BANKRUPTCY PREDICTION USING MACHINE LEARNING AND DEEP LEARNING MODELS

Sami Mestiri[1)*] , Manel Hamdi[2)]
[1]*Applied Economics and Simulation*
Faculty of Management and Economic Sciences of Mahdia,
University of Monastir, Tunisia. Rue Ibn Sina Hiboun, Mahdia Tunisia
[2]*International Finance Group Tunisia Lab,*
Faculty of Economic Sciences and Management of Tunis, University of Tunis El Manar, Tunis, Tunisia
ERF Research Fellow

Abstract:

In this study, we have compared the predictive power of five models namely the Linear discriminant analysis (LDA), Logistic regression (LR), Decision trees (DT), Support Vector Machine (SVM) and Random Forest (RF) to forecast the bankruptcy of Tunisian companies. A more advanced deep learning model, the Deep Neural Network (DNN) model, is also applied to conduct a prediction performance comparison with other statistical and machine learning algorithms. The data used for this empirical investigation covering 26 financial ratios for a large sample of 528 Tunisian firms. To interpret the prediction results, three performance measures have been employed; the accuracy rate, the F1 score and the Area Under Curve (AUC). By conclusion, DNN shows higher accuracy in predicting bankruptcy compared to other conventional models. Whereas, RF model performs better than other machine learning and statistical methods.

Keywords : *Bankruptcy Prediction; Machine Learning; Deep Learning Models; Confusion Matrix; ROC Curve.*

1. INTRODUCTION

Predicting bankruptcy has always known a great importance and a huge challenge for banks and lending institutions. Therefore, financial analysts and expert credits looking for using and finding the best techniques that can help them in decision making. Since long time, the traditional approaches have been widely used for bankruptcy prediction. These approaches are mainly based on the study of financial ratios, statistical models, and expert judgement. However, these models have limitations in predicting bankruptcy accurately (Hamdi, 2012, Altman et al., 1994, Hamdi and Mestiri, 2014).

In recent years, artificial intelligence and machine learning models have emerged as a powerful tool for bankruptcy forecasting, as they can handle big datasets and capture nonlinear relationships between input variables and the output. Several are the research studies which have been focused on bankruptcy forecasting by using machine learning models. The research paper of Ravi Kumar and Ravi (2007) summarizes existing researches on bankruptcy prediction studies using statistical and intelligent techniques during 1968-2005. For the same objective, Gergely (2015) has also presented a rich bibliographic review. He summarizes the short evolution of bankruptcy prediction and presents the main critiques made on modelling process for bankruptcy prediction.

*Coressponding author. *Email address:* mestirisami2007@gmail.com

Published by AFEBI Economic and Finance Review

Furthermore, the author announces avenues of future research recommended in these studies. More recently, a systematic literature was presented by Clement (2020) to predict bankruptcy. His review was conducted based on published papers between 2016 and 2020.

A more advanced modelling is applied in this study, we talk about the deep learning concept. For a more details about the deep learning approach, we can refer to the studies of Deng and Yu (2014) and Le Cun et al., 2015. Deep learning has been widely applied in computer vision (Chai et al., 2021), speech recognition (Roy et al., 2021), natural language processing (Xie et al., 2018), and medical image analysis (Suganyadevi, 2022). However, few studies have focused on the application of deep learning in finance (Qu et al., 2019).

The research paper is organized as follows: We provide a pertinent literature review related to bankruptcy prediction in Section 2. Section 3 presents the different statistical and artificial intelligence techniques used in this study. In section 4, we identify the data. The fifth section is devoted to the empirical investigation to predict the bankruptcy of Tunisian companies. And finally, we conclude in section 6.

2. LITERATURE STUDY

In past decades, the statistical models such as discriminant analysis (Beaver, 1966; Altman, 1968, Deakin, 1972) and logistic regression (Ohlson, 1980, Pang, 2006) were the two well-known and the most popular methods for predicting corporate bankruptcy. Slightly more recently, Mestiri & Hamdi (2012) used the logistic regression with random effect by integrating the sectoral effect in the logistic regression model in order to predict the credit risk of Tunisian banks. To forecast bankruptcy, several more developed methods have been employed. Some authors apply decision trees (Aoki and Hosonuma, 2004; Zibanezhad et al., 2011; Begović and Bonić, 2020), some others utilize a variety of machine learning models such as support vector machine (Shin et al., 2005; Härdle et al., 2005; Dellepiane et al., 2015), genetic algorithm (Shin and Lee, 2002; Kim and Han, 2003; Davalos et al., 2014) and random forest (Joshi et al., 2018; Ptak-Chmielewska and Matuszyk, 2020, Gurnani et al., 2021).

As a matter of fact and with the invasion of the artificial intelligence modeling algorithms since 1990s, in diverse domain, the artificial neural network was the most popular machine learning technique used in predicting financial distress (Odom and Sharda, 1999; Nasir et al., 2000; Atiya, 2001; Anandarajan et al. 2004; Hamdi, 2012). However, and despite the good forecasting results observed by applying this tool, deep learning models are the most applied today. This comes down to the ability of deep learning approach to overcome the some limitations in training the artificial neural network, with an important number of layers, such as the vanishing gradient, overfitting problem and the computational load (Kim, 2017).

