Hindawi
Mathematical Problems in Engineering
Volume 2019, Article ID 3085247, 12 pages
https://doi.org/10.1155/2019/3085247

*Research Article*

# A Novel Soft Ensemble Model for Financial Distress Prediction with Different Sample Sizes

**Wei Xu**,[1] **Hongyong Fu**,[2,3] **and Yuchen Pan**[2]

[1] *School of Business, Jiangnan University, Jiangsu Wuxi 214122, China*
[2] *China Research Institute of Enterprise Governed by Law, Southwest University of Political Science and Law, Chongqing 401120, China*
[3] *School of Electrical Engineering, Computing and Mathematical Science, Curtin University, 6845 Perth, Australia*

Correspondence should be addressed to Hongyong Fu; fuhongyong@foxmail.com

Received 18 December 2018; Revised 27 February 2019; Accepted 18 March 2019; Published 4 April 2019

Academic Editor: Kemal Polat

This work presents a novel soft ensemble model (ANSEM) for financial distress prediction with different sample sizes. It integrates qualitative classifiers (expert system method, ES) and quantitative classifiers (convolutional neural network, CNN) based on the uni-int decision making method of soft set theory (UI). We introduce internet searches indices as new variables for financial distress prediction. By constructing a soft set representation of each classifier and then using the optimal decision on soft sets to identify the financial status of firms, ANSEM inherits advantages of ES, CNN, and UI. Empirical experiments with the real data set of Chinese listed firms demonstrate that the proposed ANSEM has superior predicting performance for financial distress on accuracy and stability with different sample sizes. Further discussions also show that internet searches indices can offer additional information to improve predicting performance.

## 1. Introduction

Financial distress prediction, which has been used to identify the financial status of firms in the future, is an essential work in helping investors assess their investment risks [1]. It is a good practical tool for distinguishing firms in financial distress from the healthy ones [2]. Generally, it is believed that symptoms of financial distress can be perceived before encountering a failure or crisis [3]. Many researchers and practitioners have worked on this subject with great interest over decades [4]. However, financial distress prediction practice is still a major challenge for its complexity and rapid variations, especially under the urge of big data. There are three tasks involved in predicting financial distress, as shown in Figure 1.

First of all, we need to analyze the research object. Financial distress prediction of different objects may take different methods for the heterogeneity. This is why there are various literatures focused on some specific fields, such as American firms [5, 6], Chinese firms [7, 8], and so on [9]. In this paper, we are interested in the financial distress prediction of Chinese listed firms from the Shanghai Stock Exchange and Shenzhen Stock Exchange.

Secondly, we should select optimal variables for financial distress prediction according to the object. Financial ratios have been widely used in prior literatures since Beaver [10] first adopted 6 financial ratios for bankruptcy prediction [11]. As we all know, financial ratios are historical data [12]. It is hard to reflect the finance difference of firms timely. Especially, when the operation environment changed quickly, the role of financial ratios on predicting financial distress will probably diminish [13]. Some literatures have introduced nonfinancial variables to improve the predicting performance, including country characteristics [14] and industry factors [15]. Though internet searches indices have been proved to be important variables for predicting under the era of big data [16], they have been rarely involved in financial distress prediction. Therefore, here we try to expand variables for financial distress prediction by integrating internet searches indices and other predicting variables.

Finally, we focus on the predicting model for financial distress according to the object and variables. Predicting models,

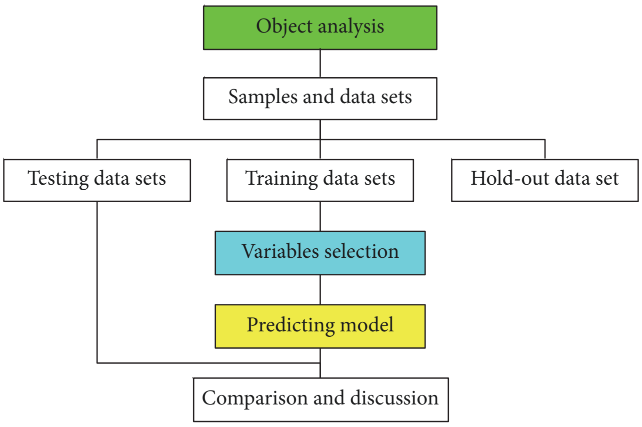


Figure 1: The process of financial distress prediction.

especially individual classifiers, have significant effect on predicting performance [17]. It motives researchers to explore predicting methods. Various statistical methods have been proposed for financial distress prediction in prior literatures, including discriminant analysis (DA) [5] and logistic regression (LR) [6] [12]. However, statistical models have disadvantages on the stringent model assumptions. After that, artificial intelligence methods have been widely applied to predict financial distress, including case-based reasoning (CBR) [18], neural networks (NN) [19], genetic algorithm (GA) [20], rough set (RS) [21], decision tree [22], and support vector machine (SVM) [23] [22]. Each individual method has advantages and disadvantages in predicting financial distress [24]. Therefore, more and more researchers have tried to employ individual methods as classifiers to construct ensemble models for financial distress prediction [25]. The main purpose is to improve predicting performance by taking full advantages of classifiers and in the same time minimizing their disadvantages. A lot of ensemble models have been proposed for financial distress prediction [25, 26]. Recently, Alaka et al. [27] investigated a systematic review of financial distress prediction models. However, there are still some deficiencies to be improved for distress prediction practice.

(1) Those models above may obtain a satisfactory performance with large sample sizes and financial ratios. However, in failure predicting practice, especially under the urge of big data, predicting financial distress with different sample sizes and other nonfinancial variables becomes increasingly frequent.

(2) For most ensemble models, it is a key point to compute the weight of individual classifiers. Meanwhile, it is still a world widely challenge task since Bates [28] first discussed the ensemble forecasting method.

(3) With classifiers increase, in most ensemble predicting models may exist serious overfitting issue and poor generalization issue. It is a contradiction that more classifiers will bring additional information to improve predicting performance [4].

(4) Qualitative methods have been rarely employed to construct ensemble predicting models. Meanwhile, some literatures have found out that qualitative approaches can play a valid role in predicting financial distress [29].

Therefore, to address the real practice circumstances, the aim of this paper is to improve a novel soft ensemble model (ANSEM) for financial distress prediction with different sample sizes. Both qualitative methods and quantitative methods are employed as classifiers to take full use of their advantages. We choose the expert system method (ES) as the qualitative classifier for its advantages [29]. For the same reason, convolutional neural network (CNN) is employed as the quantitative classifier [19]. Then the novel uni-int decision making method of soft set theory (UI), initiated by Çağman and Enginoğlu [30], is applied to integrate forecasts of each classifier. The UI has been proved theoretically as a superior nonparametric method for dealing with high dimensional and different sample sizes data [31]. In such a way, the ANSEM inherits the efficiency and flexibility of UI and takes advantages of ES and CNN at the same time.

For performance comparison, individual ES, CNN, the ensemble models with ES and CNN based on equal weights (EMEW), convolutional neural network (EMNN), rough set theory and Dempster-Shafer evidence theory (EMRD) [32] are comparative models included in this work. To demonstrate effects of different sample sizes on predicting performance of each model, we divide all real sample data from Chinese listed firms into the training data set and the testing data set randomly for percentages (25%, 75%), (50%, 50%), and (75%, 25%).

The remainder of this paper is organized as follows. Section 2 briefly reviews the classical uni-int decision making method of soft set theory and mainly introduces the proposed soft ensemble model for financial distress prediction. Section 3 provides an empirical experiment with real data sets from Chinese listed firms. Section 4 presents the empirical results and makes a comparison and discussion. In Section 5 we conclude this paper and discuss further research.

## 2. The Soft Ensemble Predicting Model

*2.1. The Uni-Int Decision Making Method of Soft Set Theory.* Soft set theory, originated by Molodtsov [33], is a novel mathematic theory for uncertain information [34]. Assuming $U$ is an initial universe of objects, $E$ is a parameters set to objects, $P(U)$ is the power set of $U$, and $A$ is a parameters subset of $E$ ($A \subseteq E$). A soft set $F_A$ can be defined by the set of ordered pairs [35], shown as

$$F_A = \{(x, f_A(x)) : x \in E, f_A(x) \in P(U)\} \tag{1}$$

where $f_A : E \longrightarrow P(U)$ such that $f_A(x) = \emptyset$ if $x \notin A$. $f_A$ is an approximate function of $F_A$.

