JOURNAL OF ECONOMIC DEVELOPMENT
Volume 51, Number 2, June 2026

MACHINE-LEARNING BASED DEFAULT PREDICTION: THE ROLE OF EXTERNAL AUDITS AND FINANCIAL CONSTRAINTS

HYUNJUN SONG [a], DOJOON PARK [b], ZOONKY LEE [c] AND YONG JOO KANG [d]

[a] *Ajou University, South Korea*
[b] *Kongju National University, South Korea*
[c] *Yonsei University, South Korea*
[d] *Thompson Rivers University, Canada*

We investigate the impact of external audits and financial constraints on predicting small and medium-sized enterprise (SME) defaults in Korea using data for both externally and non-externally audited firms. The importance of a SME-specific default model is highlighted through the performance comparison of default prediction models for SMEs and large firms. Since SMEs are more vulnerable to financial market shocks, variables related to cash reserves and financial constraints are included. The analysis by feature importance shows that variables related to cash balance and financial constraint contribute towards an improvement in default prediction for Korean SMEs.

Keywords: Default Prediction; Small and Medium Enterprises; External Audits; Financial Constraints; Machine Learning
JEL Classification: C19; C53; C55; C65; G17

1. INTRODUCTION

Firm default and default prediction models have been a topic of widespread interest in numerous existing finance and economic literature. Most of these prior studies either extend models developed for listed firms or were focused primarily on large firms. However, small and medium enterprises (SMEs) account for an overwhelming majority number of firms and provide significant number of jobs in both the United States and Korea. Despite the vital role of SMEs in the economy, SME-specific default prediction models are comparatively understudied. The difficulty in conducting default research on

SMEs lies largely in the unavailability or unreliability of the relevant data used in the prediction models, as SMEs are mostly unlisted or not externally audited (Che et al., 2020). However, investigation of SME's is crucial due to the markedly different financial, organizational, and strategic characteristics of SMEs compared to large firms, (Ciampi, 2015; Ciampi and Gordini, 2013).

Our paper fills this gap by comparing SME and large firm defaults using data on Korean firms. Given SMEs' limited access to external finance and greater financial constraints, improving default prediction may contribute to more efficient credit allocation and improved access to finance for financially viable SMEs. Korea presents a unique setting for conducting our analysis as it is one of the few countries with comparatively easier access to data on a large sample of unlisted SMEs (Kim et al., 2011). Of particular note is that the Korean data used in our study includes both non-externally audited and externally audited firms, which provides us with an advantage when documenting the effect of external audits for predicting SME defaults. In essence, our dataset consists of over 617,000 non-externally audited firms and 90,000 externally audited firms. Given this expansive and inclusive dataset, we can further divide firms by their size to examine the impact of financial constraints on default prediction. Finally, our data includes four groups of firms; externally audited large firms, externally audited medium firms, non-externally audited small firms, and non-externally audited micro firms, which allows us to determine the model's performance across several different dimensions.

Since most default prediction models rely on information contained in the financial statements, the accuracy and reliability of the financial statements is imperative. External audits examine and evaluate the financial statements to ensure that the financial records are a fair and accurate representation of the firm's financial performance (Yoo et al., 2011). However, since most SMEs are neither required by regulation[1] nor voluntarily choose to be subject to an external audit, the resulting opaqueness and non-reliability of financial information makes default prediction of SMEs challenging. In fact, the credibility and reliability of financial statements for non-externally audited firms have been a concern for decades, as their use involves a greater risk of relying on financial statements that may be manipulated to obtain higher corporate credit ratings or attract more investments (Spathis, 2002). Examples of such manipulations include overstating sales by reporting false invoices or maximizing revenues by shifting revenue from different periods. In fact, our empirical results using the full dataset show that the log odds of default are higher for the non-externally audited firms and the probability of default is significantly lower for externally audited firms.

[1] Prior to the amendment on November 1, 2018, under Article 2 of the Act on External Audit of Stock Companies, firms subject to an external audit are those that, as of the end of the immediately preceding business year, have (1) total assets of at least KRW 12 billion; or (2) total liabilities and total assets of at least KRW 7 billion; or (3) firms whose total number of employees exceeds 300 and has total assets of at least KRW 7 billion.

To properly assess the risk of default for SMEs, default prediction models need to incorporate variables that are tailored to the unique features and characteristics of SMEs. Most SMEs are relatively young, lack sufficient collateral, and have limited access to financial markets for external financing (Gertler and Gilchrist, 1994). SMEs are also more likely to be financially constrained and often need to rely on loans from financial institutions for most of their required funds. Hence, insufficient cash reserves of financially constrained firms increase the possibility of default in the event of a temporary deterioration in liquidity. Risker firms tend to accumulate higher cash reserves to mitigate default risk, as firms with greater cash reserves are less likely to default (Acharya et al., 2012), and cash holdings are particularly important for SMEs in reducing financial frictions (Martínez-Sola et al., 2018).

Our study is differentiated from previous studies given that it compares and analyzes the predictive power of the base Altman (1968) model with an extended version of the Altman model, which includes five additional variables related to the cash balance and the degree of financial constraints for SMEs. Since SMEs are vulnerable to financial market shocks, it is highly likely that variables related to cash reserves and financial constraints would be more salient in SME default prediction models. To address concerns caused by multicollinearity among the accounting variables, our study evaluates the performance of the default prediction models using six methodologies, which includes machine learning-based approaches that can better handle complex, interactive, and nonlinear relationships within the data. Using six evaluation metrics, including the geometric mean (G-mean) of the correctly classified proportions of actual default and non-default samples, we assess and compare the various methodologies.

We find that the average G-mean of the models for large and medium externally audited firms is 0.7389 and 0.7134, respectively, whereas the corresponding values for small and micro non-externally audited firms are 0.6239 and 0.6138, respectively. This result confirms the findings of previous studies that predicting default for non-externally audited SMEs is relatively more difficult than for externally audited large firms. Comparison of the Altman model and the extended model shows that the average increase in G-mean for large externally audited firms is 0.025, while the average improvement in G-mean for non-externally audited small and micro firms is 0.058 and 0.038, respectively. The results of our study show that the performance of the default prediction model is significantly improved for SMEs when the additional cash balance and financial constraint related variables are included, particularly for small non-externally audited firms.

To summarize, our results provide valuable insights into the impact of external audits and financial constraints on predicting SME defaults by including both externally audited and non-externally audited firms in the dataset. This paper makes four main contributions: First, we find that external audits are significantly negatively associated with firm default, emphasizing the critical role of external auditing. Second, we compare the performance of the Altman default prediction model and show that its application is more challenging and displays lower performance for SMEs than for the widely used

large firm samples. Third, we show that the performance of the default prediction model is lower for non-externally audited SMEs than for externally audited SMEs. Fourth, we include five additional variables related to the cash balance and degree of financial constraints in an extended model and show that it leads to an improvement in model performance, particularly for SMEs. Overall, the results of our study highlight not only the importance of financial statement credibility and reliability, but also the potential improvement in SME default prediction when SME-specific variables are included.

The rest of the paper proceeds as follows. Section 2 provides a brief overview of prior SME default research. Section 3 provides a discussion on the methodology and data used in our study. Section 4 presents the results on model performance and variable feature importance, and Section 5 concludes.

2. RELATED LITERATURE

Default prediction has been a critical issue in finance, not only for classifying firms with a high probability of default but also for managing default risk, and it has been widely studied. In the late 1960s, Altman (1968) introduced the Z-score model, a multivariate discriminant analysis model, for default prediction that utilized five significant financial ratios. Later, Ohlson (1980) introduced the O-score model, a logistic regression model with nine financial ratios. Although many other models have been proposed since then, the Z-score and logistic regression models remain the standard models for default prediction (Altman, 2018).

Early default prediction studies suffered from a lack of data and absence of standardized financial ratios. According to Altman et al. (2019), the amount of data used for default prediction studies was so crucial that he claimed, “Data is King”. With the advancement in computer technology and larger amounts of collected data, researchers started to focus on different methodologies, such as machine learning and deep learning methods, to overcome the limitations of traditional models. Barboza et al. (2017) provided a comparison of traditional models and computational models using yearly data for more than 10,000 firms. Their study confirmed that machine learning models were about 10% more accurate than statistical models.

Although smaller in comparison, research attention on default prediction has not been restricted to large firms, as SMEs have also been studied by several researchers. This is not surprising given that SMEs account for the vast majority of firms in many countries, and they contribute significantly to the economic growth, employment, and technological innovation of a country. For example, SMEs make up around 97% and 99.7% of all firms and provides 75% and 81.9% of all jobs in the United States and Korea, respectively (Altman and Sabato, 2013; Gregory et al., 2002). Altman and Sabato (2013) used multivariate discriminant analysis and logistic regression to develop a SME-focused model that utilized over 2,000 U.S. firms with sales less than $65 million. However, since SMEs have fewer legal obligations to disclose accounting data than

large firms, the information provided by these non-externally audited SMEs is by nature more opaque than those of externally audited larger firms. Hence, default prediction modeling of non-listed or non-externally audited SMEs is inevitably limited by data availability and reliability. Consequently, the performance of prediction models, which are based mainly on samples of large firms, has been shown to be quite low when applied to SMEs, particularly given their unique financial characteristics (Ciampi, 2015). Furthermore, the construction of default prediction models based on financial ratios is particularly complicated and the results are usually far less accurate for SMEs than large firms (Ciampi and Gordini, 2013).