Until now, few are the works which have been focused on the use of deep learning models to predict bankruptcy. Addo et al. (2018) build binary classifiers based on machine and deep learning techniques to predict loan default probability. More precisely, they used seven methods (the logistic regression, the random forest, the boosting approach and four deep learning models). In terms of the AUC and RMSE performance criteria, they concluded that the gradient boosting model outperforms the other models in solving binary classification problem. In another study, Hosaka (2019) proposed a convolutional neural network to predict bankruptcy. This model is specifically effective for image recognition, therefore the author has converted the set of financial ratios derived from the financial statements of Japanese companies before training and testing the network. The prediction performance results of the test set shown higher performance of the deep neural network compared to other employed methods such as decision trees, linear discriminant analysis, support vector machines, multi-layer perceptron, AdaBoost classifier.

For the same purpose, Noviantoro and Huang (2021) used machine learning and deep learning approaches to forecast bankruptcy of Taiwanese companies between 1999 and 2009.

They compared the best prediction performance of decision tree, random forest, neural network, support vector machine, Naïve bayes, k-nearest neighbour algorithm, logistic regression, rule induction and deep neural network. To evaluate the classifier's performance of these models, they computed the accuracy rate, F score and AUC of each techniques. As results, they found that random forest demonstrate the highest accuracy and AUC as well as the highest F score, and followed by the deep learning approach.

More recently, Shetty et al. (2022) utilized three machine learning methods, including, a deep neural network, extreme gradient boosted tree and support vector machine in order to predict the bankruptcy of 3728 Belgian small and medium firms for the period from 2002 to 2012. The authors concluded that the use of these different techniques yields roughly the same bankruptcy prediction accuracy rate of approximately 82-83%. Elhoseny et al. (2022) applied a new financial distress prediction tool by using an adaptive whale optimization algorithm with deep learning. They evaluated the ability of the proposed new approach, to predict the failure of any company compared to other models such as the logistic regression, the RBF Network, the Teaching Learning Based Optimization-DL and the deep neural network. The empirical results shown that the new approach based deep learning allows to reach better predictions.

3. RESEARCH METHODOLOGY

3.1 Linear Discriminant Analysis

Ronald Fisher (1933) pioneered work on discriminant analysis. In his work, he developed a statistical technique for defaults prediction, by developing a linear combination of quantitative predictor variables. This linear combination of descriptors is called discriminant function. The output of ADL is a score that is consists of classify a data observation between the good and bad classes.

$$Score = \sum_{i=0}^{p} a_i X_i \tag{1}$$

Where a_i are the weights associated with the quantitative input variables X_i.

The study of Altman (1968) is considered as the reference work that use the LDA to classify solvent and insolvent companies based on their financial statement data. More precisely, he utilized 5 ratios to create a linear discriminant function.

3.2 Logistic Regression

Logistic regression is a statistical method used for binary classification tasks (e.g. 0 or 1, bad or good, health or default, etc). Corresponding to Ohlson (1980), the outcome of LR model can be written as:

$$P(y = 1|X) = sigmoid(z) = \frac{1}{1+exp(-z)} \tag{2}$$

where $P(y = 1|X)$ is the probability of y being 1, given the input variables X, z is a linear combination of X: $z = a_0 + a_1X_1 + a_2X_2 +..+ a_pX_p$ where a_0 is the intercept term, $a_1, a_2, \dots, a_p$ are the weights, and $X_1, X_2, \dots, X_p$ are the input variables.

3.3 Decision Trees

Decision trees proceed a recursively partitioning of the data into subsets based on the values of the input variables, with each partition represented by a branch in the tree (Quinlan, J.R, 1986). The functionning of decision trees aimed to train a sequence of binary decisions that can be used to predict the value of the output for a new observation. Each decision node in the tree corresponds to a test of the value of one of the input variables, and the branches correspond to the possible outcomes of the test. The leaves of the tree correspond to the predicted values of the output variable for each combination of input values. The decision tree algorithm works by recursively partitioning the data into subsets based on the values of the input variables. At each step, the algorithm selects the input variable that provides the best split of the data into two subsets that are as homogeneous as possible with respect to the output variable. The quality of a split is typically measured using a criterion such as information gain or Gini impurity, which quantifies the reduction in uncertainty about the output variable achieved by the split.

Decision trees are typically not formulated in terms of mathematical equations, but rather as a sequence of logical rules that describe how the input variables are used to predict the output variable. However, the splitting criterion used to select the best split at each decision node can be expressed mathematically. Suppose we have a dataset with n observations and m input variables, denoted by $X_1, X_2, \ldots, X_p$, and a binary output variable y that takes values in 0,1. Let S be a subset of the data at a particular decision node, and let p_i be the proportion of observations in S that belong to class i. The Gini impurity of S is defined as:

$$G(S) = 1 - \sum_i (p_i)^2 \tag{3}$$

The Gini impurity measures the probability of misclassifying an observation in S if we randomly assign it to a class based on the proportion of observations in each class (Gelfandet al., 1991). A small value of G(S) indicates that the observations in S are well-separated by the input variables.