Based on the definition above, Çağman and Enginoğlu [30] redefined the product operation of soft sets as the binary operation to take full information of soft sets as follows. Assuming $\wedge(U)$ is the set of all $\wedge$ products (and products) of soft sets over $U$. If $F_A \wedge F_B \in \wedge(U)$, the uni-int operation denoted by $uni_x int_y$ and $uni_y int_x$ are defined, respectively, as

$$
\begin{gathered}
uni_x int_y \colon \wedge(U) \longrightarrow P(U), \\
uni_x int_y (F_A \wedge F_B) = \bigcup_{x \in A} \left( \bigcap_{y \in B} (f_{A \wedge B}(x, y)) \right)
\end{gathered} \quad (2)
$$

$$
\begin{gathered}
uni_y int_x \colon \wedge(U) \longrightarrow P(U), \\
uni_y int_x (F_A \wedge F_B) = \bigcup_{y \in B} \left( \bigcap_{x \in A} (f_{A \wedge B}(x, y)) \right)
\end{gathered} \quad (3)
$$

$F_A \wedge F_B$ is a new soft set defined by the function $f_{A \wedge B} : A \times B \longrightarrow P(U)$, $f_{A \wedge B}(x, y) = f_A(x) \cap f_B(y)$. The uni-int decision set is the union of two uni-int operation sets, as

$$
\begin{aligned}
uni - int (F_A \wedge F_B) = {} & uni_x int_y (F_A \wedge F_B) \\
& \cup uni_y int_x (F_A \wedge F_B)
\end{aligned} \quad (4)
$$

The uni-int decision making method of soft set theory (UI) is an effective integrated tool to exploit information of soft sets [11]. However, researchers working in this field mainly focused on theoretical researches. It is rarely applied to practice. This paper contributes to UI by filling this gap. We take UI as a novel integrated method for financial distress prediction to achieve a better performance.

*2.2. The Soft Ensemble Predicting Model.* Assume there are $n$ $(n = (1, \ldots, N))$ samples and $m$ $(m = (1, \ldots, M)$ predicting classifiers. $Y$ is the original status matrix of samples. $Y^m$ is the individual predicting results matrix of the *mth* classifier about the *nth* samples, $Y^u$ is the integrated results matrix of $y_{nm}$ using UI. $Y$, $Y^m$, and $Y^u$ are defined as

$$
\begin{gathered}
Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \\ \vdots \\ y_N \end{bmatrix}, \\
Y^m = \begin{bmatrix} y_{11} & \cdots & y_{1m} & \cdots & y_{1M} \\ \vdots & & & & \vdots \\ y_{n1} & \ddots & y_{nm} & \ddots & y_{nM} \\ \vdots & & & & \vdots \\ y_{N1} & \cdots & y_{Nm} & \cdots & y_{NM} \end{bmatrix}, \\
Y^u = \begin{bmatrix} y_{1u} \\ \vdots \\ y_{nu} \\ \vdots \\ y_{Nu} \end{bmatrix}
\end{gathered} \quad (5)
$$

where

$$
\begin{gathered}
y_n = \begin{cases} 1 \\ 0, \end{cases} \\
y_{nm} = \begin{cases} 1 \\ 0, \end{cases} \\
y_{nu} = \begin{cases} 1 \\ 0 \end{cases}
\end{gathered} \quad (6)
$$

$y_n = 0$ represents the *nth* sample is in the actually financial distress status; $y_n = 1$ means the *nth* sample is in the actually financial normal status. $y_{nm} = 0$ represents the *mth* predicting classifier predicts the *nth* sample will be in the financial distress status; $y_{nm} = 1$ means the *mth* predicting classifier predicts the *nth* sample will be in the financial normal status. $y_{nu}$ is the integrated result using UI for the *nth* sample. $y_{nu} = 0$ represents the *nth* sample will be in the financial distress status; $y_{nu} = 1$ means the *nth* sample will be in the financial normal status. A concise illustration of the novel soft ensemble model (ANSEM) for financial distress prediction can be shown as in Figure 2. Obviously, three key points are involved in constructing the novel soft ensemble model: the individual classifier, the integrated method, and the algorithm.

*2.2.1. Individual Predicting Classifiers.* Individual predicting classifiers play a significant role in performance of ensemble predicting models [28]. As mentioned above, many qualitative methods and quantitative methods, which can be used as classifiers, have been proposed for financial distress prediction in prior literatures [36]. While each classifier has advantages and disadvantages, according to the research object and variables, we need to select some appropriate classifiers from them. On one side, we want to employ more classifiers as components of ANSEM to take advantages of them. On the other side, it is a contradictory because the complexity of ANSEM will be a seriously problem. The computing power will decrease. According to prior literatures, two classifiers may bring a nice balance of the performance and the complexity [8]. We also want to employ both qualitative classifiers and quantitative classifiers to construct ANSEM. Therefore, expert system method (ES) and convolutional neural network (CNN) are employed as individual classifiers for their advantages. ES is a good qualitative classifier for financial distress prediction [29]; CNN is an excellent quantitative classifier too [37].

*2.2.2. Integrated Method.* The proposed ANSEM for financial distress prediction can be redefined as

$$
y_{nu} = g(y_{nm}) \quad (7)
$$

where $g(\cdot)$ is the integrated method. To overcome disadvantages of ensemble models on measuring weight coefficients, here we take the novel uni-int decision making method of

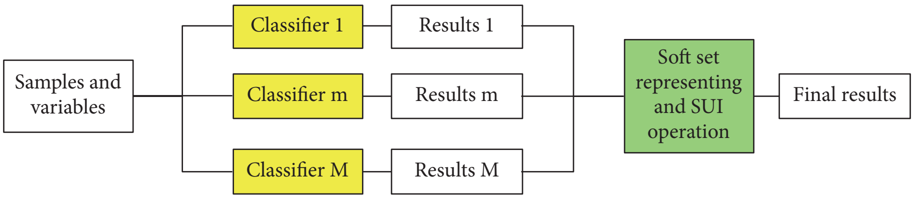


Figure 2: The principle of ANSEM. Three key points for the novel soft ensemble model: the individual classifier, the integrated method, and the algorithm.

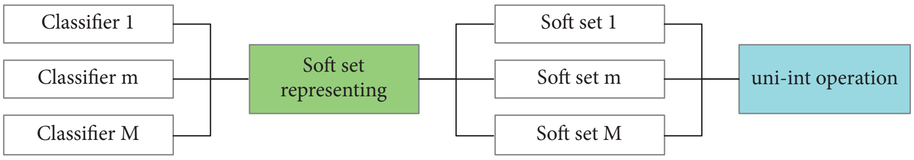


Figure 3: The integration process of ANSEM. There are two key points in the integration process: soft set representation and uni-int operation.

soft set theory (UI) as the integrated method. Soft set theory is a parameterized family of subsets of the universe $U$ [33]. The increase of $E$ will let UI perform better. That means that contradictions of the complexity and the performance can be well solved. It also does not need measure the weight coefficient of each individual classifier. Therefore, we use UI as the integrated method.

However, the UI is an operation rule of soft sets. First of all, we need to represent the process of each predicting classifier in the style of soft set theory. Each classifier translates into a typical soft set. Then we are able to run the uni-int operation and find the uni-int decision set. The integration process can be briefly showed as in Figure 3. Details are clearly illustrated as follows.

*Step 1.* We can take the samples set as the nonempty initial universe $U$ of soft sets. The variable set can be treated as the parameters set $E$ to objects of $U$.

