IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 3 Ver. II (May. - Jun. 2016), PP 31-39
www.iosrjournals.org

Classification of Business Failures in Morocco: A Comparative Analysis between a Classical Linear Model And an Artificial Neural Networks

Aomar IBOURK, Abdelkrim AAZZAB[2]
[1]*(Quantitative Economics/ Cadi Ayyad University,Morocco)*
[2]*(PhD student in economics / Cadi Ayyad University,Morocco)*

Abstract: *The failure risk analysis has for objective to highlight telltale signs of a potential failure. These signals can be alerted on the accounting elements and / or financial (Dimitras et al., 1995). Also, in discussing the question of the factors leading to the deterioration of a company, most of the studies consider that any cause leading to the failure can be embodied in the accounting records (Conan and Holder, 1979). So, the financial ratios have relevant information content that allows them to predict the risk of business failure.The statistical classification techniques used for this purpose can be divided into parametric and non-parametric techniques including discriminant analysis and neural networks are used the most. Then, based on the accounting records, this article aims to compare the discriminating power of these two classification methods on a sample of Moroccan companies based on financial indicators.*
Keywords: *Artificial neural networks, Classification of business failures, Financial indicators, Linear discriminant analysis*

I. Introduction

The methods of analysis and forecasting of business failures are diverse and are divided between explanatory, exploratory and modeling. Indeed, the first methods of analysis of business failure are explanatory nature that are divided into static method, with an interest in the study of funding cycles, investment and operating to meet such defect [15] [18], and which includes dynamic analysis methods by flows [19] [20].

In addition, other methods called exploratory and modeling are based on ratios for a systematic and comparative analysis [21]. Analysis ratios led to the development of functions called "scores" for determining the probability of failure of a business [22].

The method of "US credit-men", developed in the thirties, opened the way for the credit scoring method and therefore other models for the day such as Beaver’s model (1966), Altman’s model (1968), Deakin’s model (1972) [10], the model of Altman, Haldeman and Narayanan (1977), the model of Conan and Holder (1979) and the model of Altman and Lavallée (1980).

The statistical classification techniques can be divided into parametric and non-parametric techniques including discriminant analysis and neural networks are truly dominant. Indeed, Fisher linear discriminant analysis is the most conventional method used in the analysis of business failures [2] [23]. However, given the disadvantages of parametric techniques that requires strict statistical conditions, other authors have used the nonparametric techniques that are more robust because they do not imply any assumption about the distribution of variables whose neural networks belongs [3] [4].

The use of neural networks for the prediction of business failures really began in the 1990s with the work of Odom and Sharda (1990) [14]. This method, which is based on the information processing performed by the human brain, is to develop a learning algorithm that processes a set of information to get a result. Multiple studies and research works on business failure have practiced this technique which are found Bell and *al.* (1990), Keasey and Watson (1991), Dimitras et *al.* (1996), Altman and Narayanan (1997), Wong et *al.* (1997), Zhang et *al.* (1998), Coakley and Brown (2000), Aziz and Dar (2004), Ooghe and Balcaen (2004, 2006), Ravi Kumar and Ravi (2007) and Lin (2009).

Notwithstanding the diversity of techniques used for the explanation and prediction of business failure, the general principle underlying the various studies is similar. Indeed, the authors select firstly two classes of firms by non-defaulting defaulting character, secondly they choose a set of explanatory variables and finally they seek to establish a statistical relationship between these variables and the dichotomous state to be or not to be faulty. The quality of the model developed depends on the rate of correct classification of a business in the corresponding class.

II. Presentation Of Artificial Neural Networks

2.1 The neural network architecture

An artificial neural network is a process consisting of simple processing units, parallel distributed for accumulating experimental knowledge and make it operational [1][7]. A network is a set of formal neurons associated layers (input, hidden and output) and operating in parallel. Each artificial neuron is a set of simple mathematical operators. This is a nonlinear algebraic function which is parameterized and with a values bounded.

The architecture of a neural network is the way neurons are arranged and interconnected in a network [16]. It is then a mesh of several neurons, usually organized in layers: the input layer, hidden layer and output layer:

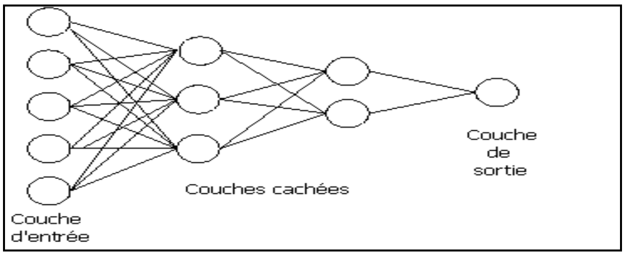


Figure 1 : Simple example of neural network

An artificial or formal neuron is then considered a device that receives, from other neurons or from the outside, n input stimuli, and weights them through a real values called synaptic coefficients or synaptic weights. These weights can be positive, it is called excitatory synapses, or negative and in this case it is said inhibitory synapses.

