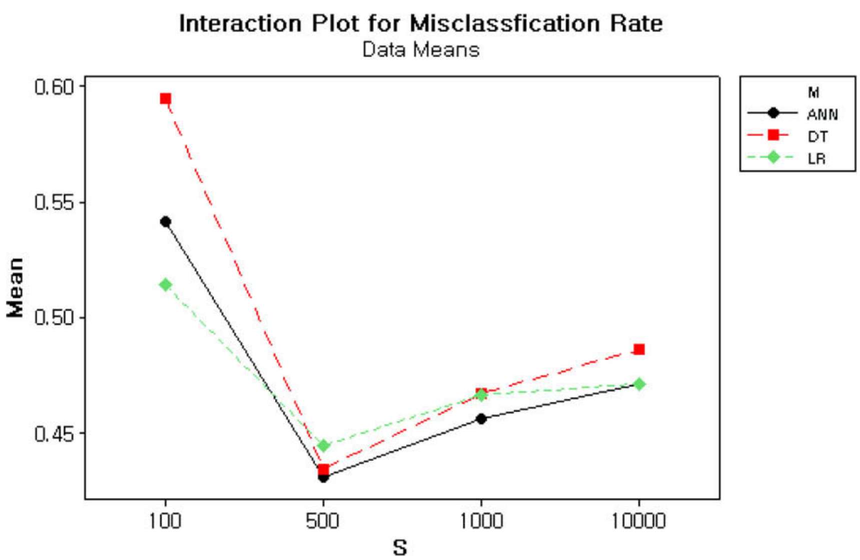

Fig. 6. Interaction plot of S and M for the three independent variables, including categorical variables.

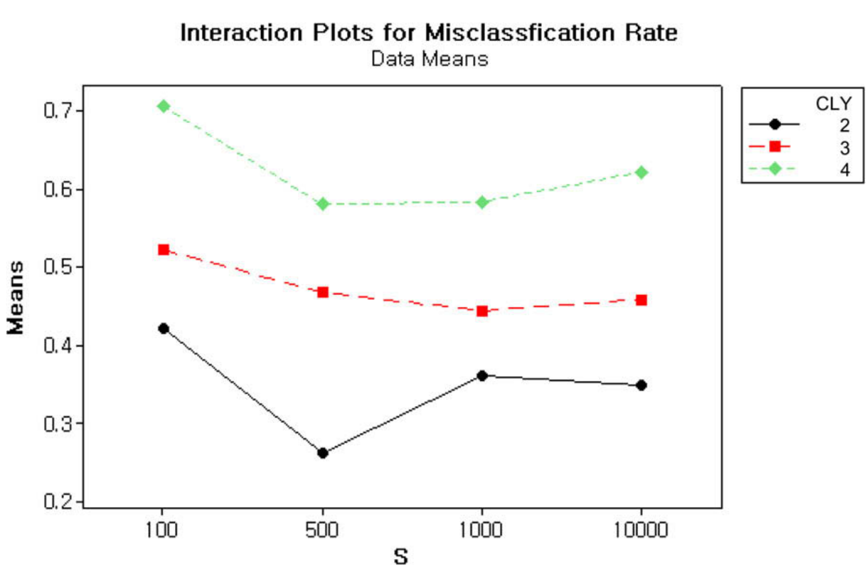

Fig. 7. Interaction plot of S and CL_Y for the three independent variables, including categorical variables.

