Expert Systems with Applications 42 (2015) 2857–2869

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks

Félix J. López Iturriaga *, Iván Pastor Sanz

University of Valladolid, Spain

ARTICLE INFO

Article history:
Available online 25 November 2014

Keywords:
Bankruptcy prediction
Financial crisis
Multilayer perceptron
Neural networks
Self-organizing maps

ABSTRACT

We develop a model of neural networks to study the bankruptcy of U.S. banks, taking into account the specific features of the recent financial crisis. We combine multilayer perceptrons and self-organizing maps to provide a tool that displays the probability of distress up to three years before bankruptcy occurs. Based on data from the Federal Deposit Insurance Corporation between 2002 and 2012, our results show that failed banks are more concentrated in real estate loans and have more provisions. Their situation is partially due to risky expansion, which results in less equity and interest income. After drawing the profile of distressed banks, we develop a model to detect failures and a tool to assess bank risk in the short, medium and long term using bankruptcies that occurred from May 2012 to December 2013 in U.S. banks. The model can detect 96.15% of the failures in this period and outperforms traditional models of bankruptcy prediction.

1. Introduction

The recent financial crisis, the generalization and propagation of systemic risk in a more and more global financial environment, and the high social costs of bank failures have drawn attention to the mechanisms of control of banks solvency (Wang, Ma, & Yang, 2014). The 2009 Basel Committee on Banking Supervision papers, widely known as the Basel III Accord,[1] advise banking regulators to develop capital and liquidity rules sufficiently rigorous to allow financial firms to withstand future downturns in the global financial system.

The Basel III Accord follows the capital agreement found in the 1988 accord, commonly known as Basel I. Basel I is enforced by law in the G10 and adopted by over 100 other countries. The goal of this 1988 roundtable was to minimize credit risk. However, innovation and financial changes in the world led to the need in 2004 for a more comprehensive set of guidelines known as Basel II. The purpose of this new framework was to promote greater stability in the financial system and reduce the social costs of financial instability. To fulfill this aim, the accord requires banks to classify their loan portfolio and identify the risks that they may face through their lending and investment practices to ensure that they hold enough capital reserves. Basel II put in place a broader view of financial risk that incorporated the differences among credit, operational, and market risk. In addition, the accord gave both supervisors and markets a wider range of action.

The collapse of a number of financial institutions in the United States at the beginning of the crisis in 2007 due to the emergence of new financial products and risks, the fall of real estate prices, and biased pricing methods of real estate premises exposed Basel II's shortcomings. The industry clearly required new standards for supervision of financial intermediaries and new metrics of financial risk. Our paper joins the stream of analysis that examines the failures of U.S. banks in recent years. In so doing, we follow the recommendations of the G20 Finance Ministers and Central Bank Governors who met June 3–5, 2010, in Busan, Korea, and "[welcomed] the progress on the quantitative and macroeconomic impact studies which will inform the calibration of…new rules.".

We develop a hybrid neural network model to study the bankruptcy of U.S. banks by combining a multilayer perceptron (MLP) network and self-organizing maps (SOMs). With this contribution, we complement previous evidence and update the methods of risk assessment (Oreski & Oreski, 2014). Our aim is twofold: descriptive and predictive. First, we describe the main characteristics of U.S. distressed banks and how bank failures have evolved from the onset of the financial crisis in 2007. The implementation of our model and the analysis of the most descriptive variables provide

* Corresponding author at: University of Valladolid, School of Business and Economics, Avda. Valle del Esgueva 6, 47011 Valladolid, Spain. Tel.: +34 983 184 395; fax: +34 983 423 299.

E-mail addresses: flopez@eco.uva.es (F.J. López Iturriaga), ivan.pastor@alumnos.uva.es (I.P. Sanz).

[1] In December 2009, the Basel Committee on Banking Supervision published two consultative documents entitled "Strengthening the Resilience of the Banking Sector" and "International Framework for Liquidity Risk Measurement, Standards and Monitoring." Although these papers retain no other official designation, they have been widely dubbed Basel III.

http://dx.doi.org/10.1016/j.eswa.2014.11.025
0957-4174/

interesting insights about the most critical features of distressed banks relative to nondistressed banks. Second, we provide a tool to predict the probability of bank failures some time before they happen. In so doing, we define three different models that are conditional on the period of time before the failure. These two objectives lead to the development of a visual tool that can assess the strengths and weaknesses of a bank in the short, medium and long term by combining the outputs of the three models in a bi-dimensional map using SOMs (Kohonen, 1993). This tool offers not only a method to detect failures but also a visual representation of when weaknesses can arise. This procedure also provides a dynamic perspective as it can assess the probability of bank failure along a period of time, unlike most previous models that are limited to a single point in time.

Reliability is concern for models of failure prediction when the time horizon goes beyond the near short term because few models achieve stable results over the time. Our work makes advances in three directions relative to previous research. First, we widen the selection of variables and implement better selection criteria based on the experience and performance of previous research. In this way we avoid the loss of predictive power due to single-period data. Second, we explicitly take into account the specific features of the recent crisis and measure credit risk in conjunction with the Basel accords. Finally, we provide three different time-horizon scenarios for failure prediction (up to three years before bankruptcy). We then combine the joint likelihood of default from each model into a visual SOM that allows us to create different profiles of risk and to extend the time horizon to evaluate the potential risk of bank failure.

We test the predictive power of our model on a sample of U.S. banks that went bankrupt between May 2012 and December 2013. We also compare our MLP-SOM results with both traditional classification models (the discriminant and logit analyses) and with more recent techniques such as support vector machine and random forest procedures. Our model, which exhibits high predictive power, predicted one year ahead 96.15% of the 52 banks that failed between May 2012 and December 2013.

Our results show that distressed banks were heavily concentrated on the real estate at the explosion of the mortgage bubble. Distressed banks carried out a strategy of quick expansion and had to pay back higher interest rates to raise enough money. The business downturn and the fall of the prices of real estate collateral resulted in a growing default rate. The poor quality of the loan portfolios of distressed banks relative to their counterparts required higher provisions. The liquidity crisis constrained the possibility of improving the solvency of the banks through equity issuance and led to a negative margin. This vicious circle could not be maintained for long, and finally financial authorities intervened.

The paper is divided into five sections. Section 2 reviews previous research on models of bankruptcy prediction. Section 3 describes the main characteristics of our hybrid NN model: the MLP and SOM methodologies. In Section 4, we provide the results. We provide descriptive results and the storyline of the failed banks. In this section we also compare the predictive power of the NN models with the output from traditional techniques and from other more recent approaches. Finally, in Section 5, we draw some conclusions from our results and offer some directions for future research.

2. Review of bankruptcy prediction models

Corporate and specially banks bankruptcy prediction is an important and widely studied topic in the business intelligence field (Chen, 2011b; Serrano-Cinca & Gutiérrez-Nieto, 2013; Sun, Li, Huang, & He, 2014; Yu, Miche, Séverin, & Lendasse, 2014; Zhou, 2013). Models of prediction have become more sophisticated to account for the effects of financial crises or other outstanding business episodes (Mokhatab Rafiei, Manzari, & Bostanian, 2011; Nassirtoussi, Aghabozorgi, Wah, & Ngo, 2014). Although a sharp line cannot be easily drawn, broadly speaking, two approaches exist to bankruptcy prediction: structural and empirical (Angelini, di Tollo, & Roli, 2007). The structural approach is based on modeling the dynamics of firm characteristics and derives the default probability based on these dynamics. The empirical approach, rather than modeling bankruptcy on firm characteristics, gleans the default relation from the data (Atiya, 2001).

The foundations of the empirical approach can be traced to Davies and Bouldin (1979) and Ingaramo, Leguizamón, and Errecalde (2005). Beaver (1966) pioneered the prediction of bankruptcy using financial statement data. His univariate analysis focused on the evolution of certain financial ratios such as financial leverage, return on assets, and liquidity and showed how these ratios worsened as long as firms faced bankruptcy. Altman (1968) and Ohlson (1980) use linear models that classify firms using financial ratios as inputs. These authors widened the scope of the model by introducing a multiple discriminant credit scoring analysis. Their models identify financial variables that have statistical explanatory power. They also introduced the logistic regression approach and used a novel set of financial ratios as inputs. Yin (2005) contributed further to the empirical approach by using a step-wise multiple discriminant analysis to distinguish between corporate future failures and successes.

From the late 1980s, artificial intelligence techniques, particularly rule-based expert systems, case-based reasoning systems, and machine learning techniques such as artificial neural networks (ANNs) have been successfully applied to bankruptcy prediction (Angelini et al., 2007; Atiya, 2001; Mukta & Kumar, 2009). Empirical approaches have been improved along the time, from the simplest and the most restrictive models to more flexible and recent ones. Some of these approaches are the statistical techniques, the neural networks, the random forest, the support vector machine, the genetic algorithm and, even, some hybrids techniques.

Compared to other empirical approaches to bankruptcy prediction, NNs have some advantages. First, as previously stated, NNs do not make assumptions about the distribution of the data. Second and interestingly, NNs allow a nonlinear set of relations. This allowance is especially important for bankruptcy predictions because the relation between the likelihood of default and the explanatory variables do not have to be linear. NNs are quite powerful and flexible modeling devices that do not make restrictive assumptions on the data-generating process or the statistical law relating variables of interest. Oreski and Oreski (2014) argued that a nonlinear approach outperforms linear models for two reasons: first, saturation effects can occur in the relation between financial ratios and the prediction of default. Second, multiplicative factors may become problematic.

