Intelligent Systems with Applications 16 (2022) 200136

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications

An intelligent bankruptcy prediction model using a multilayer perceptron

Raffael Förch Brenes [a], Arne Johannssen [b,*], Nataliya Chukhrova [c]

[a] Technical University of Berlin, Berlin, Germany
[b] University of Hamburg, Hamburg, Germany
[c] HafenCity University of Hamburg, Hamburg, Germany

ARTICLE INFO

Keywords:
Artificial Intelligence (AI)
Artificial Neural Networks (ANN)
Business failure
Data Science
Deep Learning
Machine Learning

ABSTRACT

High bankruptcy rates can lead to the collapse of economic systems. Therefore, having accurate and reliable models to predict firms in financial distress allows for proper management of the economic losses helping to prevent such crises. Since the 1930s, more than 500 studies have been published in the field of bankruptcy prediction models. In this paper, we firstly give a comprehensive literature review on the topic of statistical and intelligent models to predict firms failure. Then, we closely examine the discriminatory power of a Multilayer Perceptron (MLP) in the context of bankruptcy prediction. For this purpose, we consider different setups of optimization algorithms, activation functions, number of neurons, and number of layers. To find the parameter setup that achieves the best results, we use various evaluation metrics such as average accuracy, specificity, sensitivity, and precision. The case study is based on a data set of Taiwanese firms and includes comprehensive comparative analysis. The proposed MLPs show superior performance, and we critically examine the differences between the methodologies to explain the discrepancies in the results.

1. Introduction

Business failure can have extreme repercussions on firm owners, employees, shareholders, and creditors. High rates of bankruptcy can be severe for the economic system and society. In particular, the last financial crisis after 2008 led 389 US banks into bankruptcy and produced serious instability in the European economy, primarily in peripheral countries. For example, in 2009, enterprise bankruptcies in Portugal increased by 49% (Kirkos, 2015). Even after more than a decade, appropriate models that can effectively predict bankruptcy are highly needed.

The development of bankruptcy prediction models is a topic that has been exhaustively studied in the past 90 years. Initially, researchers focused on univariate analysis until the 1960's where the first multivariate study appeared (Bellovary et al., 2007). Ever since, over 500 studies have been published in this field using near 20 different statistical and machine learning techniques (Shi and Li, 2019). For instance, one machine learning technique that has shown itself to be resilient to continuous technological advances is given by Artificial Neural Networks (ANNs). The first applications of ANNs date back to the late 1980s, and in the 1990's they rapidly managed to become the primary method of study (Bellovary et al., 2007). After Logit Analysis (LA), ANNs are the second most widely used method to predict financial distress (Shi and Li, 2019).

An ANN is a computational model inspired by the functionality of the nerve cells and their capacity to receive, analyze and convey information. ANNs have a unique feature that makes them stand out from other statistical and approximation methods. In particular neural networks, unlike the rest, can learn the distinctive patterns that define the relationship between input and output variables. This unique feature is the reason why ANNs are such powerful tools.

The overall performance of machine learning techniques, also known as intelligent techniques, in bankruptcy prediction is more than satisfying. Machine learning has outperformed classical statistical techniques for multiple reasons, not only in bankruptcy prediction but also beyond, for instance, in the fields of energy and buildings (see, e.g., Esen et al. (2008, 2009a, 2009b, 2017)) and statistical quality control (see, e.g., Yeganeh and Shadman (2021), Yeganeh et al. (2022a, 2022b)). In particular, machine learning techniques have great flexibility and are better at handling large data sets. Their non-parametric nature makes them more suitable to deal with complex non-linear problems, such as the prediction of bankruptcy. They do not require further assumptions

* Corresponding author.
E-mail addresses: raffael.forch@outlook.com (R.F. Brenes), arne.johannssen@uni-hamburg.de (A. Johannssen), nataliya.chukhrova@hcu-hamburg.de (N. Chukhrova).

https://doi.org/10.1016/j.iswa.2022.200136
Received 18 July 2022; Received in revised form 7 September 2022; Accepted 5 October 2022
Available online 7 October 2022

on the data and are also very robust when dealing with noisy data. However, they are also highly dependent on the characteristics of the data and output preferences, which is one reason why there is no consensus on which method is the best to predict the bankruptcy of firms.