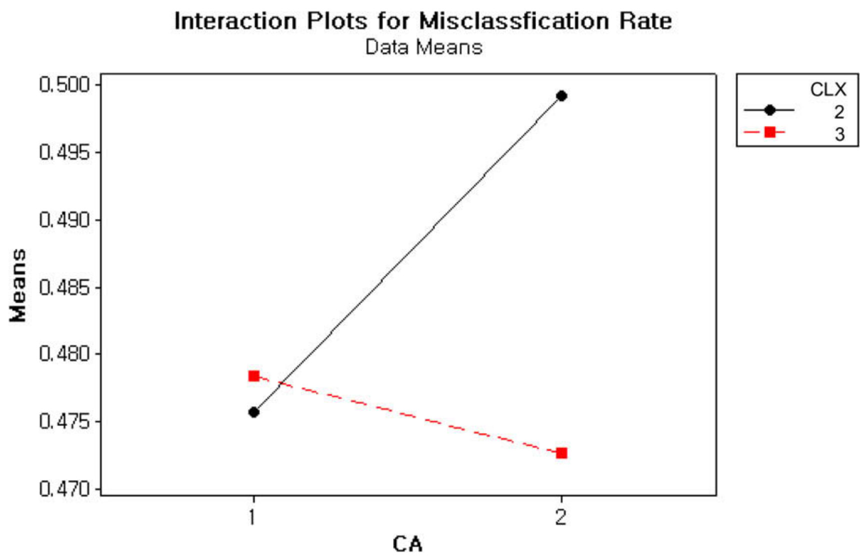

Fig. 8. Interaction plot of *CA* and *CL_X* for the three independent variables, including categorical variables.

Table 9
ANOVA table for five independent variables, including categorical variables.

Source	Degrees of freedom	Sum of squares ($\times 10^2$)	Mean squares ($\times 10^2$)	F	p-Value
CL_Y	2	368.4380	184.2190	1065.97	0.000
CL_X	1	0.5877	0.5877	3.40	0.066
S	3	12.6954	4.2318	24.49	0.000
CA	3	19.6515	6.5505	37.90	0.000
M	2	3.2776	1.6388	9.48	0.000
CL_Y × *M*	4	2.0980	0.5245	3.03	0.018
CL_Y × *CL_X*	2	2.2990	1.1495	6.65	0.000
CL_Y × *S*	6	6.5286	1.0881	6.30	0.000
*CL_* × *CA*	6	1.5399	0.2556	1.48	0.186
CL_X × *S*	3	1.5755	0.5252	3.04	0.030
*CL_* × *CA*	3	1.1457	0.3819	2.21	0.088
*CL_* × *M*	2	0.1786	0.0893	0.52	0.597
S × *CA*	9	6.7543	0.7505	4.34	0.000
S × *M*	6	0.7085	0.1181	0.68	0.663
CA × *M*	6	0.7797	0.1299	0.75	0.608
Error	229	39.5755			
Total	287	467.8276			

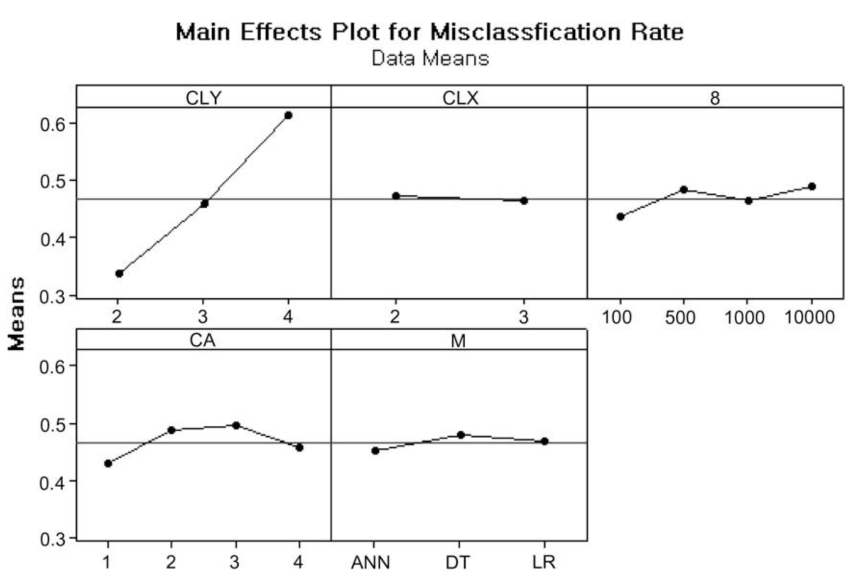

Fig. 9. Main effect plots for the five independent variables, including categorical variables.