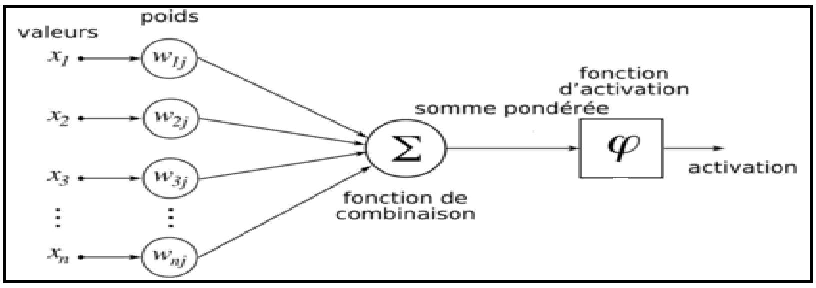


Figure 2 : Structure of an artificial neuron (Haykin, S. 1994: Neural Networks)[28]

One neuron j calculates a potential P_j, equal to the sum of its inputs (x_1, x_2, ... , x_n) weighted by the respective synaptic coefficients (w_1, w_2, etc. ., w_n), to which we add a constant term bj. The value of the potential P_j is given by the following equation:

$$P_J = \sum_{i=1}^{n} w_{ij} x_i + b_j \quad (1)$$

At this potential, the neuron applies an activation function Φ, so that the output y_j calculated by the neuron is equal to $\Phi(P_j)$, such as:

$$y_j = \Phi(P_j) = \phi\left(\sum_{i=1}^{n} w_{ij} x_i + b_j\right) \quad (2)$$

The use of the bias is such as to apply an affine transformation potential. In fact, the bias is an external parameter of the neuron j, it can be integrated in the equation of potential, as the signal x_0 is set to 1, weighted by w_{0j} weight, whose value is equal to the bias b_j.

The output value y_j is emitted by the neuron to other neurons or to the outside. So a neuron is characterized by three concepts: its internal state (its potential), its connections with other neurons and its transfer function [29].

2.2 Learning Algorithm

Neural networks are generally optimized by learning which is the fundamental property of neural networks. This is a phase where the network behavior changes until the acquisition of the desired result. In the area of forecasting failures, the gradient back propagation algorithm is the best known.

Is often called back-propagation technique of the gradient algorithm to correct errors based on the gradient of the calculation and using the back-propagation. Learning by the gradient back propagation algorithm is an application of a statistical method known as "stochastic approximation" which was proposed by Robbins and Monro (1951). This algorithm was developed by several teams of researchers [24], and its basic principle is to minimize a dependent function of the error. That is, for each configuration of the weight corresponding to a cost, to seek a minimum cost over a cost surface. However, minimizing the gradient may have local minima instead of the global minimum.

Before start learning, the network weights are first initialized with random values. Then, we consider a learning sample that will be used for the learning phase. Finally, the neural network is then called to provide and therefore to generate, from this sample, similar values as desired.

The algorithm looks like this[1] [12]:

- Either a sample $\vec{x}$ That represents the entry of the network and $\vec{t}$ the corresponding output desired ;
- The signal is propagated in front from the input layer to the output layer in passing by the hidden(s) layer(s): $x_k^{(n-1)} \mapsto x_j^n$;
- The calculation of the forward propagation is done by an activation function g, by an aggregation function[2] h and using synaptic weight $\vec{w}_{jk}$ between the neuron $x_k^{(n-1)}$ and the neuron x_j^n ($\vec{w}_{jk}$ Specifies a weight of k to j).

$$x_j^{(n)} = g^{(n)}(h_j^{(n)}) = g^{(n)}\left(\sum_k w_{jk}^{(n)} x_k^{(n-1)}\right) \qquad (3)$$

- Then, after the propagation, the result $\vec{y}$ is calculated at the output of the network.
- The output error, which represents the difference between $\vec{y}$ and $\vec{t}$, is then calculated for each neuron in the output layer:

$$e_i^{output} = g'(h_i^{output})\left[t_i - y_i\right] \qquad (4)$$

- This error is then propagated backwards $e_i^{(n)} \mapsto e_j^{(n-1)}$ through the following formula:

$$e_j^{(n-1)} = g'^{(n-1)}(h^{(n-1)})\sum_i w_{ij} e_i^{(n)} \qquad (5)$$

- After, an update of the weight in all layers is done by the following formula:

$$\Delta w_{ij}^{(n)} = \lambda e_i^{(n)} x_j^{(n-1)} \qquad (6)$$

where λ indicates the learning rate (less than 1)

