

## Research Article

# Financial Early Warning System Model Combining Hybrid Semantic Hierarchy with Group Method of Data Handling Neural Network for Detection of Banks' Risks

Guangju Li 

*School of Finance and Trade, Zhengzhou Shengda University of Economics, Business & Management, Zhengzhou 451191, China*

Correspondence should be addressed to Guangju Li; [guangjuli\\_edu@yeah.net](mailto:guangjuli_edu@yeah.net)

Received 7 July 2021; Revised 29 August 2021; Accepted 11 September 2021; Published 19 October 2021

Academic Editor: Ahmed Farouk

Copyright © 2021 Guangju Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Banks, financial, and credit institutions encountering the weakening financial system and increased risk factors cause high inflation and great losses for an economy. Detecting financial risks in advance could help financial institutions avoid losses, and the financial system could be eventually affected less. Early warning systems for banks could be helpful to identify financial risks and take measures to deal with hazardous situations. Various approaches have already been put forward. However, inaccuracy issues in risk detection are one of the main issues. Combining semantic hierarchy with the GMDH neural network to predict financial risks is proposed. A semantic hierarchy approach based on converting risk-related values and picking influential variables could be practical in risk detection. Besides, the GMDH algorithm utilizing neural networks based on available data has the capability of predicting possible risks that could occur in the future. The outcomes of the proposed method when compared to non-data mining methods suggest that it improves accuracy by almost 20%.

## 1. Introduction

With the globalization of economies and intensifying competition among banks, profit margins decreased and the risk factors increased [1]. Weaknesses in the banking system could occur due to possible reasons of boycotts imposed by the international community towards banking system, reduced government support, crisis risk, losses from transactions, exchange rate fluctuations and debt, increased market uncertainty, excessive asset holdings, cash balances on balance sheets, rising bankruptcy trends of large credit companies, and innovations of monetary and financial instruments [2]. Both risk management and forecasting play significant roles. Therefore, the risk detection of financial institutions has received much attention [3]. Financial alerting systems to assess risks for financial institutions have been a research area [4].

The current financial crisis has fueled the initiatives of policymakers to construct alert systems to predict financial crises in advance, assuming that these systems could

function and provide signals based on models and forecasts utilizing some indicators before a crisis could occur. The past crises proved to be very costly for both advanced and emerging economies [5]. Besides, the world economic system could be vulnerable and fragile. Moreover, each crisis having a unique character caused by such situations ranging from political unrest to shock of the trade market could lead to two consequences. While the first one is characterized by not being predictable, the second one is characterized by the complexity of its nature. On the other hand, convincing policymakers to take precautionary measures against any possible crisis is so difficult. Economically and politically costly measures do not take effect until a crisis develops. Precautionary measures have to be taken to prevent economic damage, and policies should be implemented to address it. Any warning system is expected to alert policymakers before a crisis occurs, so measures that are implemented could prevent it from spreading through the economy. Various methods to reduce damage have been proposed, but a series of measures could hardly be

welcomed. For example, although homeowners prefer their home values to rise rapidly, financial institutions are reluctant to put more capital aside. How crises could spread through economies is reported in [6]. Consequently, any warning system could inform financial institutions about imminent dangers and help strengthen them. Hence, it allows policymakers to take precautions before a crisis.

Various methods have been implemented to forecast crises. However, these approaches have drawbacks such as inadequate accuracy and high-risk prediction. Due to the issues that existed in the previous methods, this manuscript proposes a novel method combining the GMDH deep neural network with semantic technique. Therefore, the proposed method is utilized to generate a financial alert system. The rest of the paper is organized as follows: Section 2 summarizes various models dealing with financial warning systems and risk-based management. The GMDH neural network is presented in Section 3. The proposed method is introduced in Section 4. The results are presented in Section 5, and Section 6 presents the conclusion.

## 2. Financial Warning Systems

An early warning system (EWS), which is a monitoring, reporting, and alerting system, forecasts quantitatively success levels, probable anomalies, issues, risks, potential bankruptcies, and transactions affairs [4]. Hence, avoiding or mitigating potential issues could be possible by EWSs. Financial statements such as balance sheets and income tables providing financial information are utilized in EWSs that are, in essence, analysis techniques identifying current states of financial intuitions.

In the construction of an EWS after a crisis is defined, which types of events are related to a crisis should be identified. Then, constructing an analytical model could systematically determine possible economic damage. Afterward, it is expected that the paths from which a crisis begins and penetrates other sectors are determined. Finally, favorable conditions in the financial sector should be determined. An effective and efficient alert system includes several features such as the ability to chart the outlook for global finance and macroeconomics, assess the economic impact of countries and sectors, and analyze paths of impacts in sectors. Various financials are noted in countries [7]. Policymakers must comprehend how these EWSs function to prevent crises. However, constructing such systems is overlooked since policymakers doubt their efficiency. In addition to operating these systems, many appropriate and effective policies should be considered to reduce risks. Global communication, for instance, is very important in this regard. Admittedly, no such systems perfectly signal before any crisis occurs. False alarms could be possible [8, 9]. Besides, factors such as efficient leading-edge analysis, precrisis scrutiny, and widespread global coordination are essentials [10]. Recently, research has been conducted on risk pacing by EWSs.

A novel random forest-based alert system investigating bank-level financial statements to determine patterns that

put banks at risk was proposed, and it performs better than the conventional methods [11].

To determine the monetary misery of individual banks, AI was proposed in [12]. A decision tree involving subjective financial area and macroeconomic provisions was constructed by Quinlan's C5.0 computation to assess a model, whose final structure contains 12 elements and 19 hubs when compared to the benchmark model of a logit.

The hidden Markov model, Dynamic Linear Switching System model, and Simple Linear Dynamics System model were employed under the title of Dynamic Bayesian Networks utilizing the Markovian framework. The likelihood of an imminent crisis based on given dynamics could be computed. When compared with other methods of signal extraction and logit, it could be better utilized as EWS [13].

Monetary emergencies pertinent to small-scale and open economies employ large-scale powerful factor models that extract useful data out of bulk datasets and provide a mixed recurrence index merging worldwide and homegrown monetary policies with monetary markers. Thailand, South Korea, Singapore, Malaysia, Philippines, and Indonesia were examined by the model. The extracted elements and other driving markers utilized by the logit replace the model in foreseeing foundational occasions that have a great impact. The usefulness of the extracted components on provident control over a model employing sufficiently long history pointers is shown by in-test and out-of-test results [14].