*Step 2.* The approximate mapping function $f$ of soft sets can be each classifier for financial distress prediction. Then we get $m$ (the number of individual classifiers) different soft sets, shown as

$$F_m = \{(x, f_m(x)) : x \in E_m, f_m(x) \in P(U)\} \tag{8}$$

where $U$ is the set of samples, $E_m$ is the set of variables for the $mth$ classifier, $x$ is a variable of $E$, and $f_m$ is each classifier for financial distress prediction. Here, the parameter set $E_m$ $(m = 1, 2, \ldots, M)$ may include the same or different variables.

*Step 3.* Based on soft sets $F_m$, we can run the uni-int operation and find the uni-int decision set using (2)–(4). The uni-int decision set is the finally predicting results.

*2.2.3. The Algorithm of ANSEM.* The algorithm of ANSEM for financial distress prediction can be briefly shown as in Figure 4. Each step is clearly illustrated as follows.

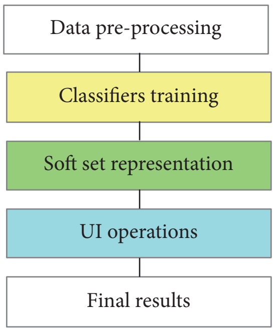


Figure 4: The algorithm of ANSEM.

*Step 1* (data preprocessing). The measuring unit of variables may be different. We need to normalize all collected data to decrease the difference at first. The function is showed as

$$x'_{ni} = \frac{x_{ni} - \min_i}{\max_i - \min_i} \tag{9}$$

where $x_{ni}$ is the $ith$ variable's original value for the $nth$ sample, and $\min_i$, $\max_i$ are the minimal value and maximal value of the $ith$ variables for all samples, respectively.

*Step 2* (classifiers training). For expert system method, unlike previous literatures, we use real financial institutions as experts, especially the security firms in China. It is a common feature for those institutions to issue research reports periodically to give investment recommendations for listed firms' stock, such as "buy", "keep" or "sell". "Buy" means the firm may have a good financial development in future and bring investors an impressive return, in the institution's opinion. "Keep" presents the opinion that the firm may not suffer financial distress in future and bring investors a normal return. "Sell" means that the firm may suffer financial distress in future and bring investors a loss. Recommendations are made based on professional analysis. It is useful to improve predicting performance.

To increase the reliability, we employ $j\ (j = 1, \ldots, J)$ financial institutions as experts. $J$ is an odd number and is bigger than 1. Predicting results can be shown as

$$Y^e = \begin{bmatrix} y_1^J \\ \vdots \\ y_n^J \\ \vdots \\ y_N^J \end{bmatrix} = \begin{bmatrix} y_{11} & \cdots & y_{1j} & \cdots & y_{1J} \\ \vdots & & \vdots & & \vdots \\ y_{n1} & \ddots & y_{nj} & \ddots & y_{nJ} \\ \vdots & & \vdots & & \vdots \\ y_{N1} & \cdots & y_{Nj} & \cdots & y_{NJ} \end{bmatrix} \quad (10)$$

where $y_{nj}$ is the latest recommendation of the $jth$ expert on the $nth$ sample. $y_n^J$ is the final predicting result about the $nth$ sample. $Y^e$ is the predicting results matrix. $y_{nj}$ equals to "buy", "keep" or "sell". Here we code "buy" as 1, "keep" as 0, and "sell" as -1. In other words, $y_n^J$ can be defined as

$$y_n^J = \begin{cases} 1, & \sum_{j=1}^{J} y_{nj} \geq 0 \\ 0, & \sum_{j=1}^{J} y_{nj} < 0 \end{cases} \quad (11)$$

That means that if more institutions recommend buying than selling, the expert system method predicts the firm will not suffer the financial distress.

For convolutional neural network, it is a key point to choose an appropriate structure [19]. Here, according to prior literatures, we employ the basic convolutional neural network as CNN algorithm. One can refer the literature of Hosaka [24] for details about CNN.

*Step 3* (soft set representing). As demonstrated in Section 2.2.2, we obtain two soft sets: $F_{es}$. and $F_{nn}$. $F_{es}$ is the soft set of ES. $F_{nn}$ is the soft set of CNN. The $U$ is the same initial universe set of samples for both $F_{es}$ and $F_{nn}$. $E_{es}$ and $E_{nn}$ are parameter sets for $F_{es}$ and $F_{nn}$, respectively. They may include the same or different variables. $E_{es}$, $E_{nn} \subseteq E$. $F_{es}$ and $F_{nn}$ can be shown as

$$F_{es} = \{(x, f_{es}(x)) : x \in E_{es}, f_{es}(x) \in P(U)\} \quad (12)$$

$$F_{nn} = \{(x, f_{nn}(x')) : x' \in E_{nn}, f_{nn}(x') \in P(U)\} \quad (13)$$

*Step 4* (uni-int operations). Based on soft sets $F_{es}$ and $F_{nn}$, (2) and (3) can be represented as (14) and (15).

$$\begin{aligned} & uni_x int_{x'} \colon \wedge(U) \longrightarrow P(U), \\ & uni_x int_{x'}(F_{es} \wedge F_{nn}) = \bigcup_{x \in E_{es}} \left( \bigcap_{x' \in E_{nn}} (f_{es \wedge nn}(x, x')) \right) \end{aligned} \quad (14)$$

$$\begin{aligned} & uni_{x'} int_{x} \colon \wedge(U) \longrightarrow P(U), \\ & uni_{x'} int_{x}(F_{es} \wedge F_{nn}) = \bigcup_{x' \in E_{nn}} \left( \bigcap_{x \in E_{es}} (f_{es \wedge nn}(x, x')) \right) \end{aligned} \quad (15)$$

The uni-int decision set can be represented as

$$\begin{aligned} uni - int(F_{es} \wedge F_{nn}) = {} & uni_x int_{x'}(F_{es} \wedge F_{nn}) \\ & \cup uni_{x'} int_x (F_{es} \wedge F_{nn}) \end{aligned} \quad (16)$$

*Step 5* (final results). Applying ANSEM to real data, we can obtain final predicting results.

## 3. Empirical Experiment

*3.1. Samples and Data Sets.* In China, listed firms are divided into two categories in practice. One is the Specially Treated (ST) group. The other is the Not Specially Treated (NST) group. The benchmark is either the negative net profit in recent two years or published financial reports with serious misstatements. In this paper, we take ST listed firms as financial distress samples and NST listed firms as financial normal samples. We randomly selected 100 ST listed firms and 100 NST listed firms in the seven-year period 2011-2017. It means that there are 50 training samples and 150 testing samples for the percentage (25%, 75%). It is the same mean for percentages (50%, 50%) and (75%, 25%). Obviously, we can observe the change of predicting performance with different sample sizes. Data of listed firms can be collected from the CSMAR solution.

For expert system method, the data set is collected from Chinese financial institutions in the seven-year period 2011-2017. There are more than 150 professional institutions that have recommended or are recommending for investments. According to the sustainability in recommendation, we randomly employed 3 ($j = 3$) institutions as experts with similar backgrounds and abilities. They are "HAITONG Securities", "CITIC Securities", and "SHENWAN&HONGYUAN Securities". We can download their research reports from the "WIND" database. Furthermore, each expert may publish two or more reports in a year. We choose the latest report as the outcome of experts.

*3.2. Variables Selection.* A lot of variables have been proposed for financial distress prediction. The widely popular variables for financial distress prediction are listed in Table 1 [6]. To get the optimal traditional variables, the stepwise logistic regression is applied to the training data set of the year $(t-1)$ to select variables from Table 1. We obtain 6 variables: $x_4$, $x_6$, $x_7$, $x_{11}$, $x_{14}$, and $x_{15}$.

Besides, we are trying to expand predicting variables in following two ways. First of all, we introduce internet searches indices of samples for financial distress prediction. Since January 1, 2011, we can collect the daily, weekly, monthly, and annual searches data for sample's name from "Baidu" website. Many literatures have demonstrated the significance of internet searches data on forecasting. We mark the annual searches data for sample's name as $x_{19}$. Secondly, we use the recommendations as the outcomes of the expert system method. Thus, we employ the variable "recommendation" for expert system method and mark it as $x_{20}$. All variables selected for ANSEM are listed in Table 2.