The backpropagation algorithm of the gradient thus contains two phases:

- The first phase "Forward": the function signal propagates to the front of the network, of a neuron to another, from the neurons of entry. During this phase, the synaptic weights remain unchanged. For each neuron j, the signal function is $y_j(n) = \varphi\left(P_j(n)\right)$, where $P_j(n)$ is the potential at the input of the neuron j, which is the weighted sum of the respective synaptic weights of all the neurons of the preceding layer.
- The second phase "backward": in this case it is the error signal which propagates in the opposite direction, layer by layer, from the output layer. For each neuron j, we calculate the local gradient $\psi_j(n)$ which allows to calculate the correction of synaptic weight, link by link, according to a rule known as Delta [3] (also

[1] We have adopted here the notation adopted by Byrd A. in his book entitled "Self-calibration of a sensor network of pollution", high school of engineering and management of the Canton of Vaud, 2006, pp. 21-22.

[2] The aggregation function is often a scalar product between the weight and the entries of the neuron.

[3] This method developed by B. WIDROW and HOFF M.E. (1960) under the acronym ADALINE (Adaptive linear neuron). It is a learning rule that minimizes the output error by using a gradient descent of the error approximated. After each iteration , the correction is applied to the weight proportionally to the error. The correction is calculated before the ranger (activation function). For more detail, see DAVALO E. and NAIM P. (1990), "neural networks " ,Eyrolles, Paris, p. 232.

known as the least-mean-squares rule (LMS), delta rule, or Widrow-Hoff rule). In the case of an output neuron, the local gradient is equal to the error signal multiplied by the derivative of the activation function applied to the potential at the input of this neuron:

$$\psi_j(n) = \varphi_j^{'}(p_j(n)) \sum_{k=1}^{N_Q} \omega_{kj}(n) \psi_k(n) \tag{7}$$

III. Data And Research Methodology

Firstly, in order to keep only the relevant variables, the most discriminating, with the aim of improving the model prediction quality, we started with the selection of explanatory variables among a set of variables candidates selected on the basis of previous empirical work. This is to make a kind of elimination of irrelevant variables considered keeping only those that explain properly and significantly the business failure. Thus, the procedure for selecting variables is based on 1000 bootstarps samples and variables used are those with the highest power of discrimination. This power is calculated by ranking the variables in ascending order according to Fisher statistic and in selecting frequency[4].

Secondly, we have developed models predicting failure based on the artificial neural network method. We also develop a predictive model based on linear discriminant analysis method to compare the different models developed to take the contributions of neural models over conventional forecasting models.

3.1 Selection of firms

Our approach to data collection consists of three steps: the choice of the database, the selection of the firms and the choice of indicators of failure. In effect, to constitute our sample, based on an official source of information, we have purchased the accounting synthesis documents from OMPIC. It is the departure of 160 firms whose 50% represents the failing firms. As well, to delineate our field of investigation and to ensure the maximum homogeneity of the composite sample, we selected companies operative in the sector of industry and which are of small and medium sizes. The choice of this sector lies first in the significant number of failed companies operating there, and to the ability to calculate the set of financial ratios described by the theory, a thing which is not possible for services companies, for example, who do not have some indicators.[5]

Thus, the criterion size affects the companies which have achieved, during the year that it was retained for the analysis, an annual turnover not exceeding 75MDhs or a balance sheet total not exceeding 50MAD. Our final sample consists of 132 companies, half of which has failed. This balance between the two types of companies up to empirical considerations which show that an imbalance between classes has a negative effect on the correct classification rate of each group and the overall correct classification rate [2] [25].

For the companies in good health, we have begun a choice at random without any other hypothesis. Whereas for the failing firms, we have identified them with the commercial courts prior to requesting their states of syntheses. The commercial courts chosen are those of Agadir, Marrakech and Casablanca. Our choice here is motivated by the ease of access to information and by the proximity.

Thus, for each failed company, we have requested the synthesis documents of an accounting period before the date of declaration of default. For the non-defaulting, it is also one exercise pulls randomly. Also, our sample covers a five-year period from 2006 to 2010. The choice of this period is mainly due to the difficulties related to the identification of failing companies on a shorter period. The following table summarizes the description of the businesses that make up our database according to the type and by regions.

Table 1 : Distribution of companies by regions

	Agadir	Marrakech	Casablanca	Total
Non-faulty	18	22	26	66
Faulty	13	17	36	66
Total	31	39	62	132

3.2 Choice of variables

Our database is composed of 18 financial ratios. These ratios are calculated on the basis of the documents collected in order to constitute a battery relevant and credible likely to respond to our question concerning the explanatory factors of business failure. The justification of choice of these ratios is based mainly on the theoretical and empirical literature [2] [5] [8].