Predictive methods as EWSs utilized in a large undergraduate STEM course were examined. Research has currently shed light on harnessing them as tools for identifying students at risk.

Student interaction with Learning Management Systems (LMSs) was utilized. Eight prediction methods and the optimal time of implementing an EWS were investigated. Findings pertinent to a statistics course at the university level were reported since LMSs had all resources related to the course and enabled weekly assessments that helped students find their study pattern [15].

A combination of models used as an EWS to exhibit scholarly studies of undergraduates was proposed. Moodle collaboration information, attributes, and grades of 802 undergraduates from a CO college were utilized. Four components, Access, Questionnaire, Task, and Age, were found to be significant to expect scholarly studies from undergraduates. Moodle helped identify factors composing Access such as visits to discussions and glossaries. Factors identified by visits and endeavors in polls constitute the Questionnaire. Task is composed of factors identified with counseled and submitted undertakings. Age was strikingly distinguished as a negative indicator. Furthermore, cluster examination suggested some connections with Moodle that are firmly identified [16].

Two methods, linear regression and family-based data analysis [17,18], generate better risk predictions for banks when compared to other methods.

Adjustments of financial indicators under emergency might affect the data extraction of EWSs in the state of a strategic relapse when genuine situations of foundational

banking emergencies occur. Considering possible limits of the regular EWSs and properties of the AI calculation, the EWSs utilizing experts' voting structure could likely fit into foundational banking emergencies. The irregular wood classifier taking into account experts' voting process is the most effective among AI classifiers. The experts' voting EWS orchestrating multivariate data could be more appropriate to provide alarms when shifted settings exist [19].

The credit-based variables as EWSs in the emerging economies were examined. The bank-related data such as collected total credit to the private sector was assessed as an indicator utilizing the Accuracy Criterion (AUC). The nominal credit growth and the change in the credit to GDP ratios are found to be the best beckoning attributes. The nominal credit growth significantly outperformed the credit-to-GDP gap in almost all specifications. These findings presented converse outcomes compared with the results of the East-West Institute since the credit-to-GDP gap is the best performance in the developed economies [20].

An incorporated EWS predicting the disturbance of the securities exchange was proposed in [21]. The proposed EWS initially characterizes securities exchange emergencies concerning a marker utilizing switching ARCH that separates probabilities of the great unpredictability system. A crossover calculation is employed to make everyday forecasts. In the observational assessment based on ten-year Chinese stock information, the proposed EWS produces better outcomes. The linear regression producing higher accuracy than other methods was reported by [22].

ANNs utilizing various sampling schemes were introduced. Three models were employed based on past stream scenes containing high convergences of airborne particulate matter. Model 3 was better in forecasting estuarial dust occasions when precisely taking into account the qualities of media and contamination origins. Hence, it helped avoid preemptive fiasco mitigated measures to stay away from the antagonistic wellbeing results of residue storms inside estuarial districts since the proposed model efficiently provided better estimates [23].

Precrisis periods in the Eurozone utilizing the EWS were researched, which led to a novel proposed model to deal with uncertainty. A set of macroeconomic risk indices was utilized to pick better ones. Then, four indices, as explanatory variables, were employed in all types of logit models to assess precrisis probabilities [24].

The underlying indicators triggering the Islamic banking crisis were examined by [25]. The method called Markov Switching Linear Regression (MS-VAR) utilizes monthly data gathered from the official database covering January 2004 to March 2017. It was concluded that Islamic banking is more resistant to domestic and foreign risks than conventional banking. When Islamic banks are compared to conventional banks, the former's  $z$ -score is greater ( $11.933 > 11.679$ ) [25].

Measuring the level of financial fixing, monetary market circumstances, and fundamental monetary danger based on the adequacy and exactness of monetary status pointers led the relapse strategy to be utilized so the determination of significant pointers could be obtained. A unique weighted

monetary condition based on the time-differing boundary vector autoregressive model consists of five factors, namely, loan fee, land value, cash supply, swapping scale, and stock value, which adequately mirror the real monetary circumstance. China's monetary circumstance from 2013 to 2017 was analyzed by this strategy. Changes in monetary markers help estimate the level of money fixing and monetary economic situations in advance [26].

The available research specific to China examining mechanisms of the industry environment found a profound impact on corporate financial risk and investigated seven dimensions of industry environmental risk. An index system was constructed employing Structural Equation Model. The development of the financial risk was found to be highly related to the industry environment. EWSs classify seven subsystems and are capable of identifying potential financial risks, which is conducive to developing effective financial policies concerning external changes [27].

Banking and money emergencies were an impressive expense for homegrown and worldwide monetary frameworks. EWSs were particularly investigated in the turn of events and a combination of conditions in monetary fields. Foreseeing and recognizing antagonistic and risky occasions in the fields of business, money, and finance were contemplated [28].

Prospective CoVaR as a market-based measurement of systemic risk was proposed in emerging markets. The presented  $\Delta$ CoVaR values represented a network of asset exposures. Two novel network-based indicators were proposed based on exposure networks. Besides, they were utilized as EWS to predict recession and crisis at the enterprise and entire market levels utilizing the relevance of new indicators. The outcomes of the EWS for market returns suggest that models predict a market crisis, which provides an alarm signal with up to 7 periods before the crisis. The implementation of those helped policymakers develops appropriate macro and micro policies to prevent systemic risks [29].

A logistic model predicting bank failures based on most of the downward trends of NIC banks was investigated. G8 banks performed well in growth phases, although the performances of the NIC banks were poor in downturns [30].

### 3. GMDH Neural Network

Deep learning neural network implemented in financial risk forecasting has four types of layers, as shown in Figure 1.

**3.1. Input Layer.** The parameters are normalized before the semantic aspect is applied. The input layer represents the input data obtained from feature extraction.