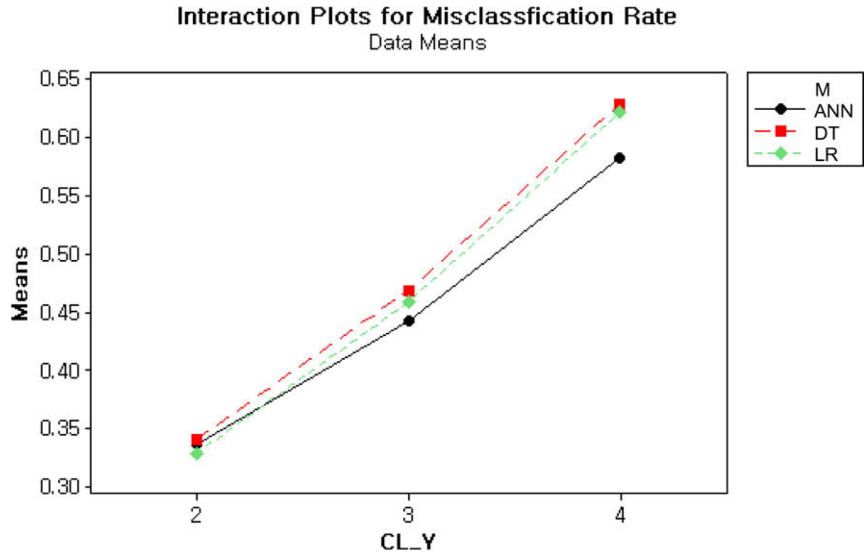

Fig. 10. Interaction plot of *CL_Y* and *M* for the five independent variables, including categorical variables.

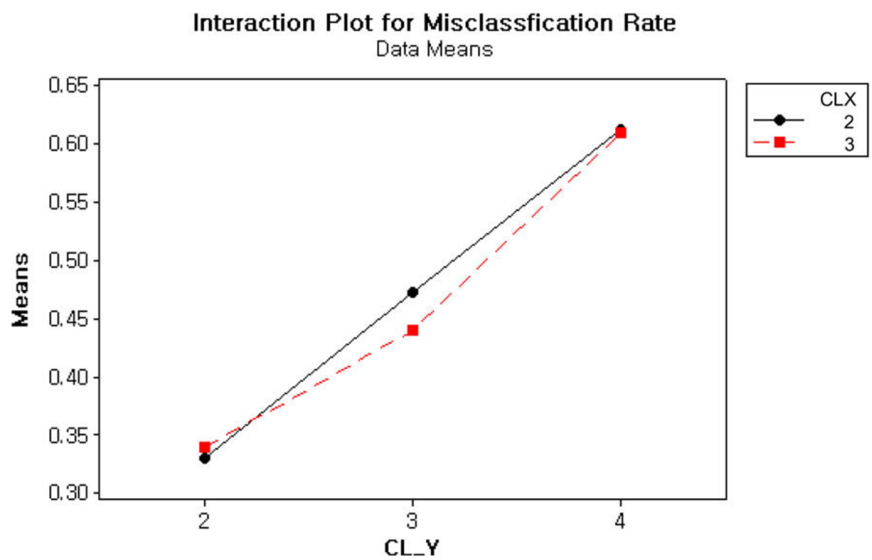

Fig. 11. Interaction plot of *CL_Y* and *CL_X* for the five independent variables, including categorical variables.

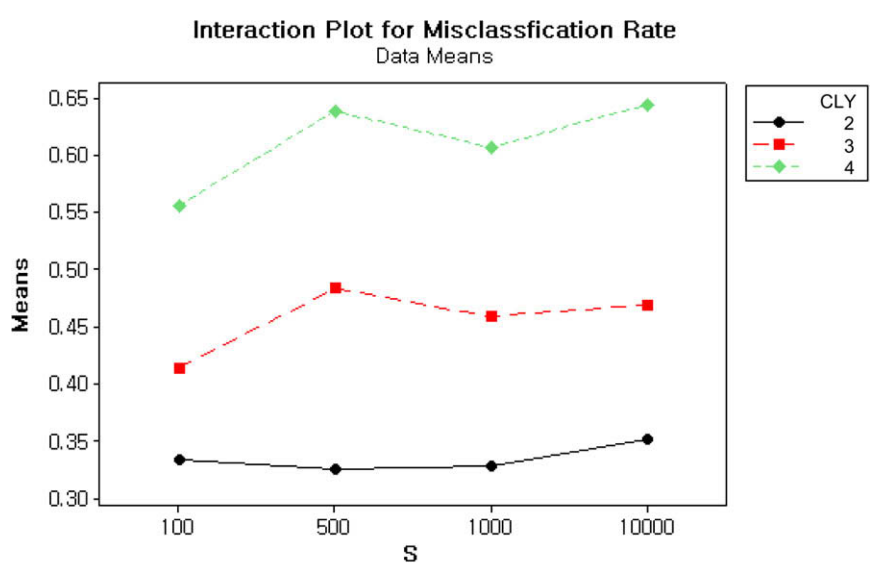

Fig. 12. Interaction plot of *CL_Y* and *S* for the five independent variables, including categorical variables.