In this paper, we firstly provide a comprehensive overview of the literature on bankruptcy prediction. We summarize and describe the main theoretical, statistical, and machine learning techniques developed in the last 30 years. Moreover, this paper is focused on the development of a Multilayer Perceptron (MLP), a type of ANN, to predict business failure. The objective is to determine how accurately an MLP can predict firms' bankruptcy. To this aim, we examine various setups of optimization algorithms, activation functions, number of neurons, and number of layers. In order to find the parameter setup that leads to the best results, we propose different evaluation metrics. The discriminatory power of MLPs for bankruptcy prediction is closely examined in the framework of an extensive case study based on a real data set using the financial data of Taiwanese firms. In this case study, twelve different MLPs are computed, and their performance is compared to the results of a past study conducted by Liang et al. (2016), i.e., we perform comprehensive comparative analysis.

In summary, the main contributions of this paper are:

- Comprehensive literature review of intelligent models to predict firms failure
- Investigation of the discriminatory power of MLPs for bankruptcy prediction
- Consideration of numerous parameter setups and various evaluation metrics
- Extensive case study including comparative analysis based on a real data set
- Critical examination of methodologies and discrepancies in the results

The rest of this paper is organized as follows. Section 2 provides an overview of the literature on bankruptcy prediction. Section 3 presents our proposed methodology for developing the MLPs. Section 4 gives insights into the real data set and data pre-processing steps. In Section 5, we report the results and conduct a detailed comparison with the results presented in Liang et al. (2016). In Section 6, we conclude by summarizing the main findings of this paper, and we give a detailed outlook on promising future directions of the topic.

2. Literature overview

The prediction of financially distressed firms is a topic that has been exhaustively explored in the financial field. The first studies date back to the 1930s, where the prediction of future bankruptcy was conducted using univariate analysis. The prediction was widely based on the comparison of individual ratios between failed companies and sound companies. The most commonly referred univariate study was proposed by Beaver (1996), and shortly after, in 1968, Altman presented the first multivariate study (Altman (1968)). Altman's work paved the way for the development of bankruptcy prediction models, and since then, over 400 studies have been published in this area (Shi and Li (2019)).

There are many factors that have driven the research of bankruptcy prediction models. On the one hand, financial institutions have a constant need to predict the possibility of a company's default. Recognizing in an early stage warning signals of bankruptcy in the quantitative and qualitative data of a firm allows better management of the economic losses (Omatu et al. (2021)). The other reason is subject to the development of more accurate prediction models. Two fundamental issues condition corporate failure prediction:

- Variable selection (see Section 2.1)
- Classification methodology (see Section 2.2)

2.1. Variable selection

Before developing a model to predict bankruptcy, it is crucial to establish the causes of failure. Financial problems can originate from numerous reasons. For instance, Bradley and Cowdery (2004) identified seven leading causes of financial default, such as outside business conditions, financing, inside business conditions, tax, dispute with a particular creditor, personal, and calamities. Finding the suitable parameters to measure the effect of each cause is a more difficult task. Researchers have relied on financial variables found in balance sheets and income statements or non-financial variables that describe the firm's structure, product, and management. Other familiar sources come from the economic sector of the company and evaluations of financial markets that measure the risk of failure. Du Jardin (2012) analyzed nearly 200 bankruptcy prediction studies and categorized the variables' typology into six different groups, such as financial ratios and statistical, variation, non-financial, market and financial market variables. He concluded that financial ratios are the most commonly used type of variables in the bankruptcy prediction literature, being used in 93% of the cases as the primary indicator of financial health.