**3.2. Latent Layers.** Training of neural networks in a supervised manner needs initial weights assigned by a Gaussian distribution, and backpropagation algorithm updates them, which is a slow process and is trapped in a local minimum. Pretraining a layer by a nonsupervised method

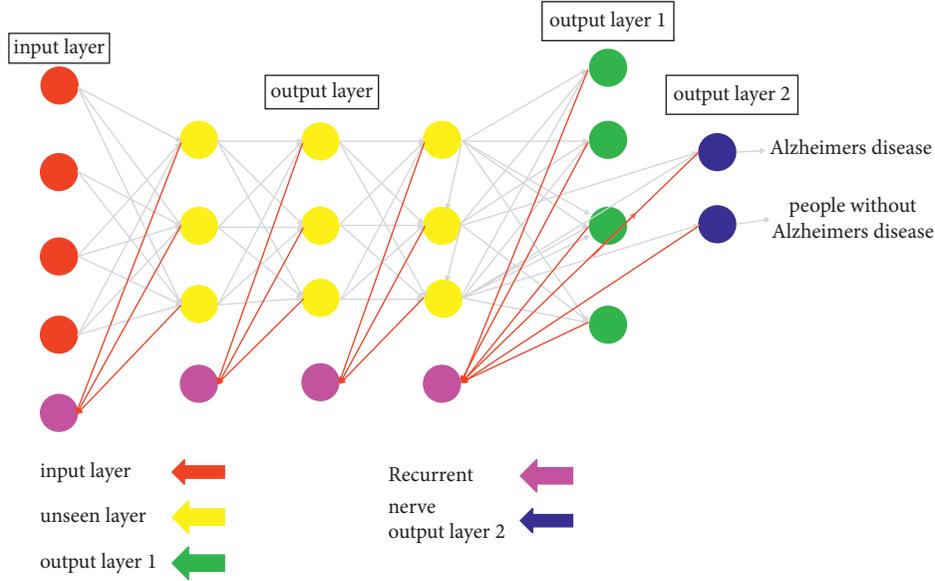


FIGURE 1: Deep learning neural network.

could resolve this issue. Therefore, deep learning neural network supplemented with this feature will be constructed.

**3.3. Output Layer.** Two types of output layers are utilized, which are called Output Layer 1 and 2. The output layer called Output Layer 1 is employed to rebuild a network, called input and utilized for training; extract information from the last layer of latent layers; and employ backward error propagation. On the other hand, Output Layer 2, called a softmax layer, is embedded in a softmax approach to conduct a regression. Probabilities of belonging to each class are obtained.

**3.4. Softmax Regression Function.** The logistic regression has a nonlinear structure that combines the attributes in a linear form called the probit function. The response variable takes binary values represented by

$$y_i \in \{-1, +1\}. \quad (1)$$

The response variable is a Bernoulli-type random variable  $Y$  whose probability is denoted by  $\eta$ . The probability of success depends on the predictor; i.e.,  $\eta = \eta(x)$ . If  $X$  is assumed to be a regularizing predictor random variable,  $\eta(x)$  is a conditional expectation of  $Y$  defined by

$$E\{Y | X\} = \eta(x). \quad (2)$$

However, a unitary binding function in (3) transforms the expectation into the linear combination of the predictions called generalized linear models.

$$g(a) = \ln \frac{a}{1-a}. \quad (3)$$

The inverse of the logarithmic function is the logistic function defined by

$$h_{\theta}(x) = \frac{1}{1 + \exp(-\theta^T \cdot x)}. \quad (4)$$

$\theta$  is optimized by the cost function defined by

$$J(\theta) = - \left[ \sum_{i=1}^m \{y^{(i)} \cdot \ln(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \cdot \ln(1 - h_{\theta}(x^{(i)}))\} \right]. \quad (5)$$

## 4. The Proposed Method

The proposed method is described and presented concerning the available trend. Figure 2 depicts the proposed financial smoothing system that utilizes the data collected from past samples. A filter is utilized to eliminate corrupt data, which is called preprocessing. Better data processing time and model robustness are reached. In other words, data should be prepared for both simulation and analysis. Semantic hierarchy helps determine salient features utilizing semantic logic. The selection of semantic features enables the model to be more accurate in risk prediction with the lowest error rate. The dataset is split into two categories, called training and test datasets. While the training data is employed to construct a model, the test data is utilized to assess the model.

The balanced sampling scheme splitting data into training and test samples is implemented. The GMDH deep neural network is constructed with salient features employing 70 to 80% of the data. The rest, 20 to 30%, is allocated to forecasting. The efficiency and performance of the proposed method are finally assessed.

**4.1. Data Preprocessing.** Preprocessing could be conducted utilizing various methods: data cleansing, data collection, data transmission, and data reduction. The proposed

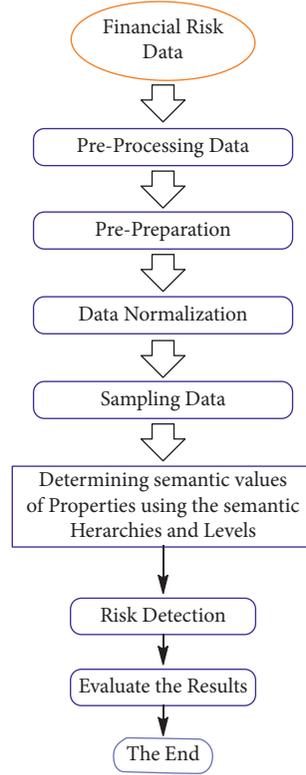


FIGURE 2: Flowchart of the proposed method.

strategy is to examine the data and to predict if the rows or columns have missing values that are replaced by calculating the mean. The steps are depicted in Figure 3.

**4.2. Data Preparation.** Data preparation is related to converting the data into a usable form for cloud computing servers.

**4.3. Normalizing Metadata.** Normalizing values of each attribute converts values to the range between 0 and 1, then the row of the data matrix is rotated to arrange. All datasets are then transformed into a matrix. Normalization is defined by

$$\text{normalize}(x) = \frac{(x - X_{\min})}{(X_{\max} - X_{\min})}. \quad (6)$$

Maximum and minimum values are represented by  $X_{\max}$  and  $X_{\min}$ , respectively. Hence, all data is contained in  $[0, 1]$ . Existing samples are preprocessed, and the normalization process is separately conducted on each block. Then, the semantic hierarchy utilizes those outputs.