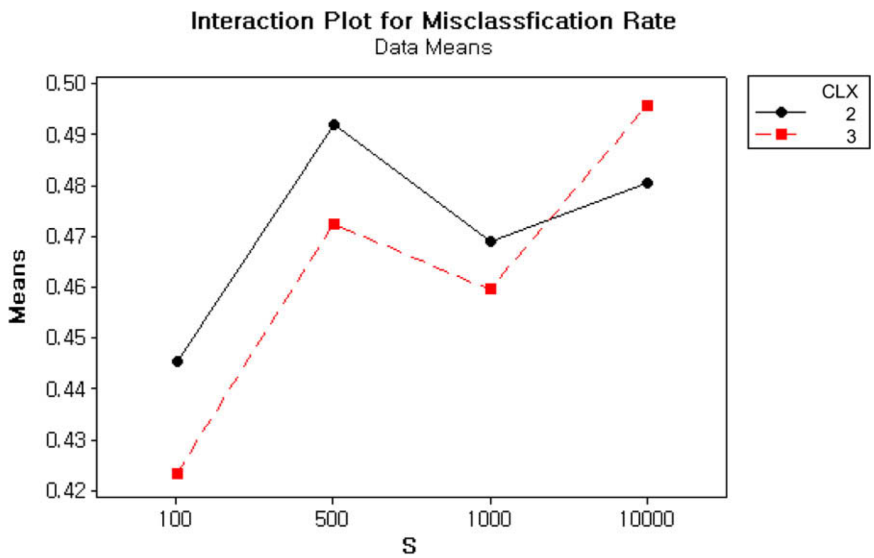

Fig. 13. Interaction plot of *CL_X* and *S* for the five independent variables, including categorical variables.

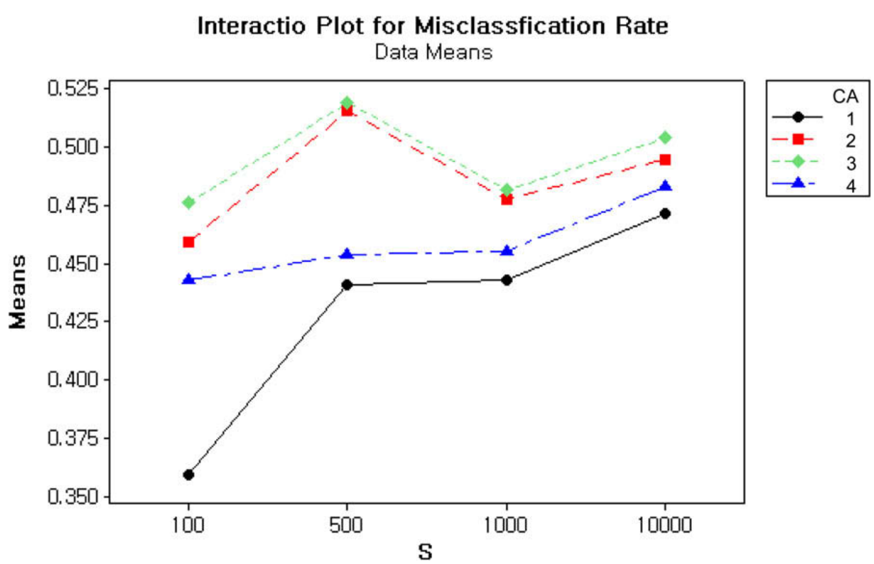

Fig. 14. Interaction plot of *CA* and *S* for the five independent variables, including categorical variables.

Table 10
ANOVA table for seven independent variables, including categorical variables.