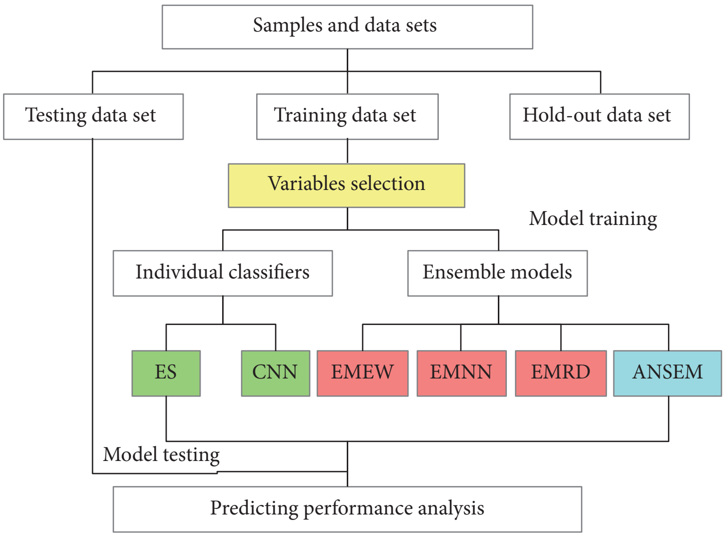


Figure 5: The framework of the empirical experiment. ES is the expert system method. CNN is the convolutional neural network. EMEW is the ensemble model based on equal weight. EMNN is the ensemble model based on the convolutional neural network. EMRD is the ensemble model based on the rough set theory and evidence theory. ANSEM is the soft ensemble model based on the uni-int decision making method.

Table 1: The most popular variables for financial distress prediction.

| No. | Variables | No. | Variables |
|---|---|---|---|
| $x_1$ | Current ratio | $x_2$ | Cash flow / total assets |
| $x_3$ | Cash flow / total debt | $x_4$ | Cash flow / sales |
| $x_5$ | Debt ratio | $x_6$ | Market value equity / total debt |
| $x_7$ | Working capital / total asset | $x_8$ | Working capital / sales |
| $x_9$ | Quick asset / total asset | $x_{10}$ | Quick asset / sales |
| $x_{11}$ | Current debt / sales | $x_{12}$ | Current assets / total asset |
| $x_{13}$ | No-credit interval | $x_{14}$ | Net income / total asset |
| $x_{15}$ | Retained earnings / total asset | $x_{16}$ | Sales / total asset |
| $x_{17}$ | log(total assets / GNP price-level index) | | |
| $x_{18}$ | Earnings before interest and taxes / total asset | | |

Table 2: All variables selected for ANSEM.

| No. | Variables | No. | Variables |
|---|---|---|---|
| $x_7$ | Working capital/ total asset | $x_4$ | Cash flow/ sales |
| $x_{11}$ | Current debt/ sales | $x_6$ | Market value equity/ total debt |
| $x_{15}$ | Retained earnings/ total asset | $x_{14}$ | Net income/ total asset |
| $x_{19}$ | The annual searches data for sample's name | $x_{20}$ | recommendation |

*3.3. Experiment Design.* Li and Sun [38] have found out that financial distress prediction of the year $t$ using data sets of the year $(t-2)$ or $(t-3)$ is more difficult than using data sets of the year $(t-1)$. Here, we tackle the challenge. The framework of the empirical experiment is briefly shown in Figure 5. Details are clearly illustrated as follows.

*Step 1* (collect samples and data sets). We randomly divide samples into three data sets by ten times' split method. One is the training data set; the other one is the testing data set and the last one is the hold-out data set.

*Step 2* (select variables). We apply the stepwise logistic regression to the training data set of the year $(t-1)$ to select optimal traditional variables from Table 1. Then we integrate optimal traditional variables, recommendation, and the annual number of searches for sample's name into the variable set of ANSEM.

*Step 3* (obtain the predicting results). We employ ES, NN, EMEW, EMNN, EMRD, and ANSEM to the testing data sets of the year $(t-2)$ or $(t-3)$ to obtain the predicting results with the 5-fold cross validation method.

*Step 4.* Compare and discuss the predicting performance.

## 4. Results and Discussion

*4.1. Empirical Results.* In this paper, the 5-fold cross validation method is employed to perform the empirical experiment [39]. MATLAB (2016) is used to obtain optimal

Table 3: Predicting results of 5-fold cross-validation and summaries of ACC on mean, variance, and variance coefficient using data sets of the year $(t-2)$ for the percentage (25%, 75%).

| | ES | CNN | EMEW | EMNN | EMRD | ANSEM |
|---|---|---|---|---|---|---|
| 1 | 0.800 | 0.733 | 0.700 | 0.667 | 0.867 | 0.967 |
| 2 | 0.933 | 0.967 | 0.967 | 0.933 | 0.867 | 0.867 |
| 3 | 0.867 | 0.933 | 0.933 | 0.800 | 0.967 | 0.867 |
| 4 | 0.767 | 0.733 | 0.867 | 0.967 | 0.733 | 0.800 |
| 5 | 0.900 | 0.933 | 0.800 | 0.800 | 0.900 | 0.933 |
| Mean | 0.853 | 0.860 | 0.853 | 0.833 | 0.867 | 0.887 |
| Variance | 0.005 | 0.014 | 0.011 | 0.014 | 0.007 | 0.004 |
| Coefficient of variation | 0.006 | 0.016 | 0.013 | 0.017 | 0.008 | 0.005 |

Table 4: Predicting results of 5-fold cross-validation and summaries of ACC on mean, variance, and variance coefficient using data sets of the year $(t-2)$ for the percentage (50%, 50%).

| | ES | CNN | EMEW | EMNN | EMRD | ANSEM |
|---|---|---|---|---|---|---|
| 1 | 0.900 | 0.900 | 0.900 | 0.950 | 0.950 | 0.950 |
| 2 | 0.850 | 0.750 | 0.800 | 0.900 | 0.700 | 0.800 |
| 3 | 0.900 | 0.950 | 0.800 | 0.700 | 0.950 | 0.950 |
| 4 | 0.800 | 0.650 | 0.650 | 0.800 | 0.800 | 0.750 |
| 5 | 0.750 | 0.900 | 0.950 | 0.600 | 0.750 | 0.900 |
| Mean | 0.840 | 0.830 | 0.820 | 0.790 | 0.830 | 0.870 |
| Variance | 0.004 | 0.016 | 0.013 | 0.021 | 0.013 | 0.008 |
| Coefficient of variation | 0.005 | 0.019 | 0.016 | 0.026 | 0.016 | 0.009 |

Table 5: Predicting results of 5-fold cross-validation and summaries of ACC on mean, variance, and variance coefficient using data sets of the year $(t-2)$ for the percentage (75%, 25%).

| | ES | CNN | EMEW | EMNN | EMRD | ANSEM |
|---|---|---|---|---|---|---|
| 1 | 1.000 | 0.900 | 0.900 | 0.900 | 0.900 | 1.000 |
| 2 | 0.800 | 0.700 | 0.700 | 0.600 | 0.800 | 0.900 |
| 3 | 0.900 | 0.900 | 0.900 | 0.800 | 0.900 | 0.900 |
| 4 | 0.800 | 0.900 | 0.900 | 0.800 | 0.600 | 0.900 |
| 5 | 0.800 | 0.600 | 0.700 | 0.600 | 0.700 | 0.700 |
| Mean | 0.860 | 0.800 | 0.820 | 0.740 | 0.780 | 0.880 |
| Variance | 0.008 | 0.020 | 0.012 | 0.018 | 0.017 | 0.012 |
| Coefficient of variation | 0.009 | 0.025 | 0.015 | 0.024 | 0.022 | 0.014 |

variables and predicting results of each model with different sample sizes.