To split the data at a decision node, we consider all possible splits of each input variable into two subsets, and choose the split that minimizes the weighted sum of the Gini impurities of the resulting subsets. The weighted sum is given by:

$$\Delta G = G(S) - \left(\frac{|S_1|}{|S|}\right). G(S_1) - \left(\frac{|S_2|}{|S|}\right). G(S_2) \tag{4}$$

where S_1 and S_2 are the subsets of S resulting from the split, and $|S_1|$ and $|S_2|$ are their respective sizes. The split with the smallest value of ΔG is chosen as the best split. The decision tree algorithm proceeds recursively, splitting the data at each decision node based on the best split, until a stopping criterion is met, such as reaching a maximum depth or minimum number of observations at a leaf node.

3.4 Support Vector Machine

Support vector machine (SVM), developed by Vapnik (1998), is a supervised learning algorithm used for classification, regression, and outlier detection. The basic idea of this technique is to find the best separating hyperplane between the two classes in a given dataset. The mathematical formulation of SVM can be divided into two parts: the optimization problem and the decision function (Hearst et al., 1998).

Given a training set (x_i, y_i) where x_i is the ith input vector and y_i is the corresponding output label $y_i = (-1,1)$, SVM seeks to find the best separating hyperplane defined by:

$$w.x + b = 0 \tag{5}$$

where w is the weight vector, b is the bias term, and x is the input vector.

SVM algorithm aims to find the optimal w and b that maximize the margin between the two classes. The margin is defined as the distance between the hyperplane and the closest data point from either class. Then, SVM optimization problem can be formulated as:

minimize $\frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n} \xi_i$ subject to $y_i(w^t x_i + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$

where $||w||^2$ is the L2-norm of the weight vector, C is a hyperparameter that controls the trade off between maximizing the margin and minimizing the classification error, ξ_i is the slack variable that allows for some misclassifications, and the two constraints enforce that all data points lie on the correct side of the hyperplane with a margin of at least $1 - \xi_i$.

The optimization problem can be solved using convex optimization techniques, such as quadratic programming. Once the optimization problem is solved, the decision function can be defined as:

$$f(x) = sign(w \cdot x + b) \tag{6}$$

Where sign is the sign function that returns +1 or -1 depending on the sign of the argument. The decision function takes an input vector x and returns its predicted class label based on whether the output of the hyperplane is positive or negative. The details of the optimization process are discussed in (Chang and Lin, 2004; Cristianini and Shawe-Taylor, 2000; Gunn, 1998).

Thereafter, SVM finds the best separating hyperplane by solving an optimization problem that maximizes the margin between the two classes, subject to constraints that ensure all data points are correctly classified with a margin of at least $1 - \xi_i$. The decision function then predicts the class label of new data points based on the output of the hyperplane.

3.5 Random Forests

Random Forest is an ensemble of learning algorithm developed by Breiman (2001). It is a type of ensemble learning method that combines multiple decision trees for making predictions. The algorithm is called "random" because it uses random subsets of the features and random samples of the data to build the individual decision trees. The data is split into training and testing sets. The training set is used to build the model, and the testing set is used to evaluate its performance. At each node of a decision tree, the algorithm selects a random subset of the features to consider when making a split. This helps to reduce overfitting and increase the diversity of the individual decision trees.

A decision tree is built using the selected features and a subset of the training data. The tree is grown until it reaches a pre-defined depth or until all the data in a node belongs to the same class. Suppose we have a dataset with n observations and p features. Let X be the matrix of predictor variables and Y be the vector of target variables.

To build a Random Forest model, we first create multiple decision trees using a bootstrap sample of the original data. This means that we randomly sample n observations from the dataset with replacement to create a new dataset, and this process is repeated k times to create k bootstrap samples. For each bootstrap sample, we then create a decision tree using a random subset of p features. At each node of the tree, we select the best feature and threshold value to split the data based on a criterion such as information gain or Gini impurity. We repeat the above steps k times to create k decision trees. To make a prediction for a new observation, we pass it through each of the k decision trees and therefore obtain k predictions. For a more details about the technical analysis of random forests, see Baiu (2012).

3.6 Deep Neural Network

In deep learning, a deep neural network (DNN) is an enhanced version of the conventional artificial neural network with at least two hidden layers (Schmidhuber, 2015). Figure 1 illustrates the standard architecture of deep neural network.

