

# Data depth based support vector machines for predicting corporate bankruptcy

Sungdo Kim<sup>1</sup> · Byeong Min Mun<sup>1</sup> · Suk Joo Bae<sup>1</sup> 

Published online: 3 August 2017  
© Springer Science+Business Media, LLC 2017

**Abstract** In financial distress analysis, the diagnosis of firms at risk for bankruptcy is crucial in preparing to hedge against any financial damage the at-risk firms stand to inflict. Some pre-alarm signals that indicate a potential financial crisis exist when a firm faces a default risk. Early studies on corporate bankruptcy prediction include parametric and nonparametric approaches, such as artificial intelligence (AI), for detecting pre-alarm signals. Among nonparametric techniques, the methods involving support vector machine (SVM) have shown potential in predicting corporate bankruptcy. We propose a hybrid method that combines data depths and nonlinear SVM for the prediction of corporate bankruptcy. We employed data depth functions to condense multivariate financial data with nonlinear and non-normal characteristics into one-dimensional space. The SVM method was introduced to classify the data points on a depth versus depth plot (*DD*-plot). Based on data set that records failed and non-failed manufacturing firms in Korea over 10 years, the empirical results demonstrated that the proposed method offers a higher level of accuracy in corporate bankruptcy prediction than existing methods. The proposed method is expected to provide a guidance in corporate investing for investors or other interested parties.

**Keywords** Artificial neural network · Bankruptcy prediction · Classification model · Data depth · *DD*-plot · Support vector machine

## 1 Introduction

External uncertainties that pose a threat to corporate survival have increased drastically due to the protean environment of the global economy. In the face of such uncertainties, financial institutions are endeavoring to prepare sophisticated countermeasures against delays or defaults on liability fulfillment on the part of firms. The prediction of default risks in firms by companies that are liable to pay is the aim of financial institutions in order to minimize their capital exposure to risks and to mitigate their own default risks. The prediction of bankruptcy has been extensively studied in finance and management literature for the last two decades (Lee et al. [17]; Salcedo-Sanz et al. [32]; Min and Lee [26]; Li and Sun [18]; Tsai [35], to name a few). Bankruptcy prediction has become even more important since the Basel Committee on Banking Supervision (Basel II) established borrowers' risk rating as a key criterion for minimum capital requirements of banks. In general, when a firm faces an inevitable default on liability fulfillment, there may exist some symptoms or pre-alarm signals indicating a financial crisis for the firm itself.

Early studies of bankruptcy prediction exploited parametric statistical techniques such as multiple discriminant analysis (MDA) [2], logit model [29] and probit model [41]. Strict assumptions for the statistical approaches (e.g., linearity, normality, and pre-existing functional forms relating criterion variables to predictor variables), however, have limited the applications of these techniques in the area of finance. To overcome this obstacle, nonparametric

---

✉ Suk Joo Bae  
sjbae@hanyang.ac.kr

Sungdo Kim  
sddream@hanyang.ac.kr

Byeong Min Mun  
stat@psm.hanyang.ac.kr

<sup>1</sup> Department of Industrial Engineering,  
Hanyang University, Seoul, Korea

techniques such as artificial intelligence (AI) have been employed since the late 1980s for the prediction of corporate bankruptcy or financial distress. The AI techniques include decision trees (DTs), artificial neural networks (ANNs), genetic algorithms (GAs), and back-propagation networks (BPNs). Odom and Sharda [28] first introduced an ANN model to predict corporate bankruptcy. Tam and Kiang [34] compared the performance of a neural network model with the performances of linear discriminant models, logit models, the iterative dichotomiser 3 (ID3) algorithm, and the  $k$ -nearest neighbor approach based on insolvent data of commercial banks. They showed that the ANN provides the most accurate and robust prediction results among the various methods. While a number of application studies have reported the outstanding performance of ANNs, the models have difficulty in clearly explaining prediction results due to the lack of explanatory power and capacity for generalization because of overfitting problems. Additionally, ANNs require significant time and efforts to construct a best architecture through multiple layers [16, 33].

As an alternative to ANNs, a method based on support vector machine (SVM) has recently attracted special attention in the area of financial distress modeling. The SVM has reported better classification results than parametric statistical methods or other nonparametric techniques such as ANNs and BPNs. Moreover, SVM can overcome the overfitting problem via the concept of structural risk minimization. Häardle et al. [13] first introduced the SVM to corporate bankruptcy prediction, comparing its performance with ANN and MDA methods, as well as a learning vector quantization proposed by Fan and Palaniswami [10]. By mapping input variables onto a high-dimensional feature space, Min and Lee [26] showed that SVM transforms complex problems of corporate bankruptcy prediction into simpler ones, to which linear discriminant functions can subsequently be applied. Häardle et al. [14] explored the suitability of smooth SVMs in predicting corporate default risk. They investigated how key factors such as selection of appropriate accounting ratios, length of training period, and structure of training samples influence prediction precision for corporate bankruptcy. While SVM achieves an excellent classification accuracy, a main disadvantage of the method is difficulty in interpreting the modeling results.

In the last decade, a number of researchers have actively developed hybrid approaches to predict corporate bankruptcy. Hybrid approaches combine several classification methods to secure greater accuracy than individual (parametric or nonparametric) models. Min et al. [27] and Ahn et al. [1] employed a genetic algorithm to design an SVM-based technique for corporate bankruptcy prediction. The selection of both SVM hyper-parameters and input features was integrated into one learning process in their genetic algorithm. Van Gestel et al. [38] applied

the Bayesian evidence framework [24, 37] to find hyper-parameters for the least squares SVM.

In this article, we propose a hybrid method based on data depth (DD) and SVM to improve the accuracy of bankruptcy prediction for Korean firms. As a nonparametric multivariate technique, DD estimates a representative value from multivariate data which may possess nonlinear and non-normal characteristics. The hybrid method (referred to hereafter as DD-SVM) calculates the DD for annual financial ratios because the ratios are generally unlikely to follow a multivariate normal distribution. Additionally, the method applies nonlinear SVM to the DD plot, which presents the depth values of the combined sample of both failed and non-failed firms to classify a binary output variable. The performance of DD-SVM is compared with other parametric methods or AI methods, including ANNs, in terms of bankruptcy prediction accuracy.

The remainder of this paper is organized as follows. Basic modeling ideas for predicting corporate bankruptcy are presented in Section 2. The ideas are based mainly on the introduction of nonlinear SVM to DD plots to classify failed or non-failed firms. The research data and pre-analytical results are given in Section 3. Section 4 presents an empirical analysis of corporate bankruptcy prediction to demonstrate the performance of the proposed method, along with a comparison of the proposed method with other competing prediction models. Section 5 concludes this study and discusses directions for future research.

## 2 The modeling of bankruptcy prediction

### 2.1 Data depth functions

The word “depth” was first used by Tukey [36] to depict high-dimensional data, and the far reaching ramifications of depth in ordering and analyzing multivariate data have been elaborated in the works of Liu [21], Donoho and Gasko [8], Liu et al. [22] and others. Data depth characterizes the centrality of high-dimensional data with respect to a distribution or a multivariate sample. Viewed as a method of dimension reduction, DD does not rely on link functions, kernel functions, or other refined mappings, unlike related methods such as principal components.

In order to form a general definition of a depth function, Zuo and Serfling [42] defined a statistical depth function as a bounded, non-negative mapping that satisfies four desirable properties: (1) affine invariance; (2) maximality at center; (3) monotonicity relative to the deepest point; and (4) vanishing at infinity. Affine invariance means that the relative depth of a point should not depend on the underlying coordinate system or the scales of the underlying measurements. For a distribution with a uniquely defined center,

maximality at center indicates that the depth function should attain the maximum at this center. Monotonicity relative to the deepest point means that, as a point moves from the center outward, the corresponding depth should decrease monotonically. Vanishing at infinity means that the depth of a point approaches zero when its norm approaches infinity. Among a number of DD functions possessing the above properties, Mahalanobis depth, simplicial depth and Tukey’s depth functions are the most popular.