Financial distress prediction is a typical two-class classification problem. The predicting outputs are usually marked as either positive (P) or negative (N). More specifically, the outputs of financial distress prediction include four different results. One is the true positive (TP). TP means that the predicting output is positive and the real status is positive too. One is the false positive (FP). FP means that the predicting output is positive, but the real status is negative. The other one is the true negative (TN). TN means that the predicting output is negative and the real status is also negative. The last one is false negative (FN). FN means that the predicting output is negative, but the real status is positive. Thus, we can use an index called accuracy of correct classification (ACC) to measure the predicting accuracy of each method. The definition of ACC is shown as

$$ACC = \frac{TP + TN}{P + N} \tag{17}$$

*4.1.1. Empirical Results with Different Sample Sizes.* Predicting results of ES, NN, EMEW, EMNN, EMRD, and ANSEM using all variables and data sets of the year $(t-2)$ and $(t-3)$ for percentages (25%, 75%), (50%, 50%), and (75%, 25%) are respectively listed in Tables 3–8.

*4.1.2. Empirical Results with Different Variables.* To investigate whether the internet searches index can improve financial distress prediction performance under the era of big data, we run the empirical experiment again with selected variables except the internet searches index. And we employ data sets for the percentage (50%, 50%) as experiment data sets. Predicting results of ES, NN, EMEW, EMNN, EMRD, and ANSEM with all variables except the internet searches

Table 6: Predicting results of 5-fold cross-validation and summaries of ACC on mean, variance, and variance coefficient using data sets of the year $(t-3)$ for the percentage (25%, 75%).

| | ES | CNN | EMEW | EMNN | EMRD | ANSEM |
|---|---|---|---|---|---|---|
| 1 | 0.800 | 0.900 | 0.800 | 0.800 | 0.933 | 0.833 |
| 2 | 0.700 | 0.933 | 0.833 | 0.700 | 0.800 | 0.900 |
| 3 | 0.833 | 0.733 | 0.867 | 0.900 | 0.833 | 0.800 |
| 4 | 0.667 | 0.667 | 0.633 | 0.567 | 0.633 | 0.833 |
| 5 | 0.667 | 0.667 | 0.567 | 0.600 | 0.733 | 0.633 |
| Mean | 0.733 | 0.780 | 0.740 | 0.713 | 0.787 | 0.800 |
| Variance | 0.006 | 0.016 | 0.017 | 0.019 | 0.013 | 0.010 |
| Coefficient of variation | 0.008 | 0.021 | 0.024 | 0.027 | 0.016 | 0.013 |

Table 7: Predicting results of 5-fold cross-validation and summaries of ACC on mean, variance, and variance coefficient using data sets of the year $(t-3)$ for the percentage (50%, 50%).

| | ES | CNN | EMEW | EMNN | EMRD | ANSEM |
|---|---|---|---|---|---|---|
| 1 | 0.800 | 0.900 | 0.800 | 0.900 | 0.700 | 0.900 |
| 2 | 0.750 | 0.850 | 0.900 | 0.700 | 0.900 | 0.700 |
| 3 | 0.800 | 0.650 | 0.750 | 0.650 | 0.850 | 0.850 |
| 4 | 0.700 | 0.800 | 0.650 | 0.900 | 0.800 | 0.800 |
| 5 | 0.650 | 0.600 | 0.600 | 0.600 | 0.600 | 0.650 |
| Mean | 0.740 | 0.760 | 0.740 | 0.750 | 0.770 | 0.780 |
| Variance | 0.004 | 0.017 | 0.014 | 0.020 | 0.015 | 0.011 |
| Coefficient of variation | 0.006 | 0.022 | 0.019 | 0.027 | 0.019 | 0.014 |

Table 8: Predicting results of 5-fold cross-validation and summaries of ACC on mean, variance, and variance coefficient using data sets of the year $(t-3)$ for the percentage (75%, 25%).

| | ES | CNN | EMEW | EMNN | EMRD | ANSEM |
|---|---|---|---|---|---|---|
| 1 | 0.900 | 0.900 | 0.900 | 0.600 | 0.900 | 0.900 |
| 2 | 0.700 | 0.600 | 0.600 | 0.800 | 0.800 | 0.900 |
| 3 | 0.700 | 0.800 | 0.800 | 0.500 | 0.500 | 0.600 |
| 4 | 0.700 | 0.500 | 0.600 | 0.400 | 0.600 | 0.800 |
| 5 | 0.800 | 0.600 | 0.600 | 0.900 | 0.700 | 0.800 |
| Mean | 0.760 | 0.680 | 0.700 | 0.640 | 0.700 | 0.800 |
| Variance | 0.008 | 0.027 | 0.020 | 0.043 | 0.025 | 0.015 |
| Coefficient of variation | 0.011 | 0.040 | 0.029 | 0.067 | 0.036 | 0.019 |

index and data sets of the year $(t-2)$ and $(t-3)$ are listed in Tables 9 and 10.

*4.2. Comparison and Discussion.* In this paper, we employ three statistical indices of ACC from the 5-fold cross validation procedure to evaluate the performance of each model. The three statistical indices are mean, variance, and coefficient of variation. The mean is critical on evaluating the predicting accuracy of each model. The variance and coefficient of variation are critical on evaluating the predicting stability of each model.

*4.2.1. Results Comparison and Discussion with Different Sample Sizes.* Based on Tables 3–8, we can demonstrate the mean clearly in Figure 6. Meanwhile, we can illustrate the variance and coefficient of variance in Figures 7 and 8.

From Tables 3–8 and Figure 6, we can find out that the proposed novel soft ensemble model (ANSEM) for financial distress prediction has the highest mean accuracy no matter which year of data sets or which percentage of data sets is employed for predicting. No matter how small or big the sample size is, the predicting accuracy of ANSEM does not have a lot of changes. However, other predicting models are different, especially the NN and EWNN. With changes in sample sizes, means of NN and EWNN are significantly different. Because NN is not good at dealing with small sample sizes [29], the mean of ES also does not change a lot. Because ESs are real practitioners, they have to pay attention on the risk.

Moreover, all predicting methods except ES have a worse predicting accuracy using data sets of the year $(t-3)$ than those using data sets of the year $(t-2)$. Predicting financial distress on a long term is much more complicated than a short term prediction.

Similar as the conclusion of predicting accuracy, from Tables 3–8 and Figures 7-8, we can find out that ANSEM has

Table 9: Predicting results of 5-fold cross-validation and summaries of ACC on mean, variance, and variance coefficient using data sets of the year $(t - 2)$ for the percentage (50%, 50%) and all variables except the internet searches index.

| | ES | CNN | EMEW | EMNN | EMRD | ANSEM |
|---|---|---|---|---|---|---|
| 1 | 0.900 | 0.850 | 0.850 | 0.800 | 0.950 | 0.700 |
| 2 | 0.850 | 0.950 | 0.900 | 0.750 | 0.700 | 0.950 |
| 3 | 0.900 | 0.850 | 0.950 | 0.850 | 0.950 | 0.950 |
| 4 | 0.800 | 0.550 | 0.600 | 0.500 | 0.650 | 0.750 |
| 5 | 0.750 | 0.800 | 0.750 | 0.950 | 0.800 | 0.850 |
| Mean | 0.840 | 0.800 | 0.810 | 0.770 | 0.810 | 0.840 |
| Variance | 0.004 | 0.023 | 0.019 | 0.028 | 0.019 | 0.013 |
| Coefficient of variation | 0.005 | 0.028 | 0.024 | 0.037 | 0.024 | 0.015 |

Table 10: Predicting results of 5-fold cross-validation and summaries of ACC on mean, variance, and variance coefficient using data sets of the year $(t - 3)$ for the percentage (50%, 50%) and all variables except the internet searches index.

| | ES | CNN | EMEW | EMNN | EMRD | ANSEM |
|---|---|---|---|---|---|---|
| 1 | 0.800 | 0.700 | 0.750 | 0.850 | 0.750 | 0.750 |
| 2 | 0.750 | 0.900 | 0.900 | 0.950 | 0.900 | 0.850 |
| 3 | 0.800 | 0.600 | 0.650 | 0.600 | 0.600 | 0.650 |
| 4 | 0.700 | 0.850 | 0.800 | 0.550 | 0.850 | 0.900 |
| 5 | 0.650 | 0.500 | 0.500 | 0.550 | 0.550 | 0.600 |
| Mean | 0.740 | 0.710 | 0.720 | 0.700 | 0.730 | 0.750 |
| Variance | 0.004 | 0.028 | 0.023 | 0.035 | 0.023 | 0.016 |
| Coefficient of variation | 0.006 | 0.039 | 0.032 | 0.050 | 0.032 | 0.022 |