The comparison between traditional models and NNs remains an open question within the literature and has led to mixed results (Bernhardsen, 2001; Fulmer, Moon, Gavin, & Erwin, 1984; Ohlson, 1980). Furthermore, the choice of one method over another is usually based on several heterogeneous criteria, such as data availability. Nevertheless, in recent years, some authors have shown the superiority of NNs relative to other techniques (Hol, 2006; Lee & Choi, 2013; Levy-Yeyati, Martínez Pernía, & Schmukler, 2010; Mokhatab Rafiei et al., 2011). In particular, du Jardin (2010) who uses more than 500 ratios taken from approximately 200 previous papers, shows that an NN-based model that uses a set of variables selected with a criterion specifically adapted to the network leads to better results than a set chosen with criteria used in the financial literature.

A number of papers have shown that NNs provide a better performance than the purely linear models or the heuristic systems based on expert rules of thumb. Likewise, Piramuthu (1999) find that the bankruptcy predictions based on NNs are more precise that the ones based on the discriminant analysis. These studies compare the prediction accuracy of the back propagation methods with others and show the outperformance of NNs.

In the case of bank failures, Martínez (1996) compares back propagation methods with discriminant analysis, logit analysis and the *k*-nearest neighbor for a sample of Texas banks and concludes that the first set of methods outperforms. Similarly, the results of Vellido, Lisboa, and Vaughan (1999) suggest that NNs are more able to predict commercial bank failures than the logit model. According to Wu and Wang (2000), the back propagation method outperforms discriminant analysis and human judgment in predicting bank failures. Also, Miguel, Revilla, Rodríguez, and Cano (1993) show that the MLP is more successful than multivariate discriminant analysis, *k*-means cluster analysis, and logistic regression analysis in predicting the financial failure of Turkish banks.

Nevertheless, like some other approaches, NN also display some weaknesses: sometimes they can be seen like black boxes, being difficult the explanation of the prediction results and, on the other hand, NN can suffer from difficulties with generalization because of over fitting and they need a lot of time to train the models and to obtain the most adequate configuration.

Recently some new models have emerged, such as the random forest (Adusumilli, Bhatt, Wang, Bhattacharya, & Devabhaktuni, 2013; Booth, Gerding, & McGroarty, 2014; Calderoni, Ferrara, Franco, & Maio, 2015) and the support vector machines (Czarnecki & Tabor, 2014; Harris, 2015; Horta & Camanho, 2013; Kurtulmuş & Kavdir, 2014). Random forest (RF) is a general data mining tool proposed by Breiman (2001), in which a set of decision trees is generated on bootstrap samples of the data and then combined by majority voting. Random forest tries to create different decision trees to get the best classification among different classes of data according to the dependent variable. In the financial domain, RF has been successfully implemented for credit card fraud detection (Whitrow, Hand, Juszczak, Weston, & Adams, 2009) and banks customer churn prediction (Xie, Li, Ngai, & Ying, 2009). But, similar to NN and simple decision trees, the random forest technique can suffer from over fitting and requires a great deal of data in order to deliver reliable predictions.

Support vector machines (SVM) use a linear model to implement nonlinear class boundaries by mapping input vectors into a high-dimensional feature space. In the new space, an optimal separating hyperplane is constructed. The maximum margin hyperplane gives the maximum separation between decision classes. The training examples that are the closest to the maximum margin hyperplane are called support vectors.

SVM has the advantage of being simple enough to be analyzed mathematically. As suggested by Min and Lee (2005), SVM may serve as a promising alternative by combining the strengths of theory-driven conventional statistical methods and data-driven machine learning methods. SVM has proved its usefulness in some financial domains such as credit ratings, the detection of insurance claim fraud, and corporate failure prediction (Angelini et al., 2007; Chen, 2011a; Chen & Li, 2014; Erdogan, 2012; Harris, 2013; Wu & Liu, 2007).

Despite this good performance, the question about whether this approach is the most suitable one to separate healthy and unhealthy banks still remains. Boyacioglu, Kara, and Baykan (2009) evaluate four different neural network models, support vector machines and three multivariate statistical methods to the problem of predicting bank failures. Although SVMs provide satisfying prediction performance, these authors corroborate the superiority of MLP in prediction problems given the difficulties in the selection of the kernel and the slow behavior in the test phases. These assertions are consistent with Lee, Booth, and Alam (2005).

Recent research has developed a hybrid intelligent model to combine the advantages of individual models and avoid their weaknesses (Sánchez-Lasheras, de Andrés, Lorca, & de Cos Juez, 2012; Tsai, Hsu, & Yen, 2014; Xu, Xiao, Dang, Yang, & Yang, 2014; Zhou, Lai, & Yen, 2012). A technique is called hybrid if several soft computing approaches are applied in the analysis and only one predictor is used to make the final prediction, or outputs of several predictors are combined, to obtain an ensemble-based prediction. The result of hybrid models might be more accurate than either of the techniques used separately (Kainulainen et al., 2014). Consistent with this approach we also combine two kinds of networks (a MLP and a SOM) to predict banking failures. Detailed information on both networks is presented in Section 3.

3. Bankruptcy prediction model: empirical design

3.1. The model for bankruptcy prediction

The methodology of NN is an efficient way to develop dynamic models for bankruptcy prediction because this approach takes into consideration the financial environment of firms and it offers a good tool for early warning system models (Altman, Marco, & Varetto, 1994; Davis & Karim, 2008). That previous models proved unable to predict the recent wave of bank failures, primarily in the United States, does not invalidate the NN approach. Rather, it suggests the need to improve the models by considering the specific issues related to the recent global financial crisis.

In a NN application, some main issues must be defined, including the network structure and the learning and the optimization algorithm, among others. Whereas the most used network structures are layered and completely connected, the three typologies of learning mechanisms are supervised learning, unsupervised learning, and reinforced learning. Supervised learning is applied when the network must learn to generalize some given examples. Unsupervised learning is used for tasks in which regularities are sought from a large amount of data. Reinforced learning algorithms are applied to train adaptive systems that perform a task composed of a sequence of actions.

We use two kinds of networks: a MLP and a SOM. The MLP networks, which have shown their ability to predict financial distress better than other methods (Vellido et al., 1999), are supervised networks that assign a probability of failure at one, two, and three years ahead of the current balance information. MLPs, which usually consist of one input, one hidden layer, and one output layer, create a classificatory and pattern detection process. For the hidden layer to serve any useful function, multilayer networks must have nonlinear activation functions for the multiple layers. The backpropagation algorithm is one of the most widely applied methods to layered feedforward networks (Huang, Li, & Xiao, 2015; Wang, Zeng, & Chen, 2015). Learning occurs in the perceptron by changing connection weights after each piece of data is processed, based on the amount of error in the output compared to the expected result.

The SOM, which is an unsupervised network, creates a visual representation of banks according to their profile of risk in different periods of time. In this kind of network, neurons learn in an unsupervised way because the network is not required to provide a specific objective. Thus, the network must discover the common patterns among the inputs; in other words, the neurons must self-organize conditional on the outside data. SOMs are competitive networks so that the neurons compete to provide the right answer, with only one neuron (or one node of neurons) becoming activated when a data pattern is presented.

As shown in Fig. 1, in the SOM the neurons of the input layer are connected with all the neurons of the output layer through synaptic weights. Thus, it is possible to establish in a bi-dimensional map different zones such as failing and non-failing regions and create bankruptcy trajectories (Chen, Ribeiro, Vieira, & Chen, 2013; Chen, Ribeiro, Vieira, Duarte, & Neves, 2011; Kiviluoto, 1998; Serrano-Cinca, 1996). Nour (1994) summarizes a SOM learning algorithm as follows:

1. As the first step, in $t = 0$ $W_i(t)$ is set randomly. In this moment the maximum number or possible iterations in the training phase of the network (T) is defined.
2. Present an input vector X to the network, and compute the distance (similarity) D using the Euclidean metric to find the closest matching unit c, to each input vector.

$$d_{i,j,(t)} = \sqrt{\sum_{h=1}^{k} (W_{i,j,h} - X_k)^2}$$

3. Update the weight vector according to the following rule:

$$W_{jik}(t+1) = W_{jik}(t) + \alpha \cdot \lfloor X_k(t) - W_{jik}(t) \rfloor$$

where α is the learning ratio, $X_k(t)$ is the input pattern in t and $W_{jik}(t)$ is the synaptic weight that connects the k input with the (j,i) neuron in t. The neighborhood function allows actualizing the weights of the winning neuron and of the neighbor neurons to localize similar patterns too. The neighborhood radio decreases with the number of iterations of the model to achieve a better specialization for each neuron.

4. The process goes on an iterative way until t reaches the maximum number of iterations T and then it jumps back to step 2.

We use the MLP to create three different prediction models of failure depending on the time horizon that we cover. First we develop a model with the most predictive ratios to detect failures one year before the failure happens. Then, we iterate the procedure with information from two and three years before the bankruptcy, respectively. We combine the output of all these models (i.e., the default likelihood of each bank in each year) to build a SOM. Our aim is to identify failing and non-failing areas to create a bi-dimensional map that enhances the study of different risk profiles. This map provides a visual representation of the risk profile of each bank and can be a useful tool for supervisor authorities who may take preventive measures in a weak bank some time before the failure arises.

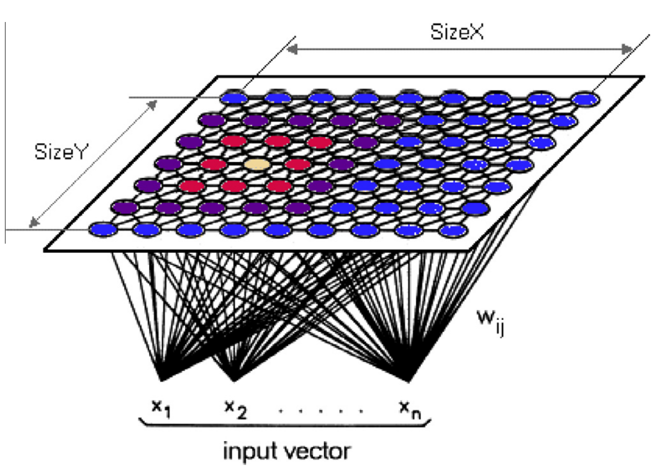

Fig. 1. SOM graphical representation.