### 2.1.1 Mahalanobis depth

Mahalanobis [25] introduced a distance function based on Hotelling’s  $T^2$  statistic (now called “Mahalanobis depth”). Serving as the first DD concept, the Mahalanobis depth measures how deep a point  $\mathbf{x} \in \mathbb{R}^p$  is with respect to a given distribution  $G$ . The Mahalanobis depth function is given by

$$\mathcal{MD}(G; \mathbf{x}) = \frac{1}{1 + (\mathbf{x} - \boldsymbol{\mu}_G)^T \boldsymbol{\Sigma}_G^{-1} (\mathbf{x} - \boldsymbol{\mu}_G)}, \tag{1}$$

where  $\boldsymbol{\mu}_G$  and  $\boldsymbol{\Sigma}_G$  denote the mean vector and the covariance of the reference distribution  $G$ , respectively. In general, because  $G$  is unknown, the sample version of the Mahalanobis depth is obtained by replacing  $\boldsymbol{\mu}_G$  and  $\boldsymbol{\Sigma}_G$  with their sample estimates  $\bar{\mathbf{x}}_n$  and  $\mathbf{S}_n$ , respectively, for multivariate data set  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ .

### 2.1.2 Simplicial depth

Liu [21] introduced the concept of “simplicial depth”, which is determined by counting simplices derived from  $n$  data points. For a reference distribution  $G$  on  $\mathbb{R}^p$ , the simplicial depth of a data point  $\mathbf{x}$  with respect to  $G$  is defined by

$$\mathcal{SD}(G; \mathbf{x}) = P_G\{\mathbf{x} \in S[\mathbf{x}_1, \dots, \mathbf{x}_{p+1}]\}, \tag{2}$$

where  $\mathbf{x}_1, \dots, \mathbf{x}_{p+1}$  are independent observations from  $G$  and  $S[\mathbf{x}_1, \dots, \mathbf{x}_{p+1}]$  is the simplex with vertices  $\mathbf{x}_1, \dots, \mathbf{x}_{p+1}$ . In other words,  $S[\mathbf{x}_1, \dots, \mathbf{x}_{p+1}]$  is the set of all points in  $\mathbb{R}^p$  that are convex combinations of  $\{\mathbf{x}_1, \dots, \mathbf{x}_{p+1}\}$ . The sample version of  $\mathcal{SD}$  is obtained by replacing  $G$  in  $\mathcal{SD}(G; \mathbf{x})$  with  $G_n$ , or alternatively, by computing the fraction of the sample random simplices containing the point  $\mathbf{x}$  as

$$\mathcal{SD}(G_n; \mathbf{x}) = \binom{n}{p+1}^{-1} \sum_{1 \leq i_1 < \dots < i_{p+1} \leq n} I_{(\mathbf{x} \in S[\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p+1}}])}, \tag{3}$$

where  $I_{(\cdot)}$  is the indicator function. Liu [21] showed that  $\mathcal{SD}(G; \mathbf{x})$  is affine invariant, and that if  $G$  is absolutely continuous, then  $\mathcal{SD}(G_n; \mathbf{x})$  converges uniformly and strongly to  $\mathcal{SD}(G; \mathbf{x})$  as  $n \rightarrow \infty$ . we can confirm that  $\mathbf{x}$  is contained

in the simplex  $S[\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p+1}}]$  if  $\mathbf{x}$  can be expressed as a convex combination of  $\{\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p+1}}\}$ .

### 2.1.3 Tukey depth

Tukey [36] proposed a half-space depth which is now commonly called “Tukey depth”. The half-space depth is the smallest proportion of data points contained on one side of any hyperplane passing through a data point  $\mathbf{x}$ , including points lying on the hyperplane. That is, the Tukey depth is defined as

$$\mathcal{TD}(G; \mathbf{x}) = \inf_H \{P_G(H)\} \\ : H \text{ is a closed half-space in } \mathbb{R}^p \text{ for } \mathbf{x} \in H$$

with respect to the reference distribution  $G$ .

In bivariate data, for instance, the Tukey depth calculates the smallest proportion of data points contained on one side of any line ( $L$ ) passing through  $\mathbf{x}$ , including points lying on the line itself. Following a method proposed by Rousseeuw and Ruts [30], for example, two-dimensional Tukey depth requires a vector connecting a fixed  $\mathbf{x}$  to each member of the reference sample  $\mathbf{x}_1, \dots, \mathbf{x}_n$  and then measures the angles of these vectors with the positive  $x$ -axis. Instead of counting the minimum number of points lying on one side of the line passing through  $\mathbf{x}$  and a reference sample, we can count the minimum number of angles between the angle of  $L$  and its opposite angle. With that, the empirical formula for the Tukey depth of  $\mathbf{x}$  is

$$\mathcal{TD}(G_n; \mathbf{x}) = \frac{1}{n} \min_i \{\min(k_i, n - k_i)\},$$

where  $k_i = \psi_1(i) - \psi_2(i)$  for  $\psi_1(i) = \#\{j : 0 \leq \theta_j < \theta_i + \pi\}$ , and  $\psi_2(i) = \#\{j : 0 \leq \theta_j < \theta_i\}$ . Here,  $\theta_i$  is the angle of  $u_i = (\mathbf{x}_i - \mathbf{x}) / \|\mathbf{x}_i - \mathbf{x}\|$  for  $i = 1, \dots, n$ . We can assume  $0 = \theta_1 \leq \dots \leq \theta_n < 2\pi$  and  $\theta_{n+1} = \theta_1 + 2\pi, \theta_{n+2} = \theta_2 + 2\pi$ , and so on. See Bae et al. [3] for the details on calculation of the Tukey depth, along with simple examples.

In addition to these three popular data depths, there are several other DD metrics, e.g., “convex hull peeling depth” by Barnett [4], “likelihood depth” by Fraiman and Meloche [11], “regression depth” by Rousseeuw and Jubert [31], and “Lens depth” by Liu and Moddares [23], to name a few.

## 2.2 Depth vs. depth plot

The depth vs. depth plot ( $DD$ -plot), which was first introduced by Liu et al. [22], is a useful analytical tool for graphical comparisons of two multivariate distributions or samples based on DD. For any two given multivariate samples, the  $DD$ -plot represents the depth values of the combined sample under the two corresponding empirical distributions. The tool transforms the two multivariate samples in any dimension to a simple two-dimensional scatter

plot. Li et al. [19] addressed some advantages of the *DD*-plot in classification problems, including that the best separating curve in the *DD*-plot is determined automatically by the underlying probabilistic geometry of the data, and that classification outcomes can be easily visualized in two-dimensional *DD*-plots. This is much simpler than tracking classification outcomes in the original sample space of high-dimensional multivariate data. In particular, the *DD*-plot is robust against outliers and extreme values.

Let  $F$  and  $G$  be two distributions on  $\mathbb{R}^p$ , and let  $\mathcal{D}$  be an affine-invariant depth. For two random samples drawn from  $F$  and  $G$ ,  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} (\equiv \mathbf{X})$  and  $\{\mathbf{y}_1, \dots, \mathbf{y}_m\} (\equiv \mathbf{Y})$ , respectively, the *DD*-plot is defined as

$$DD(F, G) = \{(\mathcal{D}(F; \mathbf{x}), \mathcal{D}(G; \mathbf{x})) \text{ for all } \mathbf{x} \in \mathbb{R}^p\}, \tag{4}$$

when  $F$  and  $G$  are known. If both  $F$  and  $G$  are unknown, then the sample version of *DD*-plot is given by

$$DD(F_n, G_m) = \{(\mathcal{D}(F_n; \mathbf{x}), \mathcal{D}(G_m; \mathbf{x})), \mathbf{x} \in \{\mathbf{X} \cup \mathbf{Y}\}\}. \tag{5}$$

Note that both  $DD(F, G)$  and  $DD(F_n, G_m)$  are always subsets of  $\mathbb{R}^2$ , no matter the size of the dimension  $p$  of the data. If the two given distributions are identical (that is,  $F \equiv G$ ), then the resulting  $DD(F, G)$  is simply a line segment on the  $45^\circ$  line in the *DD*-plot. Deviation patterns from this straight line indicate a specific type of difference between the two underlying distributions.