**4.4. Deletion of Outlier Samples Utilizing DBSCAN.** DBSCAN clustering algorithm groups data; identifies categories; and labels data points as prime, boundary, and anomalous points. The advantage of the DBSCAN over other clustering methods is that it specifies and predicts samples well, which helps deep learning models be refined and

improves the processing speed. The DBSCAN, a density-based spatial model, requires two user-defined parameters called the epsilon (eps) neighborhood and the minimum number of points (minPts). Points within the eps distance are called neighbors for a given point. If the number of neighboring points is greater than minPts, this is called the cluster of points.

Boundary points are defined as nonprime points but are neighborhoods of principal points. The DBSCAN differs from the conventional clustering approaches since it could define anomalous points not suitable for clusters. Figure 4 presents the results of clustering utilizing the eps distance threshold as a clustering criterion. The DBSCAN also requires a minimum number of points belonging to a cluster. When minPts is 3, groups 1, 2, 4, and 6 are designated as clusters. In contrast, DBSCAN discriminates groups 3 and 5 as outliers due to not having enough points. When minPts is assigned as 5, groups 3, 5, and 6 are found as outliers.

The inputs are the user-defined datasets and parameter values of eps and minPts.

**Input:** data collection, eps neighborhood distance, minPts (minimum number of points).

**Output:** distant spots and clusters determined.

**Variables:**  $M$ ,  $n$ : the values of row and column of  $D$ , distance: distance-vector, indicators: indicating a distance less than eps, class no: displaying clusters by default.

**4.5. Applying the Semantic Hierarchy Algorithm.** Defining semantic values of attributes and picking salient attributes are performed in two steps. Semantic hierarchy is employed

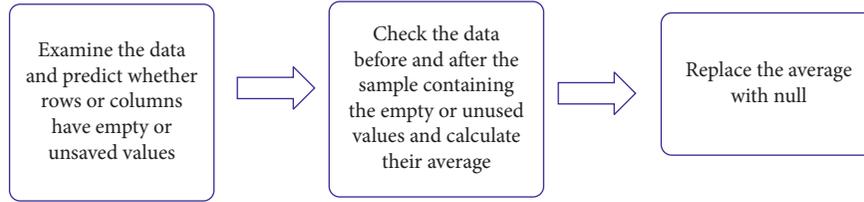


FIGURE 3: The stages of data cleansing.

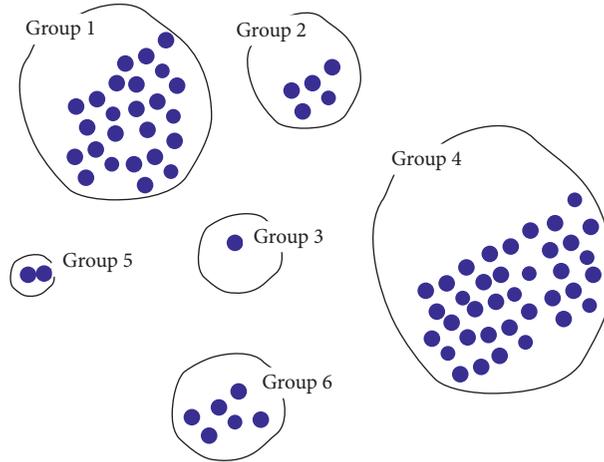


FIGURE 4: A representation of distance-based risk prediction.

to separate data values by type. Defining different scenarios and generating different outcomes are possible based on semantic hierarchy. For instance, a dataset having property of the shelf life of the system whose values range from 1 to 100 minutes is assumed. The hierarchies changing from 1, 2, 3, up to  $n$  could be defined. Transforming the values from 1 to 100 into three words or representative numbers of low (0), medium (1), and long (2) size could be one representation when assuming the level is 3. Therefore, semantic hierarchies could be defined based on the need. The proposed semantic congruence coefficient (SCC) criteria with properties could be defined as follows:

$$0 \leq \text{SCC}(G, G') \leq 1$$

If graphs  $G$  and  $G'$  are completely uniform, then  $\text{SCC}(G, G') = 1$

If graphs  $G$  and  $G'$  do not have any semantically identical vertex pairs, then  $\text{SCC}(G, G') = 0$

The larger the homomorphic subgraph between  $G$  and  $G'$  is, the greater the number of vertices, the number of edges, or the degree of semantic similarity of vertices will be. Hence, the SCC will be larger.

To compute SCC between graphs, we need to find the simplest congruent graph with the fewest vertices possible, which is denoted by  $a \rightarrow x \rightarrow b$ . Both input and output degrees of the node  $x_1$  are utilized. The node  $x$  is removed and replaced by  $a \rightarrow b$ , which is only for offline graphs.

An algorithm similar to the largest shared subgraph algorithm needs to be conducted except for the subgraphs of two graphs overlapping. The output will provide the SCC

values between the two graphs. Comparing the SCC values of the two graphs with a threshold value helps determine a sample of whether a financial risk could appear.

If  $\text{SSHC}(G, G') > \text{threshold}$ , then Label = "RISK"

Else

Label = "NONE RISK"

Besides, the amount of uncertainty in the nasal passage could be determined based on the distance between SCC and threshold.

**4.6. Separation of Training and Test Samples.** The GMDH deep neural network utilizes training and test datasets. Some sampling schemes could be utilized to extract training dataset.

- (i) Random sampling
- (ii) Classified sampling
- (iii) Balanced sampling

Random sampling is one of the simplest schemes that functions randomly and splits samples into training and test datasets. On the other hand, the main disadvantage is that it may not be sampled from a particular batch and ultimately results in reduced accuracy of data classification and validation. Classified sampling is another method that performs based on a probability-based scheme and selects the samples with some drawbacks. Balanced sampling is one of the methods that balance the required samples from the available classes, which finally picks balanced data.

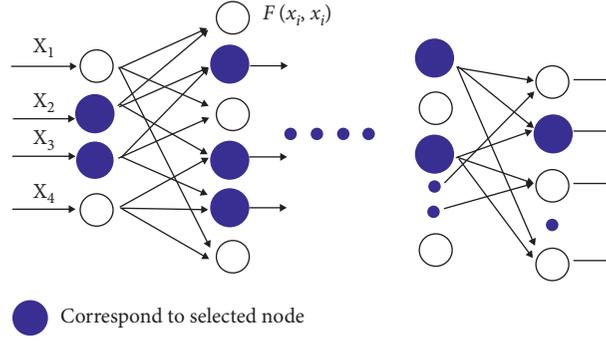


FIGURE 5: GMDH network structure of the proposed model.