Source	Degrees of freedom	Sum of squares ($\times 10^2$)	Mean squares ($\times 10^2$)	*F*	*p*-Value
CL_Y	2	527.2152	263.6076	3395.68	0.000
CL_X	1	6.1863	6.1863	79.69	0.000
S	3	15.9868	5.3289	68.64	0.000
CA	5	0.4330	0.0886	1.12	0.352
M	2	15.0311	7.5155	96.81	0.000
$CL\_Y \times CL\_X$	2	0.0175	0.0088	0.11	0.893
$CL\_Y \times M$	4	2.8143	0.7036	9.06	0.000
$CL\_Y \times S$	6	5.3866	0.8978	11.56	0.000
$CL\_Y \times CA$	10	0.8371	0.0837	1.08	0.378
$CL\_X \times S$	3	1.1571	0.3857	4.97	0.002
$CL\_X \times CA$	5	0.2703	0.0541	0.70	0.627
$CL\_X \times M$	2	1.5215	0.7607	9.80	0.000
$S \times CA$	15	1.9543	0.1303	1.68	0.053
$S \times M$	6	5.6047	0.9341	12.03	0.000
$CA \times M$	10	0.7066	0.0707	0.91	0.524
Error	355	27.5588			
Total	431	612.6812			

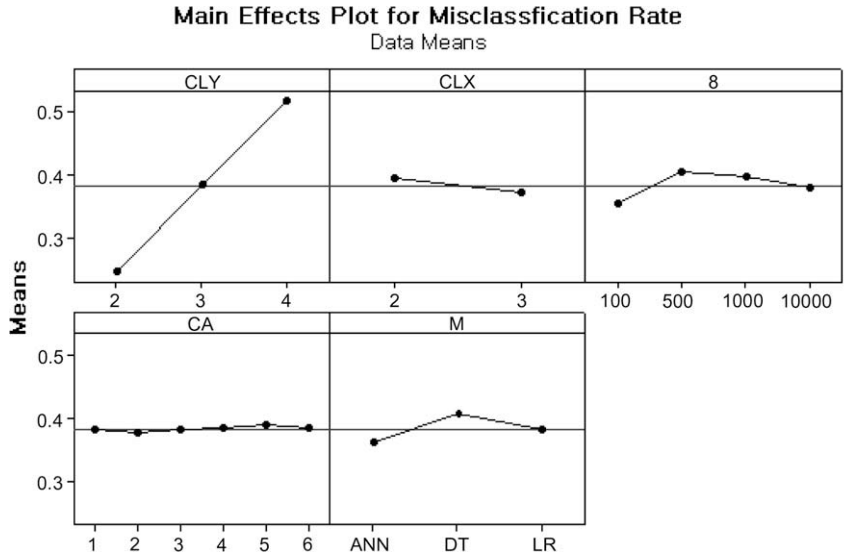

Fig. 15. Main effect plots for the five independent variables, including categorical variables.

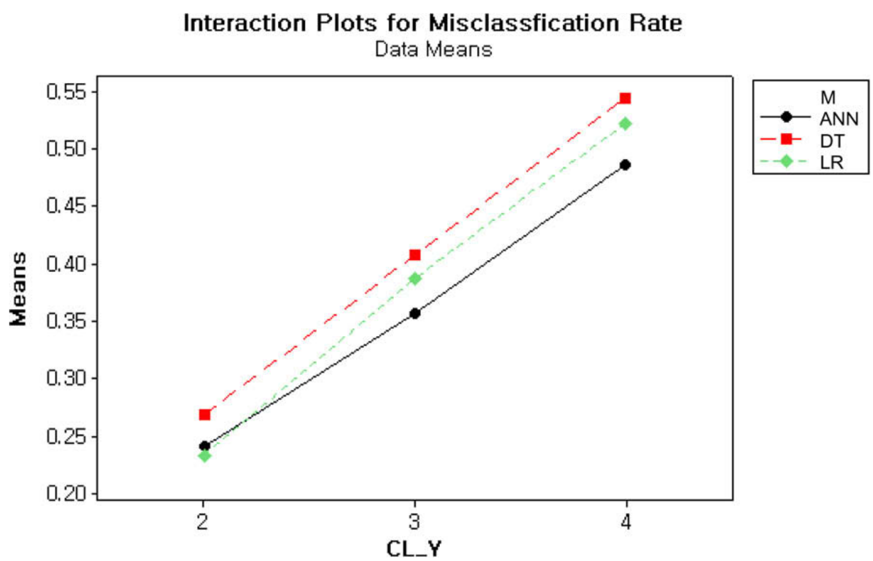

Fig. 16. Interaction plot of *CL_Y* and *M* for the seven independent variables, including categorical variables.