Figure 1 illustrates the *DD*-plot for simulated multivariate data. Figure 1a shows the *DD*-plot for two samples ( $n = m = 500$ ) drawn from the standard bivariate normal distribution. It can be observed that the data are scattered around the  $45^\circ$  line in the plot. Figure 1b presents the *DD*-plot for two samples ( $n = m = 500$ ) with one drawn from the standard bivariate normal distribution and the other drawn from the bivariate normal with mean  $(2, 0)^T$ . All of the *DD*-plots are constructed using the Mahalanobis depth. The *DD*-plot shows quite clearly that the observations from two different distributions scatter around the  $45^\circ$  line in a manner that

is almost symmetric. The  $45^\circ$  line can thus be used as the separating line for two different samples in the *DD*-plot. Its corresponding classification rule is simple in that we assign  $\mathbf{x}$  to  $F$  if  $\mathcal{D}(F_n; \mathbf{x}) > \mathcal{D}(G_m; \mathbf{x})$ , and we assign  $\mathbf{x}$  to  $G$  otherwise. Note that the classifier in the *DD*-plot is conceptually the same as the maximum depth classifier in the work of Ghosh and Chaudhuri [12].

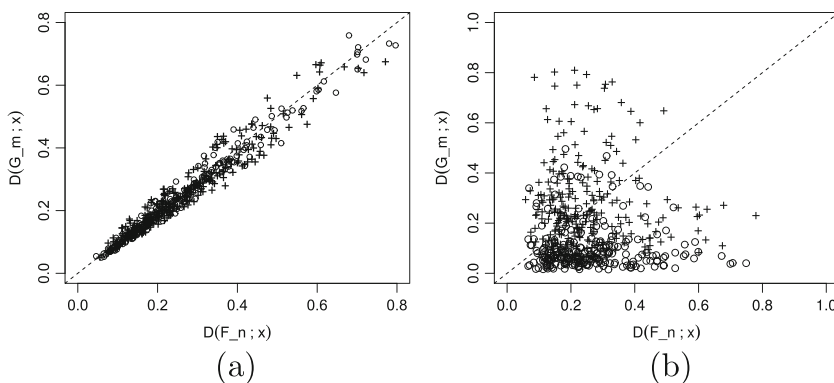
In general, however, high-dimensional observations do not scatter symmetrically along a  $45^\circ$  line because they have different dispersion structures as well as different locations. Hence, the linear classifier does not perform well. In this study, we introduce a nonlinear support vector machine (SVM) to classify failed firms and non-failed firms based on the *DD*-plot.

### 2.3 The nonlinear support vector machine

The prediction of bankruptcy can be formulated as a two-class classification problem. We apply the SVM approach to bankruptcy prediction using real-life data for Korean manufacturing companies, and compare the empirical results of the SVM-based method with the results from other prediction models.

A support vector machine, which was introduced from statistical learning theory by Vapnik [39], is a powerful classification method that provides better solutions to decision boundaries than can be obtained via traditional neural networks. SVM uses a linear model to construct nonlinear class boundaries through nonlinear mapping of input vectors into a high-dimensional feature space. In general, the linear model in the new space represents a nonlinear decision boundary in the original space. The SVM approach implements the principle of structural risk minimization, which aims to reduce the bounds of misclassification errors by constructing an optimal separating hyperplane in a high-dimensional feature space using quadratic programming to find unique solutions. The application areas of SVMs include text categorization, digital image identification, handwriting recognition, function approximation and regression, and time series forecasting.

**Fig. 1** *DD*-plots for two random samples drawn from (a) an identical distribution and (b) two different distributions (The *DD*-plot is re-produced from the example by Li et al. [19])



Assuming that there are two-dimensional  $DD$ -predictors  $(\mathcal{D}(F_n; \mathbf{x}), \mathcal{D}(G_m; \mathbf{x})) \equiv \mathbf{y}$  based on  $p$  financial ratios in our bankruptcy classifier, the (bivariate) data-depth data of predictor variables for the  $i$ th firm in the  $DD$ -plot can be represented by the vector  $\mathbf{y}_i$ . The financial status of the  $i$ th firm is denoted by  $z_i \in \{+1, -1\}$ , where  $+1$  represents a non-failed firm and  $-1$  represents a failed firm. Given a training set  $D = \{\mathbf{y}_i, z_i\}_{i=1}^N$  for constructing two parallel bounding hyperplanes at opposite sides of a separating hyperplane  $\mathbf{w}^T \mathbf{y} + b = 0$ , where  $\mathbf{w}$  is the weight vector and  $b$  is the bias, the decision rule of SVM, which finds the optimal hyperplane separating the binary decision classes, is given for the linearly separable case as

$$z(\mathbf{y}) = \text{sign} \left( \sum_{i=1}^N z_i (\mathbf{y}_i^T \mathbf{y}) + b \right). \tag{6}$$

For the nonlinearly separable case, (6) is modified as follows:

$$z(\mathbf{y}) = \text{sign} \left( \sum_{i=1}^N z_i \mathcal{K}(\mathbf{y}_i, \mathbf{y}) + b \right), \tag{7}$$

where  $\mathcal{K}(\mathbf{y}_i, \mathbf{y})$  is defined as the kernel function which performs nonlinear mapping between input space and a (high-dimensional) feature space. Popular kernel functions to construct the decision rules include the radial basis function (RBF) kernel  $\mathcal{K}(\mathbf{y}_i, \mathbf{y}) = \exp(-\sigma \|\mathbf{y} - \mathbf{y}_i\|^2)$ , where  $\sigma$  is a tuning parameter; the linear kernel  $\mathcal{K}(\mathbf{y}_i, \mathbf{y}) = \mathbf{y}_i^T \mathbf{y}$ ; the polynomial kernel  $\mathcal{K}(\mathbf{y}_i, \mathbf{y}) = (\gamma + \mathbf{y}_i^T \mathbf{y})^d$  with degree  $d$  and a tuning parameter  $\gamma (\geq 0)$ ; and the multilayer perceptron (MLP) kernel  $\mathcal{K}(\mathbf{y}_i, \mathbf{y}) = \tanh(\gamma_1 + \gamma_2 \mathbf{y}_i^T \mathbf{y})$ . Note that the MLP kernel is not positive semi-definite for all choices of the tuning parameter  $\gamma_1$  and  $\gamma_2$ .

Most of classification problems are, however, linearly (or nonlinearly) non-separable, thus it is generally not possible to find a hyperplane that can differentiate between “failed” and “non-failed” examples without mistakes. By allowing a certain level of misclassification, the soft margin method introduces slack variables measuring the degree of misclassification of  $\mathbf{y}_i$ , and a cost constant  $C$  which weights the importance of classification errors vis-à-vis the margin width. The solution of the primal problem to find the optimal separating hyperplane is generally obtained after constructing the Lagrangian. The dual problem of finding an optimal separating hyperplane, along with the kernel function  $\mathcal{K}(\mathbf{y}_i, \mathbf{y})$ , is re-written as the following quadratic programming problem:

$$\begin{aligned} \arg \min_{\alpha} \quad & Q(\alpha) = \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j z_i z_j \mathcal{K}(\mathbf{y}_i, \mathbf{y}_j) - \sum_{i=1}^N \alpha_i, \\ \text{s.t.} \quad & \sum_{i=1}^N \alpha_i z_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N, \end{aligned}$$

for the constant  $C$  with respect to the Lagrange multipliers  $\alpha \equiv (\alpha_1, \dots, \alpha_N)^T$ . Under the dual formulation, the two-class classification problem via implementation of the optimal separating hyperplane in the feature space is determined by the nonlinear SVM classifier as

$$z(\mathbf{y}) = \text{sign} \left( \sum_{i=1}^N \alpha_i z_i \mathcal{K}(\mathbf{y}_i, \mathbf{y}) + b \right). \tag{8}$$

Note that data instances corresponding to non-zero  $\alpha_i$ 's are called *support vectors*. If the bias term  $b$  is implicitly a part of the kernel function, as in the case of the RBF kernel, then (8) is reduced to

$$z(\mathbf{y}) = \text{sign} \left( \sum_{j=1}^{\text{number of SVs}} \alpha_j z_j \mathcal{K}(\mathbf{y}, \mathbf{y}_j) \right).$$

Among the big issues in SVM, the selection of appropriate values of parameters plays an important role in building a bankruptcy prediction model with high prediction accuracy and stability. The choice of the parameter  $C$  is very important. A large value of  $C$  permits the optimization to find a large margin, whereas a small value of  $C$ , which allows a small margin, results in a large number of support vectors. If the value of  $C$  is chosen to be unnecessarily large, the margin becomes narrow and the constructed classification model may fail to classify new objects properly, although the training set is separated well. Banz et al. [6] used a value of  $C = 5$ , however, there are no general rules to guarantee the best parameter values for application problems. Lin [20] provided a systematic method for selecting parameter values for SVMs by adapting the sampling theory concept into a Gaussian filter. In corporate bankruptcy prediction, Min and Lee [26] proposed a grid-search technique using five-fold cross-validation to determine optimal parameter values of  $C$  and  $\sigma$  in the RBF kernel function of the SVM. Min et al. [27] and Wu et al. [40] used genetic algorithm to optimize the values of the parameters  $C$  and  $\sigma$  with the RBF kernel in SVM. Van Gestel et al. [38] introduced Bayes' formula to the inference of the RBF kernel parameter  $\sigma$  in the least squares SVM.