One of the main reasons for the popularity of financial ratios is how easy it is to collect and analyze compared to non-financial data or financial markets data. Non-financial indicators and financial markets data are also not available for all firms. Financial ratios' forecast ability has been widely tested and has provided excellent results. However, Salmi and Martikainen (1994) pointed out that to accurately interpret the effect of financial ratios, it is also necessary to evaluate the company's size. Gupta et al. (2018) showed that a direct consequence of not accounting for the company's size is the parallel decrease of leverage and activity ratios and the increase of profitability and liquidity ratios.

Over the years, more than 500 different ratios have been used in bankruptcy prediction models. It is essential to identify the most valuable variables for the model. For example, some variables may be noisy, highly correlated, others redundant, or even irrelevant. Therefore, variable selection methods intend to find a representative subset of variables as independent as possible and its size sufficiently big to capture the features of the sample. There are three main approaches regarding variable selection (Du Jardin (2012)):

- their performance on a univariate statistical test,
- the result of selection techniques based on parametric tests,
- their ”reputation” in the financial literature.

As there are some problems related to these approaches, such as the association between variables and linearity assumptions, we refer the reader to Du Jardin (2012) and John et al. (1998) for discussions.

Further variable selection techniques depend on automatic search procedures and criterion evaluation. The mostly used criteria are *Wilks Lambda criterion* and *Likelihood ratio test*. Both criteria have been proven beneficial when using discriminant analysis and logistic regression, respectively. However, to what extent such selection methods can still deliver effective results when used with other modeling techniques, such as neural networks, is still unclear (Du Jardin (2012)).

Another issue is the instability of the variable performance through time. There is no guarantee that the variable performance will be stable over time (Du Jardin (2012)). Thus, bankruptcy prediction variables are contingent and not entirely capable of describing the permanent influences that lead to corporate failure (Bardos (1995)).

In general, variable selection needs to be in accordance with the characteristics of the modeling technique. A good fit between the variable selection procedure and the modeling method can help to increase the performance (Du Jardin (2012)). Finally, to give a perspective on variable selection, the number of different variables used from the earliest models to the latest exceeds 750. Despite the high number of variables, the average number of variables remains relatively constant, around 8 to 10 (Bellovary et al. (2007)).

2.2. Classification methodology

The nature of bankruptcy prediction can best be explained as a binary classification problem. Advanced statistics, machine learning algorithms, and theoretical models provide a broad spectrum of suitable methodologies to deal with this problem. The first two methodology types focus on using quantitative variables to indicate the firm's health, and the third relies on the qualitative causes of failure. Despite each model type's differences, they all rely on a similar statistical background (Adnan Aziz and Dar (2006)). In this section, we review the main methods used to predict insolvency with a focus on machine learning techniques as this study is mainly concerned on these methods.

2.2.1. Theoretical models

The studies done on bankruptcy prediction are mere of an empirical manner. Without a proper theoretical framework oriented on explaining the causes of bankruptcy, the research has focused on developing better empirical testing models. For instance, bankruptcy prediction models have disagreed with the definition of failure. Lim et al. (2015) examined 172 studies and noticed a significant discrepancy in the definitions used. Sometimes failure was defined as a form of underperformance where the company suffered financial stress, causing an inability to pay financial obligations. Other times the definition used was more in alliance with the legal terms of insolvency, where the firm is filing for bankruptcy or liquidation. Different definitions hinder the model comparison. New empirical models would benefit tremendously from a solid theoretical background that could lead the path. In this way, the repetition of econometric models for insolvency prediction could be reduced (Lim et al. (2015)).

Theoretical models deal mainly with qualitative data, with which they try to determine from a theoretical point of view the causes of default. The data used is selected to satisfy the arguments for the firm's failure proposed by the theory (Adnan Aziz and Dar (2006)). Onakoya and Olotu (2017) made a review of the six main bankruptcy theories. The exposed theories view the problem of bankruptcy as an entitlement problem. Namely, it is unclear how an inadequate amount of resources is distributed among an undue entity of claim holders. The foundation of these six theories relies on identifying the cause of default. They distinguish two causes of default: *economic distress* and *financial distress*. Financial distress was defined as the situation where a company without accounting for its financial debt would have reported positive earnings. Moreover, economic distress is a result of an insufficient revenue generation to cover costs, besides the cost of financing operations (Onakoya and Olotu (2017)).