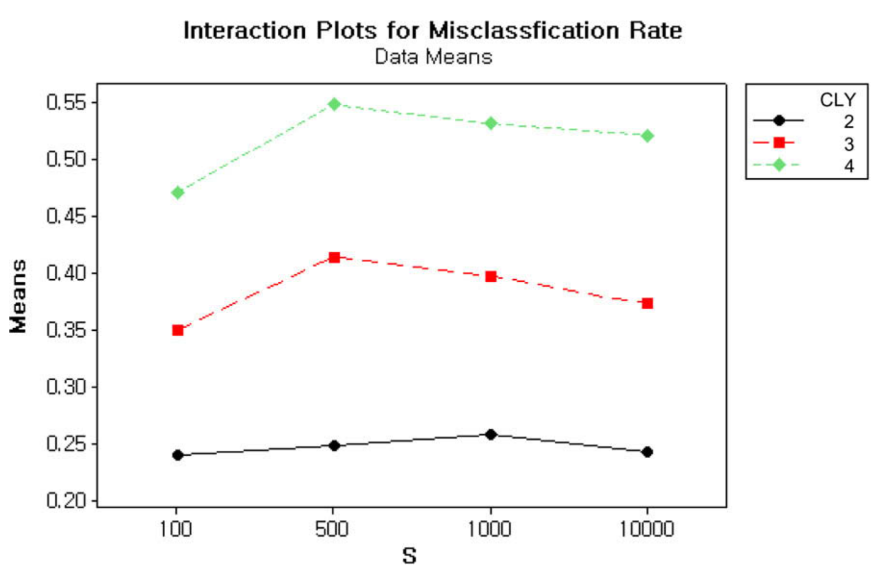

Fig. 17. Interaction plot of *CL_Y* and *S* for the seven independent variables, including categorical variables.

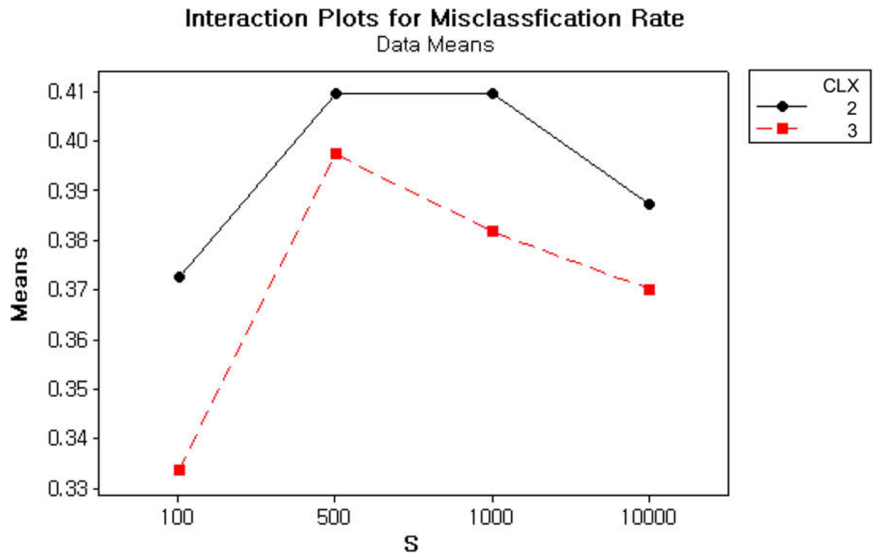

Fig. 18. Interaction plot of *CL_X* and *S* for the seven independent variables, including categorical variables.