**4.7. GMDH Deep Neural Network.** The proposed algorithm is based on deep neural network learning. Training dataset is utilized to construct the GMDH model. Then, outputs for each sample are produced with two outcomes, yes or no. Group Method of Data Handling (GMDH), a family of induction algorithms employing computer-based mathematical models dealing with multiparameter datasets, optimizes parameters of the models in areas such as data mining, knowledge discovery, prediction of complex systems, optimization, and pattern recognition. The GMDH neural network performs better than Single Exponential Smoothing, Double Exponential Smoothing, ARIMA, and Open Neural Network. The inductive method, which sorts gradually complex polynomial models and selects the best solution with external criterion, is employed in the GMDH algorithms [19].

Figure 5 presents a set of neurons for the GMDH algorithms. Different pairs of nerve cells are represented in each layer and are connected by quadratic polynomials. A connection of new neurons appears in the next layer, which maps inputs into outputs. The rhythm problem is formally defined as a method of finding a function  $f$  that could be estimated by the input-output dataset. Therefore, the single-input data pairs define the following equations.

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}), \quad i = 1, 2, \dots, M, \quad (7)$$

$$X = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}), \quad (8)$$

$$\hat{y}_i = \hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}), \quad i = 1, 2, \dots, M. \quad (9)$$

Equation (9) represents the predicted output. To minimize the squared difference between the predicted and actual outputs, a minimization problem is defined by

$$\sum_{k=1}^M \left[ (\hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) - y_i)^2 \right] \rightarrow \text{Min}. \quad (10)$$

A generic equation could be represented by complex polynomials known as Ivanchenko polynomial defined by

$$\hat{y} = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \dots \quad (11)$$

However, two-variable representation to predict  $y$  is defined by

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_3 x_i^2 + a_4 x_j^2 + a_5 x_i x_j. \quad (12)$$

The coefficients  $a_i$  in (16) are computed to minimize the difference between the observed and predicted response variable based on each pair  $(x_i, x_j)$  by utilizing the least squares. Then, the polynomial tree is constructed. Therefore, the coefficients of each quadratic function  $G_i$  are determined [19].

$$r^2 = \frac{\sum_{i=1}^M (y_i - G_i)}{\sum_{i=1}^M y_i^2} \rightarrow \text{Min}. \quad (13)$$

The polynomial structure of the regression can depend on the samples  $p, q \in \{1, 2, \dots, M\}$ . Since  $n(n-1)/2$  different samples could exist the least squares approximation could be utilized to find unknown parameters, it means least squares  $\{(y_i, x_{ip}, x_{iq}), (i = 1, 2, \dots, M)\}$ .

$$\begin{bmatrix} x_{1p} & x_{1p} & \vdots & y_1 \\ x_{2p} & x_{2q} & \vdots & y_2 \\ \dots & \dots & \dots & \dots \\ x_{Mp} & x_{Mp} & \vdots & y_M \end{bmatrix}. \quad (14)$$

The following matrix relation could be obtained by the quadratic subunit in (16) for each row of  $M$  data.

$$Aa = Y, \quad (15)$$

where the vector of unknown coefficients of the fourth-degree polynomial is represented by

$$a = \{a_0, a_1, a_2, a_3, a_4, a_5\}, \quad (16)$$

$$Y = \{y_1, y_2, y_3, \dots, y_M\}^T. \quad (17)$$

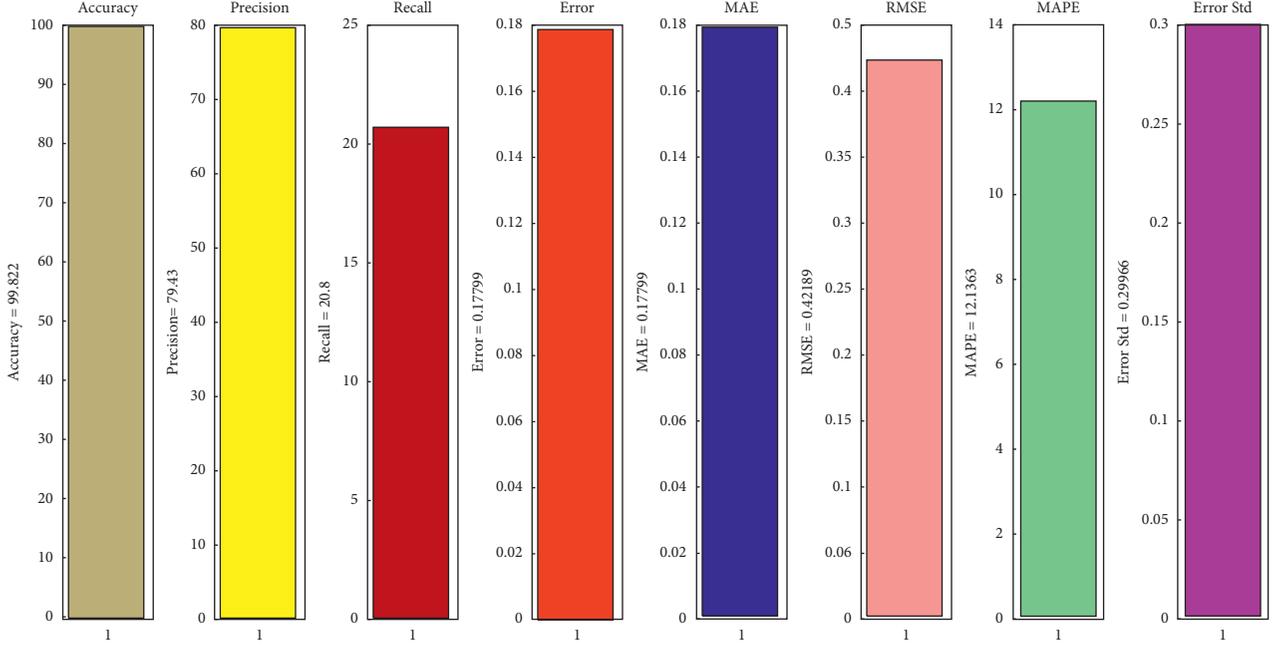


FIGURE 6: Results of employed criterion with no semantic techniques.