[4] For Sauerbrei and SCHUMACHER (1992), the variables to be considered are those that appeared in at least 70% of cases.

[5] OMPIC (Moroccan Office of Industrial and Commercial Property) provides, through the service Directinfo, an access to financial statements of enterprises with a price of 60dhs/state. Five types of documents are available: the Balance Sheet (BL), the account of Products and Charges (CPC), the status of the balances of Management (GSS), the Table of Funding (TF) and the status of the Additional Information (ETIC)

Thus, the variable to explain is dichotomous, it takes the value 1 if the firm is faulty and the value 0 if the firm is non-faulty. For the explanatory variables, and to the extent that there is no unifying theory defining the failure of businesses, our work is also included in the same way that most of the empirical models that begin with a high number of factors and reduce in order to keep only a few judges as the most explanatory of the risk of failure. Then, we are therefore limited to a basic battery consisting of 18 ratios according to their popularity and their performance in the previous studies. Annex 1 summarizes the ratios of our study which represent the set of financial indicators chosen.

3.3 Variables selection methods

We opted for the automatic variables selection methods by comparing between two methods to finally select the one which presents more precision. In effect, we have proceeded to the selection of variables using the method called "*Stepwise discriminant analysis SDA"*. This method, which is based on the criterion of the Lambda of Wilks as evaluation criterion, is to find the sub-space of representation that allows a maximum difference between the clouds of points associated with each value of the variable to predict, that is to say between the centers of gravity of the clouds of points conditional. Then, we compared between the Forward approach and the so-called Backward. The first is to choose the variable inducing the best improvement of the lambda, and the select if the improvement is statistically significant. It is in this case an iterative procedure based on adding one by one variable until the addition of a variable does not bring more improvement. The second start from the set of candidate variables and search the variable whose withdrawal would lead to the degradation the more low of the lambda, and the permanently remove it if this degradation is not statistically significant.

For these methods of variables selection, and to assess the significant role of a variable, we use the statistic F of Fisher. Therefore, it would be sufficient to compare the p-value calculated for the variable to assess and compare with the level of significance chosen. As well, the Wilks lambda, which varies between 0 and 1, represents the preferred indicator for the statistical evaluation of the model [30]. It indicates to what extent the centers of classes are separate from each other in the space of representation. As long as it tends to 0 the model will be good because the clouds are quite distinct.[6]

3.4 Construction of the neuronal model

The network of neuron developed is of type "multi-layers Perceptrons" with the simple gradient descent based on the error backpropagation algorithm [24] as optimization technique. Thus, we have retained the hyperbolic tangent as activation function and the error of least squares as a cost function. Moreover, for the modification of the weight of the network, we have opted for a term of time and each layer is begun of a bias and a term of regularization. Finally, we have retained the sum of quadratic errors (SSE) as a performance measurement function.

For network setup, we adopted supervised learning for a layered network, not curly, fully connected, with a hidden layer and a linear output.

For the input layer, it is the vector of variables selected candidates for learning. For the number of neurons to introduce in the hidden layer, it is to test the different configurations which led to a level of learning high. For the output layer, the variable to explain is dichotomous. It is a vector that takes the value 1 if the company is faulty and the value 0 if the company is not-faulty. As well, to ensure a better learning and to stabilize the process of selection of variables, we have employed *bootstrap*[7] techniques of resampling.

Too, we resorted to the definition of a random generator by creation of a variable partitioning in order to recreate exactly the samples used in the analyzes. It is a randomly Bernoulli variable generated with a probability parameter of 0.7, modified so as to take the value 1 or -1, instead of 1 or 0 (faulty or not-faulty). Then, the observations containing positive values on the variable of partitioning are assigned to the sample of learning, those with negative values are assigned to the validation sample and those with a value equal to 0 are assigned to the test sample. The latter is formed to avoid the problems of over-learning[8] in order to help the network to remain "on the right track". For the other parameters of the network (the learning step, the term of time and the terms of regularization of weights), the values are set on the basis of the empirical work found in the literature. As well, the number of iterations to retain is the one for which the error does varies almost more

[6] There is no marks to define a rule of judgment, however, we have chosen the value 3, 84 proposed by default by many software and which resembles the critical value of a test at 5% when we are working on a sample of a few thousand individuals.

[7] A sample *bootstrap* corresponds to a sample of similar size to the original sample and constitutes from this last by random with discount. This method of re-sampling is recommended by several authors (GUYON and ELISSEEFF, 2003; FERGUSON et al. , 2003; ZELLNER, 2004). By this method, the selection would be repeated on different replicas of the original sample in order to smooth the disturbances that might affect the procedure.