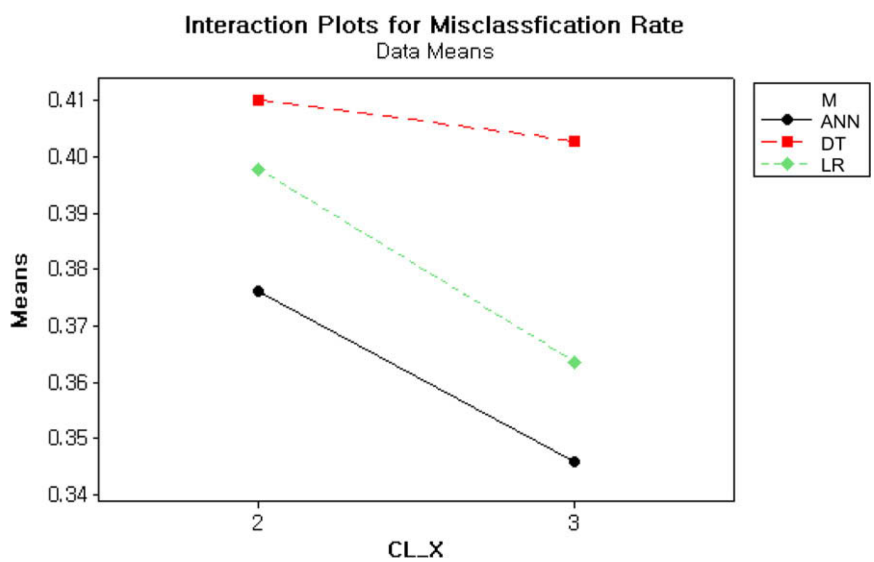

Fig. 19. Interaction plot of *CL_X* and *M* for the seven independent variables, including categorical variables.

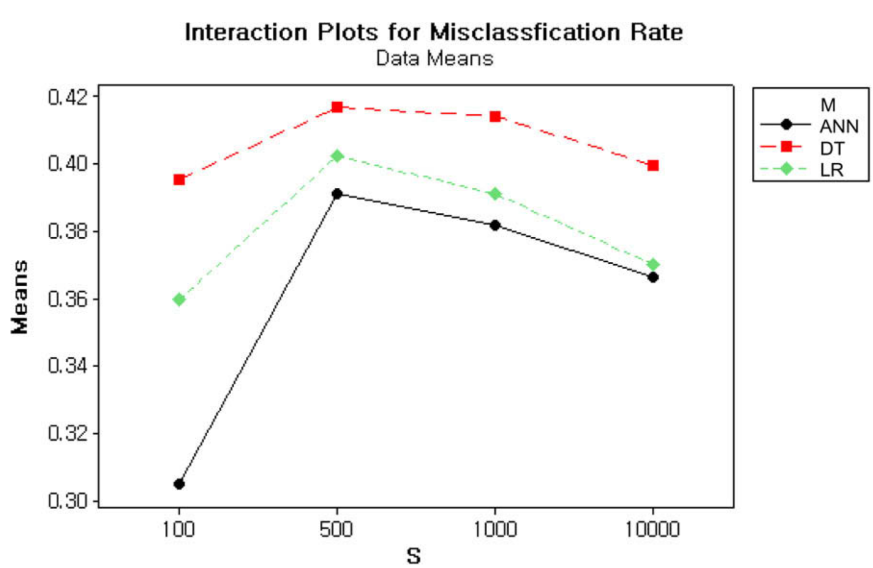

Fig. 20. Interaction plot of *S* and *M* for the seven independent variables, including categorical variables.

ertheless, the performances decreased as CL_Y increased. From Figs. 12–14, note that the prediction accuracy decreased as S increased (except for $S = 500$), and that $CL\_X = 3$ was superior to $CL\_X = 2$ (except for $S = 10{,}000$). That is, a bigger CL_X and S facilitated better performance.

Table 10 shows ANOVA results for seven independent variables, including categorical variables. The main effects CL_Y, CL_X, S and M and the interaction effects $CL\_Y \times M$, $CL\_Y \times S$, $S \times CL\_X$, $CL\_X \times M$ and $S \times M$ were statistically significant. From Figs. 15, 19 and 20, we have known that ANN outperformed LR and DT regardless of CL_X and S. However, LR performed best when CL_Y was small ($CL\_Y = 2$), while ANN was superior when CL_Y was large ($CL\_Y = 3$ and $CL\_Y = 4$) (refer to Fig. 16). This is similar to the results shown in Figs. 2 and 10, where a smaller value of CL_Y led to a better performance by LR and a larger value facilitated a better performance by ANN. Note in Figs. 17 and 18 that the performances for different S different patterns depending on CL_Y and CL_X, respectively. Nevertheless, regardless of S, the prediction accuracy consistently increased as CL_Y decreased and CL_X increased.