Despite the large body of studies on the prediction of business failures in the U.S., several researchers have extended this work to other countries, such as Korea. Altman et al. (1995) introduced the K-model, a model constructed to predict firm failures for Korean firms. The data used in the study consisted of 34 distressed firms and 61 non-default firms. The study constructed two versions of the K-model, namely the K1 and K2-models, where the accuracy of the models was around 70% to 80%. Jo et al. (1997) applied discriminant analysis, case-based forecasting, and neural networks for default prediction of Korean firms using financial data for 271 default and 271 non-default firms from 1991 to 1993. Lee, et al. (1996) introduced hybrid neural network models using sample data for 166 Korean firms. Kim and Kang (2010) applied ensemble methods, bagging and boosting, to improve the performance of neural networks using data from 1,458 externally audited manufacturing firms. Similarly, Jeong et al. (2012) applied a new tuning method for the neural network model to examine default prediction for externally audited manufacturing SMEs in Korea from 2001 to 2004. Additional studies have also utilized data obtained from one of the major banks (Ahn and Kim, 2009; Kim and Kang, 2012).

While traditional statistical methods, such as the discriminant and logistic regression methods, remain the standard for default prediction, the use of machine learning methods has gained traction due to their potential to enhance model performance (Barboza et al., 2017), especially for SMEs. For example, Gordini (2014) used genetic algorithms on a sample of 3,100 Italian manufacturing SMEs. The genetic algorithms model outperformed the logistic regression and support vector machine (SVM) model. The performance of the prediction model using genetic algorithms increased when the model was applied according to size and geographic area of the firms. Tobback et al. (2017) proposed a network of SMEs creating relational data using the company's director or manager data to determine the relationship between SMEs when sharing the same manager. The study was conducted using data on more than 400,000 Belgian SMEs and 2,000,000 UK SMEs. Xu et al. (2019) showed improvement in default prediction for SMEs by adding cash flow data using logistic regression, decision tree, SVM, neural network, random forest, and extreme gradient boosting. Despite the improvement of default prediction models using state-of-the-art machine learning methods, industries are still using traditional methods due to the limitations of machine learning models, particularly where results are not interpretable. However, recent studies

have attempted to explain the default model using variable feature importance. For example, Son, et al. (2019) used the gradient boosting model and presented the feature importance of the variables.

Given the higher default risk of SMEs compared to larger firms, SME-specific default prediction models are obviously preferred when analyzing SMEs. The topic of SME default has acquired renewed interest since the 1990s with the adoption and implementation of the Basel Capital Accords. Under the Basel Capital Accords, banks are forced to compute their capital requirements based on the ratings assigned to their borrowers, including SMEs. Gupta, et al. (2018) divided U.S. SMEs into medium, small, and micro sized firms and suggested that determinants of default were different depending on the size of SMEs. A particular feature of our data is that we can discern between firms with and without an external auditor. This allows us to understand whether the financial variables of non-externally audited firms, in addition to firm size, affect the performance of default prediction models. Moreover, we explore the effect of our chosen five additional cash balance and financial constraint related ratios on improving prediction performance. Kim and Yi (2012), who utilized a database of SMEs for the period spanning 1995 to 2007, empirically found that accounting variables related to financial stability and cashflows were useful for default prediction.

Our study focuses on identifying financial variables for SMEs default prediction based on prior research related to information asymmetries and financial constraints. To our knowledge, only a few studies focusing on default prediction models designed specifically for non-externally audited firms have been published. Kim et al (2016) built a model for non-externally audited firms using artificial neural networks and Son et al. (2019) used a dataset that included both externally audited and non-externally audited firms. Although both these studies focus on applying new machine learning techniques in the field of default prediction, none have compared differences in model performance in conjunction with the existence of an external audit. Small-sized firms are financially constrained and have limited cash to prevent financial distress. This is in stark contrast to larger firms that are unconstrained and can mitigate their default risk with greater cash holdings. Essentially, cash holdings are more valuable for financially constrained firms (Denis and Sibilkov, 2010). Previous studies suggest that a firm's cash holdings provide opportunities to obtain extra funds and enhance its ability to pay debt. This is especially important for SMEs as they need to reserve cash holdings to withstand financial frictions (Martínez-Sola et al., 2018). We have found that the inclusion of the five additional variables related to cash balance and financial constraints contributes positively to model performance for SME default predictions. In addition, our comparison of model performance for firms of different sizes show that the five selected variables contribute more significantly to the enhancement of model performance for SMEs, which confirms that the characteristics of SMEs differs from large firms.

3. DATA AND METHOD

3.1. Methodology

The six quantitative methods employed in our analysis for default prediction include two statistical approaches, the linear discriminant analysis (LDA) and logistic regression (LR), three decision-tree based approaches, random forest (RF), AdaBoost (AB), and gradient boosting (GB), and one neural network approach, the multi-layer perceptron (MLP).

In order to examine whether firm default is associated with external audits, we first run univariate and multivariate logistic regressions with the accounting variables. We then construct default prediction models using the six quantitative methods. To compare the performance of default prediction models, the data is divided into four groups, based on the existence of an external audit and on total asset size.

In terms of analysis methodology, our study is meaningful in that we propose two default prediction models, the base and extended models, and employ six different classification methods to compare their predictive power for each group. Our base model utilizes modified versions of the five Altman (1968) variables while our extended model adds five cash balance and financial constraint related variables that reflect the characteristics of SMEs to our base model. The results obtained from these models and classification methods are then used to compare the performance of the default prediction models for firms based on whether they are externally audited and on total asset size using six evaluation metrics. Furthermore, we also investigate variable feature importance for the models to determine which variables exert the most influence on the model and are important to the default prediction for each group.

Due to the relatively small number of firms that default compared to solvent or non-default firms, the data used for the prediction models is imbalanced, with a default ratio of approximately 1.7% to 3.8% for each group in our dataset. This imbalance can cause prediction models to classify all firms as belonging to the majority class when minimizing the error rate (Lin et al., 2017). Therefore, balancing the imbalanced data is necessary for the development of effective prediction models (Zhou, 2013).

We use random under-sampling , a method for resampling an imbalanced dataset by removing data from the majority class to match the training set size of the minor (default firms) and major (non-default firms) classes. The models are then evaluated ten thousand times with differently sampled data to calculate the average performance of the models while minimizing the information loss due to under-sampling. For each group, the training and test sets are generated by stratified splitting of the data between default and non-default firms with a 7:3 ratio. After creating the training and test sets, we implemented the sampling method to obtain a training set with a 50:50 ratio of default and non-default firms. The test set used to evaluate the model consists of the actual ratio of default and non-default firms, without any sampling to balance it.

3.2. Model Evaluation Metrics

The general evaluation metrics typically used for balanced datasets is considered inappropriate for evaluating imbalanced datasets as the model performance will be exaggerated when prediction models classify all examples to be in the majority class (Veganzones and Séverin, 2018). Therefore, it is crucial to calculate the percentage of accurately predicted samples in both the major and minor classes. We use six evaluation metrics in our study, namely sensitivity, specificity, F1-Score, area under the receiver operating characteristic curve (AUC), geometric mean (G-mean), and the Matthews correlation coefficient (MCC). The descriptions of the metrics are as follows:

$$\begin{aligned} Sensitivity &= Percentage\ of\ accurately\ predicted\ default\ samples \\ &= TP\ /\ (TP + FN), \end{aligned}$$

$$\begin{aligned} Specificity &= Percentage\ of\ accurately\ predicted\ non-default\ samples \\ &= TN\ /\ (TN + FP), \end{aligned}$$

$$F1 - Score\ = \frac{(Sensitivity \times Specificity)}{\left\{\frac{Sensitivity\ +\ Specificity}{2}\right\}},$$

$$AUC = \frac{(Sensitivity + Specificity)}{2},$$

$$G - mean = \sqrt{Sensitivity \times Specificity},$$

$$MCC = \frac{TN\ \times\ TP\ -\ FN\ \times\ FP}{\sqrt{(TP\ \times\ FP)(TP\ \times\ FN)(TN\ \times\ FP)(TN\ \times\ FN)}},$$

where TP is default firms correctly classified, FN is misclassified default firms, TN is non-default firms correctly classified, and FP is misclassified non-default firms.

Sensitivity and specificity indicate the proportion of actual default and non-default samples that are correctly classified, respectively. $F1 - Score$, AUC, and $G - mean$ evaluate the effectiveness of a classification in terms of the ratio of sensitivity and specificity and provide an overview of the model's overall performance. The F1-Score evaluates the model using two attributes, sensitivity and specificity, and finds the harmonic mean between them. AUC and $G - mean$ are the arithmetic and geometric averages of the two attributes, respectively, and are commonly used metrics in default prediction studies (Zelenkov and Volodarskiy, 2021; Kim et al., 2016; Veganzones and Séverin, 2018). MCC is a metric used to evaluate the performance of binary classification models, particularly when the classes are imbalanced. MCC takes into account all four confusion matrix values and produces a score ranging from -1 to +1,

where +1 represents a perfect prediction, 0 is random, and -1 is a completely wrong prediction.

3.3. Data Sources

The study uses financial ratios of non-financial firms in Korea from 2010 to 2016 including the default observation period for 2017. We acquired financial ratios and financial status (default or non-default) from Korea Enterprise Data (KED).[2] This unique dataset includes both listed and unlisted firms as well as externally audited and non-externally audited firms. We split the firms into four groups based on whether the firm is externally audited and on the firm's total assets: (1) large externally audited firms with total assets of more than KRW 60 billion; (2) medium externally audited firms with total assets of more than KRW 7 billion but under KRW 60 billion; (3) small non-externally audited firms with total assets of more than KRW 1 billion but less than 7 billion; and (4) micro non-externally audited firms with total assets of less than KRW 1 billion.[3] The final sample yielded a total of 706,388 firm-year observations. Default firms are defined as firms that lack the ability to repay or are not willing to repay liabilities based on the Basel Convention.