[8] That is to say that the research network models false appearing in the training data by random variation.

beyond this number. Finally, in the aim to delete all that is modeled in order to reduce the complexity of the network and to accelerate its convergence, we performed pretreatments on the standardization of data based on the Min-Max method.

IV. Analysis of Results

4.1 Performance of classification and selection of variables

4.1.1 Confusion matrix

The confusion matrix of the two variables selection methods (Table 2) indicate a rate of misclassification of 0.0909 for the SDAFTable 2 and a rate of 0.0985 for the named SDAB. The error rates calculated on the training data are then very optimistic and the estimator of the error *bootstrap* gives the advantage to the SDAF which has a value of 0.1221 instead of 0.1279 for the SDAB.

Table 2: Classification performance of variable selection methods

SDA (FORWARD)

Error rate	0,0909		
Bootsrap error estimation	0,1221		
Confusion matrix			
	D	ND	Sum
D	61	5	66
ND	7	59	66
Sum	68	64	132

SDA (BACKWARD)

Error rate	0,0985		
Bootsrap error estimation	0,1279		
Confusion matrix			
	D	ND	Sum
D	59	7	66
ND	9	57	66
Sum	68	64	132

The first method indicates that 61 failing companies have been well reclassified and 5 have incorrectly been. Similarly, for the companies not-faulty, 7 of them have been incorrectly reclassified and 59 are well reclassified. In total, it is therefore 120 firms (60 + 57) which have been correctly reclassified with a rate of correct classification of 90.90 %.

4.1.2 The MANOVA Test

The analysis of the multivariate variance shows that's the method of SDAF which shows good results. In effect, it has the more low of Wilks lambda statistics (0.37). This result is confirmed by the transformations of Bartlett or Rao who adjudicate on the significance of deviations, and which lead to the same conclusion on the threshold of error of 5%. We then rejects the hypothesis that the centers of classes are combined (p-value=0).

Table 3 : The analysis of variance multivariate

SDA (FORWARD)			SDA (BACKWARD)		
Stat	Value	p-value	Stat	Value	p-value
Wilks' Lambda	0,3789	-	Wilks' Lambda	0,3909	-
Bartlett C(18)	123,74	0,00	Bartlett C(18)	120,22	0,00
Rao F(18, 113)	41,31	0,00	Rao F(18, 113)	49,47	0,00

Then, by marrying the result of MANOVA test with that of the confusion matrix, we understand that the proper holding of the model holds especially for the application of the method of forward stepwise discriminant analysis.

4.1.3 Selecting variables

The individual assessment of the predictor variables shows that five variables that contribute to the explanation of the failure to the SDAF and four variables for the SDAB. Thus, table 5Table 4 shows that these results also indicates that four common variables between the two methods (R3, R5, R7 and R16).

Table 4 : Individual assessment of the predictor variables

SDA (FORWARD)	SDA (BACKWARD)
R7, R16, R5, R3 et R15	R3, R5, R7 et R16

All of these variables are selected on the basis of the statistics F which is significantly different from zero, because the p-value is less than 5%.

4.2 Neural network

4.2.1 Architecture of neural models

According to the table 5, we note that, by the employment of all the candidate variables, the best architecture is the one using a hidden layer with a single neuron (Net1_1 (18 1 1). In effect, this is the architecture for which the sum of the quadratic error is minimum for the learning sample (7.68). The SSE for the sample test is the 2.89, this is not therefore the minimum value but it corresponds as even at a low value if it is compared with the other. This architecture has been used to record a rate of correct classification of 87.3% for the learning sample and a rate of 85% for the sample test (table 6).

However, the optimum architecture corresponding to the employment of selected variables by the method SDAF is composed of a hidden layer with 9 neurons (Net2_6 (6 9 1). With regard to it, this architecture has enabled us to save the lowest value of the sum square error for the learning sample with 7.86% and an error rate of 2.98% for the sample test. For this network, the rate of correct classification is that of 84.8% for the learning sample and of 85% for the sample test (table 6).

Table 5 : Summary of tests of network architectures

Employment of 18 candidate variables			Employment of 5 selection variables (SDAF)		
Architecture	SSE[9] of the learning sample	SSE[10] of the test sample	Architecture	SSE of the learning sample	SSE of the test sample
Net1_1 [18 1 1]	**7,68**	**2,89**	Net2_1 [5 1 1]	10,83	2,78
Net1_2 [18 3 1]	10,15	2,39	Net2_2 [5 3 1]	13,79	2,69
Net1_3 [18 5 1]	9,61	2,82	Net2_3 [5 5 1]	9,83	3,08
Net1_4 [18 7 1]	9,96	2,61	Net2_4 [5 7 1]	10,89	3,22
Net1_5 [18 8 1]	11,42	2,83	Net2_5 [5 8 1]	9,75	3,11
Net1_6 [18 9 1]	11,30	2,74	**Net2_6 [5 9 1]**	**7,86**	**2,98**
Net1_7 [18 10 1]	9,58	2,91	Net2_7 [5 10 1]	11,17	2,91