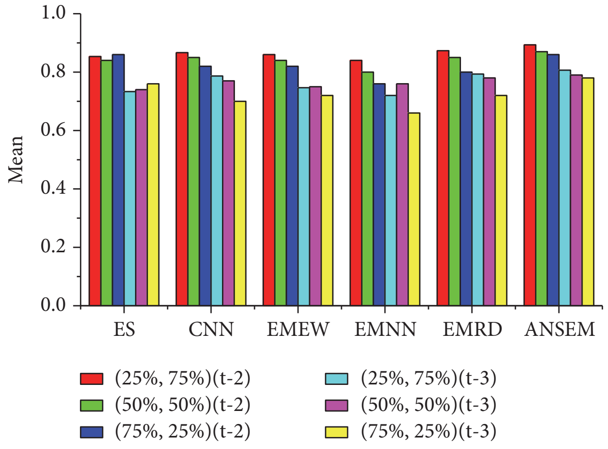


Figure 6: Mean of ACC. ES is the expert system method. CNN is the convolutional neural network. EMEW is the ensemble model based on equal weight. EMNN is the ensemble model based on the convolutional neural network. EMRD is the ensemble model based on the rough set theory and evidence theory. ANSEM is the soft ensemble model based on the uni-int decision making method.

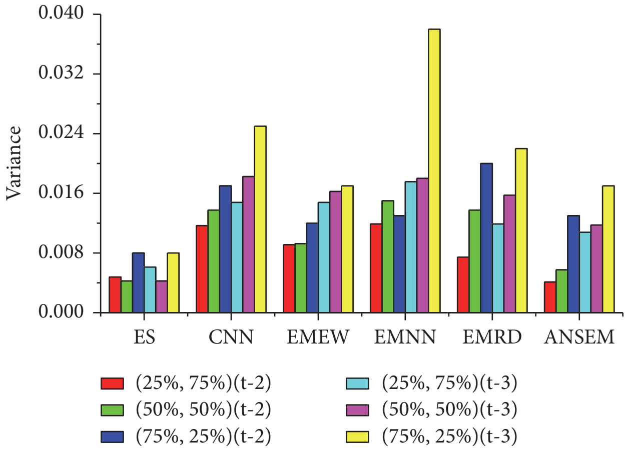


Figure 7: Variance of ACC. ES is the expert system method. CNN is the convolutional neural network. EMEW is the ensemble model based on equal weight. EMNN is the ensemble model based on the convolutional neural network. EMRD is the ensemble model based on the rough set theory and evidence theory. ANSEM is the soft ensemble model based on the uni-int decision making method.

the best predicting stability no matter which year of data sets or which percentage of data sets is employed for predicting. No matter how small or big the sample size is employed, ANSEM has the lowest variance and coefficient of variation of ACC. Besides, ES also has an excellent performance on the predicting stability. This is because ESs are real practitioners. They have to pay attention to the risk. Other predicting models are different; especially the NN and EWNN have the worst predicting stability.

Without a surprise, predicting stability of models using data sets of the year $(t - 2)$ outperforms that using data sets of the year $(t - 3)$.

*4.2.2. Results Comparison and Discussion with Different Variables.* Predicting results of ES, NN, EMEW, EMNN, EMRD, and ANSEM using different variables and data sets of the year $(t - 2)$ and $(t - 3)$ for the percentage (50%, 50%) are listed in Tables 4, 7, 9, and 10. From Tables 4, 7, 9, and 10, we can

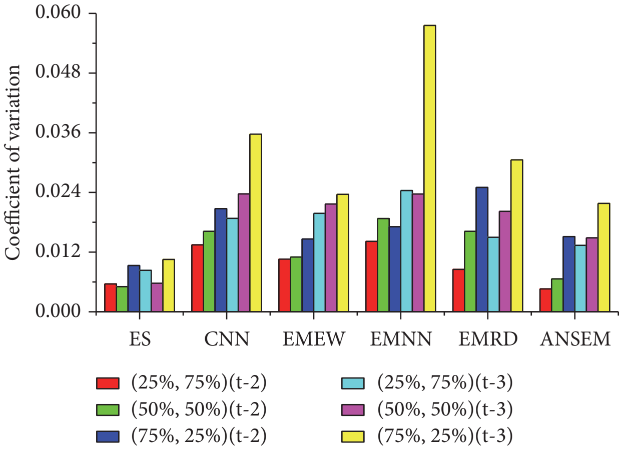


Figure 8: Coefficient of variation of ACC. ES is the expert system method. CNN is the convolutional neural network. EMEW is the ensemble model based on equal weight. EMNN is the ensemble model based on the convolutional neural network. EMRD is the ensemble model based on the rough set theory and evidence theory. ANSEM is the soft ensemble model based on the uni-int decision making method.

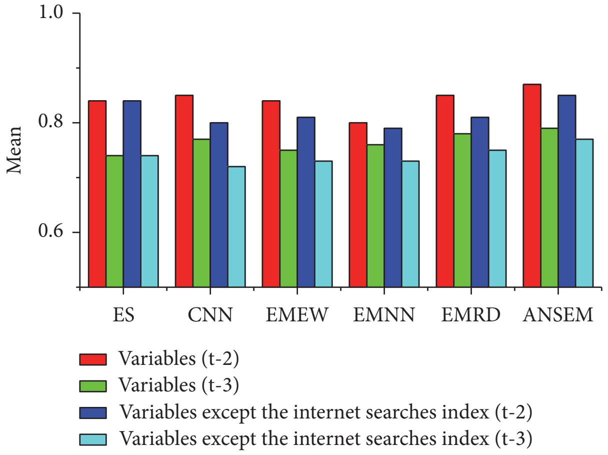


Figure 9: Mean of ACC. ES is the expert system method. CNN is the convolutional neural network. EMEW is the ensemble model based on equal weight. EMNN is the ensemble model based on the convolutional neural network. EMRD is the ensemble model based on the rough set theory and evidence theory. ANSEM is the soft ensemble model based on the uni-int decision making method.

compare the mean of ACC clearly in Figure 9. Meanwhile, we can compare the variance and coefficient of variance in Figures 10 and 11.

From Tables 4, 7, 9, and 10 and Figures 9–11, we can find out that the predicting model with all variables uniformly performs better no matter which year of data sets is used for predicting. It has a higher predicting accuracy comparing to those with all variables except the internet searches index. Also, it has a lower variance and coefficient of variation. That means the predicting model with all variables has a better predicting stability. Therefore, it is a good try to introduce the internet searches index for financial distress prediction

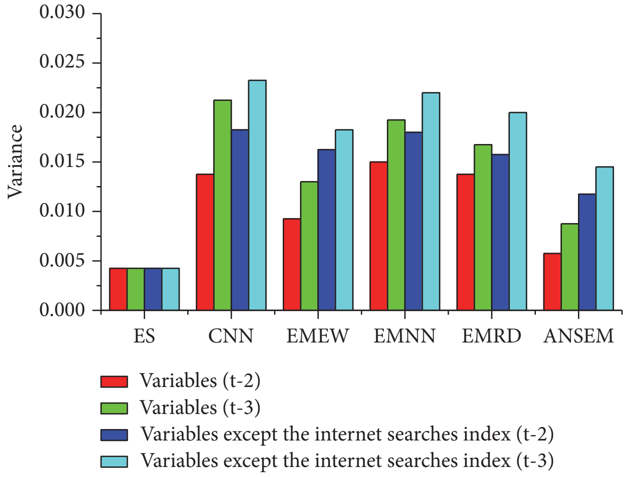


Figure 10: Variance of ACC. ES is the expert system method. CNN is the convolutional neural network. EMEW is the ensemble model based on equal weight. EMNN is the ensemble model based on the convolutional neural network. EMRD is the ensemble model based on the rough set theory and evidence theory. ANSEM is the soft ensemble model based on the uni-int decision making method.