Table 1. Sample Description

Group	Default	Non-Default	Default Ratio
Externally Audited Firms:			
(1) Large	374	21,415	1.72%
(2) Medium	1,801	66,220	2.65%
Non-Externally Audited firms:			
(3) Small	14,476	365,783	3.81%
(4) Micro	6,815	229,504	2.88%

Notes: This table provides the number of firms in each of the four groups along with their classification (default or non-default firms). The default ratio is the proportion of default firms in the sample.

[2] Korea Enterprise Data (KED) is a leading supplier of business credit reports on Korean businesses (http://www.kodata.co.kr/en/ENINT01R0.do). In addition to large sized enterprises, KED specializes in credit information on small to medium sized enterprises and has the largest database on SMEs in Korea. Through the data collected and updated over the past 30 years by the Korea Credit Guarantee Fund (KODIT), and the Korea Technology Finance Corporation (KIBO), KED has a database of 8.0 million companies, which is the largest in South Korea. Fortunately, KED provided us with data for our limited sample period.

[3] KRW 60 billion is equivalent to approximately USD 50 million, based on the prevailing exchange rate during the sample period. KRW 7 billion serves as one of the thresholds for the mandatory external audit requirement. By dividing the non-externally audited firms by the firm size of KRW 1 billion, small and micro firms exhibit a similar number of samples.

Table 1 provides the number of firms within each classification group. The default ratio is the proportion of default firms in the sample. The first two groups consist of large and medium firms that are externally audited while the remaining two groups consist of small and micro firms that are not externally audited. The group for large externally audited firms consists of 374 default and 21,415 non-default firms with a default ratio of 1.72%. The group for medium externally audited firms has 1,801 default and 66,220 non-default firms with a default ratio of 2.65%. As shown, the default ratio of non-externally audited firms is higher than that of externally audited firms where small non-externally audited firms have a default ratio of 3.81% (14,476 default and 365,783 non-default firms) and the micro non-externally audited firms have a default ratio of 2.88% (6,815 default and 229,504 non-default firms).

Table 2. List of Variables

Variables	Description
Base Model Variables:	
WC/TA	Working Capital/Total Assets
RE/TA	Retained Earnings/Total Assets
OI/TA	Operating Income/Total Assets
TL/TA	Total Liabilities/Total Assets
S/TA	Sales/Total Assets
Extended Model Additional Variables:	
QA/TA	Quick Assets/Total Assets
QA/CL	Quick Assets/Current Liability
QA/CA	Quick Assets/Current Assets
SG	Sales Growth
CF/TA	Cash Flow from Operations /Total Assets

Notes: The table lists the variables used for this study. The base prediction model uses the first five variables, and the extended model uses all ten variables.

We propose two default prediction models: the base and extended models. For the base model, we adopt and modify the five variables used in the Altman Z-score model (Altman, 1968). For the extended model, we add five variables related to the firm's cash balance and financial constraints. Table 2 lists the variables used for this study. The five variables in Altman Z-score model includes working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, market value of equity/book value of total debt and sales/total assets. Since most of the sample consists of unlisted firms, we could not include the market value of equity/book value of total debt in our base model.[4] Moreover, most non-externally audited firms did not provide information on interest expenses. For these reasons, we include only three variables in

[4] The average number of listed firms in Korea during the sample period was around 2,250.

the original Altman Z-score model in our base model: working capital/total assets (WC/TA), retained earnings/total assets (RE/TA), and sales/total assets (S/TA) and replace the remaining two variables in our base model with operating income/total assets (OI/TA) and total liabilities/total assets (TL/TA).

Small firms tend to be young, poorly collateralized, and financially constrained. Firms that have limited access to financial markets are financially vulnerable and face a high level of default risk due to these financial constraints. However, financially constrained firms with higher cash reserves are less likely to default compared to such firms with lower cash reserves. Accordingly, we select three cash related variables: quick assets/total assets (QA/TA), quick assets/current liabilities (QA/CL), and quick assets/current assets (QA/CA). The quick assets of a firm are composed of cash and cash equivalents, marketable securities, and accounts receivable, which are very liquid and can easily be converted into cash.

We further add two variables related to the financial constraints index of Whited and Wu (2006). The financial constraints index is composed of the cash flow/total assets ratio, cash dividends, the long-term debt/total assets ratio, the natural log of total assets, industry sales growth, and firm sales growth. Since most Korean SMEs do not pay dividends and their industry classifications are not available, we chose to include sales growth (SG) and the cash flow from operations/total assets ratio (CF/TA) to measure the firm's financial constraints. Hence, our base prediction model uses five variables while our extended model uses a total of ten variables.

3.4. Summary Statistics

All variables are winsorized at the top and bottom 2.5% of the datasets to limit the effect of outliers and produce more robust estimators.[5] Summary statistics and correlations of the variables are shown in Table 3. As shown, micro firms have more quick assets and lower retained earnings. Looking at the five variables in our base model, all groups display relatively high correlations between RE/TA and TL/TA, followed by WC/TA and TL/TA. The correlations between each variable are the highest for micro firms while large firms have the lowest correlations. When we look at all ten variables used in our extended model, WC/TA and QA/TA shows relatively high correlations in large, medium, and small firms, with large firms having the highest correlation.

Conversely, the correlation between WC/TA and QA/TA was the lowest for micro firms. This shows that there was a low correlation between working capital and quick assets for micro firms whereby working capital barely contains information on the firm's cash holdings. Small and micro firms show high correlations between QA/TA and

[5] We selected a winzorization level of 2.5% to prevent extreme values from skewing the data. Since the four groups were classified based on total assets, extreme values will distort the results for a specific group and cause difficulty in properly assessing the performance of the models for each group.

QA/CA.

Table 3. Summary Statistics

	Panel A: Large													
	Variables	Mean	Median	Std	1	2	3	4	5	6	7	8	9	10
1	WC/TA	0.07	0.05	0.27	1	0.51	0.33	-0.54	0.17	0.60	0.38	0.20	-0.03	0.09
2	RE/TA	0.24	0.22	0.31		1	0.42	-0.69	0.16	0.36	0.21	0.16	-0.16	0.21
3	OI/TA	0.04	0.03	0.06			1	-0.24	0.33	0.28	0.04	0.05	-0.01	0.54
4	TL/TA	0.53	0.52	0.29				1	0.04	-0.38	-0.40	-0.35	0.15	-0.11
5	S/TA	0.86	0.73	0.72					1	0.11	-0.23	-0.32	-0.12	0.19
6	QA/TA	10.47	5.50	12.55						1	0.48	0.57	-0.04	0.23
7	QA/CL	91.04	19.44	198.14							1	0.58	0.01	0.03
8	QA/CA	30.29	20.57	28.60								1	0.04	0.11
9	SG	28.94	3.68	116.85									1	-0.03
10	CF/TA	6.12	5.32	9.40										1
	Panel B: Medium													
	Variables	Mean	Median	Std	1	2	3	4	5	6	7	8	9	10
1	WC/TA	0.02	0.03	0.35	1	0.58	0.30	-0.68	0.18	0.57	0.41	0.23	-0.04	0.06
2	RE/TA	0.17	0.17	0.37		1	0.45	-0.70	0.12	0.36	0.25	0.22	-0.10	0.18
3	OI/TA	0.04	0.04	0.09			1	-0.27	0.26	0.19	0.01	0.07	0.11	0.50
4	TL/TA	0.62	0.64	0.30				1	0.02	-0.43	-0.46	-0.36	0.11	-0.08
5	S/TA	1.09	0.90	0.91					1	0.11	-0.17	-0.21	0.01	0.16
6	QA/TA	11.61	5.36	14.83						1	0.54	0.67	-0.04	0.17
7	QA/CL	86.69	13.68	219.40							1	0.56	-0.03	0.01
8	QA/CA	27.95	17.73	27.68								1	0.00	0.12
9	SG	14.69	2.30	62.83									1	0.07
10	CF/TA	6.47	5.74	11.61										1
	Panel C: Small													
	Variables	Mean	Median	Std	1	2	3	4	5	6	7	8	9	10
1	WC/TA	0.27	0.28	0.32	1	0.61	0.23	-0.68	0.18	0.42	0.42	0.15	-0.06	0.01
2	RE/TA	0.29	0.27	0.24		1	0.36	-0.78	0.16	0.36	0.30	0.23	-0.13	0.19
3	OI/TA	0.06	0.06	0.08			1	-0.21	0.33	0.16	0.02	0.08	0.20	0.40
4	TL/TA	0.56	0.59	0.24				1	-0.05	-0.46	-0.50	-0.35	0.08	-0.13
5	S/TA	1.67	1.34	1.26					1	0.13	-0.09	-0.04	0.12	0.07
6	QA/TA	13.52	7.55	15.33						1	0.58	0.84	0.02	0.23
7	QA/CL	97.03	23.03	202.40							1	0.55	-0.02	0.09
8	QA/CA	22.49	14.29	22.70								1	0.03	0.25
9	SG	20.42	4.89	66.88									1	0.05
10	CF/TA	5.64	5.42	16.62										1
	Panel D: Micro													
	Variables	Mean	Median	Std	1	2	3	4	5	6	7	8	9	10
1	WC/TA	0.23	0.33	0.53	1	0.73	0.40	-0.83	0.01	0.25	0.28	-0.01	-0.07	0.19
2	RE/TA	0.04	0.14	0.62		1	0.53	-0.81	0.02	0.07	0.09	-0.07	-0.09	0.31
3	OI/TA	0.03	0.06	0.22			1	-0.42	0.17	0.01	0.00	-0.08	0.02	0.60
4	TL/TA	0.60	0.54	0.48				1	0.07	-0.21	-0.27	-0.09	0.06	-0.26
5	S/TA	2.25	1.67	2.02					1	0.03	-0.13	-0.07	0.06	0.09
6	QA/TA	16.98	8.40	19.94						1	0.40	0.83	0.02	0.14
7	QA/CL	220.43	25.13	615.27							1	0.36	-0.01	0.04
8	QA/CA	27.30	15.16	29.19								1	0.03	0.11
9	SG	75.42	7.78	262.66									1	-0.06
10	CF/TA	2.87	4.63	30.05										1

Notes: This table includes summary statistics (mean, median, and standard deviation) and correlations of the firms by each of the four groups (large, medium, small, and micro). The definitions of variables are provided in Table 2.