4. Conclusions

In this article, we present the results of an experimental comparison study of classification techniques based on varying the number of independent variables, the types of independent variables, the number of classes of the independent variables, the number of classes of the dependent variable, and the sample size. To evaluate the performance of the different techniques, we generated various simulated problems and used the misclassification errors.

The main results include the following: when independent variables are continuous, LR performs best in the case when the number of classes of dependent variable is small (i.e., $CL\_Y = 2$), while ANN is superior to DT and LR in the case when the number of classes of dependent variable is three or more; when independent variables are continuous and categorical, LR performs better than DT and ANN in the case of small number of independent variables (i.e., $V = 3$) and small sample size ($S = 100$), while ANN was best in other cases; and ANN performance improves more relative to LR and DT performance as the number of independent variables and the number of classes of dependent variables increases. That is, in the case of complex characteristics of the condition, ANN was superior to the other methods.

The above results were derived from simulated data and need further verification using a variety of actual data. However, the results are meaningful in that this study provides the first comparison for classification techniques between statistical and data mining techniques based on the characteristics of the independent and dependent variables. In addition, the results of this study provide insight for selecting the most appropriate classification method for a problem based on characteristics of the problem's independent and dependent variables. A promising area of future research would be in applying this approach to compare the performance of other methods such as support vector machine, case-based reasoning, and so on.

References

Cao, Q., Leggio, K. B., & Schniederjans, M. J. (2005). A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market. *Computers and Operation Research, 32*(10), 2499–2512.

Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research, 95*(1), 24–37.

Gorr, W. L., Nagin, D., & Szczypula, J. (1994). Comparative study of artificial neural network and statistical models for predicting student grade point averages. *International Journal of Forecasting, 10*, 17–34.

Hardgrave, B. C., Wilson, R. L., & Walstrom, K. A. (1994). Predicting graduate student success: A comparison of neural networks and traditional techniques. *Computers and Operation Research, 21*, 249–263.

Kim, G. H., An, S. H., & Kang, K. I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment, 39*(10), 1235–1242.

Kim, Y. S. (2008). Comparison of the decision tree, artificial neural network, and linear regression methods based on the number and types of independent variables and sample size. *Expert Systems with Applications, 34*(2), 1227–1234.

Kumar, U. A. (2005). Comparison of neural networks and regression analysis: A new insight. *Expert Systems with Applications, 29*, 424–430.

Montgomery, D. C. (2000). *Design and analysis of experiments*. New York: Wiley.

Patuwo, W., Hu, M. Y., & Hung, M. S. (1993). Two-group classification using neural networks. *Decision Sciences, 24*, 824–845.

Raaymakers, W. H. M., & Weijters, A. J. M. M. (2003). Makespan estimation in batch process industries: a comparison between regression analysis and neural networks. *European Journal of Operational Research, 145*(1), 14–30.

Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures. *Decision Sciences, 23*, 899–916.

Sarle, W. S. (1994). Neural network implementation in SAS software. In *Proceedings of the nineteenth annual SAS Users Group international conference* (pp.1551–1573). Cary, NC.

Shuhui, L., Wunsch, D. D., Hair, E. O., & Giesselmann, M. G. (2001). Comparative analysis of regression and artificial neural network models for wind turbine power curve estimation. *Journal of Solar Energy Engineering, 123*, 327–332.

Subbanarasimha, P. N., Arinze, B., & Anadarajan, M. (2000). The predictive accuracy of artificial neural networks and multiple regression in the case of skewed data. Exploration of some issues. *Expert Systems with Applications, 19*, 117–123.

Tam, K. Y. (1991). Neural network models and the prediction of bankruptcy. *Omega, 19*, 429–445.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science, 38*, 926–947.

Ting, H. N., Yunus, J., & Salleh, H. (2002). Speaker-independent phonation recognition for Malay Plosives using neural networks. In *International joint conference on neural networks* (pp. 619–623). Honolulu, HI.

Wang, Y.-M., & Elhag, T. M. S. (2007). A comparison of neural network, evidential reasoning and multiple regression analysis in modeling bridge risks. *Expert Systems with Applications, 32*, 336–348.

Yeh, J. C. H., Hamey, L. G. C., & Westcott, T. (1998). Developing FENN applications using cross-validated validation training. In *Proceedings of the second IEEE international conference on intelligent processing systems* (pp. 565–569). Gold Coast.