CAN MACHINE LEARNING APPROACHES PREDICT CORPORATE BANKRUPTCY? EVIDENCE FROM A QUALITATIVE EXPERIMENTAL DESIGN

SALIM LAHMIRI [a*]   STELIOS BEKIROS [b,c †]

[a] *ESCA School of Management, Casablanca, Morocco*
[b] *European University Institute (EUI), Florence, Italy*
[c] *Athens University of Economics and Business, Athens, Greece*

ABSTRACT

Bankruptcy prediction has received a growing interest in corporate finance and risk management recently. Although numerous studies in the literature have dealt with various statistical and artificial intelligence classifiers, the performance of them in credit risk forecasting needs to be further scrutinized compared to other methods. In the spirit of Chen, Härdle and Moro (2011, *Quantitative Finance*), we design an empirical study to assess the effectiveness of variant machine learning topologies trained with big data approaches and qualitative aside from quantitative information as input variables. The experimental results from a ten-fold cross-validation methodology demonstrate that a generalized regression neural topology yields an accuracy measurement of 99.96%, a sensitivity measure of 99.91% and specificity of 100%. Indeed, this specific model outperformed multi-layer back-propagation networks, probabilistic neural networks and radial basis functions, regression trees as well as other advanced classifiers. The utilization of advanced nonlinear classifiers based on big data methodologies and machine learning training, generates outperforming results compared to traditional methods for bankruptcy forecasting and risk measurement.

JEL Classification: C32; C58; G10; G17
Keywords: credit risk; bankruptcy; neural networks; classifiers; experimental design

[*a] ESCA School of Management, 7, Abou Youssef El Kindy Street, BD Moulay Youssef, Casablanca, Morocco; E-mail address: slahmiri@esca.ma

[†] *Corresponding author:* [c] Athens University of Economics and Business (AUEB), Department of Acc. & Finance, 76 Patission Str. GR10434, Athens, Greece; [b] Department of Economics, Via delle Fontanelle 18, I-50014, Florence, Italy; Tel.: +39 055 4685 916; Fax: +39 055 4685 902; E-mail address: stelios.bekiros@eui.eu

1. INTRODUCTION

Bankruptcy inaccuracy negatively affects financial and corporate risk and yields imminent social problems such as unemployment, economic recessions and possibly a generic financial crisis. In this regard, for decades, stockholders, creditors, senior managers, auditors, and public authorities are all interested in bankruptcy prediction as it has a real direct impact on their decision making process. Hence, there is an urgent need for enhanced bankruptcy forecastability in order to lessen workload of decision makers, albeit there is no general theory of business failure.

Indeed, with regard to bankruptcy predictive system building, numerous approaches for classification of bankruptcy data have been proposed, including forecasting systems based on statistical methods or computational/artificial intelligence techniques. Most popular statistical classifiers include the approaches of discriminant analysis (Wiginton, 1980), logistic regression (Ohlson, 1980; Wiginton, 1980), naïve Bayes (Sarkar and Sriram, 2001) and hazard regressions (Gupta et al, 2015; Gupta et al, 2018). Additionally, artificial intelligence systems involve decision trees (Marais, 1984; Cho et al, 2010), fuzzy inference (Chou et al, 2017), support vector machines (Chen, 2011; Chen, 2014; Lin et al, 2014; Moro et al, 2017), case-based reasoning (Li and Sun, 2008; Li and Sun, 2009; Li et al, 2009), genetic algorithms (Etemadi et al, 2009; Gordini, 2014) and artificial neural networks (Khashman, 2010; Chen et al, 2013; Lee and Choi, 2013; Oreski and Oreski, 2014; Kim et al, 2016; Lahmiri and Gagnon, 2016; Chou et al, 2017; Wang et al, 2017; du Jardin, 2018).

The large strand of up-to-date literature acknowledges that artificial intelligence models are the most efficient toward predicting financial risk and bankruptcy (Tsai and

Wu, 2008, Olson et al, 2012). This is mainly due to strict assumptions of the traditional statistical models related to linearity, normality, independence among predictor variables and pre-existing functional forms relating to criterion and predictor variable limitations vis-à-vis their real world applications (Hua et al., 2007). On the contrary, Big data artificial intelligence based techniques do not assume specific distributions and automatically extract knowledge from training samples (Wang et al, 2012).

It is common to compare statistical versus intelligent classifiers in bankruptcy prediction. For instance, Zaini et al (2008) found that rough set theory outperforms logistic regression, Tseng and Hu (2010) who compared logit and quadratic interval logit models versus backpropagation neural network and radial basis function networks illustrated that radial basis approximators outperform the other ones. Olson et al (2012) found that decision trees are relatively more accurate compared to artificial neural networks and support vector machines. However, they concluded that decision trees generate more rule nodes than desired. In another study, Horta and Camanho (2013) found that support vector machine classifiers outperform the logistic regression. More recently, the support vector machine classifier was found to be significantly more accurate than back-propagation, radial basis neural networks, as well as linear discriminant analysis and naive Bayes classifiers (Lahmiri, 2016).

The main goal of our paper is to predict the occurrence of bankruptcy in companies as accurately as possible using mimetic intelligent emulating techniques, in particular artificial neural networks trained with qualitative information for the first time in the literature. Often, in practice, the occurrence of bankruptcy can be detected by a human expert-based knowledge, experience and judgment. Nevertheless, human personal and

affective biases cannot be eradicated and the personal judgments tend to be significantly subjective. The merits of machine learning approaches boil down to the accurate mimesis of human brain information processing. They are data-driven, have the theoretical advantage that they do not impose arbitrary assumptions on the input variables and on model specification, are robust to noise and to incomplete market information. Finally, artificial neural networks were found to be superior to various standard statistical classifiers in bankruptcy classification and prediction (Desai et al, 1996; West, 2000, Chen, 2012; Ciampi and Gordini, 2013) based on intelligent information processing.