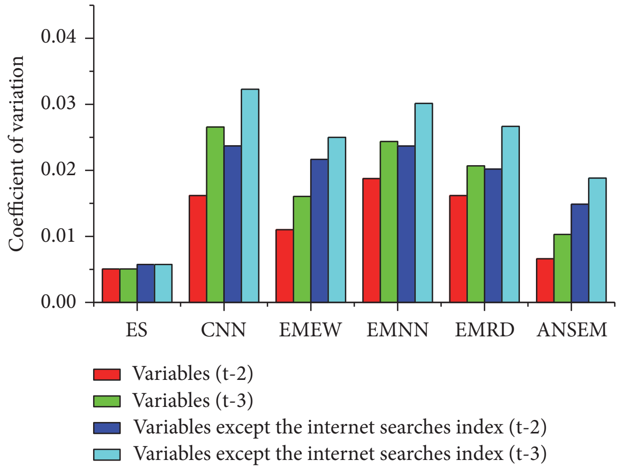


Figure 11: Coefficient of variation of ACC. ES is the expert system method. CNN is the convolutional neural network. EMEW is the ensemble model based on equal weight. EMNN is the ensemble model based on the convolutional neural network. EMRD is the ensemble model based on the rough set theory and evidence theory. ANSEM is the soft ensemble model based on the uni-int decision making method.

to improve the predicting performance under the era of big data. Also, no matter which year of data sets is used for predicting, ANSEM always obtains the best performance with all variables except the internet searches indices too.

*4.3. Summary.* Empirical results indicate that the proposed ANSEM can improve the predicting performance (accuracy and stability) of financial distress with different sample sizes, especially with the small sample sizes. This is because we use the novel uni-int decision making method of soft set theory as the integrated method and we employ real financial institutions as experts to improve the ES classifier. As a result, ANSEM has fewer restrictions and can make full use of more information when it is used to predicting practice. In other words, ANSEM can improve the practice performance.

Also, empirical results indicate that the internet searches index can offer additional information to improve the predicting performance under the era of big data because internet searches indices mainly reflect the potential demands and concerns of general public [40].

## 5. Conclusions

In this paper, we extended the model research for financial distress prediction with different sample sizes by proposing a novel soft ensemble model including qualitative classifiers (ES) and quantitative classifiers (CNN) based on the uni-int decision making method of soft set theory (UI). It constructs a soft set representation of each classifier then uses the optimal decision on soft sets to identify the status of firms. It inherits the advantage of ES, CNN, and UI. This lets ANSEM have fewer restrictions and can make full use of more information when it is used to practice. Compared with ES, CNN, EMEW, EMNN, and EMRD, our method ANSEM has demonstrated superior predicting performance for financial distress on accuracy and stability with different sample sizes. We also extend the research of predicting variables under the era of big data by introducing the internet searches index for financial distress prediction. It offers some new information to improve performance of financial distress prediction.

Though the empirical result is satisfactory, there are some inadequacies in this work. First of all, we use the annual number of searches for sample's name on "Baidu" website as the internet searches index. However, there is more than one searching engine in the world. We need to collect more internet searching data from different searching engines. Secondly, individual classifiers are simply discussed in the process of constructing the novel soft ensemble model. Because the key part of this paper is the integrated method, more attention should be paid on individual classifiers to get better performance. Finally, the proposed predicting method and predicting variables are applied to the data set of Chinese listed firms. There are no samples from unlisted firms and others regions. More samples should be included to evaluate the predicting performance.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 71801113). We also acknowledge the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (No. 18YJC630212) and the Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (No. 2017SJB0817).

## References

[1] F. De Luca and E. Meschieri, "Financial distress pre-warning indicators: a case study on italian listed companies," *The Journal of Credit Risk*, vol. 13, no. 1, pp. 73–94, 2017.

[2] C. Huang, F. Gao, and H. Jiang, "Combination of biorthogonal wavelet hybrid kernel OCSVM with feature weighted approach based on EVA and GRA in financial distress prediction," *Mathematical Problems in Engineering*, vol. 2014, Article ID 538594, 12 pages, 2014.

[3] S. F. Karabag, "Factors impacting firm failure and technological development: a study of three emerging-economy firms," *Journal of Business Research*, 2018.

[4] J. Sun, H. Li, Q. H. Huang, and K. Y. He, "Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches," *Knowledge-Based Systems*, vol. 57, pp. 41–56, 2014.

[5] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.

[6] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *Journal of Accounting Research*, vol. 18, no. 1, pp. 109–131, 1980.

[7] L. Wang and C. Wu, "Business failure prediction based on two-stage selective ensemble with manifold learning algorithm and kernel-based fuzzy self-organizing map," *Knowledge-Based Systems*, vol. 121, pp. 99–110, 2017.

[8] W. Xu, Z. Xiao, D. L. Yang, and X. L. Yang, "A novel nonlinear integrated forecasting model of logistic regression and support vector machine for business failure prediction with all sample sizes," *Journal of Testing and Evaluation*, vol. 43, no. 3, pp. 681–693, 2015.

[9] F. Ciampi, "Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms," *Journal of Business Research*, vol. 68, no. 5, pp. 1012–1025, 2015.

[10] W. H. Beaver, "Financial ratios as predictors of failure," *Journal of Accounting Research*, vol. 4, no. 4, pp. 71–111, 1966.

[11] W. Xu, Z. Xiao, X. Dang, D. L. Yang, and X. L. Yang, "Financial ratio selection for business failure prediction using soft set theory," *Knowledge-Based Systems*, vol. 63, pp. 59–67, 2014.

[12] S. Jones, D. Johnstone, and R. Wilson, "Predicting corporate bankruptcy: an evaluation of alternative statistical frameworks," *Journal of Business Finance & Accounting*, vol. 44, no. 1-2, pp. 3–34, 2017.

[13] F. Y. Lin, D. R. Liang, and W. S. Chu, "The role of non-financial features related to corporate governance in business crisis prediction," *Journal of Marine Science and Technology-Taiwan*, vol. 18, no. 4, pp. 504–5013, 2010.

[14] M. Doumpos, K. Andriosopoulos, E. Galariotis, G. Makridou, and C. Zopounidis, "Corporate failure prediction in the European energy sector: A multicriteria approach and the effect of country characteristics," *European Journal of Operational Research*, vol. 262, no. 1, pp. 347–360, 2017.

[15] S. Chava and R. A. Jarrow, "Bankruptcy prediction with industry effects," *Review of Finance*, vol. 8, no. 4, pp. 537–569, 2004.

[16] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, no. 7232, pp. 1012–1014, 2009.

[17] K. W. De Bock, "The best of two worlds: Balancing model strength and comprehensibility in business failure prediction using spline-rule ensembles," *Expert Systems with Applications*, vol. 90, pp. 23–39, 2017.

[18] H. Liu and J. Sun, "Forecasting business failure in china using case-based reasoning with hybrid case respresentation," *Journal of Forecasting*, vol. 29, no. 5, pp. 486–501, 2010.

[19] A. F. Atiya, "Bankruptcy prediction for credit risk using neural networks: a survey and new results," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 12, no. 4, pp. 929–935, 2001.

[20] Y. Zelenkov, E. Fedorova, and D. Chekrizov, "Two-step classification method based on genetic algorithm for bankruptcy forecasting," *Expert Systems with Applications*, vol. 88, pp. 393–401, 2017.

[21] M. J. Beynon and M. J. Peel, "Variable precision rough set theory and data discretisation: an application to corporate failure prediction," *Omega-International Journal of Management Science*, vol. 29, no. 6, pp. 561–576, 2001.

[22] A. Gepp, K. Kumar, and S. Bhattacharya, "Business failure prediction using decision trees," *Journal of Forecasting*, vol. 29, no. 6, pp. 536–555, 2010.