The vector is the output values of the samples.

$$A = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}x_{2q} & x_{2p}^2 & x_{2q}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{Mp} & x_{Mq} & x_{Mp}x_{Mq} & x_{Mp}^2 & x_{Mq}^2 \end{bmatrix}. \quad (18)$$

Utilizing the least-squares method leads to the solution of the normal equations defined by

$$a = (A^T A)^{-1} A^T Y. \quad (19)$$

The vector determines the best quadratic coefficients. However, a solution obtained from normal equations is almost able to bypass errors.

## 5. Results

A database with 17232 samples and 869 features was utilized implementing Mat Lab 2015a.

**5.1. Evaluation Criteria.** The criterion employed to assess the proposed method are as follows:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (20)$$

TP (true positive) and FP (false positive) denote the number of samples projected correctly and mistakenly to financial risk, respectively.

$$\text{ReCall} = \frac{TP}{TP + FN}. \quad (21)$$

FN (False negative) designates the number of samples that cause a false-positive normal error.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (22)$$

The error and average error rates are computed by (23) and (24), respectively.

$$\text{Error} = 100 - \frac{TP + TN}{TP + TN + FP + FN}, \quad (23)$$

$$\text{MSE} = \text{Mean}(\text{Error}^2), \quad (24)$$

$$\text{RMSE} = \sqrt{\text{Error}}. \quad (25)$$

**5.2. Evaluating the Results of the Proposed Model without Semantic Techniques.** Figure 6 presents the results with no semantic technique. The accuracy rate of the risk forecast, accuracy, call rate, error rate, MAE, RMSE, MAPE, and standard deviation (SD) values are 99.822%, 79.43%, 20.8%, 0.177%, 0.177, 0.421, 12.135, and 0.299, respectively.

### 5.3. Evaluation of Results with a Semantic Technique

**5.3.1. Semantic Level 2.** Figure 7 depicts the results with semantic technique and semantic level 2. The accuracy rate of the risk forecast, accuracy, call rate, error rate, MAE, RMSE, MAPE, and SD are 99.899%, 84.902%, 83.901%, 0.100%, 0.100, 0.317, 7.065, and 0.179, respectively.

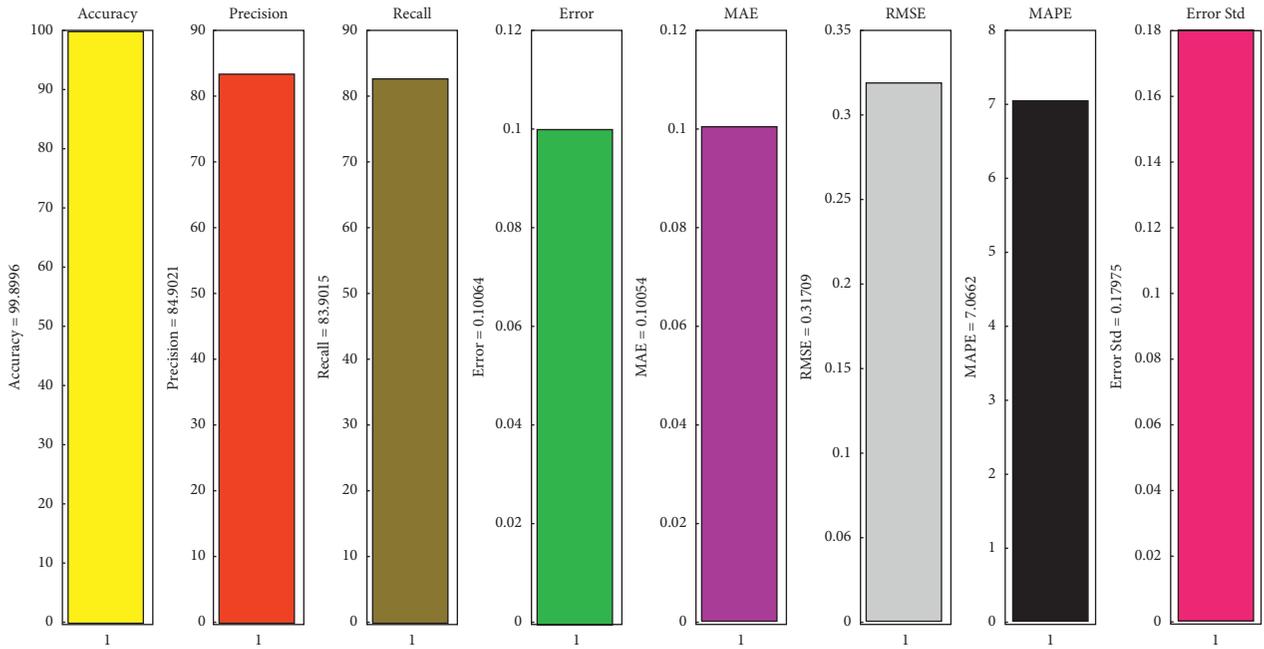


FIGURE 7: Results of employed criterion with semantic techniques and level 2.