Adnan Aziz and Dar (2006) reviewed a total of 46 different studies about insolvency prediction. Within the theoretical models, the predominant methods used were *Balance sheet decomposition measures, Gambler's ruin theory, Cash management theory* as well as *Credit risk theories*. One of the leading theoretical models was developed by Wilcox (1973). Wilcox replicated Beaver's original design and presented a theoretical framework based on Gambler's ruin theory to measure the predictive accuracy of Beavers' model. The model's accuracy was 94% outperforming most of the statistical models presented to that point.

The so called *liquidity, profitability and wealth theory* has come to be used most frequently among the empirical models. This theory is based on financial ratios to indicate the firm's health. Poor indicators can be translated into a higher probability of default (Lim et al. (2015)). The indicators can be ordered into three categories: liquidity, profitability, and wealth. This structure is considered very general, which makes it a weak and limited theory, but on the other hand, it has a certain flexibility and does not conflict with other theories (Hashi (1997)).

Despite the importance of developing a solid theoretical framework, the theory alone cannot describe the complexity of bankruptcy prediction. Empirical models are still needed but could be improved if a reliable foundation could support them.

2.2.2. Statistical models

From the 1960s until the 1990s, the trend was centered on statistical models. The both predominant techniques were *Multivariate Discriminant Analysis* (MDA) and *Logit Analysis* (LA) (Bellovary et al. (2007)).

In this direction, Altman (1968) introduced the first multivariate insolvency prediction model by suggesting an MDA. Dietrich and Kaplan (1982) constructed a linear model to classify loan risks based on three variables. Their results outperformed the models of Altman (1968) and Wilcox (1973). They used six variables found in traditional financial statements analysis to measure risk and have proven successful in explaining bond ratings. Martin (1977) developed an early warning model for bank failure using LA and compared it to a Linear Discriminant Analysis (LDA) model as an alternative to MDA. He concluded that LDA and LA models achieved similar results. Along the same line, Ohlson (1980) used LA to predict firms' default. The model used data from 1970 to 1976 and delivered accuracy rates of 96% and 95% for prediction within one and two years, respectively. A further important study was performed by Karels and Prakash (1987). They analyzed the normality assumption of financial ratios required in an MDA, transformed non-normal ratios into multivariate normal or "almost" normal, and conducted an LDA with the transformed ratios to compare the results with past studies. The model achieved a 96% prediction rate for sound firms and 54% for unsound firms. Jones and Hensher (2004) introduced a mixed LA to predict firms' distress and compared the results with a standard logit model. They obtained a better prediction rate for the mixed logit model than the standard logit model.

In addition, more innovative modeling techniques such as hybrid statistical models were proposed. Canbas et al. (2005) designed an Integrated Early Warning System (IEWS) to detect banks defaults. They analyzed the banks' financial characteristics using Principal Component Analysis and constructed the IEWS using discriminant, logit, and probit analysis. After reviewing some of the more relevant studies using statistical techniques, it is still necessary to emphasize the reliability of the models. MDA and LA were the most implemented methods, achieving a promising accuracy performance overall. However, the high prediction rates measured were mainly on a one-year horizon, and the violation of the restrictive assumption of the statistical models raises questions about the reliability of the results (Adnan Aziz and Dar (2006)). Additionally, testing the estimated model using a validation sample (generalizability) to verify the predictive claims is necessary. Unfortunately, nearly 60% of the statistical models did not use such a sample to validate their results (Bellovary et al. (2007)). All these considerations suggest that the predictive power is likely to be biased upwards (Adnan Aziz and Dar (2006)). From all statistical techniques, LA had the most reasonable generalizability (Dreiseitl and Ohno (2002)).