Table 4. Mean of Financial Ratios and Test of Mean Differences

	Default		Non-Default			
Variables	Mean	Std	Mean	Std	Mean Diff	*t*-value
Panel A: Large						
WC/TA	-0.12	0.28	0.07	0.27	-0.20***	-14.10
RE/TA	-0.06	0.26	0.24	0.31	-0.30***	-18.73
OI/TA	0.00	0.06	0.04	0.06	-0.04***	-13.47
TL/TA	0.84	0.24	0.52	0.29	0.31***	21.02
S/TA	0.73	0.67	0.86	0.72	-0.13***	-3.53
QA/TA	4.74	6.31	10.57	12.61	-5.83***	-8.92
QA/CL	14.54	58.05	92.37	199.46	-77.83***	-7.54
QA/CA	13.40	16.98	30.58	28.68	-17.18***	-11.55
SG	17.08	112.16	29.15	116.92	-12.07**	-1.98
CF/TA	2.21	9.94	6.19	9.38	-3.97***	-8.12
Panel B: Medium						
WC/TA	-0.19	0.33	0.03	0.35	-0.21***	-25.80
RE/TA	-0.11	0.35	0.18	0.37	-0.29***	-32.83
OI/TA	-0.02	0.09	0.04	0.08	-0.06***	-27.89
TL/TA	0.88	0.24	0.61	0.30	0.26***	37.08
S/TA	0.99	0.83	1.09	0.91	-0.10***	-4.77
QA/TA	3.79	6.90	11.82	14.93	-8.04***	-22.77
QA/CL	10.04	56.38	88.78	221.80	-78.74***	-15.05
QA/CA	10.03	15.62	28.44	27.77	-18.41***	-28.01
SG	10.91	77.59	14.79	62.38	-3.88***	-2.58
CF/TA	3.10	12.38	6.57	11.57	-3.46***	-12.50
Panel C: Small						
WC/TA	0.21	0.34	0.27	0.32	-0.05***	-20.12
RE/TA	0.17	0.20	0.29	0.24	-0.12***	-62.77
OI/TA	0.04	0.08	0.06	0.08	-0.02***	-33.89
TL/TA	0.67	0.21	0.56	0.24	0.11***	54.68
S/TA	1.42	1.23	1.68	1.26	-0.26***	-24.32
QA/TA	6.39	10.75	13.81	15.42	-7.42***	-57.33
QA/CL	39.14	129.42	99.32	204.41	-60.18***	-35.14
QA/CA	10.53	15.88	22.96	22.80	-12.43***	-64.98
SG	19.67	81.11	20.44	66.26	-0.78	-1.38
CF/TA	0.85	17.13	5.83	16.57	-4.98***	-35.40
Panel D: Micro						
WC/TA	0.10	0.61	0.23	0.53	-0.14***	-20.96
RE/TA	-0.14	0.73	0.04	0.62	-0.18***	-23.78
OI/TA	-0.02	0.26	0.03	0.22	-0.05***	-19.39
TL/TA	0.80	0.55	0.60	0.48	0.20***	34.26
S/TA	1.96	2.02	2.26	2.02	-0.30***	-12.07
QA/TA	8.28	14.44	17.24	20.02	-8.96***	-36.65
QA/CL	109.10	476.90	223.73	618.60	-114.64***	-15.17
QA/CA	14.80	23.33	27.67	29.26	-12.87***	-35.95
SG	98.39	324.57	74.74	260.57	23.65***	7.32
CF/TA	-3.98	33.50	3.07	29.92	-7.05***	-19.09

Notes: This table includes mean, standard deviation, and test of mean difference in financial ratios for default and non-default firms. The definitions of variables are provided in Table 2. The mean difference of all variables for default and non-default firms for each group are reported with their respective t-statistics. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4 provides mean, standard deviation, and test of mean difference in the financial ratios for the default and non-default firms. As shown, the mean value of WC/TA, RE/TA, OI/TA, and S/TA was higher for non-default firms than default firms, while the mean value of TL/TA was lower. The results are consistent with the patterns observed in Altman (1968). The mean difference in QA/CL for non-default large firms was more than six times higher than default large firms. This difference is even more pronounced for medium firms where the QA/CL of non-default firms was almost nine times more than those of default firms. In comparison, the QA/CL for non-default versus default small firms was substantially lower with non-default firms being only about two and a half times more than those of default firms. This lower difference is also seen for micro firms, where the QA/CL for non-default firms was only about twice that of default firms.

Moreover, the mean difference in SG for non-default firms was higher for large firms and medium firms, highlighting a significant difference in sales growth between non-default and default firms. However, the mean difference in SG for small and micro firms showed a different pattern. There was almost no difference in sales growth between default and non-default small firms, while the sales growth of default micro firms was higher than those of non-default micro firms.

4. RESULTS AND DATA ANALYSIS

4.1. Logistic Regression

We investigate the influence of external audits on default using logistic regressions on the full dataset. Specifically, we introduce a non-external audit dummy variable, $NonEA$, which equals one for non-externally audited firms and zero for externally audited firms. We include the dummy variable in the logistic regression analysis with the accounting variables. The following regression model is used for our analysis.

$$Default_{it} = \alpha + \beta NonEA_t + \gamma_1 X_{it} + \varepsilon_{it}. \quad (1)$$

The dependent variable, $Default_{i,t}$, equals one for default firms and zero otherwise. $NonEA$ is the dummy variable for indicating non-externally audited firms. The accounting variables $X_{i,t}$ consist of five variables for the base model (WC/TA, RE/TA, S/TA, OI/TA, and TL/TA) and ten variables for the extended model (adds QA/TA, QA/CL, QA/CA, SG, and CF/TA).

Table 5 shows the logistic regression results. First, we run univariate regressions using the individual accounting variables. As shown in Column (1) of Table 5, the signs of the estimated coefficients are consistent with the mean difference test results of Table 4, and the coefficient estimates are statistically significant at the 1% level. The mean

difference in SG for micro firms was positive, while the others are negative, and the estimated coefficient is positive.

Table 5. Logistic Regression

	Univariate Regression	Multivariate Regression			
Variables		(1)	(2)	(3)	(4)
NonEA	0.365***		0.390***		0.327***
	(16.03)		(16.64)		(13.87)
WC/TA	-0.438***	0.701***	0.631***	0.608***	0.550***
	(-30.52)	(31.34)	(27.68)	(24.51)	(21.80)
RE/TA	-0.501***	0.319***	0.346***	-0.038	-0.021
	(-48.12)	(11.61)	(12.55)	(-1.35)	(-0.76)
OI/TA	-1.204***	-0.207***	-0.183***	-0.258***	-0.251***
	(-33.41)	(-4.29)	(-3.80)	(-4.37)	(-4.27)
TL/TA	0.896***	1.884***	1.861***	1.219***	1.193***
	(64.43)	(50.09)	(49.10)	(31.81)	(30.96)
S/TA	-0.125***	-0.153***	-0.166***	-0.141***	-0.149***
	(-24.82)	(-30.11)	(-32.04)	(-26.68)	(-27.99)
QA/TA	-0.050***			-0.009***	-0.010***
	(-66.32)			(-6.94)	(-7.50)
QA/CL	-0.002***			0.000**	0.000***
	(-31.30)			(-2.38)	(-2.61)
QA/CA	-0.034***			-0.025***	-0.025***
	(-73.92)			(-31.91)	(-31.20)
SG	0.0001***			0.000***	0.000**
	(3.11)			(2.59)	(2.30)
CF/TA	-0.010***			-0.002***	-0.002***
	(-37.37)			(-5.54)	(-4.96)
McFadden R^2		2.62	2.76	5.94	6.04

Notes: This table reports estimation result for the logistic regression with the full sample. The dependent variable is set to one for default firms, and zero otherwise. $NonEA$ denotes the non-external audit indicator variable, and the definitions for the other variables are provided in Table 2. The pseudo R^2 is calculated according to McFadden (1974). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5 shows the logistic regression results. First, we run univariate regressions using the individual accounting variables. As shown in the first column of Table 5, the signs of the estimated coefficients for each variable are consistent with the mean difference test results of Table 4, and the coefficient estimates are statistically significant

at the 1% level. The mean difference in SG for micro firms was positive, while the others are negative, and the estimated coefficient is positive. Columns (1) and (3) of the multivariate regression in Table 5 show the estimation results for both the base and extended models. Given the potential issues caused by multicollinearity among the accounting variables, the signs of the estimated coefficients for the variables differ from those observed in the univariate regressions. We calculate the pseudo R^2 following the methodology of McFadden (1974), and the pseudo R^2 value for the extended model is higher than that of the base model. Columns (2) and (4) of the multivariate regression in Table 5 add $NonEA$ to the base and extended models. The coefficient estimates for the dummy variable are positive and statistically significant at the 1% level. Since $NonEA$ is a dummy variable, the log odds of default are notably higher for the non-externally audited firms by 0.39 in the base model and 0.33 in the extended model.

We estimate the variance inflation factor (VIF) in the multivariate logistic regression models to diagnose the presence of collinearity. The VIF is a measure of collinearity for a single variable in a regression model. VIF estimates ranging from 1 to 5 indicate the presence of moderate correlation between a variable and the other variables, while a value greater than 5 indicates a potentially high degree of correlation. As shown in Table 6, the estimated VIF values for RE/TA and TL/TA are greater than 5, and those for QA/TA and QA/CA are greater than 3. These results indicate the presence of multicollinearity within the multivariate logistic regression models.