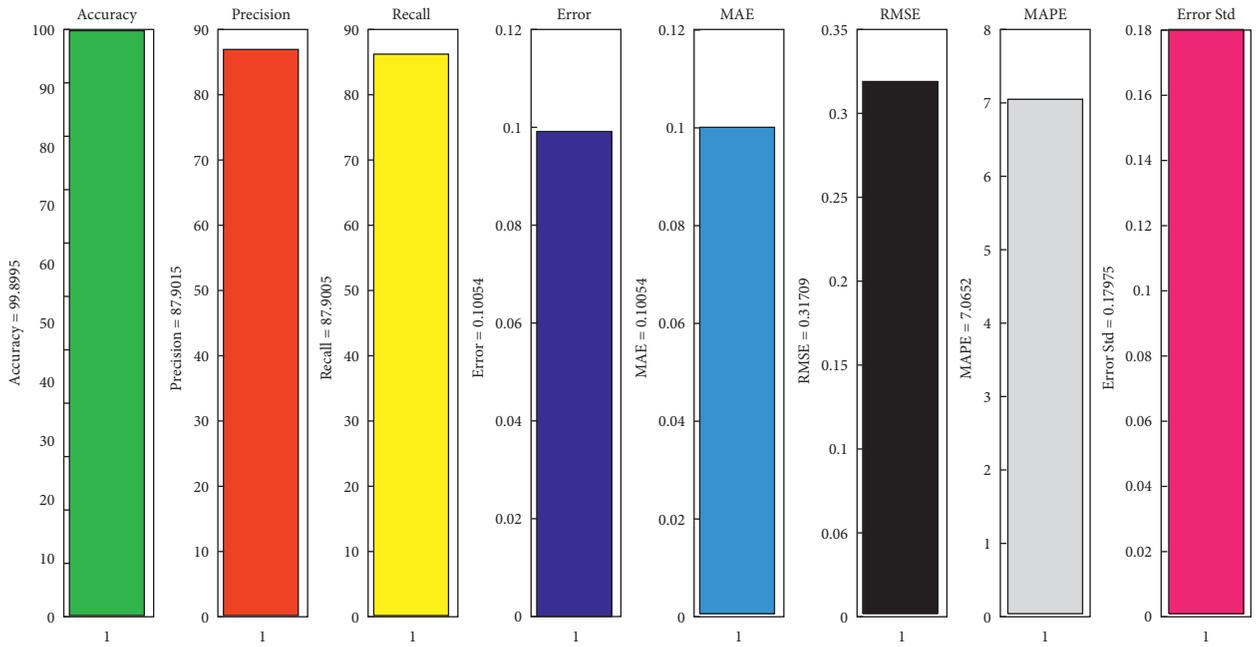


FIGURE 8: Results of the employed criterion with semantic techniques and level 3.

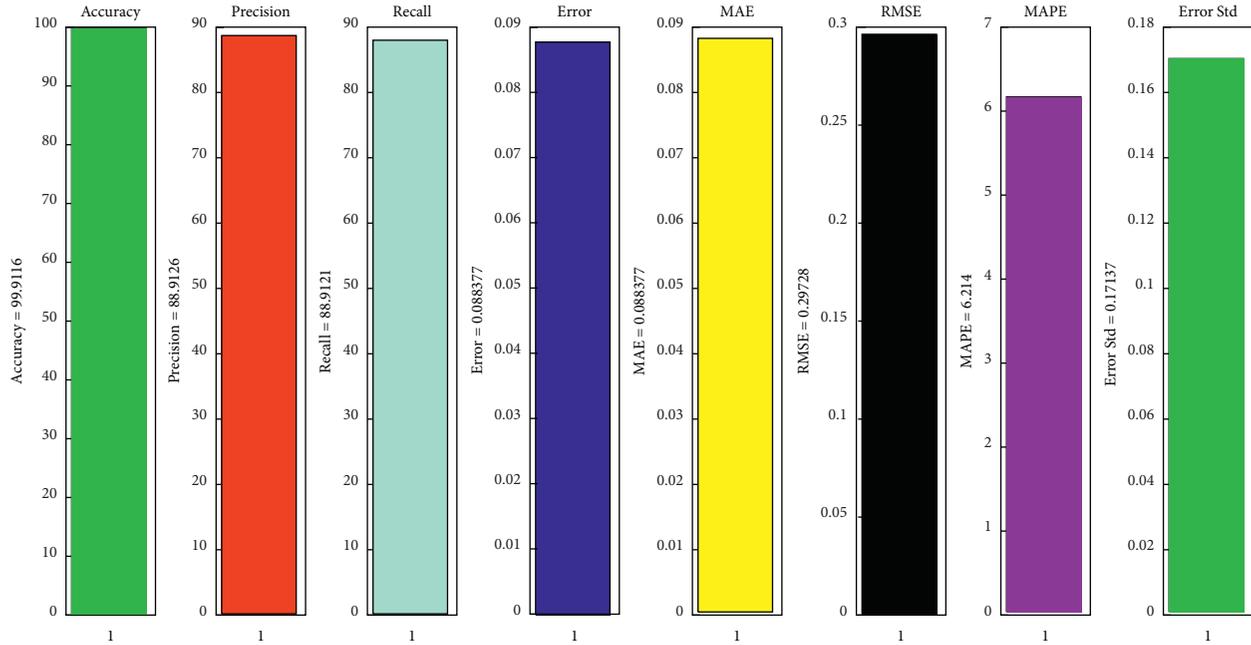


FIGURE 9: Results of the employed criterion with semantic techniques and level 5.

5.3.2. *Semantic Level 3.* Figure 8 summarizes the results with both semantic technique and level 3. The accuracy rate of the risk forecast, accuracy, call rate, error rate, MAE, RMSE, MAPE, and SD are 99.899%, 87.901%, 87.900%, 0.100%, 0.100%, 0.317, 7.065, and 0.176, respectively.

5.3.3. *Semantic Level 5.* Figure 9 reports the results with semantic technique and level 5. The accuracy rate of the risk forecast, accuracy, call rate, error rate, MAE, RMSE, MAPE, and SD are 99.911%, 88.912%, 88.912%, 0.088, 0.088, 0.297, 6.214, and 0.171, respectively.

## 6. Conclusion

This manuscript introduces a novel approach to the prediction of financial risks by constructing a model. SHT and GMDH deep learning neural networks are employed. While the SHT is employed to convert data to a specific range of values and to select salient features, the GMDH deep learning neural network is utilized to construct a model and to predict risk. By simulating the proposed model on financial data and presenting a semantic approach to risk prediction, a 99% accuracy level is reached. Comparing the results of the proposed model with those of other non-data mining methods reveals that the proposed method improved accuracy by almost 20%.

## Data Availability

All the data can be obtained from the author upon reasonable request.

## Conflicts of Interest

The author declares no conflicts of interest.

## Acknowledgments

The research was supported by the Research on Multiple Conjectural Causes and Countermeasures for the Distortion of Investor Protection Policy in China's Stock Market (No. 17BGL059).