3.2. Empirical design of the model

We obtain our data from the Federal Deposit Insurance Corporation (FDIC). All U.S. banks must report their financial statements quarterly in the Uniform Bank Performance Report so that the information then becomes publicly available. For each bank, this report includes information about the bank's loans portfolio, default rate, capital composition, liquidity, and so on. We select a set of 32 variables that are potentially explanatory for bankruptcy risk. Most of these variables have been previously used by other authors, which enhances the comparability of our results (Hammer, Kogan, & Lejeune, 2012; Lin, Liang, & Chen, 2011; Martínez, 1996; Nour, 1994). Consistent with du Jardin (2010), we also add some variables suitable to control for specific features of the current global financial crisis. These variables are selected with a criterion adapted to the network to improve the results of the model. We report the name and definition of the all initial variables in Table 1. Ratios are grouped into five different sets of variables: Ratios 1–14 measure the bank's earnings, ratios 15–21 assess the asset structure of each bank, ratios 22–25 go deeper on the assets and assess the loan portfolio, ratios 26–29 measure risk concentration, and ratios 30–32 are related to solvency. This approach is close to the CAMEL rating system established by the FDIC to conduct banking supervision.

Our training sample covers the period from December 2002 to May 2012. We test the predictive power of our model by applying it to banks that actually went bankrupt between May 2012 and December 2013. Between 2003 and 2013, the FDIC reported 516 U.S. banks failures. However, given our set of variables, the necessary information was available for only 386 banks. This sample size of distressed banks is consistent with previous research (Bernhardsen, 2001; Chen, 2011a; Ingaramo et al., 2005; Moro, Cortez, & Rita, 2015).

To isolate the main drivers of bankruptcy, we select a random sample among the remaining U.S. FDIC member banks. In bankruptcy analyses, it is a common practice to use one-to-one match of failure and non-failure cases (Davies & Bouldin, 1979; Kurtulmuş & Kavdir, 2014; Wu & Liu, 2007). Consequently, our training sample is made up of 386 failed and 386 non-failed banks. Since most of our explanatory variables are ratios, it is expected that these variables may have fat tails with large positive and negative values, so that outliers can have a negative influence on the model performance. We follow the procedure suggested by Van Gestel et al. (2006) for outliers handling. The boundaries are defined using the winsorized mean procedure, where the maximum (and minimum) allowed values are defined by median $(x) \pm 3 \cdot \sigma$, with $\sigma = IQR(x)/2 \times 0.6745$ and IQR is the interquartile range. The sample test is made up of 52 banks that went bankrupt between May 2012 and December 2013 and other 52 non-failed banks. Table 2 reports some descriptive statistics of our sample. Table 2 also displays the Kolmogorov–Smirnov normality test. Because several of the variables are not normally distributed, the Mann–Whitney U test is more reliable than the t-Student test to compare the differences between failed and non-failed banks. We denote with a number suffix if the variable is reliable to compare failed banks and non-failed banks one, two and three years before the fail. As we can see, variables (e.g., BAL and MMDA) display noticeable changes in discriminant power across time.

Once we test the ability of the initial variables to detect differences between failed and non-failed banks, we have two possibilities for the selection of the most relevant variables. First, we can choose a set of ratios with low variation in their predictive ability across time. Second, we can use the most predictive ratios for each time horizon and design different models conditional on the time horizon considered. We balance both alternatives and opt for designing specific models for each time horizon and at the same

Table 1
List of variables and definition.

Classification	Name	Calculation
Performance	INC	Interest income/average assets
Performance	EXP	Interest expense/average assets
Performance	NET_INC	Net interest income/average assets
Performance	NON_INC	Noninterest income/average assets
Performance	NON_EXP	Noninterest expense/average assets
Performance	PROV	Provision for loans and leases receivables losses/average assets
Performance	EASST	Average earning assets/average assets
Performance	LNLCR	Gross loans and lease charge-off less gross recoveries/average total loan and leases
Performance	EFFCY	Efficiency ratio (average total costs/total assets)
Performance	ASST	Average assets per employee (millions)
Performance	YLD_LNLS	Yield of total loans and lease/average total loans
Performance	YLD_DOM	Yield of interest and fees on domestic office loans secured primarily by real state/average domestic real state loans
Performance	MBS_YLD	Interest on mortgage backed securities (MBS)/average MBS
Performance	HIGH_INT	Cost of interest on deposits higher $100,000/average deposits higher $100,000
Asset structure	BAL	On average, all interest-bearing balances due from depository institutions/total assets
Asset structure	NLLA	The sum of the averages for net loans and lease-financing receivables, held-to-maturity and available-for-sale securities, interest-bearing balances due from depository institutions, federal funds sold and resold, and trading-account securities/average total assets.
Asset structure	FIX	Average of bank premises, furniture, equipment and others/total assets
Asset structure	PREM	Average real estate owned other than bank premises/average total assets
Asset structure	MMDA	Average money market deposit account/average total assets
Asset structure	DEP	Total deposits as a percent of average assets
Asset structure	TMDEP	The sum of the averages for time certificates of deposit of $100,000 or more and other time deposits in amounts of $100,000 or more/average total assets
Loan portfolio	LNLL	Gross loan and lease losses/average total loans and leases
Loan portfolio	LNLR	Gross loan and lease recoveries/average total loans and leases
Loan portfolio	OFCR	Credit to the bank's executive officers, main shareholders as of the report date/total loans
Loan portfolio	OFCR_ASST	Credit to the bank's executive officers, main shareholders/total assets
Concentration	CONS	Construction, land development and other land loans plus closed end loans secured by family residential properties first liens plus revolving open-end loans plus loans secured by farmland plus secured by nonfarm nonresidential properties as a percentage of total capital
Concentration	CMID	Commercial and industrial loans to U.S. addressees in domestic offices plus commercial and industrial loans to non-U.S. addressees in domestic offices as a percentage of total capital
Concentration	CARDO	Credit card plans in domestic offices plus other revolving credit plans in domestic offices plus other consumer loans in domestic offices as a percentage of total capital
Concentration	REAL	Real estate loans 90+ day past due
Capital	LNEQ	Number of times net loans and lease-financing receivables exceed equity capital
Capital	NET_INCEQ	Net income/average total equity capital
Capital	RISK	Total risk-based capital/risk-weighted assets

time use the ratios with a certain stable predictive power. In any case, the complexity of such a process and the fact that the variables are not normally distributed advice the implementation of complex computational methods such as NNs (Lin, Shiue, Chen, & Cheng, 2009).

We complement the information from the Mann–Whitney test with the Gini index to assess the predictive power of the variables. This index is one of the most widely used tools for calibration and a metric of the model's ability to classify correctly the dependent variable. It provides a relative metric of how close an actual model or a variable is to the ideal model of prediction. The highest score is 100%, and the lower the index, the worse the predictive power of the model. To some extent, this process is a univariate predictive analysis that allows us to accomplish one of our objectives, namely, the identification of the characteristics of the banks likely to go bankrupt. By calculating the individual predictive power of each variable, we can assess its relative impact and its evolution across time. Depending on its degree of influence, this variable can be reinforced or dropped from the calibration process.

In Table 3 we report the predictive power for each variable and model, along with the Gini test for the differences between failed and non-failed banks. We select the variables with the highest predictive power for each period of time. The number of selected variables (in grey shadow and bold font) is, to some extent, an ad-hoc decision that must balance the advantages and disadvantages of choosing many variables. Too few variables can result in poor predictive power of the model; conversely, too many variables can lead to an overly complex model, with redundant information, higher computational costs, and the loss of some observations. Accordingly, we order the variables according to the Gini index and the Mann–Whitney test and select the variables with low correlation.[2]

Each MLP model is implemented using PASW20 software. The dependent variable is a dummy variable that equals 1 if the bank has gone bankrupt, and zero otherwise. The model's output is a set of relations among variables that explains bank defaults. Each model tries to predict bankruptcy for a different time horizon prior to the event. The MLP1 model predicts bank failures one year ahead; MLP2, two years ahead; and MLP3, three years ahead. Obviously, the shorter the difference between the year analyzed and the bankruptcy, the better the performance of the model is. The next step is to gather the output of each model in a SOM to define different risk profiles of banks depending on the areas of the map.

The prediction model is a set of three MLPs (MLP1, MLP2, and MLP3, depending on the time horizon for the bankruptcy prediction). Nine different variables are included in the three models (five in MLP1, four in MLP2, and five in MLP3). In Table 4, we provide the Pearson correlation coefficients among the most predictive variables included in the MLP1 model.[3]

[2] Two variables with high predictive power cannot be chosen simultaneously if they are highly correlated.

[3] All the correlations matrix are available on request for each model.

Table 2
Descriptive analysis of the sample.