Our main contributions are threefold: firstly, four different types of artificial neural networks are compared in relation to their accuracy of bankrupt and non-bankrupt company classification. We rely on such investigations because such comparative studies are missing from current literature, to the best of our knowledge, and previous comparative studies focused exclusively on comparing different statistical and intelligent models pairwise (Zaini et al, 2008, Tseng and Hu, 2010, Olson et al, 2012, Horta and Camanho, 2013, Lahmiri, 2016). The considered neural networks in our study are the multi-layer back-propagation (BPNN) (Rumelhart et al, 1986), probabilistic neural network (PNN) (Specht, 1990), radial basis functions (RBFNN) (Cybenko, 1989), and generalized regression neural networks (GRNN) (Specht, 1991). The obtained results will help decision makers in choosing the best topology for bankruptcy data classification. Secondly, contrary to previous works on bankruptcy prediction where only accuracy was measured to assess the performance of the classifier (Khashman, 2010; Chen et al, 2013; Lee and Choi, 2013; Oreski and Oreski, 2014; Kim et al, 2016; Lahmiri and Gagnon, 2016; Chou et al, 2017; Wang et al, 2017; du Jardin, 2018), sensitivity and specificity are

now used to measure Type I error and Type II error for each machine learning model under study. Indeed, for the analyst, the most important issue is to correctly predict the risk of bankruptcy for all companies under scrutiny. Thirdly, our paper investigates the performance of artificial neural networks when they are all trained with *qualitative data*. Without a doubt, previous works all used solely quantitative data (Khashman, 2010; Chen et al, 2013; Lee and Choi, 2013; Oreski and Oreski, 2014; Kim et al, 2016; Lahmiri and Gagnon, 2016; Chou et al, 2017; Wang et al, 2017; du Jardin, 2018), but nothing is known about the performance of artificial neural networks when they are trained with bankruptcy qualitative data, to the best of our knowledge. This is a significant gap in design intelligent systems for bankruptcy prediction. Although quantitative data analysis is standard in bankruptcy risk evaluation, many bankruptcy experts in financial industry rely on qualitative data and their own experience and judgment to assess bankruptcy risk. In this regard, employing artificial neural networks used to mimic human decision making coupled with qualitative data is a most efficient way to build an artificial agent for bankruptcy prediction under a cognitive informatics point of view. To enhance robustness and for comparative purposes, regression trees (RT) with an advanced learning algorithm (Breiman et al, 1984) is used as a reference classifier. The benchmark RT training algorithm comprises a simple and effective machine learning system that imitates human decision processing.

The rest of the paper is organized as follows; section 2 provides an overview of the topologies used in the current study, whilst section 3 analyzes the performance measures employed and the protocol of experiments. Section 4 presents the empirical results. Finally, Section 5 discusses and concludes the study.

2. METHODOLOGY

The back-propagation neural network model (BPNN) (Rumelhart et al, 1986) comprises one or more hidden layers. Usually it consists of three types of layers; the first layer is the input layer and corresponds to the problem input variables with one node for each input variable. The second layer is the hidden layer used to capture non-linear relationships among variables. The third layer is the output layer used to provide predicted values. The output equation of the network is given by:

$$Output_j = S\left(\sum_{i=1}^{q} w_{ij} x_i + b_j \right) \qquad (1)$$

where w_{ij} is the weight connecting input i to neuron j, q is sample size, b is the bias, and S is a non-linear transfer function. The error E between the obtained output and the desired one is given by:

$$E = \frac{1}{2} \sum_{j=1}^{k_j} (d_j - y_j)^2 \qquad (2)$$

where y_j and d_j are the actual and the desired output in each node j respectively. The weights of the network as updated as follows:

$$\Delta w(k) = -\alpha(k) g(k) + m \Delta w(k-1) \qquad (3)$$

where $\Delta w(k)$ is the vector of weight changes, $g(k)$ is the current gradient, $\alpha(k)$ is the learning rate, which determines the length of the weight update, and m is the momentum parameter used to escape from small local minima on the error surface (Ramirez et al., 2003) and to avoid oscillations toward reducing the sensitivity of the network to fast changes in the error surface (Jang et al., 1997). Considering the learning rate, lower values help getting to the global minimum, yet over longer periods for the algorithm to

converge. Specifically, in case learning rate is too small, gradient descent can be slow, whereas if it is too large, it can overshoot the minimum and fail to converge. In other words, when the learning rate is too high the artificial neural network will either oscillate around the true solution or will diverge completely. In case it is very low, it will take a longer time to converge. Additionally, the momentum parameter plays an important role in accelerating gradient vectors towards the right direction vis-à-vis convergence while preventing the neural network from converging to a local minimum or saddle point. However, a very high value for the momentum parameter creates a risk of overshooting the minimum, which can cause the system to become unstable. On the contrary, low momentum parameters cannot consistently avoid local minima and can slow down the training of the artificial neural networks, leading to calculative overburden for convergence. In our experimental setting, the learning rate and the momentum parameter were varied across a plethora of potential values. Eventually, the most appropriate values for the learning rate and the momentum parameter are respectively 0.01 and 0.9, hence the BPNN was trained under acceptable error (misclassification) rates. The specific values provided the optimal trade-off between accuracy and computational cost. The number of epochs that are used to train the BPNN is set to 100. The training of the BPNN will stop when the error score E achieves 0.0001 or when the number of epochs reaches 100. In the current work, the nonlinear transfer function S is chosen to be the sigmoid function given by:

$$S(x) = \frac{1}{1+e^{-x}} \tag{4}$$

Finally, the number of neurons in the input and hidden layer is set to the number of inputs and one neuron is set in the output layer corresponding to the class label.