### 3 Research data

The sample of firms for analysis herein consists of manufacturing firms in the Korea Composite Stock Price Index 2000 (KOSPI 2000) from 2000 to 2013. The manufacturing firms of interest were extracted from the Korea Listed Companies Association (KLCA). The KLCA also provides financial information related to the firms through audit reports from external auditors. This research studies 144 firms that have filed bankruptcy petitions in Korea in the 21st century with the main objective of taking into account corporate

management risks, and excludes cases of irresistible liability failures from structural risks such as abrupt economic crises (e.g., the bailout from the International Monetary Fund (IMF) in 1997). Originally, corporate failure has been defined as bankruptcy, default on bonds, overdrawing of a bank account, or non-payment of a preferred stock dividend [5]. For simplicity, however, we restricted the definition of corporate failure in this study to only corporate bankruptcy. For the same period, we selected 144 ‘non-failed’ firms randomly from all solvent firms. A failed firm was paired with a non-failed firm in a similar industry, which deals with similar products, with similar capitalization, and with similar values of assets. The one-to-one matching ratio is a potential cause of oversampling problems [41]. However, in order to highlight the effects of key financial ratios on the likelihood of corporate bankruptcy, a matched sample of non-failed firms was selected. We only selected medium-sized and large-sized firms with property amounts of at least \$10 billion. The failed and non-failed data were arbitrarily split into two subsets: about 80% of the data was used for a training set and 20% for a validation set for *k*-fold cross validation. The training data was used to build the bankruptcy prediction model using the data depth and SVM. The prediction model was verified by the validation data which was not utilized to construct the model.

The KLCA reported 111 financial ratios representing profitability, stability, activity, and productivity with respect to individual firms. Out of them, we selected 53 significant ratios using two-sample *t*-test between failed firms and non-failed firms. The selected financial ratios, which are summarized in Table 1, were used to build a prediction model for classifying failed or non-failed firms. We analyzed the financial data on individual firms for 10 years preceding bankruptcy (or survival up to 2013) to examine the existence of chronological trends in corporate bankruptcy.

### 3.1 Multivariate normality test

Existing parametric distress models have been constructed under the assumption of (multivariate) normality. However, empirical results have shown that, in practice, most financial ratios violate the normality assumption, thereby justifying the introduction of nonparametric techniques. First, we assessed the normality of financial ratios using the Henze-Zirkler’s test statistic [15]. The Henze-Zirkler test statistic for multivariate normality is given by

$$T_{HZ} = \frac{1}{n} \sum_{i,j=1}^n e^{-\frac{\xi^2}{2} A_{ij}} - 2(1 + \xi^2)^{-p/2} \sum_{j=1}^n e^{-\frac{\xi^2}{2(1+\xi^2)} A_j} + n(1 + 2\xi^2)^{-p/2},$$

where *p* is the number of variables,  $\xi = \sqrt{2}^{-1} (n(2p + 1)/4)^{-(p+4)}$ ,  $A_i = (\mathbf{x}_i - \bar{\mathbf{x}})^T S^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$  is the

**Table 1** Financial ratios included in the bankruptcy prediction model

Category	Financial ratios	
Profitability	Cost of sales to net sales	
	Operating ratio	
	Non-operating income ratio	
	Financial expenses to sales	
	Ratio of expenditure to revenue	
	Financial expenses to total expenses	
	Accumulated depreciation ratio	
	ratio of interest burden	
	Interest paid ratio	
	Financial expense to total borrowings	
	Accumulated earning to stockholder’s equity	
	Additional paid in capital and retained earnings to stockholdersequity	
	Reserve ratio	
	R & D investment ratio	
	Stability	Ratio of current assets to gross assets
		Inventories to current assets
		Current assets to non-current assets
Non-current assets to total assets		
Ratio of net worth to total capital		
Ratio of net worth to borrowed capital		
Bank for international settlements capital ratio		
Non-current assets to stockholder’s equity & non-current liabilities		
Cash ratio		
Raw materials to total manufacturing costs		
Ratio of selling credit to inventories		
Receivables to merchandise & finished goods		
short-term borrowings to total borrowings		
Non-current liabilities ratio		
Net working capital to total assets		
Total borrowing & bonds payable to total assets		
Total borrowing & bonds payable to total liabilities		
Reserves to total assets		
Reserves to capital stock		
Current assets to net sales		
Operating capital to net sales		
Cash flow to liabilities		
Cash flow to total assets		
Cash flow to net sales		
Financial expenses to net income to operating income		
Activity	Net worth turnover ratio	
	Capital stock turnover	
	Turn-over ratio of borrowing capital	
	trade payable turnover period	
	Inventories to net sales	
Goods in process turnover ratio		

**Table 1** (continued)

Category	Financial ratios
Productivity	Account Receivable Collection Period
	Net working capital turnover
	Value added per employee
	Machinery & equipment per employee
	liabilities & stockholder’s equity per employee
	Ratio of value added to liabilities & stockholder’s equity
	Ratio of value added to tangible fixed assets
	Ratio of value added to net sales

Mahalanobis distance of the  $i$ th observation to the centroid,  $A_{ij} = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j)$  is the Mahalanobis distance between the  $i$ th and  $j$ th observations, and  $\xi$  is a smoothing parameter. If the data are (multivariate) normally distributed, then the test statistic is approximately log-normally distributed, with mean and variance respectively corresponding to

$$E[T_{HZ}] = 1 - (1 + 2\xi^2)^{-p/2} \times \left\{ 1 + \frac{p\xi^2}{(1 + 2\xi^2)} + \frac{p(p + 2)\xi^4}{2(1 + 2\xi^2)^2} \right\}, \quad \text{and}$$

$$\text{Var}[T_{HZ}] = 2(1 + 4\xi^2)^{-p/2} + 2(1 + 2\xi^2)^{-p} \times \left\{ 1 + \frac{2p\xi^4}{2(1 + 2\xi^2)^2} + \frac{3p(p + 2)\xi^8}{4(1 + 2\xi^2)^4} \right\} - 4\omega^{-p/2} \left\{ 1 + \frac{3p\xi^4}{2\omega} + \frac{p(p + 2)\xi^8}{2\omega^2} \right\},$$

where  $\omega = (1 + \xi^2)(1 + 3\xi^2)$ . Using the log-normal distribution parameters, the Wald test statistic can be applied to test the significance of multivariate normality. The Henze-Zirkler test was implemented by an *MVN* package in R software for assessing multivariate normality. The results of multivariate normality testing of corporate financial ratios for 10 years are summarized in Table 2. The periods represent the years prior to bankruptcy for failed firms and survival years for non-failed firms. All of the  $p$ -values derived from the Henze-Zirkler test were very close to zero, thus leading us to conclude that the data set did not satisfy the multivariate normality assumption.