## References

- [1] H. V. Greuning and B. S. Brajovic, "Analyzing and managing banking risk," *p*, World Bank Group, Washington, DC, USA, 2003.
- [2] A. Talavera, L. Cano, D. Paredes, and M. Chong, "Data mining algorithms for risk detection in bank loans," in *Proceedings of the Annual International Symposium on Information Management and Big Data*, pp. 151–159, Springer, Cham, Switzerland, September 2018.
- [3] M. Leo, S. Sharma, and K. Maddulety, "Machine learning in banking risk management: a literature review," *Risks*, vol. 7, no. 1, p. 29, 2019.
- [4] A. S. Koyuncugil and N. Ozgulbas, "Social aid fraud detection system and poverty map model suggestion based on data mining for social risk mitigation," in *Surveillance Technologies and Early Warning Systems: Data Mining Applications for Risk Detection*, pp. 173–193, IGI Global, Pennsylvania, PA, USA, 2011.
- [5] A. S. Koyuncugil and N. Ozgulbas, "Financial early warning system model and data mining application for risk detection," *Expert Systems with Applications*, vol. 39, no. 6, pp. 6238–6253, 2012, p.
- [6] L. Xu and S. Tang, "Technology innovation-oriented complex product systems R&D investment and financing risk management: an integrated review," *Advances in Intelligent Systems and Computing in Proceedings of the Tenth International Conference on Management Science and Engineering Management*, vol. p, pp. 1653–1663, Springer, Singapore, 2017.

- [7] M. Bussiere and M. Fratzscher, *Towards a New Early Warning System of Financial Crises*, European Central Bank, Frankfurt, Germany, 2002.
- [8] P. E. Davis and K. Dilruba, "Comparing early warning systems for banking crises," *Journal of Financial Stability*, vol. 4, pp. 89–120, 2007.
- [9] S. Shvets, "Early warning system: logit/probit introduction for Ukraine," *pModern Economics*, vol. 13, no. 1, pp. 266–271, 2019.
- [10] M. Fratianni and F. Marchionne, "The role of banks in the subprime financial crisis," *SSRN Electronic Journal*, vol. 1, 2009.
- [11] K. Tanaka, T. Kinkyo, and S. Hamori, "Random forests-based early warning system for bank failures," *Economics Letters*, vol. 148, pp. 118–121, 2016, p.
- [12] K. Tanaka, T. Higashide, T. Kinkyo, and S. Hamori, "Analyzing industry-level vulnerability by predicting financial bankruptcy," *Economic Inquiry*, vol. 57, no. 4, pp. 2017–2034, 2019.
- [13] J. J. Dabrowski, C. Beyers, and J. P. de Villiers, "Systemic banking crisis early warning systems using dynamic Bayesian networks," *Expert Systems with Applications*, vol. 62, pp. 225–242, 2016, p.
- [14] C. Truong, J. Sheen, S. Trück, and J. Villafuerte, "Early warning systems using dynamic factor models: an application to Asian economies," *Journal of Financial Stability*, Article ID 100885, 2021.
- [15] E. Howard, M. Meehan, and A. Parnell, "Contrasting prediction methods for early warning systems at undergraduate level," *The Internet and Higher Education*, vol. 37, pp. 66–75, 2018, p.
- [16] J. Bravo-Agapito, S. J. Romero, and S. Pamplona, "Early prediction of undergraduate student's academic performance in completely online learning: a five-year study," *Computers in Human Behavior*, vol. 115, Article ID 106595, 2021.
- [17] C. Filippopoulou and S. I. Spyrou, "An early warning system for predicting systemic banking crises in the Eurozone: a logit regression approach," *Journal of Economic Behavior & Organization*, vol. 172, pp. 344–363, 2018.
- [18] I. Aldasoro, C. E. Borio, and M. Drehmann, "Early warning indicators of banking crises: expanding the family," *BIS Quarterly Review*, 2018.
- [19] T. Wang, S. Zhao, G. Zhu, and H. Zheng, "A machine learning-based early warning system for systemic banking crises," *Applied Economics*, vol. 53, no. 26, pp. 2974–2992, 2021.
- [20] A. Geršl and M. Jašová, "Credit-based early warning indicators of banking crises in emerging markets," *Economic Systems*, vol. 42, no. 1, pp. 18–31, 2018, p.
- [21] P. Wang, L. Zong, and Y. Ma, "An integrated early warning system for stock market turbulence," *Expert Systems with Applications*, vol. 153, Article ID 113463, 2020.
- [22] J. H. Lang, T. A. Peltonen, and P. Sarlin, "A framework for early-warning modeling with an application to banks," *p*, 2018.
- [23] K. Lan Phuong Nguyen, Y. Hsun Chuang, R.-F. Yu, and H.-W. Chen, "Developing an ANN-based early warning model for airborne particulate matters in river banks areas," *Expert Systems with Applications*, vol. 183, Article ID 115421, 2021.
- [24] V. Coudert and J. Idier, "Reducing model risk in early warning systems for banking crises in the euro area," *International Economics*, vol. 156, pp. 98–116, 2018, p.
- [25] I. N. Irfan Nurfalaha, A. S. Rusydiana, N. Laila, and E. F. Cahyono, "Early warning to banking crises in the dual financial system in Indonesia: the Markov switching approach," *Journal of King Abdulaziz University Islamic Economics*, vol. 31, no. 2, pp. 133–156, 2018, p.
- [26] A. Huang, L. Qiu, and Z. Li, "Applying deep learning method in TVP-VAR model under systematic financial risk monitoring and early warning," *Journal of Computational and Applied Mathematics*, vol. 382, Article ID 113065, 2021.
- [27] Y. Zhang, Y. Peng, and A. M. Rahman, "Simulation analysis of financial early-warning systems based on industrial risk monitoring for Chinese enterprises," *Finance and Accounting Monthly*, no. 2, p. 1, 2015.
- [28] J. Lin, "Design of enterprise financial early warning model based on the complex embedded system," *Microprocessors and Microsystems*, vol. 80, Article ID 103532, 2021.
- [29] H. Dastkhan, "Network-based early warning system to predict financial crisis," *International Journal of Finance & Economics*, vol. 26, no. 1, pp. 594–616, 2021.
- [30] Y. Wang, J. Yang, and Z. Liu, "Bank failure prediction in relation to the business life cycle," *Modern Economy*, vol. 10, no. 3, pp. 757–777, 2019.