The probabilistic neural network (PNN) (Specht, 1990) is a Bayesian classifier that combines Bayesian decision-making strategy and a non-parametric estimator to obtain the probability density function within parallel intelligent systems. The architecture of the PNN is composed of numerous interconnected processing neurons organized in successive layers: the input layer, pattern layer, summation layer and decision (classification) layer. The neurons in the input layer distribute the input information to the neurons in the pattern layer. Those in the pattern layer are used to record the knowledge of classification extracted from the training pattern. Finally, the neurons in the summation layer calculate the maximum likelihood of a given pattern to be classified. Conditional on the information provided by the summation layer, the decision layer uses the Bayes's decision rule to perform the classification task. In this topology, all layers are fully interconnected and the pattern layer neurons are activated by exponential functions. Mathematically, let's consider an input layer with N neurons each corresponding to one input attribute, and a pattern layer with M neurons. The input vector x_i of size $1 \times N$ is processed by pattern layer neurons j through an exponential activation function given by:

$$a\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_j - x_i\right\|^2}{\sigma^2}\right) \tag{5}$$

where σ is a smoothing parameter that the width of the activation function $a(\cdot)$ with large values producing highly smoothing results. We use a score set to one. In particular, parameter σ is used to determine the Gaussian window width and the degree of interpolation between points (Specht, 1990). Accordingly, a small value indicates distinct modes of the distribution corresponding to the divergent training samples. As the value increases, the degree of interpolation increases too. For instance, a very large σ causes

the estimated density to be Gaussian regardless of the true underlying distribution. On the contrary, as σ decreases the flexibility of the decision boundary increases as well, yet yielding to overfitting. As we conducted various experiments with diverse values of σ, we found that the value of one is optimal, beyond which the classification rate does not change dramatically. In fact, according to Specht (1990) it is recommended to select a value of σ for which the overall training performance does not significantly alter.

Furthermore, the outputs of the pattern layer neurons are passed to the summation layer composed of K competitive neurons each corresponding to one class. Each summation layer neuron k takes the output of the pattern neuron to which it is connected and produces the output $f_k(x_i)$ as follows:

$$f_k(x_i) = \frac{1}{M_k} \sum_{\forall x_j \in y_k} a(x_i, x_j) \tag{6}$$

where y_k is the class corresponding to the summation of neuron k, and M_k is the number of training instances that belong to this class. Finally, the output of the summation layer neuron is given in terms of posterior class membership probability as:

$$\Pr(y_i = k \mid x_i) = \frac{f_k(x_i)}{\sum_{k=1}^{K} f(x_i)} \tag{7}$$

In this framework, the output with the highest value represents the most probable category.

Next, the radial basis function neural network (RBFNN) (Cybenko, 1989) is composed of input, hidden and output layers. The output of the jth unit, $H_j(x)$, in the hidden layer for an input x_i is computed as follows:

$$H_j(x) = H_j\left(\|x_i - c_j\|\right) = \exp\left(\frac{-(x_i - c_j)^2}{2\delta_j^2}\right) \tag{8}$$

where x_i is the pattern, c_j represents the position of the center of the jth Gaussian function, and δ is a Gaussian function parameter. The output y of the network is a linear combination of the K radial basis functions plus the bias w_0 as follows:

$$y(x_j) = \sum_{j=1}^{K} w_j H_j\left(\|x_i - c_j\|\right) + w_0 \tag{9}$$

In this work, the value of Gaussian width δ is set to one. Similarly to parameter σ and PNNs, in order to enhance robustness we conducted various experiments by varying the value of δ. The specific value minimizes misclassification rate.

The generalized regression neural network (GRNN) (Specht, 1991) is a parallel and memory-based system that estimates the regression surface of a continuous variable. The GRNN consists of four layers, including the input layer, pattern layer, summation layer, and output layer that yield the predicted value corresponding to an unknown input vector. Let's consider the system input be the vector X, and the desired estimate of the system output be the vector Y. Assume that the joint continuous probability density function of X and Y is represented by an unknown function $f(x,y)$. Then, the regression of y on x is represented by:

$$E[y \mid X = x] = \frac{\int_{-\infty}^{+\infty} y\, f(x,y)\, dy}{\int_{-\infty}^{+\infty} f(x,y)\, dy} \tag{10}$$

The unknown joint continuous probability density function $f(x,y)$ is estimated as follows:

$$\hat{f}(X,y) = \frac{1}{2\pi^{(p+1)/2}\sigma^{(p+1)}} \times n^{-1} \times \left\{ \sum_{i=1}^{n} \exp\left[-\frac{(X - X_i)^T(X - X_i)}{2\sigma^2}\right] \times \exp\left[-\frac{(Y - Y_i)^2}{2\sigma^2}\right] \right\} \tag{11}$$

where $\sigma < 1$ is the width of the Gaussian kernel, p is the dimension of X, and n is the sample size. When $\hat{f}(X, y)$ is substituted in $E[y \mid X = x]$, then the output function $Y(X)$ is given by:

$$Y(X) = \frac{\sum_{i=1}^{n} Y_i \exp\left[-\frac{D_i^2}{2\sigma^2}\right]}{\sum_{i=1}^{n} \exp\left[-\frac{D_i^2}{2\sigma^2}\right]} \tag{12}$$

where D_i^2 is expressed as follows:

$$D_i^2 = (X - X_i)^T (X - X_i) \tag{13}$$

Finally, we utilized as benchmark the approach of regression tree (RT) learning algorithms, also known as Classification and Regression Trees (CART) (Breiman et al, 1984). The RTs are nonparametric methods in estimating regression functions, which comprise a set of *if-then* rules minimizing the misclassification cost by considering both misclassification rate and variance. The major advantages of the RT technique include waiving of any assumptions regarding the distribution of predictors and gripping highly skewed numerical data and categorical inputs by using ordinal or non-ordinal tree construction (Breiman et al, 1984). Basically, RTs are designed to solve binary tasks, employ the Gini index to ranking tests, and prune trees by a cost-complexity model. The classification tree is represented graphically using nodes and branches, where each node indicates a decision about one of the attributes and gives rise to two branches. As the final edge, there is a terminal leaf node whereby homogeneity is obtained and the decision about the assigned class is taken. The tree is built by recursively partitioning the learning dataset into two subsets by binary split until the terminal nodes are achieved. When a new case is presented to the tree, it undergoes a test in the nodes, wherein each

test has exclusive and exhaustive outputs. As recommended by Breiman et al (1984), the Gini index is employed to diminish impurities in tree construction. The Gini index $G(t)$ of impurity of a node t is given as follows:

$$G(t) = \sum_{j \neq i} p(j|t) p(i|t) \qquad (14)$$

where i and j are diverse classes of the outputs, and $p(t)$ refers to the relative frequency of the first class. For instance, the goodness of splitting a data set D into subsets D_1 and D_2 is defined as:

$$G_{split}(D) = \frac{n_1}{n(G(D_1))} + \frac{n_2}{n(G(D_2))} \qquad (15)$$

where n, n_1 and n_2 are the sizes of D, D_1 and D_2, respectively.

3. PERFORMANCE MEASURES AND PROTOCOL OF EXPERIMENTS

In order to assess the performance of each methodology, three standard performance measures used in patter recognition and classification problems are calculated; namely accuracy (correct classification rate), sensitivity, and specificity. These are given by:

$$Accuracy = \frac{Classified\ Samples}{Total\ Number\ of\ Samples} \qquad (16)$$

$$Sensitivity = \frac{Correctly\ Classified\ Positive\ Samples}{True\ Positive\ Samples} \qquad (17)$$

$$Specificity = \frac{Correctly\ Classified\ Negative\ Samples}{True\ Negative\ Samples} \qquad (18)$$

where positive samples and negative samples respectively correspond to bankrupt and non-bankrupt cases in our experiments. The Type I error i.e., cost of misclassifying a

bankrupt company, is equivalent to one minus sensitivity, whilst Type II which represents

the cost of misclassifying a non-bankrupt company is equivalent to one minus specificity.

For robustness and generalization of the results, all experiments are performed

following the standard ten-fold cross-validation method. Accordingly, the data is

randomly divided into ten equally sized subsamples where subsets are put together to

form a training set and the remaining subset is used for testing. In this framework, every

data point gets to be in a validation set exactly once, and gets to be in a training set 9

times. Therefore, the bias and variance are significantly reduced as most of the data is

used for learning and for testing. Hence, ten out-of-sub-samples are used for testing the

effectiveness of each neural network and decision tree. Each performance measure is

averaged over all ten trials to get total effectiveness for each neural network classifier.

Subsequently, for each performance measure the mean and its standard deviation are

calculated.

4. EMPIRICAL RESULTS

We used the dataset from the University of California Irvine Machine Learning Database

Repository. The main reason is its accessibility due to the fact that it is freely available. It

is composed of 250 instances and 6 attributes each corresponding to qualitative

parameters in bankruptcy (i.e., the 7th attribute included in the dataset)[1]. The attributes are

industrial risk, management risk, financial flexibility, credibility, competitiveness, and

operating risk. Each attribute can assume one of the following three categorizations:

positive, average, and negative. The output label can take only one of the following two

---

[1] The specific freely available dataset can be found under the corresponding folder of the repository of the UCL
"*Center for Machine Learning and Intelligent Systems*": https://archive.ics.uci.edu/ml/datasets/Qualitative_Bankruptcy

values: non-bankruptcy or bankruptcy. There are 143 instances for non-bankruptcy and 107 instances for bankruptcy. The description of each attribute constituents is shown in Table 1. In our work, the parameters of BPNN, PNN and RBFNN were not optimized by a particular heuristic technique as this task is out of the scope of the current study. However, their respective key parameters were set to their standard values. In addition, it is worth to mention that a features reduction/selection scheme was not considered in our study as the number of attributes is limited. The classification results from a ten-fold cross validation approach are provided in Table 2 where the average and standard deviation accuracy, sensitivity, and specificity are presented for each artificial neural network architecture. Accordingly, the PNN and GRNN respectively yielded to the lowest (84.81%±0.0591) and highest accuracy (99.96%±0.0013), the RBFNN and PNN respectively achieved the lowest (51.78%±0.0626) and the highest sensitivity (100%), and the PNN and GRNN respectively produced the lowest (58.94%±0.0715) and highest specificity (100%). The RBFNN performed better than the BPNN in terms of accuracy (98.41%±0.0089 versus 84.81%±0.0591), sensitivity (100%±0.000 versus 73.03%±0.0842), and specificity (97.26%±0.0154 versus 81.77%±0.0722). Overall, the GRNN outperforms, followed by RBFNN, BPNN, and PNN. Moreover, the benchmark namely the RT learning algorithm which is a non-artificial neural network system, achieved 97.42%±0.0114 accuracy, 96.74%±0.0197 sensitivity, and 97.12±0.0101 specificity. Therefore, it performed better than BPNN and PNN in terms of all three performance metrics, yet underperformed vis-à-vis the RBFNN and GRNN in terms of all measures. The decision rules generated by the RT learning algorithm for the

classification of bankruptcy (B) versus non-bankruptcy are provided in Table 3 and displayed in a tree format in Figure 1.