**Table 2** Multivariate normality tests for financial ratios of the sampling firms in Korea

Period	1	2	3	4	5	6	7	8	9	10
Non-failed firms	1.002	1.001	1.001	1.002	1.001	1.001	1.001	1.004	1.002	1.011
Failed firms	1.023	1.055	1.031	1.101	1.011	1.013	1.009	1.095	1.072	1.072

## 4 Analytical results

To predict the bankruptcy of manufacturing firms in Korea, this study was conducted according to the following steps:

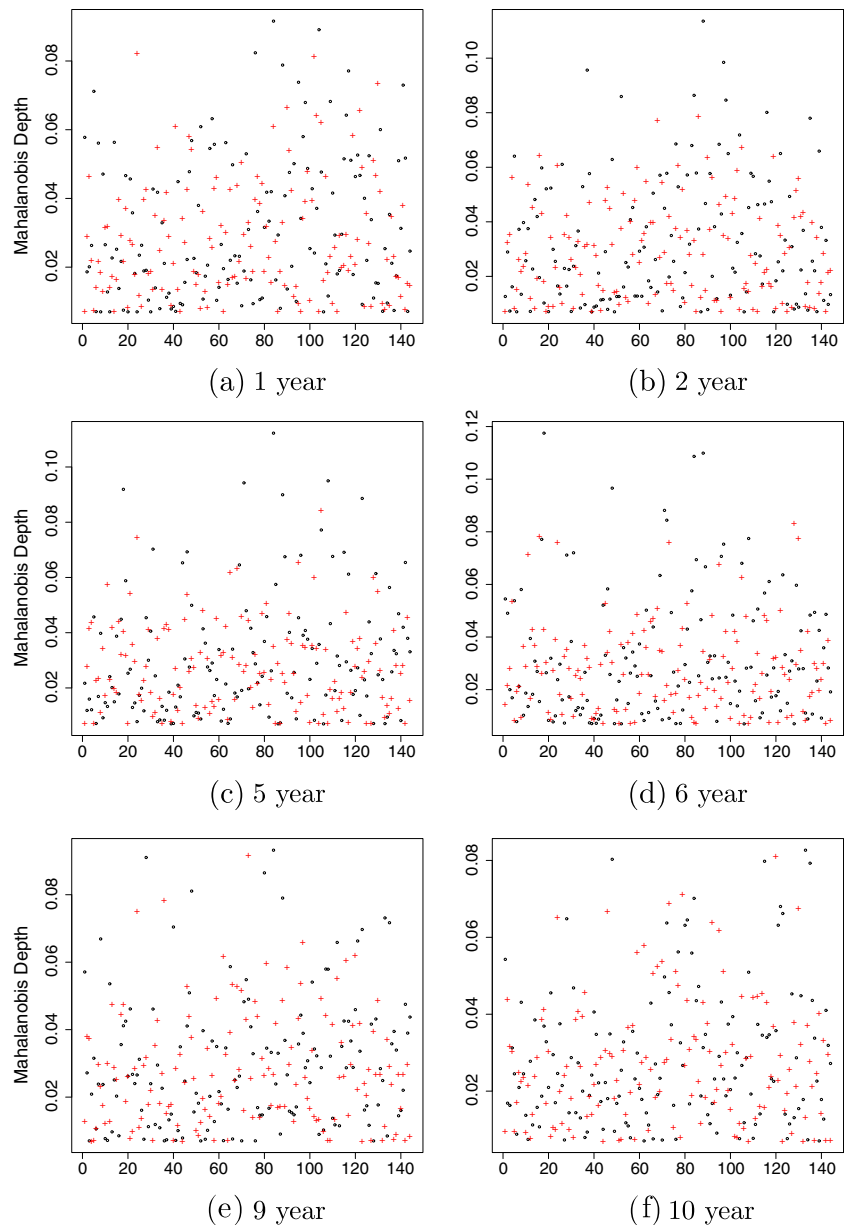
- 1) Reduction of the number of multi-dimensional financial ratios by DD.
- 2) Plotting of the values of DD into a *DD*-plot.
- 3) Classification of the DD in the *DD*-plot using nonlinear SVM.

We compared the performance of the proposed method with other bankruptcy prediction models in the existing literature. The hit ratio of classification was used as an indicator to evaluate the predictive accuracy of each models. In addition, Type I error (defined as the probability that a firm predicted not to fail will, in fact, fail) and Type II error (defined as the probability that a firm predicted to fail will not, in fact, fail) were also included in the evaluation criteria.

### 4.1 Computing data depths

The preliminary analysis shows that the financial ratios of Korean manufacturing firms deviate from the multivariate normality assumption. Instead of parametric approaches, therefore, we introduce nonparametric methods based mainly on DD functions to predict bankruptcy among sample manufacturing firms in Korea. Three types of DD were applied to reduce the high-dimensional financial ratios: the Mahalanobis depth, the simplicial depth, and the Tukey depth. For failed firms  $\mathbf{x}_i$  ( $i = 1, 2, \dots, 144$ ) and non-failed firms  $\mathbf{y}_i$  ( $i = 1, 2, \dots, 144$ ), 53 financial ratios were condensed into one-dimensional measure of DD, without any distributional assumption. For example, Fig. 2 presents the depth values resulting from the Mahalanobis measure for both failed and non-failed firms for 10 years. Figure 2 does not show any remarkable difference between the two groups in a scatter plot. We could find similar trend with respect to the other two depth measures. In general, because computing the Tukey depth for dimensions higher than three is intractable using the method by Rousseeuw and Ruts [30], we can use the random approximation algorithm introduced by Cuesta-Albertos and Nieto-Reyes [7], which is computationally efficient in high-dimensional data-sets.

**Fig. 2** Plot of the Mahalanobis depth values (+: failed firms, o: non-failed firms)



#### 4.2 Plotting via *DD*-plot

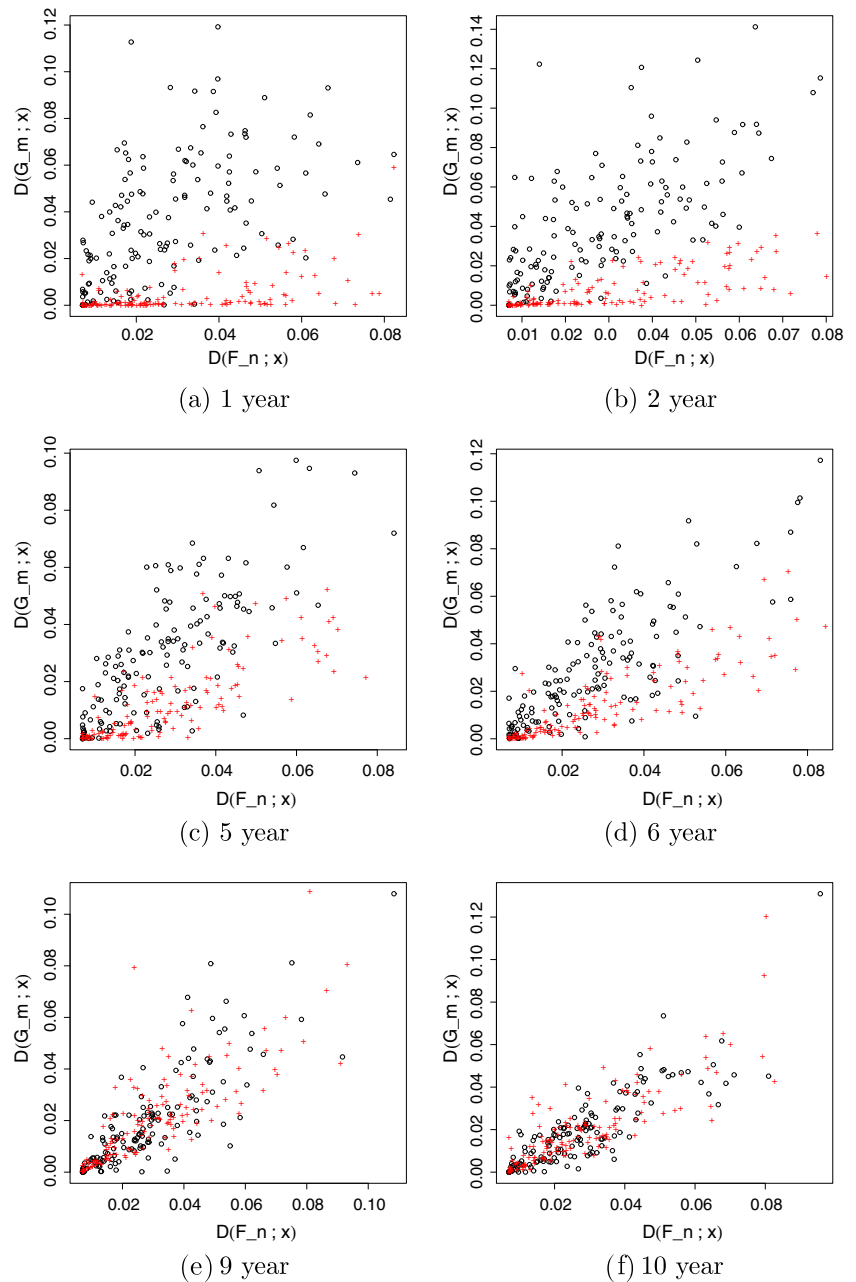
A *DD*-plot can serve as a simple diagnostic tool for visual comparison between two samples of any dimension. Distributional differences (e.g., changes of location, scale, skewness or kurtosis) may present different graphical patterns in *DD*-plots. We drew the *DD*-plots based on the three measures of DD. For example, the *DD*-plots based on the Mahalanobis depth for 10 years are given in Fig. 3. In comparing failed and non-failed firms, the scatter plots in Fig. 3a and b present any separable differences between the two groups. Note that the failed firms scatter at the bottom of the *DD*-plot at two years prior to bankruptcy, representing a