E

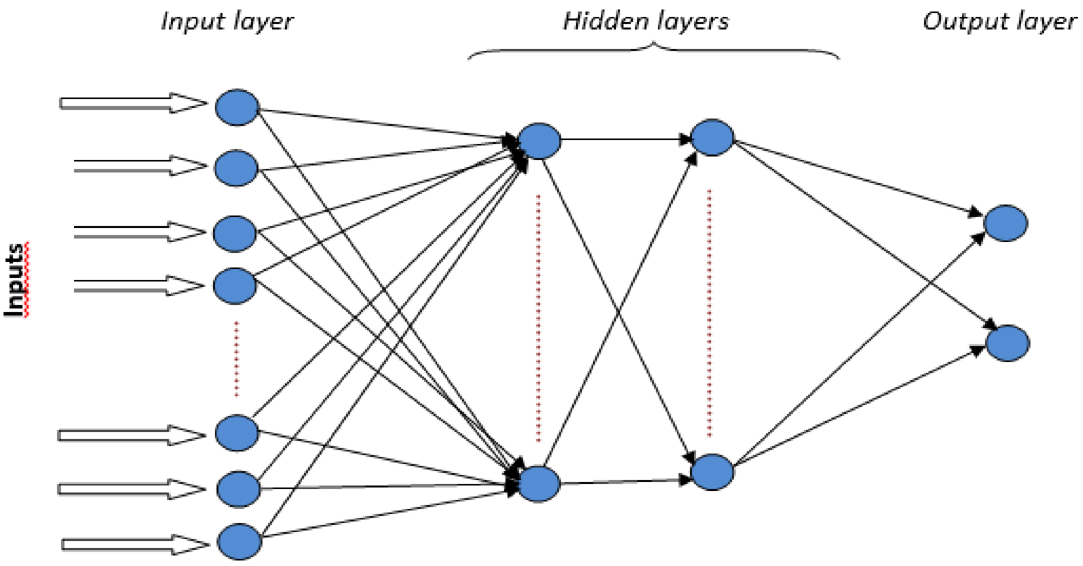

Figure 1: The Standard architecture of DNN

To fully understand how DNN works, a thorough knowledge of the basics of artificial neural network is then necessary. For more information, we can look at the studies of Walczak & Cerpa (2003) and Zou et al., 2008. According to Addo et al., 2018, the DNN output is computed as:

$$y(t) = \sum_{k=1}^{L} f(w_k + x_k(t)) + \epsilon(t) \quad (7)$$

Where W_K are the weights of the layer trained by backpropagation. $X_K(K = 1, \ldots, L)$ is the total number of sequence of real values called events, during an epoch. f is the activation function.

3.7 Data

A series of financial ratios was calculated from balance sheets and income statements of 528 firms in different sectors of activity for the period between 1999-2006. Our database is composed on 3065 credit files provided by the Central Bank of Tunisia. Table 1 presents the 26 ratios used as input variables of the proposed models.

Table 1: The series of financial ratios

Code	Variable	Code	Variable
R_1	Raw stock / Total assets	R_{14}	Rate of return on equity
R_2	Duration credit to the customer	R_{15}	Permanent capital turnover
R_3	Gross margin rate	R_{16}	Return on permanent capital
R_4	Operating margin rate	R_{17}	Rate of long-term debt
R_5	Ratio of personnel expenses	R_{18}	Ratio of financial independence
R_6	Net margin rate	R_{19}	Total debt ratio
R_7	Asset turnover	R_{20}	Immobilisation coverage by equity capital
R_8	Equity turnover	R_{21}	The long and medium term debt capacity
R_9	Economic profitability	R_{22}	Ratio of financial expenses
R_{10}	rate of return on assets	R_{23}	Financial expenses/total debt
R_{11}	Operating profitability of total	R_{24}	Working capital ratio

Code	Variable	Code	Variable
	assets		
R_{12}	Gross economic profitability	R_{25}	Relative liquidity ratio
R_{13}	Net economic profitability	R_{26}	Quick ratio

On the other hand, the estimated output (Y) can be written as a binary values:

$$Y = \begin{cases} 1 \ for \ default \ firm \\ 0 \ for \ healthy \ firm \end{cases} \tag{8}$$

Following this classification criterion, the out-of-sample test is composed on 448 healthy companies and 80 are bunkrapt companies.

4. RESULT AND DISCUSSION

4.1 Predictive Performance Measures

There are several criteria that can be utilized to compare and evaluate the predictive ability of the employed techniques including accuracy rate, F1 score and AUC.

4.1.1 Accuracy Rate

The accuracy rate is the most famous performance metric, deduced from the matix confusion (see. Table 2) and calculated following this formula:

$$Accuracy \ rate = \frac{(T_0+T_1)}{(T_0+F_1)+(F_0+T_1)} \tag{9}$$

Table 2: Confusion matrix

	Predicted class "0"	Predicted class "1"
Actual class "0"	True positive (T_0)	False positive (F_1)
Actual class "1"	False negative (F_0)	True negative (T_1)

4.1.2 F1 Score

The F1 score is also computed from the confusion matrix. The value of F1 score varies between 0 and 1, since 1 is the best possible score. A high F1-score indicates that the model shows both high precision and high recall, meaning it can correctly identify positive and negative cases.