Table 6. Variance Inflation Factor for Multivariate Logistic Regressions

	Multivariate Regression			
Variables	(1)	(2)	(3)	(4)
Non-EA		1.05		1.06
WC/TA	2.72	2.87	2.97	3.11
RE/TA	6.12	6.22	5.32	5.41
OI/TA	1.73	1.73	2.24	2.25
TL/TA	6.56	6.71	5.63	5.76
S/TA	1.04	1.06	1.12	1.13
QA/TA			3.40	3.39
QA/CL			1.16	1.16
QA/CA			3.22	3.23
SG			1.05	1.05
CF/TA			1.54	1.55

Notes: This table reports variance inflation factor (VIF) for the multivariate logistic regressions using the full sample. The dependent variable is set to one for default firms, and zero otherwise. $NonEA$ denotes the non-external audit indicator variable, and the definitions for the other variables are provided in Table 2.

The results of logistic regression analysis on the full dataset suggest that the probability of default is significantly lower for externally audited firms, underscoring the importance of external audits. Due to the multicollinearity associated with the

accounting variables, we employ machine learning methods in the subsequent section to address such concerns. Machine learning methods also have the advantage of properly dealing with complex, interactive, and nonlinear relationships within the data.

4.2. Model Performance

To clearly explore the difference in model performance between large firms and SMEs, which includes the medium, small, and micro firms, we train our base model with five variables and our extended model with ten variables, respectively. Using our extended model, we explore the effect of the additional five variables in improving model prediction performance. The performance of our six default prediction classification methods is then examined for each of the four groups using the six evaluation metrics (Sensitivity, Specificity, F1-Score, AUC, G-mean, and MCC).[6]

Table 7 provides the performance results of our base and extended models for each of the four groups. The results show that the overall performance of the models for large firms in terms of F1-score, AUC, and G-mean is superior compared to the other three groups. Moreover, the model performance for medium and small firms is relatively low, whereby the performance for micro firms is the lowest. Looking at the values of G-mean, which is the geometric mean of the proportion of correctly classified default and non-default samples, the average G-mean of the models for large and median externally audited firms is 0.7389 and 0.7134, respectively, while it is 0.6239 and 0.6138 for small and micro non-externally audited firms. This result reflects the fact that it is more difficult to predict default for non-externally audited SMEs compared to externally audited large firms.

Looking at the results of our base model for large firms with five variables, the GB method outperforms the other selected methods in terms of the F1-score, while the RF method has the highest AUC, G-mean, and MCC. In terms of F1-score, the best performing method for medium and micro firms is the LDA method and the best performing method for small firms is the LR method. The GB method has the highest AUC, G-mean, and MCC values for medium and small firms. For micro firms, both the GB and MLP methods have similarly high AUC, G-mean, and MCC values. The performance of our extended model with ten variables for large firms shows similar patterns to the base model with the GB method displaying the highest F1-score and the RF method displaying the highest AUC, G-mean, and MCC values. Moreover, the values of the F1-score, AUC, G-mean, and MCC for our extended model confirm that the GB method outperforms the other selected methods for medium, small, and micro firms. In general, the results show that the overall performance of the four machine learning approaches outperforms the two statistical methods for SMEs, especially when the extended model with ten variables is used.

[6] Details on each of the six default prediction classification methods (LDA, LR, RF, AB, GB, and MLP) are provided in Appendix A. Classification Methods.

Table 7. Performance of Models on Large firms and SMEs

	Base model (5 variables)						Extended model (10 variables)					
Metric	Sensitivity	Specificity	F1-Score	AUC	G-mean	MCC	Sensitivity	Specificity	F1-Score	AUC	G-mean	MCC
Panel A: Large												
LDA	0.7543	0.7128	0.8174	0.7335	0.7331	0.1330	0.7982	0.7219	0.8239	0.7600	0.7589	0.1493
LR	0.7609	0.7093	0.8151	0.7351	0.7345	0.1334	0.8001	0.7264	0.8270	0.7633	0.7622	0.1519
RF	0.7915	0.7188	0.8217	0.7551	0.7539	0.1462	0.8130	0.7418	0.8372	0.7774	0.7764	0.1628
AB	0.7932	0.6835	0.7976	0.7384	0.7360	0.1323	0.8244	0.7039	0.8120	0.7642	0.7616	0.1491
GB	0.7595	0.7286	0.8280	0.7440	0.7436	0.1413	0.7864	0.7448	0.8389	0.7656	0.7651	0.1565
MLP	0.7853	0.6836	0.7977	0.7345	0.7325	0.1301	0.7844	0.7368	0.8336	0.7606	0.7599	0.1521
Average	0.7741	0.7061	0.8129	0.7401	0.7389	0.1360	0.8011	0.7293	0.8288	0.7652	0.7640	0.1536
Panel B: Medium												
LDA	0.6904	0.7082	0.8062	0.6993	0.6992	0.1393	0.8202	0.6799	0.7889	0.7500	0.7467	0.1703
LR	0.7038	0.6997	0.8006	0.7018	0.7017	0.1399	0.8071	0.6822	0.7903	0.7447	0.7419	0.1670
RF	0.8441	0.5947	0.7267	0.7194	0.7082	0.1430	0.8134	0.6831	0.7910	0.7483	0.7453	0.1696
AB	0.8495	0.5897	0.7230	0.7196	0.7076	0.1428	0.8143	0.6872	0.7939	0.7507	0.7480	0.1717
GB	0.8192	0.6665	0.7794	0.7428	0.7388	0.1639	0.8205	0.7075	0.8079	0.7640	0.7619	0.1839
MLP	0.8150	0.6449	0.7638	0.7299	0.7248	0.1533	0.8306	0.6869	0.7939	0.7587	0.7552	0.1772
Average	0.7870	0.6506	0.7666	0.7188	0.7134	0.1470	0.8177	0.6878	0.7943	0.7527	0.7498	0.1733
Panel C: Small												
LDA	0.6500	0.6090	0.7262	0.6295	0.6291	0.1011	0.7523	0.6029	0.7241	0.6776	0.6735	0.1382
LR	0.6468	0.6113	0.7279	0.6290	0.6288	0.1009	0.7368	0.6076	0.7272	0.6722	0.6689	0.1343
RF	0.7785	0.4483	0.5956	0.6134	0.5906	0.0875	0.7270	0.6078	0.7271	0.6674	0.6647	0.1305
AB	0.7307	0.5373	0.6719	0.6340	0.6265	0.1027	0.7272	0.6562	0.7626	0.6917	0.6908	0.1530
GB	0.7262	0.5617	0.6915	0.6440	0.6386	0.1108	0.7360	0.6632	0.7678	0.6996	0.6986	0.1599
MLP	0.6926	0.5733	0.6997	0.6330	0.6299	0.1027	0.7261	0.6611	0.7659	0.6936	0.6924	0.1553
Average	0.7041	0.5568	0.6855	0.6305	0.6239	0.1009	0.7342	0.6331	0.7458	0.6837	0.6815	0.1452
Panel D: Micro												
LDA	0.5797	0.6458	0.7588	0.6127	0.6118	0.0786	0.7287	0.5619	0.6977	0.6453	0.6399	0.0979
LR	0.5829	0.6429	0.7568	0.6129	0.6121	0.0786	0.7429	0.5331	0.6743	0.6380	0.6291	0.0925
RF	0.7229	0.4800	0.6282	0.6014	0.5884	0.0681	0.7266	0.5779	0.7102	0.6522	0.6479	0.1030
AB	0.6689	0.5740	0.7063	0.6215	0.6195	0.0821	0.6971	0.6364	0.7541	0.6668	0.6660	0.1154
GB	0.6539	0.5980	0.7247	0.6260	0.6252	0.0858	0.7027	0.6440	0.7597	0.6734	0.6727	0.1205
MLP	0.6229	0.6288	0.7470	0.6259	0.6254	0.0870	0.6807	0.6341	0.7519	0.6574	0.6566	0.1090
Average	0.6385	0.5949	0.7203	0.6167	0.6138	0.0800	0.7131	0.5979	0.7247	0.6555	0.6520	0.1064

Notes: This table reports the performance of the base and extended prediction models by each of the four groups (large, medium, small, and micro). For each model, the six evaluation metrics (Sensitivity, Specificity, F1 Score, AUC, G-mean and MCC) are shown for comparison purposes.

Although the models for large firms outperform the models for SMEs, the performance improvement of the models for large firms with five additional variables is relatively small. The average increase in G-mean and AUC for large firms across all six default prediction classification methods is 0.025 for both metrics. The method with the best improvement was the LR method, with an increase in G-mean and AUC of 0.028. On the other hand, the SME groups show different patterns compared to large firms with the performance of all six default prediction classification methods improving in terms of AUC, G-mean, and MCC when the five additional variables are added. The overall G-mean for medium firms increases by an average of 0.037 while the overall AUC increases by an average of 0.034. The LDA method shows the highest increase in G-mean and AUC with a value of 0.048 and 0.051, respectively. For small firms, the overall G-mean and AUC improves by an average of 0.058 and 0.053, respectively. The overall average G-mean and AUC improvement for micro firms is 0.038 and 0.039, respectively.