[23] Y. S. Ding, X. P. Song, and Y. M. Zen, "Forecasting financial condition of Chinese listed companies based on support vector machine," *Expert Systems with Applications*, vol. 34, no. 4, pp. 3081–3089, 2008.

[24] T. Hosaka, "Bankruptcy prediction using imaged financial ratios and convolutional neural networks," *Expert Systems with Applications*, vol. 117, pp. 287–299, 2019.

[25] F. Antunes, B. Ribeiro, and F. Pereira, "Probabilistic modeling and visualization for bankruptcy prediction," *Applied Soft Computing*, vol. 60, pp. 831–843, 2017.

[26] D. Liang, C. Tsai, A. Dai, and W. Eberle, "A novel classifier ensemble approach for financial distress prediction," *Knowledge and Information Systems*, vol. 54, no. 2, pp. 437–462, 2018.

[27] H. A. Alaka, L. O. Oyedele, H. A. Owolabi et al., "Systematic review of bankruptcy prediction models: Towards a framework for tool selection," *Expert Systems with Applications*, vol. 94, pp. 164–184, 2018.

[28] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *Operational Research Quarterly*, vol. 20, no. 4, pp. 451–468, 1969.

[29] K. Boratyńska and E. Grzegorzewska, "Bankruptcy prediction in the agribusiness sector: Lessons from quantitative and qualitative approaches," *Journal of Business Research*, vol. 89, pp. 175–181, 2018.

[30] N. Çağman and S. Enginoğlu, "Soft set theory and *uni-int* decision making," *European Journal of Operational Research*, vol. 207, no. 2, pp. 848–855, 2010.

[31] F. Feng, Y. Li, and N. Çağman, "Generalized *uni-int* decision making schemes based on choice value soft sets," *European Journal of Operational Research*, vol. 220, no. 1, pp. 162–170, 2012.

[32] Z. Xiao, X. Yang, Y. Pang, and X. Dang, "The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster-Shafer evidence theory," *Knowledge-Based Systems*, vol. 26, pp. 196–206, 2012.

[33] D. Molodtsov, "Soft set theory - first results," *Computers & Mathematics with Applications*, vol. 37, no. 4-5, pp. 19–31, 1999.

[34] S. Danjuma, T. Herawan, M. A. Ismail, H. Chiroma, A. I. Abubakar, and A. M. Zeki, "A review on soft set-based parameter reduction and decision making," *IEEE Access*, vol. 5, pp. 4671–4689, 2017.

[35] P. K. Maji, R. Biswas, and A. R. Roy, "Soft set theory," *Computers & Mathematics with Applications*, vol. 45, no. 4-5, pp. 555–562, 2003.

[36] A. Amendola, F. Giordano, M. L. Parrella, and M. Restaino, "Variable selection in high-dimensional regression: a nonparametric procedure for business failure prediction," *Applied Stochastic Models in Business and Industry*, vol. 33, no. 4, pp. 355–368, 2017.

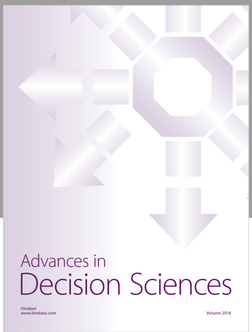
[37] J. Sun, H. Fujita, P. Chen, and H. Li, "Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble," *Knowledge-Based Systems*, vol. 120, pp. 4–14, 2017.

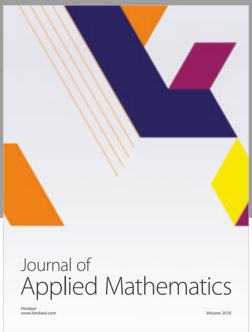
[38] H. Li and J. Sun, "Forecasting business failure:the use of nearest-neighbour support vectors and correcting imbalanced samples – evidence from the chinese hotel industry," *Tourism Management*, vol. 33, no. 3, pp. 622–634, 2012.

[39] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th international joint conference on Artificial intelligence*, vol. 2, pp. 1137–1143, 1995, https://www.mendeley.com/catalogue/study-crossvalidation-bootstrap-accuracy-estimation-model-selection/.

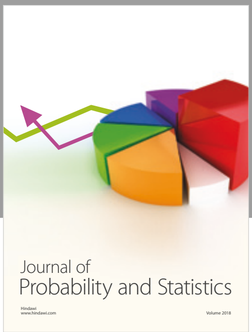
[40] I. Bertschek, W. Briglauer, K. Hüschelrath, B. Kauf, and T. Niebel, "The Economic Impacts of Broadband Internet: A Survey," *Review of Network Economics*, vol. 14, no. 4, pp. 201–227, 2015.

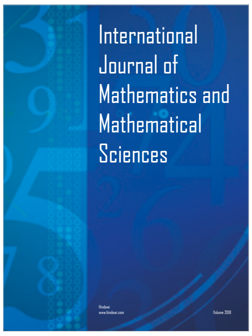
Advances in
Operations Research


Advances in
Decision Sciences

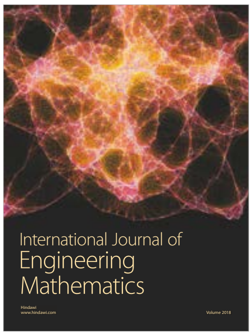

Journal of
Applied Mathematics

The Scientific
World Journal

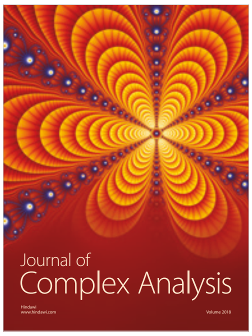

Journal of
Probability and Statistics

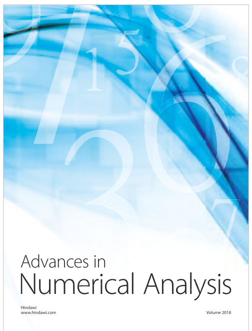

International
Journal of
Mathematics and
Mathematical
Sciences

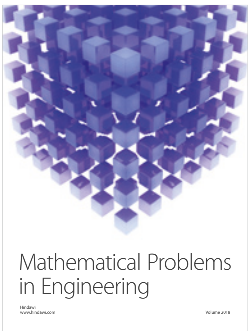
Journal of
Optimization

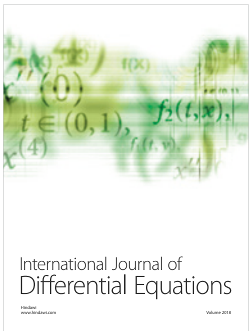

International Journal of
Engineering
Mathematics

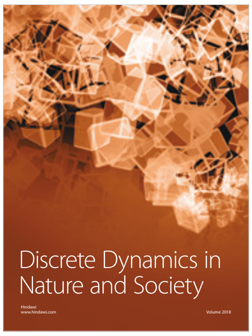
International Journal of
Analysis

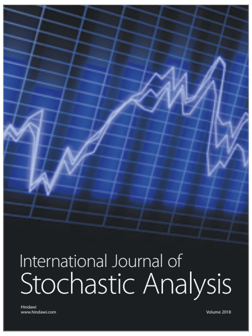

Journal of
Complex Analysis


Advances in
Numerical Analysis


Mathematical Problems
in Engineering

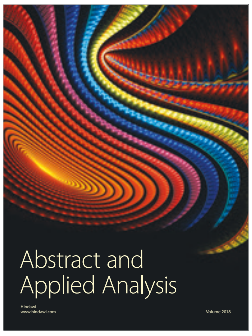

International Journal of
Differential Equations

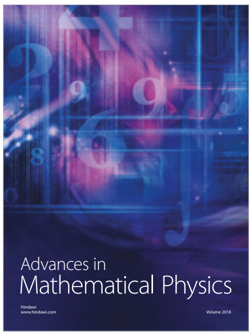

Discrete Dynamics in
Nature and Society


International Journal of
Stochastic Analysis

Journal of
Mathematics

Journal of
Function Spaces


Abstract and
Applied Analysis


Advances in
Mathematical Physics