$$F1 \ score = 2 * \frac{(Precision*Recall)}{(Precision+Recall)} \tag{10}$$

Where $Recall = \frac{T_0}{T_0+F_0}$ and $Precision = \frac{T_0}{T_0+F_1}$

4.1.3 AUC

Area Under Curve (AUC) is a synthetic indicator derived from the ROC curve. This curve is a graphical indicator used to assess the forecasting accuracy of the model (Pepe, 2000, Vuk and Curk, 2006). The ROC curve is based on two relevant indicators that are specificity and sensitivity (see Zweig & Campbell, 1993 and Mestiri & Hamdi (2012) for further details). This curve is caracterized by the 1- specificity rate on the *x* axis and by sensitivity on the *y* axis Where

$$= True\ positive\ rate = \frac{T_0}{Positives} = \frac{T_0}{T_0+F_1} \quad (11)$$

and

$$Specificity = True\ negative\ rate = \frac{T_1}{Negatives} = \frac{T_1}{T_1+F_0} \quad (12)$$

Moreover, the AUC measure reflects the quality of the model classification between heath and default firms. In the ideal case, AUC is equal to 1, i.e. the model makes it possible to completely separate all the positives from the negatives, without false positives or false negatives.

4.2 Results and Discussion

Table 3 presents the empirical results of the accuracy rate, F1 score and AUC criteria used to evaluate the classifier's performance of the applied models.

Table 3: Prediction results and models accuracy

Models	Accuracy Rate	F1- score	AUC	Rank
Linear Discriminant Analysis (LDA)	80.9%	0.890	0.574	5
Logistic Regression (LR)	85.8%	0.922	0.633	3
Decision Trees (DT)	74.3%	0.838	0.675	6
Random Forest (RF)	88.2%	0.933	0.815	2
Support Vector Machine (SVM)	84.8%	0.910	0.563	4
Deep Neural Network (DNN)	93.6%	0.964	0.888	1

According to Table 3, the deep neural network outperforms the other techniques in terms of all prediction performance metrics. DNN shows the highest accuracy rate with 93.6% whereas 88.2% for RF and 85.8% for LR. The lowest rate of prediction accuracy was found by the use of DT (74.3%). For the same objective to assess the predictive ability of the proposed algorithms, F1-score equal to 0.964 proves DNN's ability to identify with a great precision healthy companies from bankrupt companies. Since 1 is the best desired F1 score, DNN reaches the highest score while F1 score value was equal to 0.933, 0.922, 0.910, 0.890 and 0.838 for RF, LR, SVM, LDA and DT respectively.

Other graphical indicator was also used to evaluate the quality of classification of the models under study, we talk about the ROC curve (see. Figure 2). From this curve, we deduce the AUC measure. More the AUC value is near to unity more the model shows high quality of classification between health and default firms. Based on Table 3, the AUC of DNN yields 0.888. In the second rank, we found the RF with AUC equals to 0.815. The RL and ADL models present the worst classification results as the AUC is 0.633 and 0.574, respectively, in the testing sample.