[Please insert Table 1 here]

[Please insert Figure 1 here]

[Please insert Table 2 here]

[Please insert Table 3 here]

The experimental results from a ten-fold cross validation protocol indicated that GRNN and RBFNN demonstrated stronger predictability than the BPNN and PNN. Indeed, the best accuracy (overall correct classification rate) and specificity (correct classification of non-bankrupt companies) were achieved by the GRNN, whilst the RBFNN yielded the best sensitivity (correct classification of bankrupt companies). Specifically, the GRNN achieved the measure of accuracy 99.96%±0.0013, sensitivity of 99.91%±0.0030 and specificity of 100%. On the other hand, the RBFNN yielded 98.41%±0.0089 accuracy, 100% sensitivity, and 97.26%±0.0154 respectively. Additionally RBFNNs and GRNNs outperformed RTs, whilst the latter yielded better results compared to BPNNs and PNNs.

As RBFNN and GRNN obtained the highest sensitivity measurement, this point is of particular importance. Without a doubt the cost of misclassifying a bankruptcy for a company (Type I error) is far greater than the cost of misclassifying a non-bankrupt company (Type II error). For instance, the type I error involves for an investor or creditor the loss of an investment or debt that will not be reimbursed. Contrarily, a type II error implies loss of a potential profit. Therefore, from an investment perspective a robust prediction system should minimize the type I error. In this regard, the RBFNN is

privileged over the GRNN as it generated 100% correct classification (detection) rates of

bankrupt companies. However, the GRNN is definitively the best choice as it offers a

compromise between Type I and Type II error alongside with the highest overall

accuracy, i.e., 99.96%±0.0013.

To further examine the effectiveness of the GRNN and RBFNN in bankruptcy

qualitative data classification, we performed receiver operating characteristic (ROC)

curve analyses, under different data partition schemes whereby 50%, 60%, 70%, and 80%

of the entire sample is randomly used for testing. Such data partitions are conservative

ones as only a small data subset is employed for training the artificial neural network

which corresponds to real world financial applications. A ROC curve is a graph plotting

the combination of sensitivity (true positive rate, TPR) and the complement to specificity

(1− specificity = false positive rate, FPR). A test with high discrimination power has a

ROC curve approaching the upper left corner of the graph. Consequently, the closer the

ROC plot is to the upper left corner, the higher the accuracy is of the test. Fig.2 displays

the ROC for the GRNN and RBFNN under each random data partition scheme.

Accordingly, the comparisons in Fig. 2 depict better the ROC curve performance with

GRNN under all data partition schemes than with RBFNN. The ROC-based results

confirm that the GRNN is the optimal model as it provides a good compromise between

sensitivity and specificity.

[Please insert Figure 2 here]

Both RBFNN and GRNN are highly parallel networks with Gaussian kernels

allowing local approximations better than the BPNN and PNN. The high accuracy of the

GRNN over the remaining artificial neural networks examined in the current work can be

explained by its attractive features. Indeed, the GRNN has the ability to converge to optimal nonlinear regression surfaces as the number of sample becomes very large (Specht, 1991) by using a multivariate probability density function which is automatically determined from the underlying data during the learning phase. Therefore, the parallel structure of the GRNN automatically learns and generalizes immediately (Specht, 1991). Accordingly, the algorithm provides smooth transitions from one observed value to another. Also, it has the advantage of precise training with Gaussian functions and along with smooth interpolation in case of small samples.

In order to improve the performance of the BPNN is it essential to optimize its initial starting weights and parameters using a particular heuristic optimization technique which is left for future work as it is not the focus in the present study. Similarly, to improve the performance of the RBFNN and PNN, it is recommended to optimize the width of their respective Gaussian kernels. Those issues are our main future research directions. Without a doubt, by optimizing the parameters of the BPNN, PNN and RBFNN topologies, and by using larger datasets whenever they are available, all aforementioned networks are all expected to perform better to reach at least a comparable performance with the GRNN. In practice, the decision maker could start with the proposed GRNN when the data sample is not very large or comprehensive. It is worth indicating that in the current study, features selection was not pursued as the initial number of features was very limited. Although there are numerous and various feature selection techniques that could be explored, this issue is out of scope and was left for future investigation. Indeed, the GRNN topology achieved accuracy very close to 100% without any features selection

procedure, followed by the RBFNN. Such performance was achieved without additional computational costs associated with features selection techniques.

5. Conclusions

Financial risks are uncertainties commonly associated with credit risk, business risk, investment risk, and operational risk. Therefore, risk forecastability is an important topic for practitioners and academics as it could facilitate companies in taking suitable actions in advance to minimize defaults. Specifically, bankruptcy prediction and credit risk assessment are crucial for financial institutions in order to optimize future cash flows. In this regard, bankruptcy forecasting with high accuracy promotes institutional stability and profitability. A plethora of approaches used to classify companies into bankrupt and non-bankrupt ones include multivariate statistical models and artificial intelligence techniques including machine learning. The problem with applying statistical methods to bankruptcy evaluation is that some assumptions, such as the multivariate normality for independent variables, are actually violated. This renders them theoretically invalid for finite samples and less effective in real-world problems. In contrast to statistical modeling, artificial intelligence based methods do not involve any assumptions upon data distribution specificities. In fact, they automatically extract knowledge from training datasets and perform optimal classification. In addition, they are robust to noisy or incomplete data.

In our paper, we examined the performance of various artificial neural network topologies trained with *qualitative information* in bankruptcy prediction. Employing machine learning models used to mimic human brain decision making processing with

qualitative data, proved to be a more efficient way to build artificial agents for bankruptcy prediction taken from a cognitive informatics point of view.

We tested multi-layer back-propagation, probabilistic, radial basis function neural networks, and generalized regression neural topologies. They are all chosen based on their popularity in classification problems faced in engineering, economics and finance. They are also selected with respect to their inherent diversity in terms of information processing. Each one of these neural networks processes information differently to yield a final output. More advanced systems such utilizing deep learning are discarded in the present work as they require enriched datasets for better convergence and generalization of the results at the cost of serious high computational burden. In this regard, the artificial neural networks we used in our study are appropriate to the database we considered, despite the fact that in principle they are data consuming. Certainly, more advanced neural network topologies maybe encountered in the literature, such as extreme learning and deep learning, however they are not considered in this work as they require big datasets for convergence and generalization of the results at the cost of serious high computational burden. The artificial neural networks we explore are suitable for the relatively small database we utilized. To enhance robustness, the key parameters of all architectures were varied to select those with the best trade-off between accuracy and computational cost.