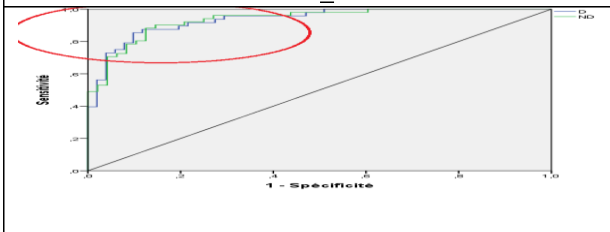
Table 6 : The confusion matrix neural models

Modele_1:Employment of 18 candidate variables (Net_1) (Net_1)					Modele_2: Employment of 5 selected variables (Net_2)				
Sample		Forecasts			Sample		Forecasts		
		D	ND	% Correct			D	ND	%
Learning	D	34	6	85.0 %	Learning	D	34	6	85.0 %
	ND	4	35	89.7 %		ND	6	33	84.6 %
	% Global	48.1 %	51.9 %	**87.3 %**		% Global	50.6 %	49.4 %	**84.8 %**
Test	D	7	1	87.5 %	Test	D	7	1	87.5 %
	ND	2	10	83.3 %		ND	2	10	83.3 %
	% Global	45.0 %	55.0 %	**85.0 %**		% Global	45.0 %	55.0 %	**85.0 %**
Validation	D	15	3	83.3 %	Validation	D	16	2	88.9 %
	ND	1	13	92.9 %		ND	1	13	92.9 %
	% Global	50.0 %	50.0 %	**87.5 %**		% Global	53.1 %	46.9 %	**90.6 %**

4.2.2 Validation of neural models

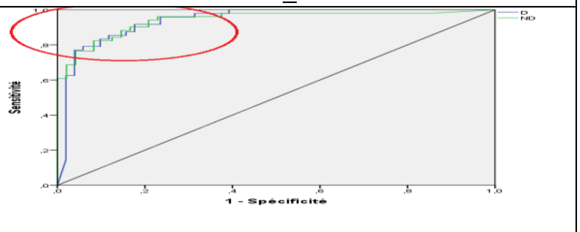
As for the validation sample, that evaluates the final neural network and thus validates the model, the percentage of correct classification shows that 88.9% of the failing companies its well classified by the network Net_2 whereas the network Net_1 arrives only to properly classifying 83.3% of these companies. For the companies non-defaulting, the both networks have correctly classified 92.9% of them. Therefore, the overall rate of correct classification of the Net_2 displays is of 90.6% and that of Net_1 is only 87.5%. The validation of neural models can be strengthened by the analysis of ROC curves[11].

Net_1



Area under the curve	
Faulty	0,936
Not-faulty	0,936

Net_2



Area under the curve	
Faulty	0,941
Not-faulty	0,941

[9] The sum of the quadratic errors (Sum squared error SSE) committed at the time of the classification of firms in the sample of learning.

[10] The sum of the quadratic errors (Sum squared error) committed during the classification of firms from the test sample.

[11] An ROC (Receiver Operating Characteristic) curve displays the modalities for each dependent variable qualitative. It presents a visual display of *sensitivity* and *specificity* for all possible hyphenation should in a unique diagram, which constitutes a tool more clear and more powerful than a series of tables.

Figure 3: the ROC curves of the two neural networks

We find for the two networks that the ROC curves are a little close to the top corner-left. Then, the performance of discrimination factors is acceptable. This means that the probability that the Score function, developed by the neural model, place a failing company before a company non-defaulting is almost close to 1 for the two neural models. Thus, for the network of neurons Net_2, for a random choice of a failing firm and a company not-faulty, there is a probability of 94.1% that the pseudo-probability of breach provided by the model is higher for failed company. That is to say that the probability that the network place a failing firm before a company not-faulty is of 94.1%. This rate is 93.6% for the network Net_1.

It is apparent that, from the analysis of the whole of these elements of validation, the two neural models are valid and record of good results. As well, the rates recorded by the two models are very optimistic that this is for the learning sample, the test sample or the validation sample. However, to decide between them, we can say that, without doubt, the neural model based on 6 variables is the most powerful on all levels. In effect, with the exception of the learning sample, this network has recorded the highest rates for the sample test and for the validation sample. More, with a reduced number of variables, the second model has been used to record the results more salient compared to the first model based on 18 variables. Then, our model of neural networks chosen is the one based on 6 explanatory variables with an architecture consisting of a hidden layer with 9 neurons.