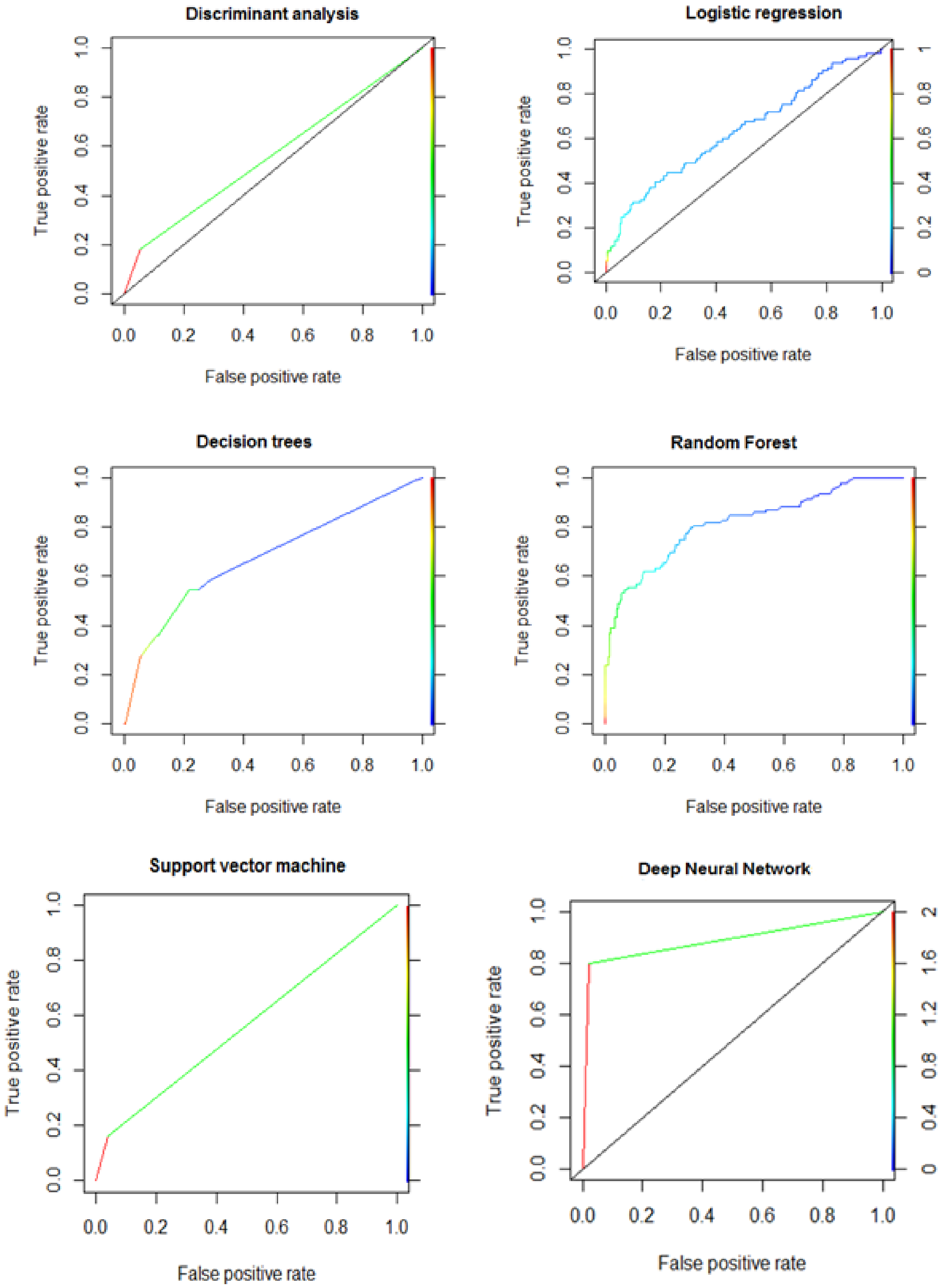

Figure 2: ROC curve for the five machine learning models and DNN

Similar conclusions were provided by Hosaka (2019). The study's findings of the author indicate that the deep neural network model specifically the convolutional neural network has superior prediction performance compared to statistical and conventional machine learning methods. Besides the work of Efron (1975) who prove the robustness of the LR model compared to the LDA. In a more research study, Barboza et al. (2017) obtained similar results in predicting bankruptcy of North American firms. Their empirical findings indicate that random forest is the best prediction model compared to LR and ADL. They found that RF reaches 87% of accuracy, whereas LR and LDA led to 69% and 50% of accuracy rate, respectively.

As a final conclusion, the ability of DNN outperforms the statistical and conventional machine learning models in forecasting bankruptcy. In the second rank we found that RF has a significantly higher prediction accuracy compared to other employed techniques. Based on our empirical investigation, the deep neural network can be considered as the best technique to detect company's financial distress and therefore can help to make managerial decisions.

To note, in this empirical task, we have used 20% of the sample (613 firms) are used for testing the prediction accuracy and classifier's quality of the models. For the training process, a deep feed-forward network with six hidden layers is adopted, a sigmoid activation function for the hidden layers and linear activation function for the output layer.

5. CONCLUSION

Bankruptcy prediction has always been of great concern to the credit financial institutions to make appropriate lending decisions. The purpose of this article is to develop an effective model that can be used in Tunisian banking sector to make a good forecast of the financial difficulties of companies. In this research study, we applied statistical, machine learning and deep learning models such as the ADL, LR, DT, SVM, RF and DNN in order to predict the financial distress of 528 Tunisian firms from different activity sectors. The empirical findings showed that DNN is a highly suitable tool for studying financial distress in Tunisian credit institutions. Compared to past work, this study is distinguished from other references in predicting bankruptcy as we have employed an interesting number of input features (26 ratios) as well as a large sample of firms in training phase (2452 ≈ 80% of total sample of firms). Wilson and Sharda (1994) used only five ratios (same input ratios employed by Altman, 1968) to predict the bankruptcy of 169 firms. The machine learning models applied in their work are the shallow neural network and multi-discriminant analysis.

In a related study, Chen (2011) utilized a set of eight selected features as inputs of machine learning models and an evolutionary computation approaches are used for predicting financial distress of 200 Taiwanese companies. The total sample of firms are divided into a training set and a testing set, with a distribution ratio near 2:1. To forecast the bankruptcy of Korean construction companies, Heo and Yang (2014) are used a total of 2762 samples and 12 ratios for training several models such as adaptive boosting with DT, SVM, DT and ANN. For future research studies, we can apply hybrid learning techniques by combining the DNN with other machine learning model which can provide higher performance than when using single model. In this context, Ben Jabeur and Serret (2023) proposed a combined method for bankruptcy prediction, the fuzzy convolutional neural networks. The present work as well as previous research supports the idea that artificial intelligence models performs better than traditional methods. However, it will be interesting for further researches to diversify the data sources and no use only classical numerical data (e.g. financial ratio data), by adding textual data (e.g. news, public report of the companies, notes and comments from experts, auditors' reports and managements' statements) that can improve the accuracy of financial distress prediction (Mai et al., 2019; Matin et al.,2019).