Contrary to previous works in the literature, the four artificial neural networks were exclusively trained with *qualitative data* where each attribute was designed to represent relevant qualitative information common deducted by the financial and banking industry for credit risk evaluation assessment and bankruptcy prediction. This was attempted in

accordance with domain field experts to find out the reasons for accepting or rejecting a financing inquiry and also to take corrective actions during the decision making process. The problem we tackled is similar to a supervised segmentation task applied to qualitative and non-continuous financial data. Generally, artificial neural networks can be promising machine learning tools in clustering and segmentation of continuous financial datasets. For instance, they are appropriate to study financial network connectedness as opposed to vector autoregression modeling adopted to measure such linkages (Diebold and Yilmaz 2014, 2016, Demirer et al 2018). The fact that artificial neural networks are data-driven and assumption-free intelligent systems makes them highly capable to analyze and simultaneously model interactions of noisy, nonlinear, and nonstationary data.

To summarize, our study demonstrated that the GRNN architecture may assist effectively the decision maker to accurately establish claim profiles which potentially may bring profits to the company under investigation.

REFERENCES

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., *Classification and regression trees*. Belmont CA: Wadsworth; 1984.

Chen S., Härdle, W.K., and Moro, R.A., Modeling default risk with support vector machines. *Quantitative Finance*, 2011, 11, 135-154.

Chen, J.-H., Developing SFNN models to predict financial distress of construction companies. *Expert Systems with Applications*, 2012, 39, 823-827.

Chen, M.-Y., Using a hybrid evolution approach to forecast financial failures for Taiwan-listed companies. *Quantitative Finance*, 2014, 14, 1047-1058.

Chen, N., Ribeiro, B., Vieira, A., and Chen A., Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert Systems with Applications*, 2013, 40, 385-393.

Cho, S., Hong, H. and Ha, B.-C., A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: for bankruptcy prediction. *Expert Systems with Applications*, 2010, 37, 3482-3488.

Chou, C.-H., Hsieh, S.-C., and Qiu, C.-J., Hybrid genetic algorithm and fuzzy clustering for bankruptcy prediction. *Applied Soft Computing*, 2017, 56, 298-316.

Ciampi, F., and Gordini, N., Small enterprise default prediction modeling through artificial neural networks: An empirical analysis of Italian small enterprises. *Journal of Small Business Management*, 2013, 51, 23-45.

Cybenko, G., Approximations by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems,* 1989, 2, 303-314.

Demirer, M., Diebold, F.X., Liu, L., and Yilmaz, K., Estimating global bank network connectedness. *Journal of Applied Econometrics*, 2018, 33, 1-15.

Desai, V., Crook, J., and Overstreet, G., A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operations Research*, 1996, 95, 24-37.

Diebold, F.X. and Yilmaz, K., On the network topology of variance decompositions: measuring the connectedness of financial firms, *Journal of Econometrics*, 2014, 182, 119-134.

Diebold, F.X. and Yilmaz, K., Trans-Atlantic equity volatility connectedness: U.S. and European financial institutions, 2004-2014, *Journal of Financial Econometrics*, 2016, 14, 81-127.

du Jardin P., Failure pattern-based ensembles applied to bankruptcy forecasting. *Decision Support Systems*, 2018, 107, 64-77.

Etemadi, H., Rostamy, A.A.A., and Dehkordi, H.F., A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. *Expert System with Applications*, 2009, 36, 3199-3207.

Gordini N., A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. *Expert Systems with Applications*, 2014, 41, 6433-6445.

Gupta, J., Gregoriou, A., and Ebrahimi, T., Empirical comparison of hazard models in predicting SMEs failure. *Quantitative Finance*, 2018, 18, 437-466.

Gupta, J., Gregoriou, A., and Healy, J., Forecasting bankruptcy for SMEs using hazard function: To what extent does size matter? *Review of Quantitative Finance and Accounting*, 2015, 45, 845-869.

Horta, I.M., and Camanho, A.S., Company failure prediction in the construction industry. *Expert Systems with Applications*, 2013, 40, 6253-6257.

Hua, Z., Wang, Y., Xu, X., Zhang, B., and Liang, L., Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications*, 2007, 33, 434-440.

Jang, J.-S.R., Sun, C.-T., and Mizutani, E., *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. PrenticeHall, NJ, 1997.

Khashman A., Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications*, 2010, 37, 6233-6239.

Kim, H.-J., Jo, N.-O., and Shin K.-S., Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. *Expert Systems with Applications*, 2016, 59, 226-234.

Lahmiri, S., and Gagnon, S., A Sequential Probabilistic System for Bankruptcy Data Classification. In D. Jakóbczak (Ed.), *Analyzing Risk through Probabilistic Modeling in Operations Research* (pp. 138-147). Hershey, PA: IGI Global, 2016.

Lahmiri, S., Features selection, data mining and finacial risk classification: a comparative study. *Intelligent Systems in Accounting, Finance and Management*, 2016, 23, 265-275.

Lee, S. and Choi W.S., A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*, 2013, 40, 2941-2946.

Li, H., and Sun, J., Predicting business failure using multiple case-based reasoning combined with support vector machine. *Expert Systems with Applications*, 2009, 36, 10085-10096.

Li, H., and Sun, J., Ranking-order case-based reasoning for financial distress prediction. *Knowledge-Based Systems*, 2008, 21, 868-878.