4.3 The linear discriminant analysis

For the discriminant analysis, our methodology is to develop two models discriminating characteristics which the prime is based on the set of candidate variables and the second is based only on the most discriminant variables selected by the SDAF in respecting the partition of the sample database, composed of 132 observations, in learning sample is account 79 comments (59.84%), in test sample is account 20 comments (15.15%) and in validation sample is account 33 observations (25%). Table 8 presents the results obtained.

Table 7 : The confusion matrix of linear discriminant analysis

Modele_1: Employment of 18 candidate variables					Modele_2 : Employment of 5 selected variables				
Sample		Forecasts			Sample		Forecasts		
		D	ND	% Correct			D	ND	% Correct
Learning	D	37	3	92.5 %	Learning	D	37	3	92.5 %
	ND	1	38	97.43 %		ND	6	33	84.6 %
	% Global	48.1 %	51.9 %	**95.0 %**		% Global	54.4 %	45.6 %	**88.5 %**
Test	D	7	0	100%	Test	D	7	0	100%
	ND	2	11	84.6 %		ND	1	12	92.3 %
	% Global	45%	55%	**92.3 %**		% Global	40%	60%	**96.15 %**
Validatio n	D	19	0	100%	Validation	D	18	1	94.7 %
	ND	1	13	92.9 %		ND	2	12	85.7 %
	% Global	60.6 %	39.4 %	**96.4 %**		% Global	60.6 %	39.4 %	**90.2 %**

After the reading of this table, it is apparent that the discriminant analysis designed on the basis of the set of candidate variables is the more efficient. In effect, he recorded an overall rate of correct classification of 95% for the learning sample, of 92.3% for the test sample and of 96.4% for the validation sample. For the discriminant model based only on the 5 selected variables, these rates are respectively 88.5%, 96.15% and 90.2%. Nevertheless, the rates recorded by the latter model are also efficient and are close to those recorded by the first model discriminant. As well, we note also that the rate of correct classification of the test sample of the second model is higher than that recorded by the first.

Then we can say that the model with 5 variables enabled us to say almost the same thing that the model developed by 18 variables. Thus, the goal of the modeling in general is to make a simple model with a reduced number of explanatory variables which allows on one side, a better understanding of the phenomena being studied and on the other side a possible action to correct such a situation. Consequently, the model that appears most useful among the two is the second based on the five selected variables.

V. Conclusion

The results obtained show that the analysis of financial variables (ratios) has allowed us to detect those more revealing of the failure. It is five variables from four dimensions of the financial analysis that are the basis of the explanation of the failure of companies to know the financial structure, activity, liquidity and management.

Also, the results show that the analysis of financial variables (ratios) has allowed us to detect those most indicative of the failure. These are the variables from four of the five dimensions of financial analysis that are the basis of the companies failure explanation to know the financial structure, activity, liquidity and management. This result then confirms the successful outcomes of Conan and Holder (1979) or Combier and Blazy (1997) [6].

The comparison of the two classification methods in terms of predictability shows in our case the performance of conventional models over than the neural networks networks. In fact, the percentage of correct classification measured by the linear discriminant analysis is better than artificial neural networks on the learning samples and test sample, with the exception of the validation sample or the neural networks show a slight superiority.This result thus cripples those already established empirical studies that have shown the success of these nonparametric methods in predicting business failure [26] [27] [11] [13].

Note finally that this study has some limitations in the frame where the models developed are based on a small number of observations and multicollinearity tests, multi-normality tests and homoscedasticity tests are not checked.