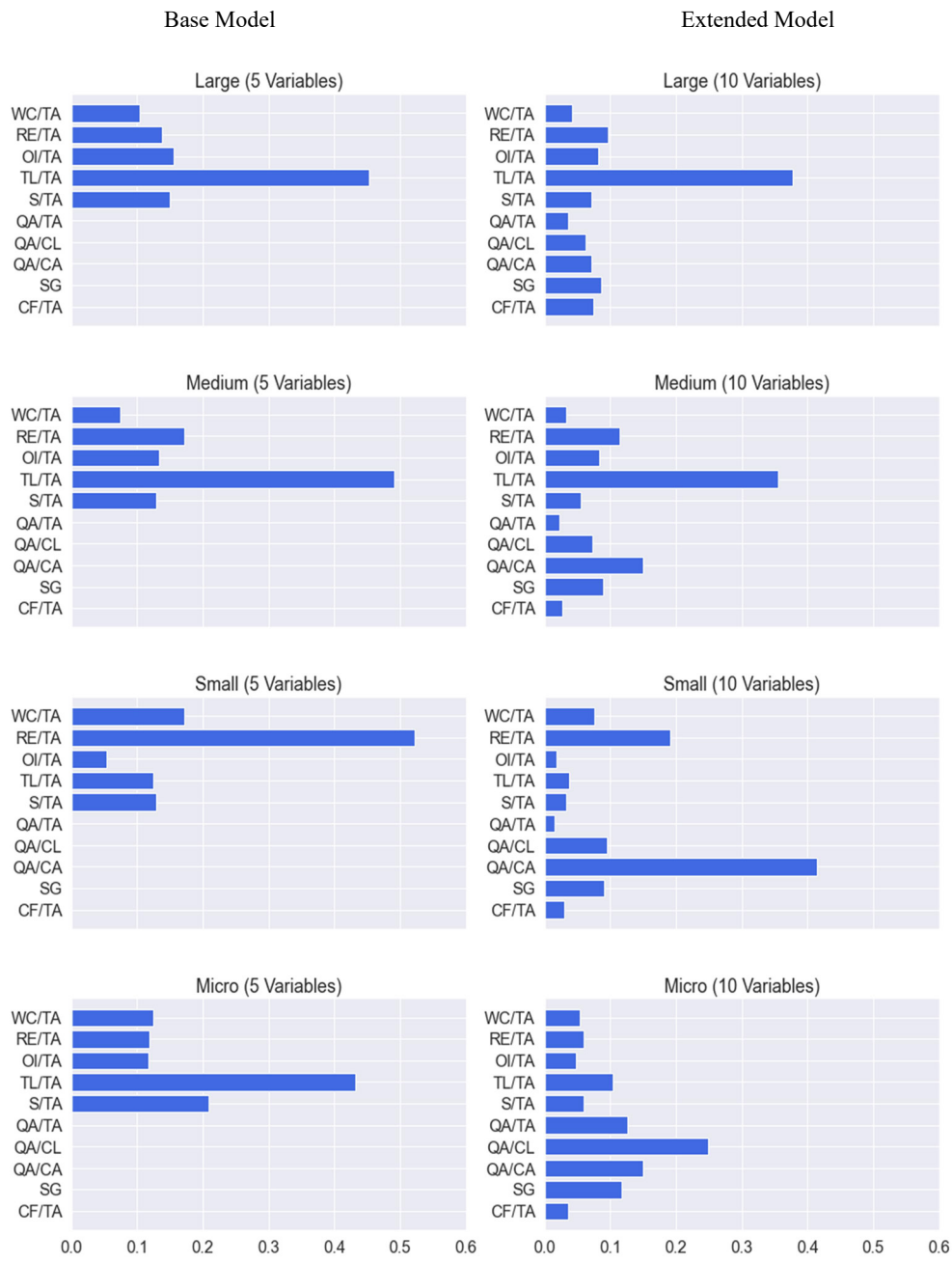
The method with the highest increase in G-mean for small firms is the RF method with an increase of 0.074 while the model with the highest AUC improvement for small firms is the MLP method with an increase of 0.061. The method with the highest AUC and G-mean improvement for micro firms is the RF method with an increase of 0.051 and 0.060, respectively. As a result, all classification methods for SMEs including medium, small, and micro firms display an improvement in model performance when the five variables related to cash balance and financial constraints are added, with the GB model showing the best performance across all SMEs in terms of the F1-Score, AUC, G-mean, and MCC metrics.

Overall, the results obtained show that the default prediction model performance differs depending on whether the firms are externally audited or not. When firms are externally audited, the models seem to have relatively good predictive power compared to firms that are not externally audited. Moreover, the results show that the predictive power of the extended model increases more substantially for medium, small, and micro (financially constrained) firms than those of large firms.

4.3. Feature Importance

The decision tree-based mythologies (RF, AB, and GB) have the advantage of extracting important variables from the trained models, which can be valuable for understanding the variables related to default risk. Given that the GB method has outperformed the other selected methods for SMEs, we examine the feature importance of this method.[7]

[7] Additional feature importance analysis using the Random Forest (RF) and AdaBoost (AB) methods are provided in Appendix B. Additional Feature Importance Analysis.



Notes: The figures below show the feature importance for the variables used in the base model (left-hand side figures) and the extended model (right-hand side figures) when the gradient boosting (GB) method is used.

Figure 1. Feature Importance of the Gradient Boosting (GB) Method

Figure 1 provides the feature importance of the variables used in the base and extended models when the gradient boosting (GB) method is employed for large firms and SMEs. The figures for the base model (left-hand side figures) show that the feature importance of large, medium, and micro firms is similar, where the TL/TA has the highest contribution. However, for small firms, the variable with the highest contribution is RE/TA. When the five additional variables related to cash balance and financial constraints are included using the extended model (right-hand side figures), the variable with the highest contribution for large firms remains TL/TA, which is then distantly followed by RE/TA. Both these variables are those used in the base model and are not one of the additional variables in the extended model. Looking at extended model for medium firms, the variable with the highest contribution is also TL/TA and the variable with the second highest contribution is QA/CA, which is one of the additional variables related to cash balance.

However, the results for small and micro firms display different results from those of large and medium firms. For the small and micro firms, the variable with the highest contribution are the cash balance related variables. The variable with the highest contribution for small firms is QA/CA, which is distantly followed by RE/TA. Furthermore, the third and fourth largest contributions for small firms are QA/CL and SG, which are variables related to cash balance and financial constraints, respectively. This result suggests that the additional variables related to cash balance and financial constraints are important to the performance enhancement of the default prediction model for small firms. Turning to micro firms, the variable with the highest contribution is QA/CL followed by QA/CA, QA/TA, and SG. This suggests that the additional variables related to cash balance and financial constraints are important to the performance enhancement of the default prediction model for micro firms as well.

5. SUMMARY AND DISCUSSION

The vast literature on default prediction has primarily focused on creating and testing models using data on large firms, neglecting the fact that most established firms are SMEs. Furthermore, while a few prior default prediction studies have been conducted on externally audited SMEs, only a handful have included non-externally audited SMEs due to data access difficulties or reliability issues with regards to financial statements for non-externally audited firms. Using data provided by Korea Enterprise Data, our study differs from these prior studies given its focus on SMEs, including non-externally audited firms, and compares the performance of the default prediction model for firms based on the existence of an external audit and on total asset size.

Specifically, we split firms into four groups, namely large externally audited, medium externally audited, small non-externally audited, and micro non-externally audited firms and compare the default model performance for each group separately.

Our results show that the default probability is significantly lower for externally audited firms. Moreover, we show that predicting default based on the financial statements of non-externally audited firms is more challenging, underscoring the important role of the external auditor in ensuring the credibility and reliability of financial statements. To properly evaluate the default probability of Korean SMEs, our study incorporates well-known financial constraints and additional cash related variables for default prediction and analyzes the effects of these variables on firm default predictions. In addition, while traditional statistical methods, such as the discriminant and logistic regression methods, are still considered to be the standard for default prediction, machine learning methods have the potential to improve model performance. In fact, several prior applications of machine learning models in finance have successfully demonstrated its ability to enhance model performance. Furthermore, machine learning methods enable the use and incorporation of various firm characteristics that not only could help improve the performance of default prediction models but are also potentially essential for better capturing the differences between firms.

For the purposes of our study, we propose two default prediction models, the base and extended models, and employ six different classification methods, two traditional statistical and four machine learning methods. The results obtained from these models and classification methods are then used to compare the performance of the default prediction models using six evaluation metrics. Our base model adopts and modifies the five financial variables used in the Altman Z-score model (Altman, 1968) while our extended model adds five cash balance and financial constraint related variables to our base model. A comparison of model performance for the four different groups is then conducted to determine whether the extended model, with the additional cash balance and financial constraint related variables, and the machine learning classification methods improves default prediction model performance. Furthermore, we investigate variable feature importance in the models to identify variables that contributed more significantly to prediction improvement for each group.

The results obtained suggest that the addition of the five cash balance and financial constraint related variables could contribute significantly to improving Korean SMEs default prediction, particularly for small and micro non-externally audited firms. The highest average increase of G-mean was 0.058 and the highest average increase of AUC was 0.053 for the small non-externally audited group. The micro non-externally audited group had an average increase of 0.038 and 0.039 for G-mean and AUC, respectively. Moreover, the medium externally audited group had an average increase in G-mean and AUC of 0.037 and 0.034, respectively. However, the improvement in the prediction models for large firms was the lowest with the average increase in G-mean and AUC of only 0.025 for both metrics.

These results highlight the significant contribution that the use of machine learning methods and addition of the five selected cash balance and financial constraint-related variables have on improving the performance of the Korean SME default prediction models with little or no effect on large firms. Observation of the feature importance for

the GB method provides additional evidence that the selected cash balance and financial constraint related variables contributed to improving default prediction model performance for Korean SMEs, particularly the small and micro non-externally audited firms. Future studies should extend the analysis to different countries and use more or different financial variables in machine learning classification methods to enhance default prediction and default risk management of SMEs.

APPENDIX

Appendix A. Classification Methods

The six quantitative methods employed in our analysis include linear discriminant analysis (LDA), logistic regression (LR), random forest (RF), AdaBoost (AB), gradient boosting (GB), and multi-layer perceptron (MLP) for default prediction. Initial default prediction models used LDA and LR statistical methods (Altman, 1968; Ohlson, 1980). Although these methods have been widely used in academia and industry, newly developed machine learning approaches have gained popularity for default prediction as they are more accurate than the statistical methods.