location shift from the group of non-failed firms. However, this separation trend decreased gradually at five (Fig. 3c) and six years (Fig. 3d) prior to bankruptcy or survival. At nine (Fig. 3e) and ten years (Fig. 3f) (prior to bankruptcy or survival), the resulting  $DD(F, G)$  were scattered on the  $45^\circ$  line in the *DD*-plot. This means that the distributions of the two groups are almost the same.

In summary, the use of *DD*-plots is likely to separate any firms encountering imminent bankruptcy from financially healthy firms. Accordingly, in the next step, we tried to classify failed firms and non-failed firms via the nonlinear SVM, based on the values of DD on the *DD*-plot. Because similar trends were observed when we employed different depth



**Fig. 3** *DD*-plot based on the Mahalanobis depths (+: failed firms,  $\circ$ : non-failed firms)



measures (i.e., simplicial and Tukey depths), we illustrate the classification results of nonlinear SVM based mainly on the Mahalanobis depth.

### 4.3 Classifying via the nonlinear SVM

In the context of the corporate bankruptcy classification problem, we introduced nonlinear SVM to a two-dimensional data-set on the *DD*-plot (called “*DD*-SVM”). This study employed three kernel functions for nonlinear SVM, including the RBF, polynomial, and MLP kernel functions. The effectiveness of SVM techniques depends on the proper

selection of a kernel function, the parameters of the selected kernel function, and the soft margin parameter  $C$ .

For example, there are two parameters associated with the RBF kernels, namely, a tuning parameter  $\sigma$  and  $C$ . In general, it is not known beforehand which values of  $C$  and  $\sigma$  are the best for any given problem. Consequently, some kinds of model selection approaches must be employed. After conducting a grid-search for the training set according to an approach by Erdogan [9], the best combination of  $\sigma$  and  $C$  values was selected by the search algorithm with exponentially growing sequences using five-fold cross-validation such as  $\sigma \in \{2^{-15}, 2^{-13}, \dots, 2^1, 2^3\}$

**Table 3** Grid-search results using the RBF kernel function for the first year prior to bankruptcy or survival

$\sigma$										
$C$	$2^3$	$2^1$	$2^{-1}$	$2^{-3}$	$2^{-5}$	$2^{-7}$	$2^{-9}$	$2^{-11}$	$2^{-13}$	$2^{-15}$
$2^{-5}$	0.4655	0.8620	0.8448	0.74137	0.4655	0.4655	0.4655	0.4655	0.4655	0.4655
$2^{-3}$	0.9310	0.9310	0.9137	0.8620	0.7586	0.4827	0.4655	0.4655	0.4655	0.4655
$2^{-1}$	0.9482	0.9310	0.9310	0.9137	0.8793	0.7586	0.4827	0.4655	0.4655	0.4655
$2^1$	0.9310	0.9482	0.9310	0.9310	0.9137	0.8793	0.7586	0.4827	0.4655	0.4655
$2^3$	0.8965	0.9482	0.9655	0.9310	0.9310	0.9137	0.8620	0.7586	0.4827	0.4655
$2^5$	0.8793	0.9310	0.9482	0.9655	0.9310	0.9310	0.9137	0.8448	0.7586	0.4827
$2^7$	0.8793	0.9137	0.9137	0.9655	0.9310	0.9310	0.9310	0.9137	0.8448	0.7586
$2^9$	0.8448	0.8965	0.9310	0.9655	0.9482	0.9310	0.9310	0.9310	0.9137	0.8448
$2^{11}$	0.8793	0.9137	0.9137	0.9482	0.9655	0.9482	0.9310	0.9310	0.9310	0.9137
$2^{13}$	0.8793	0.9137	0.9137	0.9310	0.9655	0.9482	0.9310	0.9310	0.9310	0.9310
$2^{15}$	0.8103	0.8793	0.8965	0.9310	0.9655	0.9655	0.9310	0.9310	0.9310	0.9310

and  $C \in \{2^{-5}, 2^{-3}, \dots, 2^{13}, 2^{15}\}$ . For the first year prior to bankruptcy (or the first year of survival), Table 3 shows that the optimal values of the parameters ( $\sigma, C$ ) in the RBF kernel (even if they are not unique) report a prediction accuracy of 96.55% for the validation set. In the polynomial kernel function, the best prediction performance for the first year prior to bankruptcy (or survival) was obtained at values of  $d = 3, \gamma = 2^{-3}$ , and  $C = 2^5$ , yielding a prediction accuracy of 96.55%. For the MLP kernel function, the best prediction results for the first year prior to bankruptcy (or survival) were obtained at parameter ( $\gamma_1, \gamma_2, C$ ) values of  $(2^{-5}, 1, 2^5)$  and  $(2^{-7}, 1, 2^7)$ , yielding a prediction accuracy of 91.38%. With the results of our analysis, we confirmed that the prediction performance of nonlinear SVM is sensitive to the kernel parameters and the upper bound  $C$ .

Table 4 summarizes the results of the prediction of corporate bankruptcy in Korea over 10 years. We compared the prediction powers of the DD-SVM with the prediction powers of nonlinear SVM based on 53 financial ratios. Overall, the DD-SVM outperforms nonlinear SVM in terms of prediction accuracy for corporate bankruptcy. As shown in Table 4, it was noted that the prediction accuracy resulting from DD-SVMs drastically decreases after seven years prior to bankruptcy. The RBF kernel of the DD-SVM produces