Variable	Mean	Std. dev.	Min	Max	25%	50%	75%	Z K-S	U1	U2	U3
INC	5.23	1.16	1.39	17.31	4.53	5.11	5.86	1.59 (0.01)	0.00	0.07	0.00
EXP	2.15	0.88	0.00	4.75	1.50	2.09	2.72	1.16 (0.13)	0.00	0.00	0.00
NET_INC	3.08	1.14	−0.24	17.30	2.41	3.11	3.75	1.74 (0.01)	0.00	0.00	0.15
NON_INC	0.88	5.61	−7.01	131.78	0.14	0.42	0.78	10.82 (0.00)	0.00	0.00	0.00
NON_EXP	3.87	5.60	0.33	123.85	2.60	3.18	3.96	8.85 (0.00)	0.00	0.92	0.92
PROV	2.17	2.40	−0.35	17.09	0.33	1.22	3.54	4.86 (0.00)	0.00	0.00	0.00
EASST	92.38	5.51	35.72	103.88	90.48	93.31	95.73	3.25 (0.00)	0.06	0.06	0.00
LNLCR	2.34	2.89	−0.53	32.00	0.27	1.25	3.69	5.48 (0.00)	0.00	0.00	0.00
EFFCY	129.20	258.94	−916.43	5,812.00	64.85	85.05	137.81	9.47 (0.00)	0.00	0.00	0.05
ASST	4.99	2.98	0.24	42.60	3.32	4.36	5.77	4.30 (0.00)	0.00	0.00	0.00
YLD_LNLS	6.46	2.57	3.24	69.20	5.55	6.24	6.97	5.88 (0.00)	0.00	0.04	0.00
YLD_DOM	6.45	2.58	3.07	69.20	5.54	6.22	6.96	5.85 (0.00)	0.00	0.05	0.00
MBS_YLD	4.40	1.72	0.00	29.72	3.80	4.54	5.12	3.77 (0.00)	0.49	0.34	0.12
HIGH_INT	3.21	1.11	0.00	6.74	2.33	3.21	4.02	1.09 (0.18)	0.00	0.00	0.00
BAL	3.68	5.37	0.00	49.12	0.16	1.61	5.03	6.85 (0.00)	0.00	0.27	0.74
NLLA	90.47	4.67	27.30	98.47	88.39	91.25	93.51	2.76 (0.00)	0.00	0.41	0.01
FIX	2.03	1.67	0.00	14.31	0.84	1.63	2.82	3.13 (0.00)	0.26	0.34	0.32
PREM	1.60	2.47	0.00	16.66	0.08	0.59	2.03	7.16 (0.00)	0.00	0.00	0.00
MMDA	12.49	10.23	0.00	66.91	5.26	9.98	17.23	3.12 (0.00)	0.16	0.74	0.05
DEP	63.18	11.93	0.00	87.74	56.67	64.17	70.81	1.70 (0.01)	0.00	0.00	0.00
TMDEP	19.33	10.21	0.00	61.53	11.73	17.97	25.21	2.02 (0.00)	0.00	0.00	0.00
LNLL	2.45	2.97	0.00	32.71	0.33	1.33	3.77	5.66 (0.00)	0.00	0.00	0.02
LNLR	0.12	0.34	0.00	7.67	0.01	0.04	0.12	10.17 (0.00)	0.00	0.01	0.00
OFCR	2.00	2.56	0.00	25.27	0.29	1.33	2.79	6.01 (0.00)	0.01	0.46	0.49
OFCR_ASST	1.34	1.54	0.00	13.12	0.18	0.89	1.95	5.32 (0.00)	0.28	0.31	0.02
CONS	2,040.21	27,258.56	−27,045.46	750,687.50	408.11	668.86	1,240.75	12.48 (0.00)	0.00	0.00	0.00
CMID	311.88	3,548.07	−2,786.49	91,968.75	50.45	97.36	181.84	12.30 (0.00)	0.00	0.00	0.09
CARDO	61.22	514.37	−761.54	13,466.67	8.76	22.23	52.76	12.05 (0.00)	0.00	0.00	0.00
REAL	9.04	9.56	0.00	53.79	1.36	5.66	14.59	4.76 (0.00)	0.00	0.00	0.00
LNEQ	35.84	413.11	0.00	11,071.42	6.39	9.19	17.80	12.76 (0.00)	0.00	0.00	0.00
NET_INCEQ	−40.67	67.34	−563.79	98.03	−74.86	−14.62	6.77	4.38 (0.00)	0.00	0.00	0.00
RISK	766,183.98	8,222,516.17	1,493.80	222,501,589.20	64,683.53	142,636.65	340,695.63	12.86 (0.00)	0.00	0.00	0.00

Note: See Table 1 for variable definitions.
Mean, standard deviation, minimum, maximum, percentiles, the Kolmogorov–Smirnov (significance) test of normality and the U Mann–Whitney tests. See Table 1 for variable definitions.

When designing a multilayer network, two decisions are key. First, we must choose the right number of layers to maximize the accuracy and the precision of the model. Lee et al. (2005) and Zhang, Hu, Patuwo, and Indro (1999) show that models with one hidden layer can deal with most of the classification problems. The second decision concerns the number of neurons or units in the hidden layer. This number is an important part of the overall NN architecture and must balance two counteracting effects: Too many units can result in a problem of over fitting, and too few units can lead to an under fitting network (Khashman, 2010). Comparable research such as Sharda and Wilson (1996) and Tam and Kiang (1992) uses one hidden layer with 10 nodes, whereas Boritz and Kennedy (1995) use a single hidden layer with 9 nodes. The optimal number of nodes must provide the highest classification accuracy in the hold-out sample. This number can be chosen by trial-and-error tests or according to the formula $0.75i$ where $i = 1,\dots,I$ represents the number of variables under consideration (Harris, 2015; Olmeda & Fernández, 1997). We implement both procedures, and both suggest using four nodes in the MLP1 model, three nodes in MLP2, and four nodes in the MLP3 model. The change of scale to activate the hidden layer and to obtain the output layer is a sigmoid function.

The network must go through a learning process, which can be either online, batch, or stochastic. Although it requires more memory capacity, we use the batch learning because it yields a much more stable descent to the optimal adjust to the patterns. The rest of parameters are similar to the ones used in previous literature, with initial lambda of 0.0000005, initial sigma of 0.00005, the center of the interval fixed around zero, and displacement of +/−5. The optimization algorithm is the scaled conjugated gradient.

4. Results

This section reports the main results of our empirical analysis. First, we analyze the evolution of failed banks before the failure according to the variables considered in this study. Second, we describe the results of the models using the MLP. Then, we combine the MLP output with the SOM model and show the results. We also compare our models with some common approaches in the literature such as the discriminant analysis, the logistic regression, support vector machines and random forest.

4.1. The storyline of distressed banks

As previously stated, in Table 3 we report the predictive power of the 32 initial variables for each year according to the Gini index. Significant differences exist among their predictive power. The table provides some interesting insights about the evolution of distressed and non-distressed banks up to three years before bankruptcy. These results contribute to our first objective concerning the identification of the main characteristics of defaulted banks prior to their bankruptcy.

The shadowed cells are the five most significant variables one, two, and three years before the failure. We can differentiate three kinds of variables. The first group includes the variables with high predictive power both in the near future (one year) and in the medium-long term (two or three years). These variables are PROV (the importance of provisions), CONS (risk concentration on the construction industry), and LNEQ (equity support to loans). According to these variables, failed banks relative to the profile of non-failed banks have higher provisions (likely as a consequence of riskier

Table 3
Power predictive of each variable and means in each model.

Variable	1 year before the failure			2 years before the failure			3 years before the failure		
	Non-failed	Failed	Gini index (%)	Non-failed	Failed	Gini index (%)	Non-failed	Failed	Gini index (%)
INC	5.24	4.87	26.74	5.90	6.09	7.33	6.31	6.86	30.49
EXP	1.70	2.41	43.48	2.12	2.88	45.71	2.34	3.09	50.22
NET_INC	3.63	2.50	66.44	3.78	3.20	39.30	3.97	3.76	12.12
NON_INC	0.92	0.41	43.72	1.04	0.59	33.67	0.87	0.59	29.70
NON_EXP	3.26	3.78	18.28	3.51	3.53	6.36	3.62	3.71	0.15
PROV	0.69	3.46	76.20	**0.53**	**1.97**	**62.42**	**0.37**	**1.06**	**45.90**
EASST	92.47	91.80	7.14	93.22	93.17	2.56	93.37	94.00	10.75
LNLCR	0.80	3.84	74.97	0.68	1.72	44.01	0.36	0.80	22.96
EFFCY	**74.46**	**179.89**	**66.60**	75.75	101.75	34.43	74.77	93.46	13.40
ASST	4.38	5.51	30.41	4.19	5.32	34.73	4.08	5.00	32.94
YLD_LNLS	6.71	5.98	43.57	7.25	7.10	8.50	7.71	7.94	12.87
YLD_DOM	6.70	5.98	42.89	7.24	7.10	8.21	7.70	7.94	13.46
MBS_YLD	4.58	4.31	31.71	4.69	4.83	38.73	4.92	5.47	39.36
HIGH_INT	2.84	3.23	17.68	3.67	4.02	18.05	4.11	4.41	22.48
BAL	3.72	4.33	10.23	2.40	1.94	3.66	1.74	1.12	1.26
NLLA	91.40	89.17	31.98	91.55	90.98	6.66	91.85	92.08	6.53
FIX	1.78	2.20	8.67	1.83	2.18	6.85	1.86	2.11	4.57
PREM	**0.62**	**3.06**	**62.27**	**0.40**	**1.43**	**43.45**	0.23	0.60	22.00
MMDA	13.41	11.98	6.47	13.09	12.70	0.82	12.85	13.56	5.65
DEP	66.37	63.12	1.38	65.94	59.86	26.72	66.15	58.15	41.00
TMDEP	16.26	21.51	44.78	16.38	21.86	34.84	15.52	22.39	40.21
LNLL	0.89	4.00	74.06	0.78	1.80	40.70	0.44	0.84	16.74
LNLR	0.09	0.15	8.08	0.09	0.08	14.86	0.08	0.04	23.40
OFCR	2.03	1.80	4.54	2.09	2.19	2.86	2.29	4.45	3.08
OFCR_ASST	1.28	1.31	0.13	1.34	1.62	7.91	1.39	1.80	10.16
CONS	**463.10**	**3675.16**	**80.14**	**450.83**	**789.10**	**66.21**	**431.64**	**658.04**	**56.11**
CMID	90.23	547.60	42.45	91.57	119.33	15.30	91.23	102.90	7.81
CARDO	**36.90**	**79.42**	**11.51**	38.66	22.40	35.52	**41.44**	**20.75**	**40.70**
REAL	**145.97**	**244.74**	**80.17**	2.33	8.11	60.30	1.49	3.92	33.32
LNEQ	**6.56**	**66.94**	**89.61**	6.44	10.48	60.09	**6.32**	**8.41**	**42.37**
NET_INCEQ	**2.42**	**−83.74**	**91.26**	5.16	−20.18	58.55	7.85	−1.55	27.64
RISK	411,159.76	506,968.28	24.01	405,422.65	545,675.16	30.55	378,166.09	540,985.14	31.96

Note: See Table 1 for variable definitions.