References

[1] ALEKSANDER I. et MORTON H., [1995], “An introduction to neural computing”, London, Chapman et Hall, 284 p.
[2] ALTMAN E. I., [1968], “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy”, Journal of Finance, vol. 23, n° 4, pp. 589-609.
[3] ALTMAN E. I., MARCO G. et VARETTO F., [1994], “Corporate Distress Diagnosis : Comparisons Using Linear Discriminant Analysis and Neural Network – The Italian Experience”, Journal of Banking and Finance, vol. 18, n° 3, pp. 505-529.
[4] ANANDARAJAN M., LEE P. et ANANDARAJAN A., [2001], “Bankruptcy Prediction of Financially Stressed firms: An Examination of the Predictive Accuracy of Artificial Neural Networks”, International Journal of Intelligent Systems in Accounting, Finance and Management, vol. 10, n° 2 pp 69-81.
[5] BARDOS M. et ZHU W. H., [1997], « Comparaison de l’analyse discriminante et des réseaux de neurones. Application à la détection des défaillances d’entreprises », Revue de statistique appliquée, tome 45, n°4, pp. 65-92.
[6] BLAZY R. et COMBIER J., [1997], « La défaillance d’entreprise : Causes économiques, traitement judiciaire et impact financier », INSEE Méthodes n° 72-73, Économica, Paris ;
[7] CLAUDE T., [1992], « Les Réseaux de Neurones Artificiels, Introduction au connexionnisme » EC2, 160 p.
[8] CONAN J. et M. HOLDER [1979], « Analyse des causes de faillite des PMI », Thèse d'Etat, Université ParisDauphine, CEREG, cités par DAIGNE, 1986.
[9] DAVALO E. et NAIM P., [1990], « Des réseaux de neurones »,Eyrolles, Paris, 232 p.
[10] DEAKIN E., [1972], “A Discriminant Analysis of Predictors of Business Failure”, Journal of Accounting Research, 10(1), pp. 167-179.
[11] DESMET P., [1996], « Comparaison de la prédictivité d’un réseau de neurone à rétropropagation avec celles des méthodes de régression linéaire, logistique et AID pour le calcul des scores en marketing direct», Recherches et Application en Marketing, vol.11, n°2, pp.17-27.
[12] DU JARDIN P., [2007], « Prévision de la défaillance et réseaux de neurones : l’apport des méthodes numériques de sélection de variables », Thèse de doctorat, Université de Nice Sophia-Antipolis, 88 p.
[13] ETHERIDGE H. L. SRIRAM R. S., [1997], “A Comparison of the Relative Costs of Financial Distress Models : Artificial Neural Networks, Logit and Multivariate Discriminant Analysis”, International Journal of Intelligent Systems in Accounting, Finance and Management, vol. 6, n° 3, pp. 235-248.
[14] KOUAM A., BADRAN F. et THIRIA S., [1992], « Approche méthodologique pour l'étude de la prévision à l'aide des réseaux de neurones », Actes du congrès Neuro-Nîmes 92 : Neural Networks and their Applications, Nîmes, 2-6 Novembre, pp. 117-127.
[15] MALECOT J.F. [1991], « Analyse historique des défaillances d’entreprises : une revue de la littérature », Revue d’économie financière, vol. 19, p. 205-226.
[16] WIRA P., [2009], « Réseaux de neurones artificiels: architectures et application », Université de Haute Alsace, Laboratoire MIPS.
[17] DIMITRAS A., ZANAKIS S. et ZOPOUDINIS C., [1996], “A Survey of Business Failures with an Emphasis on Prediction Methods and Industrial Applications”, European Journal of Operational Research, vol. 90, n° 3, pp. 487-513.
[18] LEVASSEUR M., [1979], "Gestion de trésorerie"; ECONOMICA.
[19] PONCET P. et PORTAIT R., [1978], « La théorie financière de la firme et le tableau d'emplois-ressources », Revue Française de Gestion, Mars-Avril.
[20] De MURARD G., [1977], « Le tableau pluriannuel des flux financiers », Revue Française de Gestion, Janvier-Février, n°8, pp. 78-86.
[21] LOCHARD J., [1998], « Les ratios: aide au management et au diagnostic », Paris: Editions d'Organisation.
[22] HOLDER M., LOEB J. et PORTIER J., [1984], « Le Score de l'Entreprise », Paris: Nouvelles Editions Fiduciaires.
[23] ROSE P.S. et GIROUX G.A., [1984], “Predicting corporate bankruptcy: an analytical and empirical evaluation”, Review of Business and Economic Research, 21, pp. 1-12
[24] RUMELHART D. E., HINTON, G. E. et WILLIAMS R. J., [1986b], “Learning Representation by Back-Propagation Errors”, Nature, Vol. 323, No. 9, pp. 533-536.
[25] BERG D., [2005], “Bankruptcy Prediction by Generalized Additive Models”, University of Oslo, Department of Mathematics, Statistical Report, n° 1, January, 36 p.
[26] ODOM M. D. et SHARDA R., [1990], “A neural network model for bankruptcy prediction”, In : Proceedings of the IEEE International Joint Conference on Neural Networks. San Diego, CA, pp. 163–168.
[27] BALCEAN S. et OOGHE H., [2004], “Alternative Methodologies in Studies on Business Failure : Do They Produce Better Results than the Classical Statistical Methods ?”, Ghent University, Working Paper, n° 2004-249, 40 p.
[28] HAYKIN S., [1994], "Neural Networks. A Comprehensive Foundation", Macmillan, New York.
[29] DAHMANI H. et ZAOUI A. [2005], « Application des réseaux de neurones artificiels pour la détection et la localisation des défauts de la machine asynchrone », Thèse d’ingéniorat, université de M'sila.
[30] BARDOS M. , (2001), "discriminant analysis - Application to the risk scoring and financial ", Dunod.