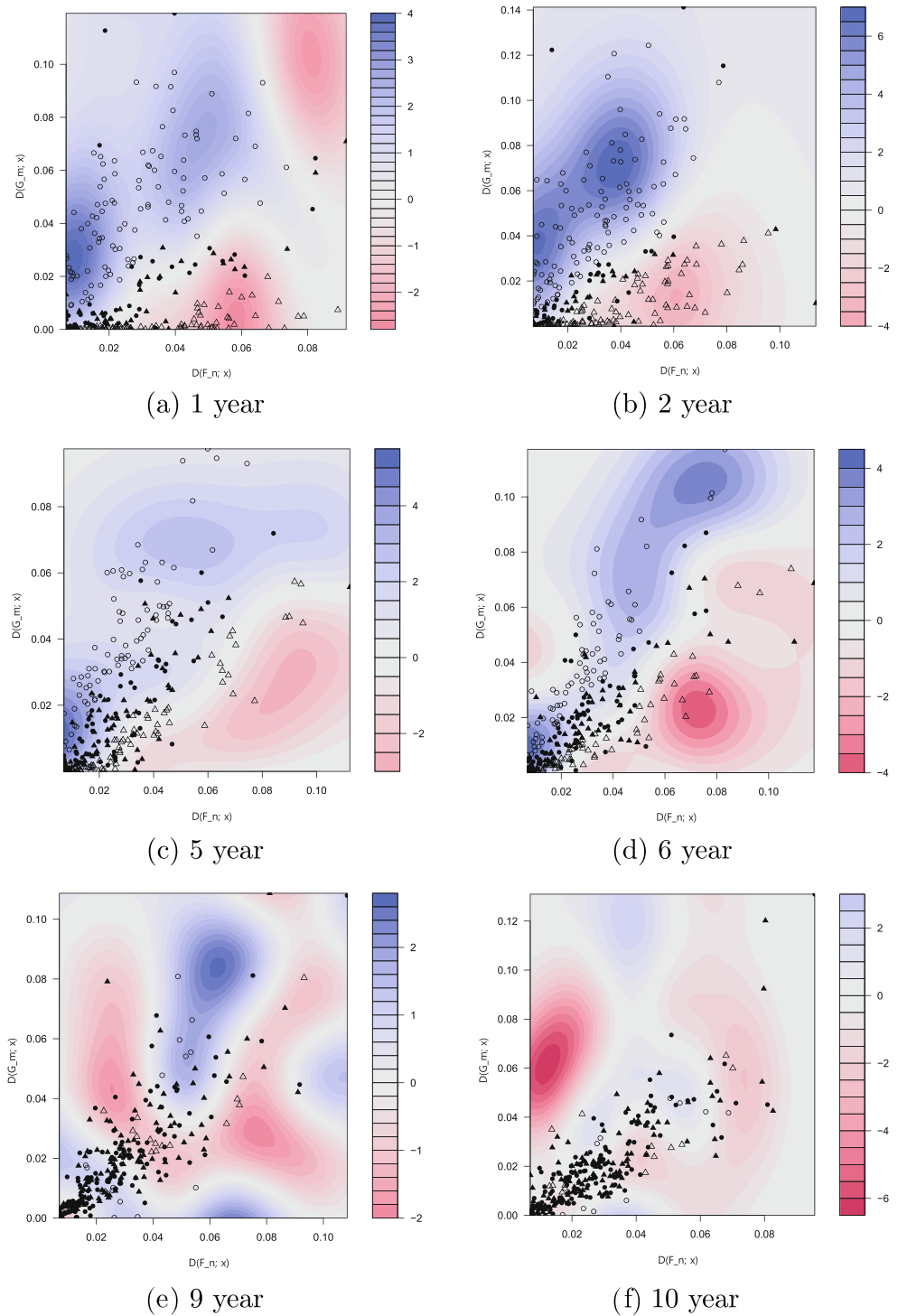
the best prediction performance for 10 years. Prediction powers from the polynomial kernel are comparable with those from the RBF kernel. However, the polynomial kernel requires more hyperparameters (the polynomial degree  $d$ ) than the RBF kernel, and under-fitting and over-fitting problems can occur when  $d$  is poorly selected. Thus, we selected the RBF kernel to compare with other traditional approaches in the prediction of bankruptcy among Korean firms. Figure 4 presents the classification results using the RBF kernel in nonlinear SVM on a  $DD$ -plot based on the Mahalanobis depth. Note that all of the SVM analyses were conducted using R *kernelab* package.

We compared the prediction powers of the DD-SVM with other traditional bankruptcy prediction methods. The bankruptcy classifiers considered in the comparison studies are the logistic regression, multiple discriminant analysis (MDA), and artificial neural network (ANN). The logit model in the logistic regression was employed to investigate the relationship between binary response and financial ratios without the assumption of multivariate normality. The regression parameters were estimated using the maximum likelihood (ML) method. The final logit models for 10 years were selected using the stepwise selection method. We introduced MDA to derive a linear combination of 53

**Table 4** Comparisons of prediction accuracy between DD-SVM and SVM

Kernel		1	2	3	4	5	6	7	8	9	10
RBF	DD-SVM	0.9655	0.9655	0.8966	0.9483	0.8793	0.9138	0.7586	0.6724	0.6034	0.6724
	SVM	0.8444	0.8103	0.7414	0.7759	0.7241	0.6724	0.6897	0.6379	0.6897	0.7069
Polynomial	DD-SVM	0.9655	0.9655	0.8793	0.9483	0.8793	0.9138	0.7586	0.6724	0.5862	0.5517
	SVM	0.8276	0.8103	0.7586	0.7586	0.7759	0.6724	0.6724	0.7069	0.6897	0.7069
MLP	DD-SVM	0.9138	0.9483	0.8621	0.9310	0.8621	0.8966	0.7586	0.6724	0.5690	0.5517
	SVM	0.8103	0.7931	0.7759	0.7586	0.7931	0.6897	0.6552	0.7069	0.7241	0.7414

**Fig. 4** SVM classification results using the RBF kernel function. *Filled circles and triangles* represent support vectors ( $\Delta$ : failed firms,  $\circ$ : non-failed firms)



financial ratios that best discriminates between failed and non-failed firms. The ANN model in this study employed a three-layer connected back-propagation. Following the approach in Min and Lee [26], after fixing the number of layers to one, we varied the number of hidden nodes (8, 12, 16, 24, 32) at learning epochs of 100, 200, and 300 and then recorded the parameter values that derived

the best prediction powers for 10 years. Table 5 reports bankruptcy prediction results for the manufacturing firms for 10 years, along with Type I and Type II errors. The logit and MDA models show poor prediction powers from one to five years prior to bankruptcy. As confirmed in the preliminary analysis, the financial ratios in this study deviate from the multivariate normality assumption, while the MDA

**Table 5** Comparative analysis of prediction accuracy for corporate bankruptcy

		1	2	3	4	5	6	7	8	9	10
Logit	Accuracy	0.6897	0.6724	0.5345	0.6724	0.6724	0.5172	0.5862	0.6379	0.5690	0.5862
	Type I err.	0.1379	0.1552	0.2586	0.1552	0.1552	0.1379	0.2241	0.1379	0.2759	0.1552
	Type II err.	0.1724	0.1724	0.2069	0.1724	0.1724	0.3448	0.1897	0.2241	0.1552	0.2586
MDA	Accuracy	0.7758	0.6552	0.6552	0.7241	0.6552	0.6379	0.5172	0.6207	0.6379	0.6897
	Type I err.	0.1034	0.2069	0.2069	0.1552	0.1897	0.1207	0.2586	0.0862	0.1379	0.0690
	Type II err.	0.1207	0.1379	0.1379	0.1207	0.1552	0.2414	0.2241	0.2931	0.2241	0.2414
ANN	Accuracy	0.8276	0.7759	0.7586	0.7069	0.6207	0.6379	0.6724	0.6724	0.6379	0.6207
	Type I err.	0.0690	0.1034	0.1552	0.1724	0.1379	0.1724	0.1379	0.1207	0.1724	0.2241
	Type II err.	0.1034	0.1207	0.0862	0.1207	0.2414	0.1897	0.1897	0.2069	0.1897	0.1552
DD-SVM	Accuracy	0.9655	0.9655	0.8966	0.9483	0.8793	0.9138	0.7586	0.6724	0.6034	0.6724
	Type I err.	0.0000	0.0000	0.0517	0.0000	0.0000	0.0517	0.1379	0.0517	0.3448	0.1897
	Type II err.	0.0345	0.0345	0.0517	0.0517	0.1207	0.0345	0.1034	0.2759	0.0517	0.1379

performs well under multivariate normality. The DD-SVM outperforms the ANN model, which is a main comparison target from one to six years prior to bankruptcy. Overall, the DD-SVM has the best prediction power among prominent bankruptcy prediction models reviewed in this study. Because of the impact on the domestic economy, Type I error is more important to diagnose in firms at risk for bankruptcy. Remarkably, the DD-SVM shows almost zero Type I errors from one to five years prior to bankruptcy.

## 5 Discussion

Because corporate bankruptcy causes heavy damage to the economy, it is essential to diagnose firms at risk for bankruptcy in order to hedge against the financial harm the at-risk firms stand to inflict. We presented a novel hybrid method that combines DD and nonlinear SVM for the prediction of corporate bankruptcy. This study pioneers the application of DD to financial distress prediction. The empirical results demonstrate that the proposed method offers the highest level of accuracy in bankruptcy prediction. By condensing various financial data from the firms through DD metrics, the DD-SVM can be utilized as part of an early warning system for corporate failure. Unless pertinent information about an underlying distribution is available, we strongly recommend the use of our procedures (detailed in Section 4) to establish a prediction model for corporate bankruptcy. The proposed method is expected to help provide a guidance on corporate investment for investors or other potentially interested parties.