First introduced by Breiman (2001), random forest (RF) is a method that uses an ensemble of decision trees. The method generates multiple versions of decision trees by selecting random samples to create new learning sets for aggregation of the results, using bagging and bootstrap aggregating methods. During the training time, this machine learning technique creates numerous decision trees based on randomly selected subsets, and classification is done by reporting the mean probability value from the trees. The classification result is then made from the predictions of multiple trees, resulting in more precise results.

Boosting is a technique that repeatedly generates a weak decision tree using training data and combines multiple decision trees to provide a single composite classifier to enhance the performance of the classification model. The adaptive boosting or AdaBoost (AB) method, introduced by Freund and Schapire (1996), is a popular boosting algorithm used for classification. The method works by iteratively applying a weak decision tree to the training data and assigning weights to each observation in the dataset based on the errors made by the previous iterations. Initially, all observations are assigned equal weights, and the algorithm generates a base decision tree on this weighted data. In each subsequent iteration, the algorithm increases the weights of the observations that were misclassified by the previous classifier and generates a new classifier on the updated weighted data. This process is repeated until the maximum number of iterations is reached. Finally, the algorithm combines the individual decision

trees into a single composite classifier that provides the final prediction based on a weighted voting scheme.

Gradient boosting (GB), introduced by Friedman (2001), is a boosting algorithm that combines multiple weak decision trees to improve classification accuracy. It creates a sequence of decision trees that fit to the residuals of the previous tree, with each tree correcting the errors of the previous one. The final classifier is obtained by combining the predictions of all the trees. Unlike AdaBoost, which updates the weights of the observations, gradient boosting uses an additive regression tree approach.

Multi-layer perceptron (MLP) is a type of artificial neural network that consists of multiple layers of nodes, including an input layer, one or more hidden layers, and an output layer. The nodes in each layer are connected to the nodes in the next layer through weighted connections. MLP is a feedforward neural network, which means that the signal flows only in one direction, from the input layer through the hidden layers to the output layer. In this study, we use a model with one hidden layer since it has demonstrated successful performance in previous studies (Guliyev and Ismailov, 2016). The MLP method estimates the connectivity weight of each neuron and computes the probability of default for each firm.

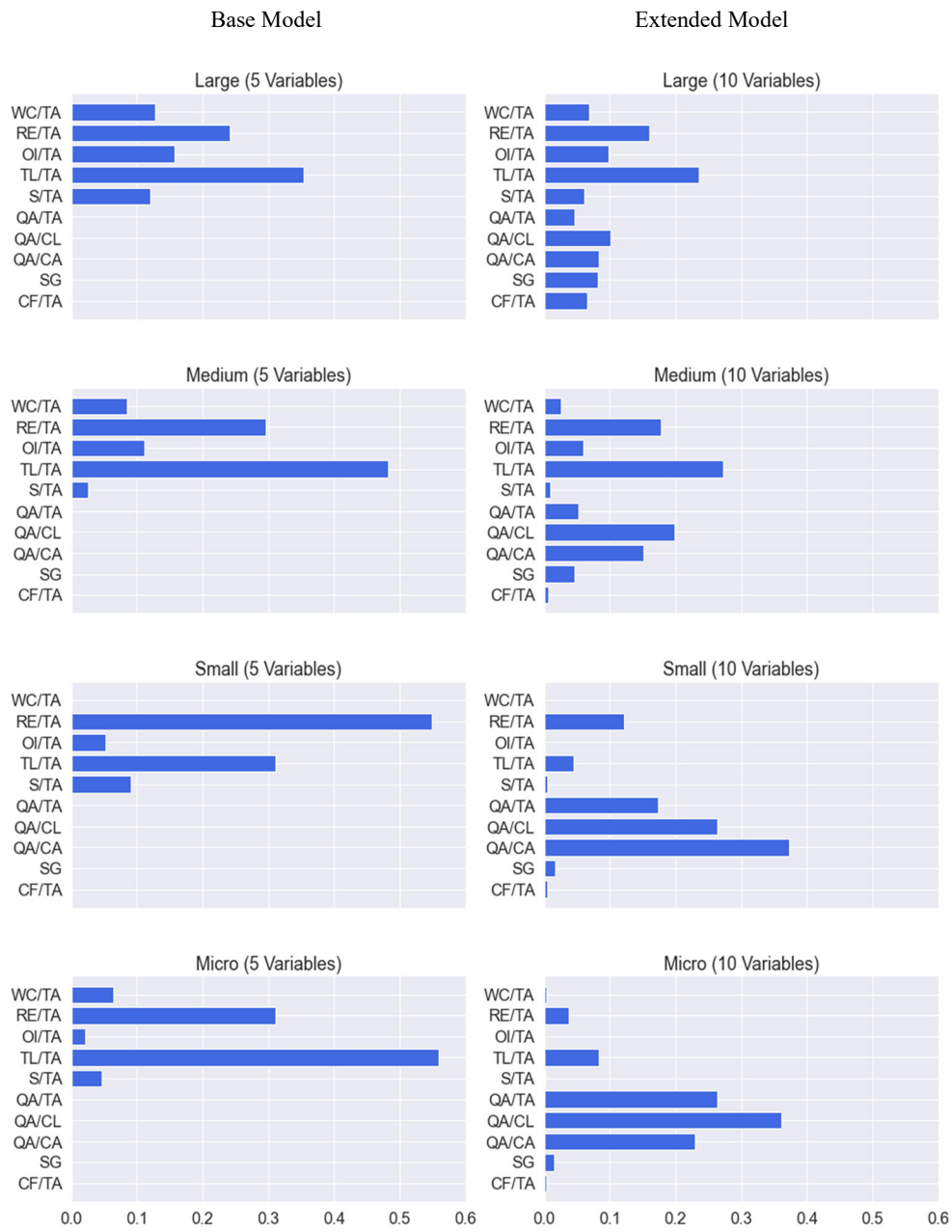
The decision tree-based machine learning methods used in our study can provide measurement of feature importance. The importance score is typically calculated based on how much each feature contributes to reducing the model's loss function or improving its performance. In the case of an ensemble of decision trees, feature importance can be measured by calculating the average of the importance scores assigned by each decision tree in the ensemble.

Appendix B. Additional Feature Importance Analysis

Figures A1 and A2 show the feature importance of the variables used in the base and extended models when the random forest (RF) and AdaBoost (AB) methods are employed.

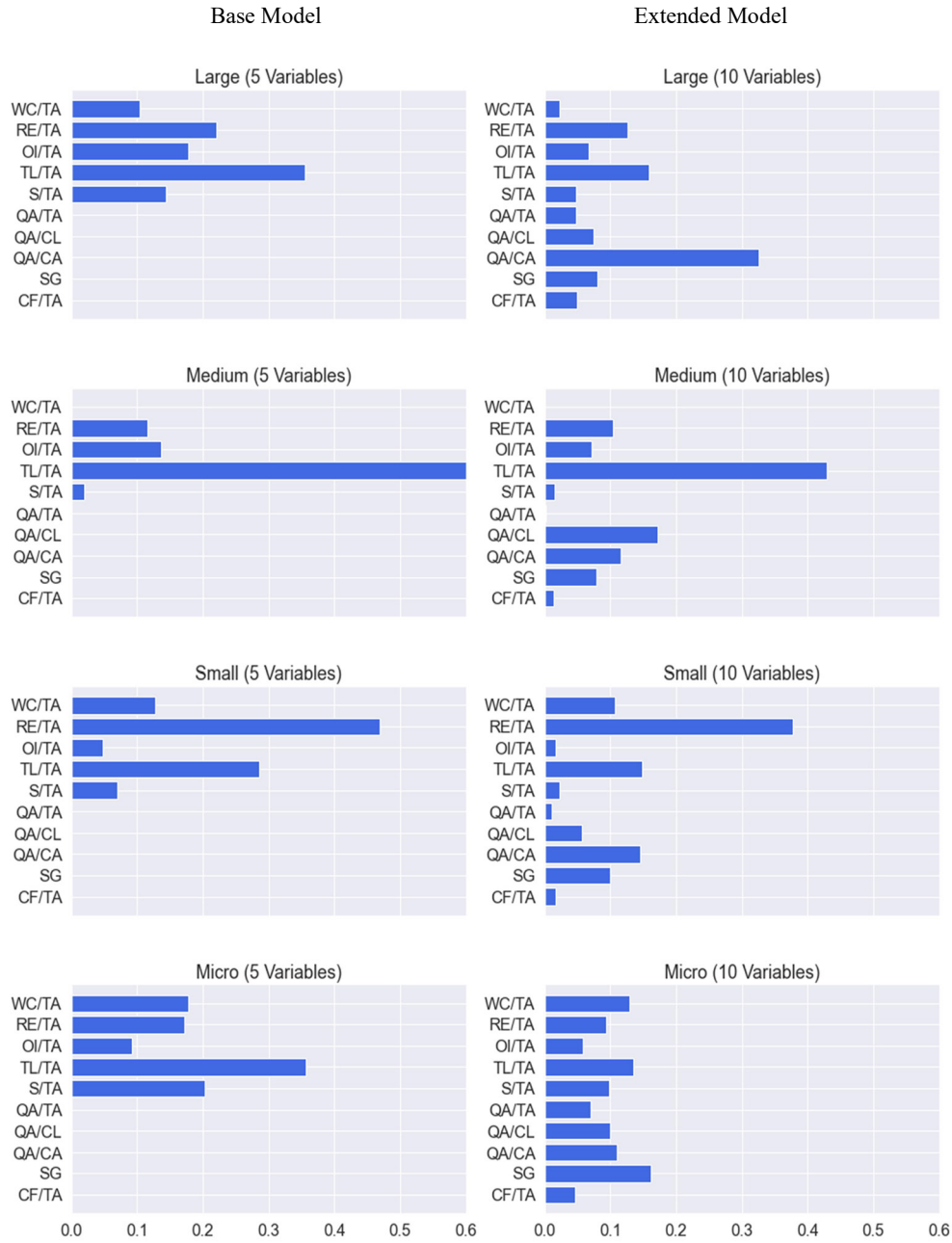
The feature importance obtained using the base model for both the RF and AB methods (left-hand side figures) are quite similar to the results obtained for the GB method, where the variable with the highest contribution is TL/TA for large, medium, and micro firms and is RE/TA for small firms.