References

Addo, P.M., Guegan, D., Hassani, B. (2018). Credit Risk Analysis Using Machine and Deep Learning Models. *Risks* 6 (38), 1-20.

Altman, E. I., Marco, G. and Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal*

of Banking and Finance 18 (3), 505-529.
Altman, E.I., (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance* 23, 589-609.
Anandarajan, M., Lee, P., Anandarajan, A. (2004). Bankruptcy Prediction Using Neural Networks. *Business Intelligence Techniques*, 117–132.
Aoki, S. and Hosonuma, Y. (2004). Bankruptcy prediction using decision tree. *The Application of Econophysics*, 299–302.
Atiya, A.F. (2001). Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results. *IEEE Transactions on Neural Networks* 12 (4), 929-935.
Barboza F, Kimura H, Altman E (2017). Machine learning models and bankruptcy prediction.*Expert Systems with Applications* 83, 405–417.
Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research* 4, 71–111.
Begović, S.V. and Bonić, L. (2020). Developing a model to predict corporate bankruptcy using decision tree in the Republic of Serbia.*Facta Universitatis, Series: Economics and Organization* 17, 127-139.
Ben Jabeur, S., Serret, V. (2023). Bankruptcy prediction using fuzzy convolutional neural networks. *Research in International Business and Finance* 64, 101-844.
Berkson, J. (1944). Application of the logistic function to bio-essay *Journal of the American Statistical Association*, 39 :357–365
Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research* 13, 1063-95
Breiman, L. (2000). Some Infinity Theory for Predictors Ensembles *Technical Report*; Berkeley: UC Berkeley, vol. 577.
Breiman, L. (2001). Random Forests. *Machine Learning* 45, 5–32.
Breiman, L. (2004). Consistency for a Sample Model of Random Forests. *Technical Report* 670; Berkeley: UC Berkeley, vol. 670.
Chai, J., Zeng, H., Li, A., Ngai, E. W. T. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning With Applications* 6 (15), 100-134.
Chang, C., Lin, C.J (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2 (27),:1-27.
Chen, M. Y. (2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers & Mathematics with Applications* 62 (12), 4514–4524.
Clement, C. (2020). Machine learning in bankruptcy prediction – a review. *Journal of Public Administration, Finance and Law* 17, 178-196.
Cox, D. R. (1972). Regression Models and Life-Tables.*Journal of the Royal Statistical Society: Series B (Methodological)*34 (2), 187-202.
Cristianini, N., & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. *Cambridge University Press*, Cambridge, U.K., 1-189.
Davalos, S., Leng, F., Feroz, E.H, Cao, Z. (2014). Designing an if-then rules-based ensemble of heterogeneous bankruptcy classifiers: a genetic algorithm approach. *Intelligent Systems in Accounting, Finance and Management* 21 (3), 129–153.
Dellepiane, U., Di Marcantonio, M., Laghi, E., Renzi, S. (2015). Bankruptcy Prediction Using Support Vector Machines and Feature Selection During the Recent Financial Crisis. International Journal of Economics and Finance 7 (8), 182-195.
Deng, L., Yu, D. (2014). Deep Learning: Methods and Applications. *Foundations and Trends*

in Signal Processing 7(3–4), 197–387.

Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research* 10 (1), 167–179.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis.*Journal American Statistical Society* 7, 892-898.

Elhoseny, M., Metawa, N., Sztano, G., El-hasnony, I.M. (2022). Deep Learning-Based Model for Financial Distress Prediction. *Annals of Operations Research*, 1-23.

Fisher, R. (1933). The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7, 179–188.

Gelfand S. B., Ravishankar C. S., and Delp E. J. (1991). An iterative growing andpruning algorithm for classification tree design. *IEEE Transaction on PatternAnalysis and Machine Intelligence*, 13 (2):163-174.

Gergely, F. k. (2015) Bankruptcy Prediction: A Survey on Evolution, Critiques, and Solutions. *Acta Universitatis Sapientiae, Economics and Business* 3(1), 93-108.

Gurnani, I., Vincent, V. Tandian, F.S., Anggreainy, M.S. (2021). Predicting Company Bankruptcy Using Random Forest Method. *Proceedings in: IEEE* 2^{nd} *International Conference on Artificial Intelligence and Data Sciences* (AiDAS).

Gunn, S.R. (1998). Support vector machines for classification and regression. *Technical Report*, University of Southampton.

Hamdi, M. (2012). Prediction Of Financial Distress For Tunisian Firms: A Comparative Study Between Financial Analysis and Neuronal Analysis. *Business Intelligence Journal* 5 (2), 374-382.

Hamdi, M., Mestiri, S. (2014). Bankruptcy Prediction For Tunisian Firms: An Application Of Semi-Parametric Logistic Regression and Neural Networks Approach. *Economics Bulletin* 34 (1), AccessEcon, 133-143.

Härdle, W., Moro, R., Schäfer, D. (2005). Predicting Bankruptcy with Support Vector Machines. *In: Statistical Tools for Finance and Insurance*. Springer, Berlin, Heidelberg.

Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent System* 13(4), 18–28.

Heo, J., Yang J. Y. (2014) AdaBoost based bankruptcy forecasting of Korean construction companies. *Applied Soft Computing* 24, 494–499.

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. Expert Systems with Applications 117, 287-299.

Joshi, S., Ramesh, R., ; Tahsildar, S. (2018). A Bankruptcy Prediction Model Using Random Forest. *Proceedings in IEEE Second International Conference on Intelligent Computing and Control Systems* (ICICCS).

Kim, P. (2017). MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence. *Apress Book*, Edition Number 1, Number of Pages, 151.

Kim, M.-J. and Han, I. (2003). The Discovery of Experts' Decision Rules from Qualitative Bankruptcy Data Using Genetic Algorithms. *Expert Systems with Application* 25, 637-646.

Le Cun, Y., Bengio, Y., and Hinton, G. E. (2015). Deep learning.*Nature* 521(7553), 436- 444.

Mai, F., Tian, S., Lee, C., Ling, M. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research* 274 (2), 743-758.

Matin, R., Hansen, C., Hansen, C., Mølgaard, P. (2019). Predicting distresses using deep learning of text segments in annual reports. Expert Systems with Applications 132, 199-208.

Mestiri, S., Hamdi, M.(2012). Credit Risk Prediction: A Comparative Study Between Logistic Regression and Logistic Regression with Random Effects *International Journal of Management Science and Engineering Management* 7 (3), Taylor & Francis, 200-204.

Nasir, M.L., John, R.I., Bennett, S.C., Russell, D.M. and Patel, A. (2000). Predicting corporate bankruptcy using artificial neural networks. *Journal of Applied Accounting Research* 5 (3), 30-52.

Noviantoro, T., Huang, J.P. (2021). Comparing machine learning algorithms to investigate company financial distress. *Review of Business, Accounting & Finance* 1(5), 454 - 479.

Odom, M., Sharda, R. (1990). A neural network model for bankruptcy prediction.*In Proceedings of the International Joint Conference on Neural Networks* Vol. 2, IEEE Press, Alamitos, CA, 163–168.

Ohlson, J.A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy . *Journal of Accounting Research* 18 (1), 109–131.

Pang, S. (2006). Application of Logistic Regression Model in Credit Risk Analysis. *Mathematics in Practice and Theory* (9), 129–137.

Pepe, M. S. (2000). Receiver operating characteristic methodology. *Journal of the American Statistical Association* 95(49), 308–311.

Ptak-Chmielewska A., Matuszyk, A. (2020). Application of the random survival forests method in the bankruptcy prediction for small and medium enterprises.*ARGUMENTA OECONOMICA* 1 (44), 127-142.

Qu, Y., Quan, P., Lei, M., Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. Procedia Computer Science 162, 895-899.

Quinlan, J.R (1986). Induction of decision trees. *Machine Learning* 1, 81-106.

Ravi Kumar, P., Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. *European Journal of Operational Research* 180 (1), 1–28.

Roy, T., Tshilidzi, M., Chakraverty, S. (2021). Chapter 12 - Speech emotion recognition using deep learning. *New Paradigms in Computational Modeling and Its Applications*, 177-187.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks* 61, 85-117.

Shetty, S., Musa, M., Brédart, X. (2022). Bankruptcy Prediction Using Machine Learning Techniques. *Journal of Risk and Financial Management* 15(35), 1-10.

Shin, K.S, and Lee, Y.J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications* 23, 321–28.

Shin, K.S., Lee, T.S, Kim, H.J. (2005). An Application of Support Vector Machines in Bankruptcy Prediction Model.*Expert Systems and Applications* 28, 127-135.

Suganyadevi, S., Seethalakshmi, V., Balasamy, K. (2022). A review on deep learning in medical image analysis. *International Journal of Multimedia Information Retrieval*, 1-20.

Vapnik, V. (1998). The nature of statistical learning theory. New York: Springer.

Vuk, M., Curk, T. (2006). Roc curve, lift chart and calibration plot. *Organization Science* 3(1), 89-108.

Walczak, S., Cerpa, N. (2003). Artificial Neural Networks.*Encyclopedia of Physical Science and Technology* (Third Edition).

Wilson, R.L. and Sharda, R. (1994). Bankruptcy Prediction Using Neural Networks*Decision Support Systems* 11, 545-557.

Xie, Y., Le, L., Zhou, Y., Raghavan, V. V. (2018). Chapter 10 - Deep Learning for Natural Language Processing. *Handbook of Statistics* 38, 317-328.

Zibanezhad, E., Foroghi, D. and Monadjemi, A. (2011). Applying decision tree to predict bankruptcy. *Proceedings IEEE International Conference on Computer Science and Automation Engineering*, 165–169.

Zou, J., Han, Y., So, SS. (2008). Overview of Artificial Neural Networks. In: Livingstone, D.J. (eds) Artificial Neural Networks. *Methods in Molecular Biology* (458), 14-22.

Zweig, M., Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 39, 561-577.