Table 4
Correlation matrix in the MLP1 model.

Variable	NET_INCEQ	LNEQ	CONS	EFFCY	PREM
NET_INCEQ	1	−0.152	−0.056	−0.233	−0.355
LNEQ	−0.152	1	−0.003	0.029	0.017
CONS	−0.056	−0.003	1	−0.062	0.163
EFFCY	−0.233	0.029	−0.062	1	0.237
PREM	−0.355	0.017	0.163	0.237	1

Note: MLP1 = multilayer perceptron model for one year ahead bank failures. See Table 1 for variable definitions.

loans), are more concentrated on financing construction and residential properties, and have less equity relative to loans. Furthermore, the differences increase as the bank failure approaches. For instance, whereas the provision ratio (PROV) for failed banks is 2.86 times the ratio for non-failed banks (1.06 to 0.37) three years ahead, this proportion increases to 5.01 (3.46 to 0.69) the year before the financial distress. The same can be said for the concentration on the construction industry and for the equity.

The second group of variables includes the ones whose relative predicting power decreases as the bank failure approaches. These variables are EXP (interest expenses) and DEP (deposits). Table 3 shows that failed banks consistently pay higher interests than their non-failed counterparts over time, despite the decrease in the predicting power. Similarly, the proportion of deposits over total assets is initially lower in failed banks so that these banks are more leveraged.

The third group of variables includes the ones whose predicting power increases one or two years before the failure, namely, REAL (overdue real estate loans) and NET_INCEQ (income to equity). Once again, remarkable differences exist between both kinds of banks: Specifically, failed banks have more overdue loans to the real estate industry and receive lower income (even scaled by equity).

Taken together, these results provide a clear portrait of the crisis of U.S. banks in recent years. As a consequence of the U.S. business upturn fueled by low interest rates, financial institutions expanded rapidly to gain market share as quick as possible. The real estate boom along with low interest rates compelled banks to grant loans to construction and land development irrespective of the credit quality. Distressed banks had to pay back higher interest rates to depositors to raise money to reinvest in the real estate industry. Due to the business downturn in 2008 and 2009, along with the fall of the prices of real estate collateral, these banks faced a growing default rate, had to create more provisions, and accumulated a troublesome portfolio of real estate. The provisions impacted earnings negatively, and the solvency of the banks worsened. The liquidity crisis constrained the possibility of improving the solvency of the banks through equity issuance, and banks had to pay higher interest rates. These growing interests on deposits along with troublesome loans led to a negative margin. This vicious circle could not be maintained for long time, and finally financial authorities intervened.

4.2. The MLP model to predict failures

The predictions about possible future bankruptcies are done through models MLP1 to MLP3, and the results are reported in Table 5. This table provides more detailed information on the comparison among models. We measure the power of the models through receiver operating characteristic (ROC) curves. The ROC curves convey the same information as the confusion matrix but

Table 5
Performance of each model.

Model	Training squared errors	Trial squared errors	Area below ROC
MLP3	106.978	7.738	0.87
MLP2	90.097	5.660	0.91
MLP1	60.403	2.500	0.95

Note: ROC = receiver operating characteristic. MLP1 (MLP2; MLP3) = multilayer perceptron model for one (two; three) year ahead bank failures.

Table 6
Models comparison.

Method	Years before failure		
	1 (%)	2 (%)	3 (%)
Non-failed banks			
DA	78.85	76.92	71.15
LR	82.69	84.62	76.92
RF	88.46	76.92	75.00
MLP	92.31	86.54	84.62
SVM	90.38	88.46	82.69
Failed banks			
DA	76.92	71.15	69.23
LR	80.77	78.86	73.08
RF	86.54	80.77	76.92
MLP	94.23	84.69	80.77
SVM	88.46	86.54	82.69

Correct classification rates for alternative models: discriminatory analysis (DA), logit regression (LR), random forest (RF), multilayer perceptron (MLP) and support vector machine (SVM).

in a more visual, intuitive, and robust way (Lee & Choi, 2013). The accuracy of the test depends on how well the model separates failed and non-failed banks. The area below the ROC curve is 1 when the model has a complete discriminatory power and 0.5 when the model has no discriminatory power. As shown in Table 5, the performance of all our models is quite acceptable, and it increases as the bank moves closer to the failure date: that is, the closer the bank is to failure, the lower the training and trial squared errors are and the higher the area is below the ROC curve.

To assess the performance of NNs as banking failure prediction methods, we compare the discriminatory power with the output of two traditional techniques (the discriminant analysis and the logistic regression) and with two more recent and flexible models (support vector machines and random forest). Table 6 shows the correct classification rates of the five methods. Three main issues arise from this table. First, there are two models (MLP and SVM) with better performance than the other ones. Both form non-failed banks and for failed banks irrespective of the time framework, MLP and SVM deliver higher correct classification rates. Second, for non-failed banks, MLP usually outperforms SVM. Third, for failed banks, MLP outperforms SVM in the short term (one year before failure) but SVM outperforms in the medium-long term (two and three years before failure). Taken together, these results corroborate the superiority of the flexible and emerging techniques over the traditional methods and a certain outperformance of MLP. These results are consistent with previous research (Chen et al., 2013).

NNs have sometimes been criticized for being like black boxes because the manner in which they receive the inputs and provide the output is not very transparent. The complexity of the mathematical and algorithmic elements makes the flow of information through the net difficult to understand. One of the most widely used methods to cope with this criticism is the so-called sensitivity analysis (Garson, 1991; Hunter, Kennedy, Henry, & Ferguson, 2000; Rambhia, Glenny, & Hwang, 1994; Zurada, Malinowski, & Cloete, 1994). The sensitivity analysis is based on measuring the observed effect on an output Y_j due to the change in an input X_i: the bigger the effect, the more sensitive. Figs. 2–4 show the results of each sensitivity analysis. The importance of an independent variable is a measure of how much this variable affects the probability of distress. As shown in Fig. 3, for the predictions one year ahead of bankruptcy, the most important variable is net income as a percentage of total capital, followed by the efficiency ratio and total loans to construction and real estate. Figs. 4 and 5 show that during the years prior to bankruptcy the most determinant variables of the possibility of banking failure are provisions and interest expenses.

4.3. SOM to monitor potential bank risk

Once we have checked the validity of the MLP method for bankruptcy prediction, we gather the output of the models to create a visual tool to detect the risk of bank failure. Similar to other prediction models that provide a dichotomous output (i.e., failing vs. non-failing entities), SOM models give a wide range of intermediate possibilities. In this sense, SOMs are an interesting tool for decision-making due to their capability to project multidimensional data onto a less dimensional output space. We first combine the output of the MLP models (i.e., the likelihood of default one, two, and three years ahead) to create a map in which healthy banks

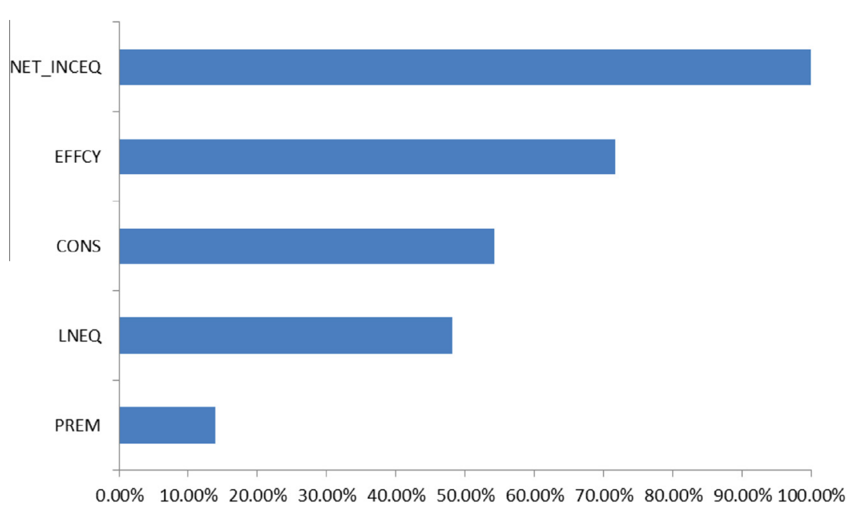

Fig. 2. Sensibility analysis one year before the failure (MLP1 model).

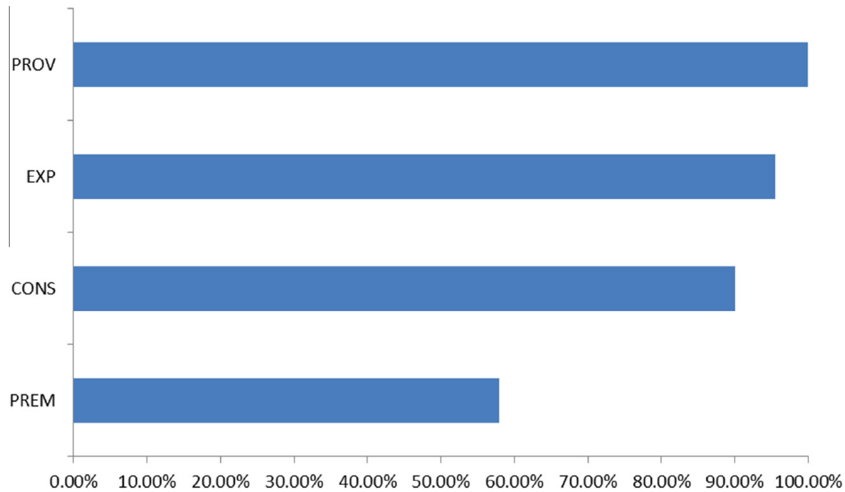

Fig. 3. Sensibility analysis two years before the failure (MLP2 model).