However, the results for the RF method show a more pronounced difference in the contribution of the five additional variables to the classification in the extended model (Figure A1, right-hand side figures). Although, the variables with the highest contribution in the RF method are identical to the GB method, the contribution of all three cash balance related variables, QA/TA, QA/CA and QA/CL, are the highest for both small and micro firms. This shows that the five variables related to cash balance and financial constraints are important to the performance enhancement of default prediction models for small and micro firms when the RF method is used.



Notes: The figures below show the feature importance for the variables used in the base model (left-hand side figures) and the extended model (right-hand side figures) when the random forest (RF) method is used.

Figure A1. Feature Importance of the Random Forest (RF) Method



Notes: The figures below show the feature importance for the variables used in the base model (left-hand side figures) and the extended model (right-hand side figures) when the AdaBoost (AB) method is used.

Figure A2. Feature Importance of the AdaBoost (AB) Method

As for the results when the AB method is used for the extended model (Figure A2, right-hand side figures), the variables with the highest contribution are different from those obtained using the GB and RF methods for all except the medium firms. For the AB method, the variable with the largest contribution for the large, small, and micro firms are QA/CA, RE/TA, and SG respectively.

REFERENCES

Acharya, V., S.A. Davydenko and I.A. Strebulaev (2012), "Cash Holdings and Credit Risk," *The Review of Financial Studies*, 25(12), 3572-3609.

Ahn, H. and K.J. Kim (2009), "Bankruptcy Prediction Modeling with Hybrid Case-Based Reasoning and Genetic Algorithms Approach," *Applied Soft Computing*, 9(2), 599-607.

Altman, E.I. (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *The Journal of Finance*, 23(4), 589-609.

Altman, E.I. (2018), "A Fifty-Year Retrospective on Credit Risk Models, the Altman Z-Score Family of Models and Their Applications to Financial Markets and Managerial Strategies," *Journal of Credit Risk*, 14(4), 1-34.

Altman, E.I., Y.H. Eom and D.W. Kim (1995), "Failure Prediction: Evidence from Korea," *Journal of International Financial Management and Accounting*, 6(3), 230-249.

Altman, E.I., E. Hotchkiss and W. Wang (2019), *Corporate Financial Distress, Restructuring, and Bankruptcy: Analyze Leveraged Finance, Distressed Debt, and Bankruptcy*, John Wiley and Sons, Hoboken, New Jersey.

Altman, E.I. and G. Sabato (2013), "Modeling Credit Risk for SMEs: Evidence from the US Market," in *Managing and Measuring Risk: Emerging Global Standards and Regulations After the Financial Crisis*.

Barboza, F., H. Kimura and E. Altman (2017), "Machine Learning Models and Bankruptcy Prediction," *Expert Systems with Applications*, 83, 405-417.

Breiman, L. (2001), "Random Forests," *Machine Learning*, 45(1), 5-32.

Che, L., O. Hope and J.C. Langli (2020), "How Big-4 Firms Improve Audit Quality," *Management Science*, 66(10), 4552-4572.

Ciampi, F. (2015), "Corporate Governance Characteristics and Default Prediction Modeling for Small Enterprises: An Empirical Analysis of Italian Firms," *Journal of Business Research*, 68(5), 1012-1025.

Ciampi, F. and N. Gordini (2013), "Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises," *Journal of Small Business Management*, 51(1), 23-45.

Denis, D.J. and V. Sibilkov (2010), "Financial Constraints, Investment, and the Value of Cash Holdings," *The Review of Financial Studies*, 23(1), 247-269.

Freund, Y. and R.E. Schapire (1996), "Experiments with a New Boosting Algorithm," *International Conference on Machine Learning*, 96, 148-156.

Friedman, J.H. (2001), "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, 29(5), 1189-1232.

Gertler, M. and S. Gilchrist (1994), "Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms," *The Quarterly Journal of Economics*, 109(2), 309-340.

Gordini, N. (2014), "A Genetic Algorithm Approach for SMEs Bankruptcy Prediction: Empirical Evidence from Italy," *Expert Systems with Applications*, 41(14), 6433-6445.

Gregory, G., C. Harvie and H.H. Lee (2002), "Korean SMEs in the Wake of the Financial Crisis: Strategies, Constraints, and Performance in a Global Economy," Working Paper 02-12, Department of Economics, University of Wollongong.

Guliyev, N.J. and V.E. Ismailov (2016), "A Single Hidden Layer Feedforward Network with Only One Neuron in the Hidden Layer Can Approximate Any Univariate Function," *Neural Computation*, 28(7), 1289-1304.

Gupta, J., M. Barzotto and A. Khorasgani (2018), "Does Size Matter in Predicting SMEs Failure?" *International Journal of Finance and Economics*, 23(4), 571-605.

Jeong, C., J.H. Min and M.S. Kim (2012), "A Tuning Method for the Architecture of Neural Network Models Incorporating GAM and GA as Applied to Bankruptcy Prediction," *Expert Systems with Applications*, 39(3), 3650-3658.

Jo, H., I. Han and H. Lee (1997), "Bankruptcy Prediction Using Case-Based Reasoning, Neural Networks, and Discriminant Analysis," *Expert Systems with Applications*, 13(2), 97-108.

Kim, H.J., N.O. Jo and K.S. Shin (2016), "Optimization of Cluster-Based Evolutionary Undersampling for the Artificial Neural Networks in Corporate Bankruptcy Prediction," *Expert Systems with Applications*, 59, 226-234.

Kim, J., D.A. Simunic, M.T. Stein and C.H. Yi (2011), "Voluntary Audits and the Cost of Debt Capital for Privately Held Firms: Korean Evidence," *Contemporary Accounting Research*, 28(2), 585-615.

Kim, M.J. and D.K. Kang (2010), "Ensemble with Neural Networks for Bankruptcy Prediction," *Expert Systems with Applications*, 37(4), 3373-3379.

Kim, M.J. and D.K. Kang (2012), "Classifiers Selection in Ensembles Using Genetic Algorithms for Bankruptcy Prediction," *Expert Systems with Applications*, 39(10), 9308-9314.

Kim, S. and H.D. Yi (2012), "The Effects of Business Cycle on the Credit Rating Model for Unlisted SMEs in Korea," *Korean Accounting Journal*, 21(6), 287-323.

Lee, K.C., I. Han and Y. Kwon (1996), "Hybrid Neural Network Models for Bankruptcy Predictions," *Decision Support Systems*, 18(1), 63-72.

Lin, W.C., C.F. Tsai, Y.H. Hu and J.S. Jhang (2017), "Clustering-Based Undersampling

in Class-Imbalanced Data," *Information Sciences*, 409, 17-26.
McFadden, D. (1974), "The Measurement of Urban Travel Demand," *Journal of Public Economics*, 3(4), 303-328.
Martínez-Sola, C., P.J. García-Teruel and P. Martínez-Solano (2018), "Cash Holdings in SMEs: Speed of Adjustment, Growth and Financing," *Small Business Economics*, 51(4), 823-842.
Ohlson, J.A. (1980), "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, 18(1), 109-131.
Son, H., C. Hyun, D. Phan and H.J. Hwang (2019), "Data Analytic Approach for Bankruptcy Prediction," *Expert Systems with Applications*, 138, 112816.
Spathis, C.T. (2002), "Detecting False Financial Statements Using Published Data: Some Evidence from Greece," *Managerial Auditing Journal*, 17(4), 179-191.
Tobback, E., T. Bellotti, J. Moeyersoms, M. Stankova and D. Martens (2017), "Bankruptcy Prediction for SMEs Using Relational Data," *Decision Support Systems*, 102, 69-81.
Veganzones, D. and E. Séverin (2018), "An Investigation of Bankruptcy Prediction in Imbalanced Datasets," *Decision Support Systems*, 112, 111-124.
Whited, T.M. and G. Wu (2006), "Financial Constraints Risk," *The Review of Financial Studies*, 19(2), 531-559.
Xu, Y., G. Kou, Y. Peng and F.E. Alsaadi (2019), "Bankruptcy Forecasting for Small and Medium-Sized Enterprises Using Cash Flow Data," in *International Conference on Data Science*, 477-487, Springer, Singapore.
Yoo, S.K., S.Y. Choi, H.L. Yim, J.H. Kim and J.J. Kim (2011), "Liquidity Crisis Factor Analysis of the General Contractor under Global Financial Crisis in Korea–Focused on Financing Structure by the Company's Scale," *Engineering and Technology*.
Zelenkov, Y. and N. Volodarskiy (2021), "Bankruptcy Prediction on the Base of the Unbalanced Data Using Multi-Objective Selection of Classifiers," *Expert Systems with Applications*, 185, 115559.
Zhou, L. (2013), "Performance of Corporate Bankruptcy Prediction Models on Imbalanced Dataset: The Effect of Sampling Methods," *Knowledge-Based Systems*, 41, 16-25.

Mailing Address: Dojoon Park, Kongju National University, Division of International Studies, 56 Gongjudeahak-ro, Gongju-si, Chungcheongnam-do 32588, Korea, Email: dojoon@kongju.ac.kr

Received May 20, 2025, Accepted June 18, 2026.