Li, H., Sun, J., and Sun, B. L., Financial distress prediction based on OR-CBR in the principle of k-nearest neighbors. *Expert Systems with Applications*, 2009, 36, 643-659.

Lin F., Liang D., Yeh C.-C., and Huang J.-C., Novel feature selection methods to financial distress prediction. *Expert Systems with Applications*, 2014, 41, 2472-2483.

Marais, M.L., Patel, J. and Wolfson, M. The experimental design of classification models: an application of recursive partitioning and bootstrapping to commercial bank loan classifications. *Journal of Accounting Research*, 1984, 22, 87-113, (Supplement).

Moro, R.A., Härdle, W.K., and Schäfer, D. Company rating with support vector machines. Statistics & Risk Modeling with Applications in Finance and Insurance *Statistics & Risk Modeling*, 2017, 34, 1-2, S. 55-67.

Ohlson, J.A., Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 1980, 18, 109-131.

Olson, D.L., Delen, D., and Meng, Y., Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, 2012, 52, 464-473.

Olson, D.L., Delen, D., and Meng, Y., Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, 2012, 52, 464-473.

Oreski, S. and Oreski, G., Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications,* 2014, 41, 2052-2064.

Ramirez, E., Castillo, O., and Soria, J., Hybrid system for cardiac arrhythmia classification with fuzzy k-nearest neighbors and multi layer perceptrons combined

by a fuzzy inference system. *The International Joint Conference on Neural Networks* (IJCNN), 18-23 July, Barcelona, Spain, 2003.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J., Learning representations by back-propagating errors. *Nature*, 1986, 323, 533-536.

Sarkar, S. and Sriram, R.S, Bayesian models for early warning of bank failures. *Management Science*, 2001, 47, 1457-1475.

Specht D.F., A general regression neural network. *IEEE Transactions on Neural Networks*, 1991, 2, 568-576.

Specht, D.F., *Probabilistic neural networks. Neural Networks*, 1990, 3, 109-118.

Tsai, C.-F., and Wu, J.-W., Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 2008, 34, 2639-2649.

Tseng, F.-M., and Hu, Y.-C., Comparing four bankruptcy prediction models: logit, quadratic interval logit, neural and fuzzy neural networks. *Expert Systems with Applications*, 2010, 37, 1846-1853.

Wang, G., Ma, J., Huang, L., and Xu, K., Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems*, 2012, 26, 61-68.

Wang, M., Chen, H., Li, H., Cai, Z., Zhao, X., Tong, C., Li, J., and Xu X., Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction. *Engineering Applications of Artificial Intelligence*, 2017, 63, 54-68.

West, D., Neural network credit scoring models. *Computers and Operations Research*, 2000, 27, 1131-1152.

Wiginton, J., A note on the comparison of logit and discriminant models of consumer credit behaviour. *Journal of Financial and Quantitative Analysis*, 1980, 15, 757-770.

Zaini, B.J., Shamsuddin, S.M., and Jaaman, S.H., Comparison between rough set theory and logistic regression for classifying firm's performance. *Journal of Quality Measurement and Analysis*, 2008, 4, 141-153.

TABLE 1. CONSTITUENTS OF ATTRIBUTES

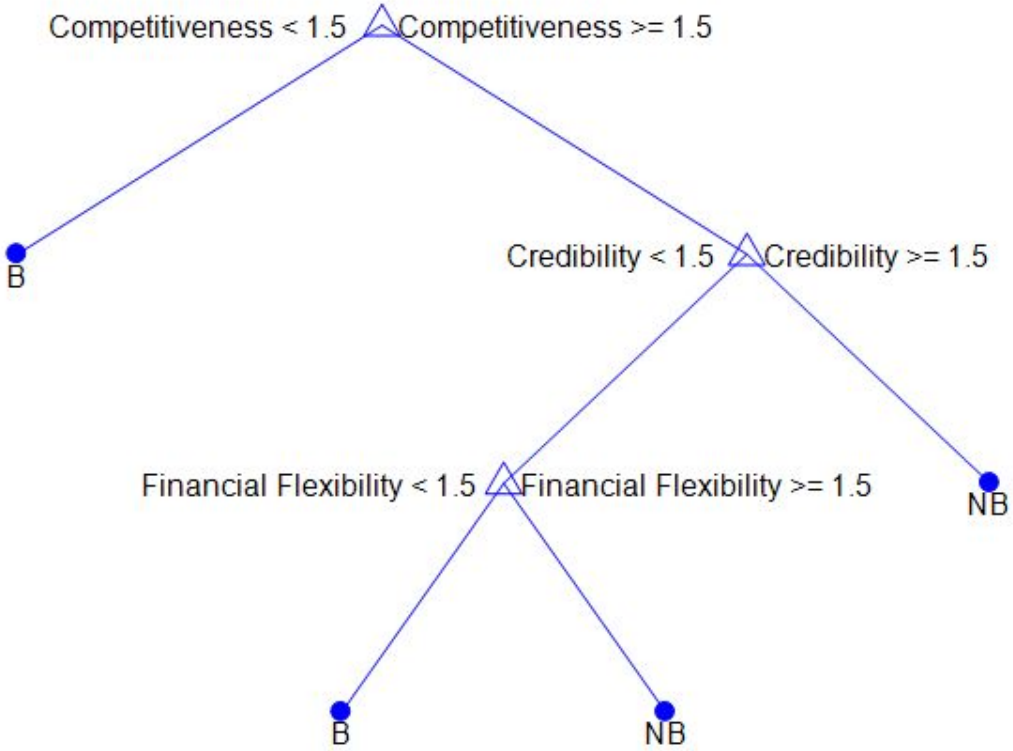| Attribute	Constituents
Industry risk	Government policies and international agreements, cyclicality, degree of competition, the price and stability of market supply, the size and growth of market demand, the sensitivity to changes in macroeconomic factors, domestic and international competitive power, product life cycle.
Management risk	Ability and competence of management, stability of management, the relationship between management/ owner, human resources management, growth process/business performance, short and long term business planning, achievement and feasibility.
Financial Flexibility	Direct financing, indirect financing, other financing.
Credibility	Credit history, reliability of information, the relationship with financial institutes.
Competitiveness	Market position, the level of core capacities, differentiated strategy.
Operating Risk	The stability and diversity of procurement, the stability of transaction, the efficiency of production, the prospects for demand for product and service, sales diversification, sales price and settlement condition, collection of A/R, effectiveness of sale network.