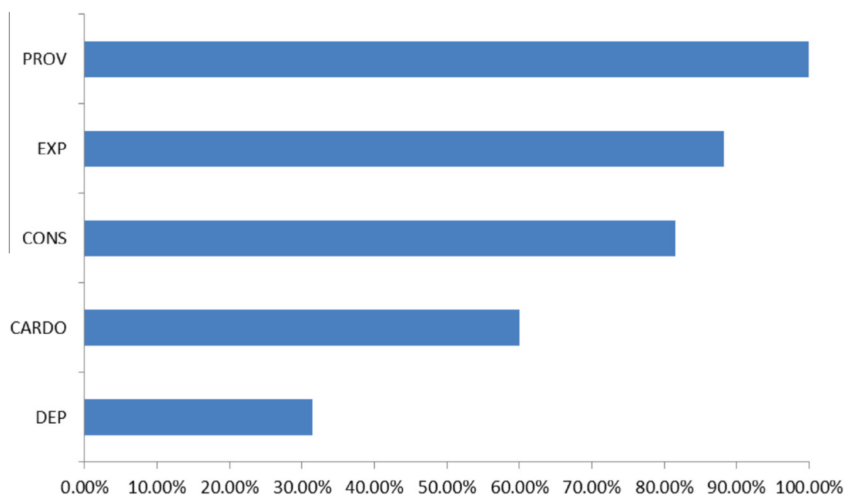

Fig. 4. Sensibility analysis three years before the failure (MLP3 model).

are displayed clearly away from unhealthy banks. We then differentiate more groups inside the unhealthy banks area to assess dynamically the solvency and the potential risks of each bank across time. In so doing, the model provides a time framework with information about how long before the bankruptcy the threats arise, so that correctional measures can be taken to reverse the situation and avoid the failure.

To build the map we use the same training and test samples as in the previously reported models. The first layer of the net has three input patterns, and the output layer is a bi-dimensional 21×7 map. The size of the map follows the recommendations of Kohonen (1993) and Kaski and Kohonen (1994). The results of the training phase are shown in Fig. 5. In the so-called U-matrix we display the distances among the neurons by training the model. Different colors represent the varying distances: light colors for short distances and dark colors for long distances. Given the bi-dimensional nature of the map, the differences and distances among neurons can proxy possible groups of neurons or clusters. Thus, the U-matrix can be a useful tool for visualizing clusters in the input data without any a priori information about these clusters.

The other three figures in Fig. 5 show the density function of the bankruptcy probability for each year. They enhance the visualization of the cluster structure and the correlation among the input variables. The bottom areas comprise the banks with the highest default probabilities, and the upper areas comprise the most solvent banks. Fig. 6 complements Fig. 5; in Fig. 6 we display the neurons distribution on the map and draw a line to differentiate the solvency zone (blue area) and the bankruptcy zone (red area). We use the proportion of failed and non-failed banks in the sample to label each neuron as failed or non-failed.

The bankruptcy area is bigger than the solvency area and accounts for 58.5% of the cells. If we apply this map to the test sample, the model predicts 98.08% of the bank failures one year ahead. This result means that the Type I error is 1.52%. In addition, the model correctly classifies 94.23% of the non-failed banks, so the Type II error is 5.77%.

Chen (2011a) suggests some advanced metrics to assess the quality and to compare different classification models. In Table 7 we report the performance of our hybrid model and the comparison with the other techniques. The overall accuracy is the percentage of correctly classified instances. Precision is defined as the number of classified positive or abnormal instances that actually are positive instances. Sensitivity measures how well a classifier can recognize abnormal records. Specificity refers to how well a classifier can recognize normal records.

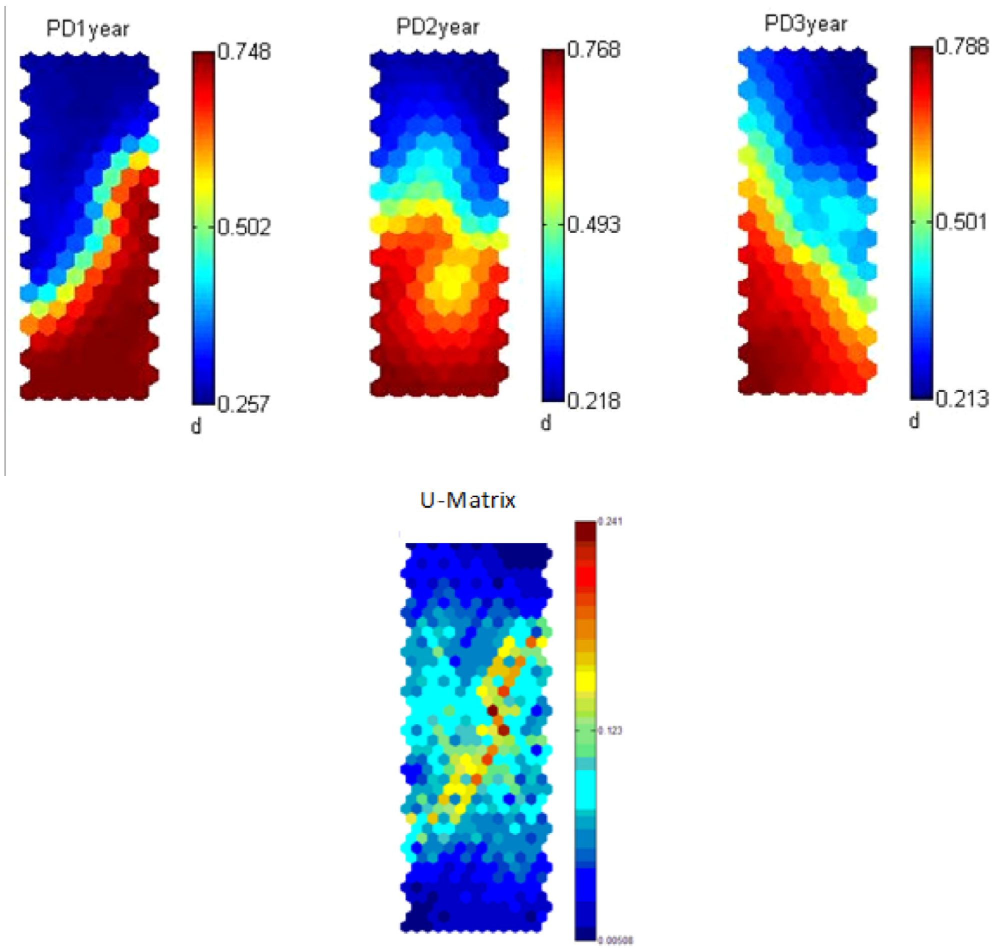

Fig. 5. Learning results of SOM.

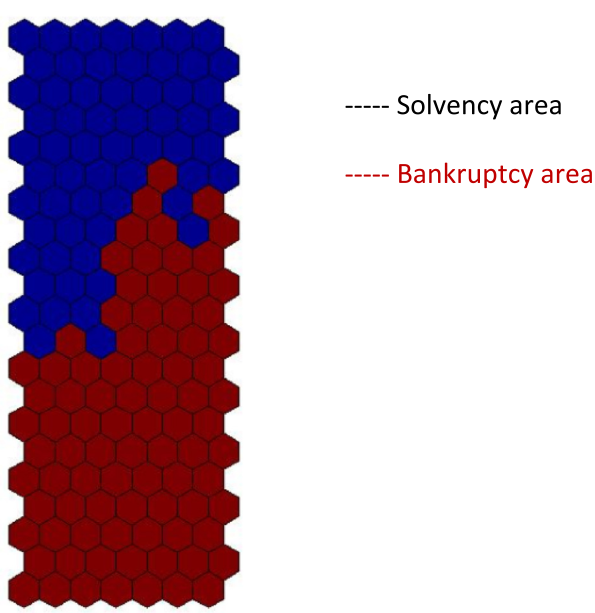

Fig. 6. Distribution of neurons in the Kohonen map and the solvency zones.

We note a gap between the performance of MLP, SVM and our hybrid MLP-SOM model and the performance of the random forest and the traditional techniques (DA and LR). As expected, the performance of all the models deteriorates as long as the time before failure increases. Nevertheless, the decline of the MLP, SVM and MLP-SOM models is lower than the alternative models[4]. More interestingly, our hybrid method usually outperforms even the MLP and SVM procedure. In addition, the MLP-SOM model has a high and stable predictive power over the time, reaching a balance between Type I and Type II errors.

4.4. Creating several groups of risk in the SOM

Along with the prediction on bankruptcy likelihood, another relevant output of the model is the time horizon in which a bank may fail. To improve the predictive power of the model, further clusters might be identified within the two initial groups of banks (failed vs non failed). We use *k*-means to create different zones in the original map displayed in Fig. 6. Although several methods can be used to determine the optimal number of groups, one of the most widely used algorithms in SOM is the Davies–Bouldin index (Lin et al., 2011). This index is a function of the within-cluster variation to the between-cluster variation ratio (Ingaramo et al., 2005). According to this index, the optimal number of groups is six, as shown in Fig. 7. Based on this classification, in Table 8 we report the mean value of the failure likelihood for each group one, two, and three years ahead.