Our comparison is based on a data-set with an equal proportion of failed and non-failed firms. In real cases, however, the number of bankrupt firms constitutes only a small portion of the whole population. To avoid modeling bias, which may be caused by the one-to-one matching process,

the entire population of both failed and non-failed firms can be used to construct a corporate bankruptcy prediction model. Financial distress is not the only area wherein DD based SVM shows potential for beneficial use. We believe that the DD-SVM can be extended to other managerial applications, particularly classification problems in various areas.

**Acknowledgements** This work was supported by the Human Resources Program in Energy Technology of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea. (No. 20154030200900), and Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (No. 201500000002571).

## References

- Ahn H, Lee K, Kim KJ (2006) Global optimization of support vector machines using genetic algorithms for bankruptcy prediction. Lecture notes in computer science, vol 4234. Heidelberg, pp 420–429
- Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J Financ* 23:589–609
- Bae SJ, Do G, Kvam PH (2016) On data depth and the application of nonparametric multivariate statistical process control charts. *Appl Stoch Models Bus Indus*. To appear
- Barnett V (1976) Ordering of multivariate data. *J R Stat Soc A* 139:318–354
- Beaver WH (1966) Financial ratios as predictors of failure. *J Account Res* 4:71–111
- Blanz V, Schölkopf B, Bühlhoff H, Burges C, Vapnik V, Vetter T (1996) Comparison of view-based object recognition algorithms using realistic 3D models. In: von der Malsburg C, von Seelen W, Vorbrüggen J C, Sendhoff B (eds) *Artificial neural networks-ICANN'96*. Springer Lecture Notes in Computer Science, Berlin, pp 251–256
- Cuesta-Albertos J, Nieto-Reyes A (2008) The random Tukey depth. *Comput Stat Data Anal* 52:4979–4988
- Donoho DL, Gasko M (1992) Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann Stat* 20:1803–1827

9. Erdogan BE (2013) Prediction of bankruptcy using support vector machines: an application to bank bankruptcy. *J Stat Comput Simul* 83:1543–1555
10. Fan A, Palaniswami M (2000) Selecting bankruptcy predictors using a support vector machine approach. In: Proceedings of the international joint conference on neural networks, vol 6, pp 354–359
11. Fraiman R, Meloche J (1999) Multivariate L-estimation. *Test* 8:255–317
12. Ghosh A, Chaudhuri P (2005) On maximum depth and related classifiers. *Scand J Stat* 32:327–350
13. Häardle W, Moro R, Schafer D (2003) Predicting corporate bankruptcy with support vector machines Working Slide. Humboldt University and the German Institute for Economic Research, pp 352–360
14. Häardle W, Lee YJ, Schafer D, Yeh YR (2009) Variable selection and oversampling in the use of smooth support vector machines for predicting the default risk of companies. *J Forecast* 28:512–534
15. Henze N, Zirkler B (1990) A class of invariant consistent tests for multivariate normality. *Commun Stat Theory Methods* 19(10):3595–3617
16. Lawrence S, Giles CL, Tsoi AC (1997) Lessons in neural network training: overfitting may be harder than expected. In: Proceedings of the 14th national conference on artificial intelligence, AAAI-97, pp 540–545
17. Lee H, Jo H, Han I (1997) Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. *Expert Syst Appl* 13:97–108
18. Li H, Sun J (2008) Ranking-order case-based reasoning for financial distress prediction. *Knowl-Based Syst* 21:868–878
19. Li J, Cuesta-Albertos JA, Liu R (2012) DD-classifier: nonparametric classification procedure based on DD-plot. *J Am Stat Assoc* 107:737–753
20. Lin PT (2001) Support vector regression: systematic design and performance analysis. Unpublished Doctoral Dissertation, Department of Electronic Engineering, National Taiwan University
21. Liu RY (1990) On a notion of data depth based on random simplices. *Ann Stat* 18:405–414
22. Liu RY, Parelius JM, Singh K (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann Stat* 27:783–858
23. Liu Z, Modarres R (2011) Lens data depth and median. *J Nonparametric Stat* 23:1063–1074
24. MacKay DJ (1992) Bayesian kernel based classification for financial distress detection. *Neural Comput* 4:448–472
25. Mahalanobis PC (1936) On the generalized distance in statistics. *Proc Nat Acad India* 12:49–55
26. Min JH, Lee YC (2005) Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Syst Appl* 28:603–614
27. Min SH, Lee J, Han I (2006) Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Syst Appl* 31:652–660
28. Odom MD, Sharda R (1990) A neural network model for bankruptcy prediction. In: Proceedings of the IEEE international joint conference on neural networks, vol 2. San Diego, pp 163–168
29. Ohlson JA (1980) Financial ratios and probabilistic prediction of bankruptcy. *J Account Res* 18:109–131
30. Rousseeuw PJ, Ruts I (1996) Bivariate location depth. *Appl Stat* 45:516–526
31. Rousseeuw PJ, Jubert M (1999) Regression depth. *J Am Stat Assoc* 94:388–402
32. Salcedo-Sanz S, Fernandez-Villacanas JL, Segovia-Vargas MJ, Bousono-Calzon C (2005) Genetic programming for the prediction of insolvency in non-life insurance companies. *Comput Oper Res* 32:749–765
33. Sarle WS (1995) Stopped training and other remedies for overfitting. In: Proceedings of the 27th symposium on the interface of computing science and statistics, pp 352–360
34. Tam KY, Kiang MY (1992) Managerial applications of neural networks: the case of bank failure predictions. *Manag Sci* 37:926–947
35. Tsai CF (2009) Feature selection in bankruptcy prediction. *Knowl Based Syst* 22:120–127
36. Tukey JW (1975) Mathematics and picturing of data. *Proc 1974 Int Congress Math* 2:523–531
37. Van Gestel T, Suykens JA, Lanckriet G, Lambrechts A, De Moor B, Vandewalle J (2002) Bayesian framework for least-squares support vector machine classifiers, Gaussian processes, and kernel Fisher discriminant analysis. *Neural Comput* 15:1115–1148
38. Van Gestel T, Baesens B, Suykens JA, Van den Poel D, Baestaens DE, Willekens M (2006) Bayesian kernel based classification for financial distress detection. *Eur J Oper Res* 172:979–1003
39. Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
40. Wu CH, Tzeng GH, Goo YJ, Fang WC (2007) A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Syst Appl* 32:397–408
41. Zmijewski ME (1984) Methodological issues related to the estimation of financial distress prediction model. *J Account Res* 22:59–82
42. Zuo Y, Serfling R (2000) General notions of statistical depth function. *Ann Stat* 28(2):461–482



**Sungdo Kim** was born in 1979, in Busan, Republic of Korea. He received the B.S. Degree in industrial engineering from the Kongju University, Chungcheongnam-do, Korea, and the M.S degree in industrial engineering from Hanyang University, Seoul, Korea, in 2005 and 2007, respectively. He received the Ph.D. degree at the Department of industrial engineering, Hanyang University, Seoul Korea in 2017. His research interests in statistical process control, bankruptcy prediction model and data mining.



**Byeong Min Mun** received the Ph.D. degree from the Department of Industrial Engineering, Hanyang University, Seoul, Korea. He is research professor in the Department of Industrial Engineering at Hanyang University, Seoul, Korea. His current research interests include reliability, data mining, and Bayesian inference.



**Suk Joo Bae** received the Ph.D. degree from the School of Industrial and Systems Engineering at the Georgia Institute of Technology, Atlanta, GA, USA, in 2003. He is a professor in the Department of Industrial Engineering at Hanyang University, Seoul, Korea. He worked as a reliability engineer at Samsung SDI, Korea, from 1996 to 1999. His research interests are centered on reliability evaluation of light displays and nano-devices via accelerated life and degrada-

tion testing, statistical robust parameter design, and process control for large-volume on-line processing data. He published about 30 papers in journals such as *Technometrics*, *Journal of Quality Technology*, *IIE Transactions*, *IEEE TRANSACTIONS ON RELIABILITY*, and *Reliability Engineering & System Safety*. Prof. Bae is an Associate Editor of the *IEEE TRANSACTIONS ON RELIABILITY*. He is a member of INFORMS and ASA.