TABLE 2. CLASSIFICATION RESULTS

	Accuracy	Sensitivity	Specificity
RT	97.42% (±0.0114)	96.74% (±0.0197)	97.12 (±0.0101)
BPNN	84.81% (±0.0591)	73.03% (±0.0842)	81.77% (±0.0722)
PNN	79.64% (±0.0000)	51.78% (±0.0626)	58.94% (±0.0715)
RBFNN	98.41% (±0.0089)	100% (±0.000)	97.26% (±0.0154)
GRNN	99.96% (±0.0013)	99.91% (±0.0030)	100% (±0.000)

TABLE 3. DECISION RULES GENERATED BY RT LEARNING ALGORITHM TO CLASSIFY
BANKRUPT (B) AND NON-BANKRUPT COMPANIES (NB)

Node	Decision rules
1	*if Competitiveness < 1.5 then node 2*
	else if Competitiveness ≥1.5 then node 3
	else NB
2	*class = B*
3	*if Credibility<1.5 then node 4*
	else if Credibility≥1.5 then node 5
	else NB
4	*if Financial Flexibility<1.5 then node 6*
	else if Financial Flexibility≥1.5 then node 7
	else NB
5	*class = NB*
6	*class = B*
7	*class = NB*

FIGURE 1. RT ALGORITHM DECISION TREE

Notes: Resulting decision tree by RT learning algorithm whereby the final output is either bankruptcy (B) of non-bankruptcy (NB)

FIGURE 2. ROC CURVES UNDER DIFFERENT RANDOM DATA PARTITION SCHEMES
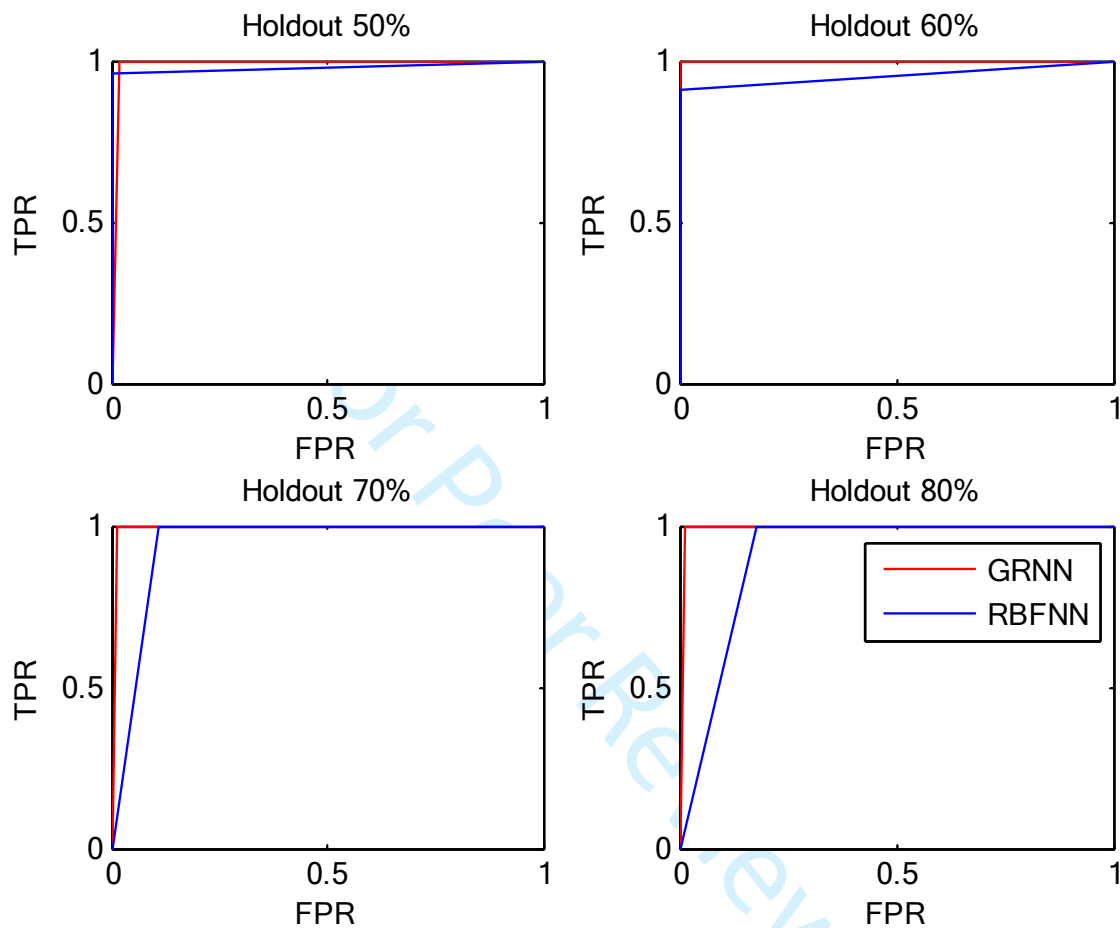
Notes: Receiver operating characteristic (ROC) curve generated under different random data partition schemes. TPR (true positive rate) is equivalent to sensitivity and FPR (false positive rate) is equivalent to one minus specificity.