[4] For instance, in terms of overall accuracy, RF correctly classifies 87.5% of banks one year before, but only 78.85% of banks two years before. Similar results hold for other metrics and models.

Table 7
Classification results by model.

	Overall accuracy			Precision			Sensitivity			Specificity		
	1 year	2 years	3 years	1 year	2 years	3 years	1 year	2 years	3 years	1 year	2 years	3 years
DA	77.88	74.04	70.19	78.43	75.51	70.59	76.92	71.15	69.23	78.85	76.92	71.15
LR	81.73	81.73	75.00	82.35	83.67	76.00	80.77	78.85	73.08	82.69	84.62	76.92
RF	87.50	78.85	75.96	88.24	77.78	75.47	86.54	80.77	76.92	88.46	76.92	75.00
MLP	93.27	85.58	82.69	92.45	86.27	84.00	94.23	84.62	80.77	92.31	86.54	84.62
SVM	89.42	87.50	82.69	90.20	88.24	82.69	88.46	86.54	82.69	90.38	88.46	82.69
MLP-SOM	96.15	90.38	84.62	94.44	92.00	87.50	98.08	88.46	80.77	94.23	92.31	88.46

Performance (in percentage) of the classification models. The overall accuracy is the percentage of correctly classified instances. Precision is the number of classified positive or abnormal instances that actually are positive instances. Sensitivity is how well a classifier can recognize abnormal records. Specificity is how well a classifier can recognize normal records. Discriminatory analysis (DA), logit regression (LR), random forest (RF), multilayer perceptron (MLP) and support vector machine (SVM) are compared.

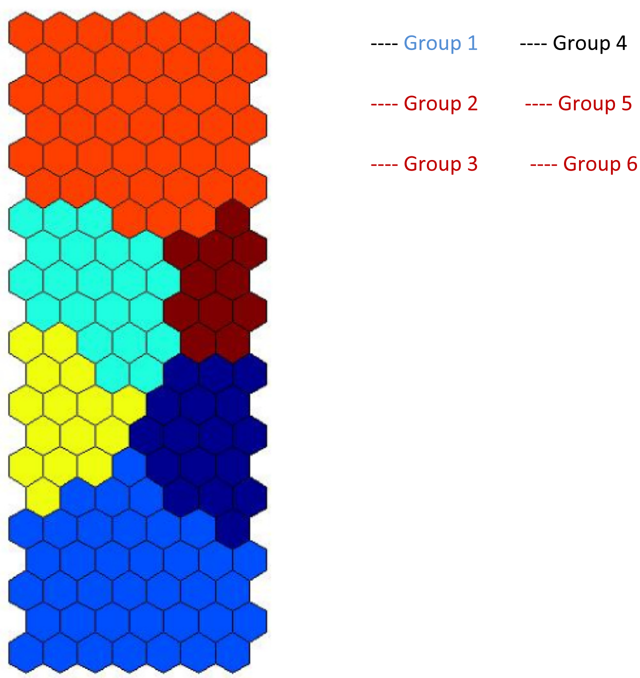

Fig. 7. Distribution of neurons in the Kohonen map according the different profiles of the banks.

Table 8
Probability of distress by group.

Group	Proportion of failed banks (%)	Mean of P.1 year (%)	Mean of P.2 years (%)	Mean of P.3years (%)
Group 1	96.28	73.83	73.55	74.56
Group 2	89.87	73.12	63.01	48.64
Group 3	86.67	71.23	32.28	36.30
Group 4	57.81	33.84	68.09	71.91
Group 5	14.29	30.83	45.66	48.70
Group 6	2.30	27.02	25.34	26.46

In the same vein, Table 9 reports the values of the main variables by group. The comparison of these probabilities with the actual rate of bankruptcy provides some interesting insights on the timing of the bankruptcy symptoms. Banks in group 1 have important weaknesses over the entire period, both in the short and long term. Banks in groups 2 and 3 are examples of short-term concerns but not long-term concerns: although they have high rates of failure, these bankruptcies are difficult to predict three years ahead. Banks in group 4 are in the opposite situation: the probability of distress decreased across time so that failures are difficult to predict one year prior to bankruptcy. Finally, banks in groups 5 and 6 have the lowest rate of predicted failure, and, in fact, their actual rate of failure is lower than their predicted rate for both the short and the long term.

Table 9
Average of the main variables by group.

Group	REAL	EXP	CONS	LNEQ	NET_INCEQ	DEP	PROV
1	95.53	2.80	2,179.76	36.69	−43.05	60.51	2.48
2	100.83	2.68	1,034.75	17.78	−31.51	62.12	2.10
3	24.59	2.51	710.85	11.33	−23.92	62.24	1.41
4	14.85	2.91	712.36	9.09	−5.01	56.96	1.26
5	3.07	2.53	575.20	7.70	6.00	60.13	0.80
6	64.31	1.88	381.71	5.84	8.55	68.08	0.29

Note: See Table 1 for variable definitions.

The proportion of bank failures inside each group is complemented by the mean of the most predictive variables as previous discussed (Table 3). The differences between groups 1, 2, and 6 are considerable for all the variables; namely, group 6 has lower level of expenses, construction exposure, provisions, and real state past due and a higher level of deposits compared to the other two groups. However, banks that are clearly identifiable as "safe" or "non-safe" are not a big challenge to regulators and policymakers. They are more likely interested in the intermediate groups—that is, groups 3, 4, and 5—because they may be a threat in a short or medium term. The first two groups (groups 3 and 4) are very close; however, in group 4 all the variables improve slightly, particularly real state past due and net income to total equity. As noted previously, these slight differences contribute to change in the risk profile across time. Threats in group 4 are derived of high levels of construction loans and an increase in the level of expenses of provisions. The probability of default is moved to two or three years ahead considering potential worsening of the economy and the increase in past due loans. Also the importance of deposits in the financial structure of the banks is lower than the other groups. These results suggest that regulators should place their highest priority on group 3 and their next level of priority on group 4.

5. Summary and conclusions

The recent financial crisis and the globalization process have accelerated the obsolescence of bankruptcy prediction models and emphasized the need of reformulation. We develop a model of neural networks to study the bankruptcy of U.S. banks, taking into account the specific features of the current financial crisis. We combine multilayer perceptrons and self-organizing maps to provide a tool that displays the probability of distress up to three years before bankruptcy occurs.

Previous research has proposed many statistical and intelligent methods to predict corporate bankruptcy, although there is no overall best method that has been used (Chen et al., 2013; Fedorova, Gilenko, & Dovzhenko, 2013; Lee & Choi, 2013; Xu et al., 2014). We build on this previous research and go a step ahead in several domains. First, our model outperforms most of the previous ones in terms of predicting ability. Second, the output of our model is compared against a wider set of alternative methods than previous papers do. Third, our model is simpler and, at the same time, provide a clearer visualization of the complex temporal behaviors.

Our procedure can be a useful tool for bank supervisors and other stakeholders to delineate the risk profile of each bank. As far as the supervisory authority is concerned, the preventive measures to correct imbalances can be different in the short, medium, or long term depending on the probability of banking failure—that is, on the group to which the bank belongs. Investors, depositors, and other participants in capital markets can assess the risk profile of their investment and, consequently, define their optimal risk-return combination.

This study has three limitations. First, as many other hybrid techniques, our model needs multiple calculations to deliver all the output and a complete visualization. Second, in spite of the fact that we consider some features of the financial crisis, we do not control for macroeconomic factors potentially affecting the banks propensity to fail. Third, we focus on commercial banks, so there may be some concerns about whether our results can be applied to large investment banks, whose failure has more far-reaching consequences.

Although our hybrid model performs better than the existing methods for bankruptcy prediction, there is still room for improvement. One direction for future research is the extension of our model to the international framework to determine to what extent different national regulations address credit risk and to identify the best performing regulations. Another application may be the assessment of whether the wave of mergers and acquisitions among banks reduces the risk of bankruptcy and reinforces the solvency of financial institutions. Given the support to hybrid models that our research lends, another meaningful future work could be combining alternative sets of techniques to achieve an optimal balance between prediction accuracy and interpretation simplicity.

Acknowledgments

The authors would like to thank the editor-in-chief, Binshan Lin, two anonymous reviewers, Alisa Larson, and David Moreno for their insightful comments and suggestions. The financial support of the Spanish Ministry of Science and Innovation (ECO2011-29144-C03-01) is gratefully acknowledged. A previous version of this research has been included as Working Paper 568 of the Spanish Saving Banks Foundation (FUNCAS). All the remaining errors are the authors' only responsibility.

References

Adusumilli, S., Bhatt, D., Wang, H., Bhattacharya, P., & Devabhaktuni, V. (2013). A low-cost INS/GPS integration methodology based on random forest regression. *Expert Systems with Applications, 40*, 4653–4659.
Altman, E. (1968). Financial ratios, discriminant analysis and prediction of corporate bankruptcy. *Journal of Finance, 23*, 589–609.
Altman, E., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks. *Journal of Banking and Finance, 18*, 505–529.
Angelini, E., di Tollo, G., & Roli, A. (2007). A neural network approach for credit risk evaluation. *Quarterly Review of Economics and Finance, 48*, 733–755.
Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks, 12*, 929–935.
Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accountant Research, 4*, 71–111.
Bernhardsen, E. (2001). A model of bankruptcy prediction. In *Working paper: Financial analysis and structure department*. Norges Bank.
Booth, A., Gerding, E., & McGroarty, F. (2014). Automated trading with performance weighted random forests and seasonality. *Expert Systems with Applications, 41*, 3651–3661.
Boritz, J. E., & Kennedy, D. B. (1995). Effectiveness of neural networks types for prediction of business failure. *Expert Systems with Applications, 9*, 503–512.
Boyacioglu, M. A., Kara, Y., & Baykan, Ö. K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications, 36*, 3355–3366.
Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.
Calderoni, L., Ferrara, M., Franco, A., & Maio, D. (2015). Indoor localization in a hospital environment using random forest classifiers. *Expert Systems with Applications, 42*, 125–134.
Czarnecki, W. M., & Tabor, J. (2014). Two ellipsoid support vector machines. *Expert Systems with Applications, 41*, 8211–8224.
Chen, C.-C., & Li, S.-T. (2014). Credit rating with a monotonicity-constrained support vector machine model. *Expert Systems with Applications, 41*, 7235–7247.
Chen, M.-Y. (2011a). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers & Mathematics with Applications, 62*, 4514–4524.
Chen, M.-Y. (2011b). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications, 38*, 11261–11272.
Chen, N., Ribeiro, B., Vieira, A., & Chen, A. (2013). Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert Systems with Applications, 40*, 385–393.
Chen, N., Ribeiro, B., Vieira, A. S., Duarte, J., & Neves, J. C. (2011). A genetic algorithm-based approach to cost-sensitive bankruptcy prediction. *Expert Systems with Applications, 38*, 12939–12945.
Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1*, 224–227.
Davis, E. P., & Karim, D. (2008). Comparing early warning systems for banking crises. *Journal of Financial Stability, 4*, 89–120.
du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing, 73*, 2047–2060.
Erdogan, B. E. (2012). Prediction of bankruptcy using support vector machines: An application to bank bankruptcy. *Journal of Statistical Computation and Simulation, 83*, 1543–1555.
Fedorova, E., Gilenko, E., & Dovzhenko, S. (2013). Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications, 40*, 7285–7293.
Fulmer, J. G., Moon, J. E., Gavin, T. A., & Erwin, J. M. (1984). A bankruptcy classification model for small firms. *Journal of Commercial Bank Lending, 66*, 25–37.
Garson, G. D. (1991). A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review, 9*, 399–434.
Hammer, P. L., Kogan, A., & Lejeune, M. A. (2012). A logical analysis of banks' financial strength ratings. *Expert Systems with Applications, 39*, 7808–7821.
Harris, T. (2013). Quantitative credit risk assessment using support vector machines: Broad versus narrow default definitions. *Expert Systems with Applications, 40*, 4404–4413.
Harris, T. (2015). Credit scoring using the clustered support vector machine. *Expert Systems with Applications, 42*, 741–750.
Hol, S. (2006). The influence of the business cycle on bankruptcy probability. In *Discussion papers statistics norway research department*.
Horta, I. M., & Camanho, A. S. (2013). Company failure prediction in the construction industry. *Expert Systems with Applications, 40*, 6253–6257.
Huang, H.-X., Li, J.-C., & Xiao, C.-L. (2015). A proposed iteration optimization approach integrating backpropagation neural network with genetic algorithm. *Expert Systems with Applications, 42*, 146–155.
Hunter, A., Kennedy, L., Henry, J., & Ferguson, I. (2000). Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Computer Methods and Programs in Biomedicine, 62*, 11–19.
Ingaramo, D. A., Leguizamón, G., & Errecalde, M. (2005). Adaptive clustering with artificial ants. *Journal of Computer Science and Technology, 5*, 264–271.
Kainulainen, L., Miche, Y., Eirola, E., Yu, Q., Frénay, B., Séverin, E., et al. (2014). Ensembles of local linear models for bankruptcy analysis and prediction. *Case Studies in Business, Industry and Government Statistics, 4*, 116–133.
Kaski, S., & Kohonen, T. (1994). Winner-take-all networks for physiological models of competitive learning. *Neural Networks, 7*, 973–984.
Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications, 37*, 6233–6239.
Kiviluoto, K. (1998). Predicting bankruptcies with the self-organizing map. *Neurocomputing, 21*, 191–201.
Kohonen, T. (1993). Physiological interpretation of the self-organizating map algorithm. *Neural Networks, 6*, 895–905.
Kurtulmuş, F., & Kavdir, İ. (2014). Detecting corn tassels using computer vision and support vector machines. *Expert Systems with Applications, 41*, 7390–7397.
Lee, K., Booth, D., & Alam, P. (2005). A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms. *Expert Systems with Applications, 29*, 1–16.

Lee, S., & Choi, W. S. (2013). A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications, 40*, 2941–2946.
Levy-Yeyati, E., Martínez Pernía, M. S., & Schmukler, S. L. (2010). Depositor behavior under macroeconomic risk: Evidence from bank runs in emerging economies. *Journal of Money, Credit and Banking, 42*, 585–614.
Lin, F., Liang, D., & Chen, E. (2011). Financial ratio selection for business crisis prediction. *Expert Systems with Applications, 38*, 15094–15102.
Lin, S. W., Shiue, Y. R., Chen, S. C., & Cheng, H. M. (2009). Applying enhanced data mining approaches in predicting bank performance. A case of Taiwanese commercial banks. *Expert Systems with Applications, 36*, 11543–11551.
Martínez, I. (1996). Forecasting company failure: Neural approach versus discriminant analysis: An application to Spanish insurance companies. In: G. Sierra Molina & E. Bonsón Ponte (Eds.), *Intelligent systems in accounting and finance* (pp. 169–185). Huelva.
Miguel, L. J., Revilla, E., Rodríguez, J. M., & Cano, J. M. (1993). A comparison between statistical and neural network based methods for predicting bank failures. In: *Proceedings of the III international workshop on artificial intelligence in economics and management*. Portland.
Min, J. H., & Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications, 28*, 603–614.
Mokhatab Rafiei, F., Manzari, S. M., & Bostanian, S. (2011). Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. *Expert Systems with Applications, 38*, 10210–10217.
Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications, 42*, 1314–1324.
Mukta, P., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications, 36*, 2–17.
Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications, 41*, 7653–7670.
Nour, M. A. (1994). *Improved clustering and classification algorithms for the Kohonen self-organizing neural network* (PhD dissertation). Kent State University, Unpublished.
Ohlson, J. (1980). Financial ratios and probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*, 109–131.
Olmeda, I., & Fernández, E. (1997). Hybrid classifier for financial multicriteria decision making: The case of bankruptcy prediction. *Computational Economics, 10*, 317–335.
Oreski, S., & Oreski, G. (2014). Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications, 41*, 2052–2064.
Piramuthu, S. (1999). Financial credit-risk evaluation with neural and neurofuzzy systems. *European Journal of Operational Research, 112*, 310–321.
Rambhia, A. H., Glenny, R., & Hwang, J. (1994). Critical input data channels selection for progressive work exercise test by neural network sensitivity analysis. In: *IEEE international conference on acoustics, speech, and signal processing* (pp. 1097–1100).
Sánchez-Lasheras, F., de Andrés, J., Lorca, P., & de Cos Juez, F. J. (2012). A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Systems with Applications, 39*, 7512–7523.
Serrano-Cinca, C. (1996). Self-organizing neural networks for financial diagnosis. *Decision Support Systems, 17*, 217–228.
Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2013). Partial least square discriminant analysis for bankruptcy prediction. *Decision Support Systems, 54*, 1245–1255.
Sharda, R., & Wilson, R. L. (1996). Neural network experiments in business-failure forecasting: predictive performance measurement issues. *International Journal of Computational Intelligence and Organization, 1*, 107–117.
Sun, J., Li, H., Huang, Q.-H., & He, K.-Y. (2014). Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems, 57*, 41–56.
Tam, K. Y., & Kiang, M. (1992). Predicting bank failures: A neural network approach. *Decision Sciences, 3*, 926–947.
Tsai, C.-F., Hsu, Y.-F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing, 24*, 977–984.
Van Gestel, T., Baesens, B., Van Dijcke, P., Garcia, J., Suykens, J. A. K., & Vanthienen, J. (2006). A process model to develop an internal rating system: Sovereign credit ratings. *Decision Support Systems, 42*, 1131–1151.
Vellido, A., Lisboa, P., & Vaughan, J. (1999). Neural networks in business: A survey of applications (1992–1998). *Expert Systems with Applications, 17*, 51–70.
Wang, G., Ma, J., & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications, 41*, 2353–2361.
Wang, L., Zeng, Y., & Chen, T. (2015). Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Systems with Applications, 42*, 855–863.
Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery, 18*, 30–55.
Wu, C., & Wang, X. M. (2000). A neural network approach for analyzing small business lending decisions. *Review of Quantitative Finance and Accounting, 15*, 259–276.
Wu, Y., & Liu, Y. (2007). Robust Truncated Hinge Loss Support Vector Machines. *Journal of the American Statistical Association, 102*, 974–983.
Xie, Y., Li, X., Ngai, E., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications, 36*, 5445–5449.
Xu, W., Xiao, Z., Dang, X., Yang, D., & Yang, X. (2014). Financial ratio selection for business failure prediction using soft set theory. *Knowledge-Based Systems, 63*, 59–67.
Yin, J. (2005). Comparison of country risk models: hybrid neural networks, logit models, discriminant analysis and cluster techniques. *Expert Systems with Applications, 28*, 137–148.
Yu, Q., Miche, Y., Séverin, E., & Lendasse, A. (2014). Bankruptcy prediction using extreme learning machine and financial expertise. *Neurocomputing, 128*, 296–302.
Zhang, H., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research, 116*, 16–32.
Zhou, L. (2013). Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems, 41*, 16–25.
Zhou, L., Lai, K. K., & Yen, J. (2012). Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation. *International Journal of Systems Science, 45*, 241–253.
Zurada, J. M., Malinowski, A., & Cloete, I., 1994. Sensitivity analysis for minimization of input data dimension for feedforward neural network. In: *Proceedings of ISCAS'1994* (vol. 10, pp. 447–450).