Moreover, MDA relies on two fundamental conditions. First, all explanatory variables are multivariate normal distributed, and second, all group dispersion matrices are equal for all the groups. Violating these conditions can affect the estimation of the discriminant function and bias the test of significance (Karels and Prakash (1987)), arguing that MDA may not be the optimal modeling option. However, MDA models' average accuracy rates are slightly below 80% (Alaka et al. (2018)), suggesting that MDA is relatively robust to assumption violation (Raghupathi et al. (1991)). Furthermore, the relation between the output variable and input variable in the case of bankruptcy prediction is often non-linear, making the choice of MDA still questionable because linear models have a hard time capturing non-linear patterns (Coakley and Brown (2000)). Consequently, an alternative to overcome the non-linear input-output relation and assumption violation was to apply LA models. LA is a generalized linear model, so it is less demanding, and it is a better predictor of non-linear functions than MDA. It is also important not to forget that LA performs best when the relationship between input and output variables follows a logistic form, which is still not necessarily given (Alaka et al. (2018)).

A downside of LA is that the implementation is more costly than MDA (Raghupathi et al. (1991)). Kristóf and Virág (2012) also found that LA is more sensitive to outliers than other methods. In general, statistical tools are also highly sensitive and reactive to multicollinearity. Consequently, multicollinearity can alter the results and impair the performance of the model. Hence, methods of choosing non-collinear variables are often employed in the literature. LA is significantly affected when compared with other statistical models (Alaka et al. (2018)).

Finally, it is important to evaluate the cost of a prediction error in the literature to analyze the accuracy rates. A prediction model can imply two types of errors: Type I and Type II. A Type I error is the misclassification of a bankrupt firm as a non-bankrupt firm, while a Type II error is the opposite, the misclassification of a non-bankrupt firm as a bankrupt firm. It is clear that from a lender's perspective classifying a bankrupt firm as a non-bankrupt firm is more costly than doing the opposite. Thus error cost should be an essential criterion for model evaluation (Adnan Aziz and Dar (2006)). Alaka et al. (2018) reported that MDA models accounted for more Type I errors than LA. The average frequency of Type I errors of MDA was around 25, while for LA it was nearly 19. Despite the relatively high accuracy of MDA estimators, the high Type I error can make us reconsider the prediction power of this tool. The LA accuracy rate is in balance with the frequency of Type I errors committed, which makes this prediction power reliable.

2.2.3. Machine learning techniques

Machine learning techniques appeared in the late 1980s as an alternative to statistical models (Bellovary et al. (2007)). There have been over 100 different papers published using machine learning techniques in the context of bankruptcy prediction, and it is not in the dimension of this paper to review them all. Instead, we focus on ANNs and Support Vector Machines (SVMs) as (1) they are the major methodologies in the context of intelligent bankruptcy prediction, (2) this study concentrates on MLPs, and (3) SVMs are an integral part of the comparative analysis performed in this paper. In brief, ANNs and SVMs can be explained as follows:

- An ANN tries to emulate the structure and functionality of the neurons in the brain. In a neural network, the processing unit is known as *artificial neuron* or *node*. The nodes are connected to each other by *weights*. Learning is achieved by adjusting the weights, depending on the learning algorithm that has been predefined (Priddy and Keller (2005)).
- An SVM is a robust supervised learning technique primarily used for classification (Clement (2006)). To make the groups separable, the data set is transformed into a high-dimensional data set using the *kernel method* (Qu et al. (2019)). Then an optimal separating *hyperplane* is derived using a linear model that maximizes the margin between the classes (Kumar and Ravi (2007)). In this way, the non-linear class boundaries are defined.

Both techniques are non-parametric in nature. Non-parametric models, despite requiring more computational power than their parametric counterparts, burst with great force into the world of forecasting models due to the development of new advanced technologies and the continuously growing size of data sets. Intelligent models have become the primary bankruptcy prediction method outperforming statistical models (Clement (2006)). We proceed by reviewing relevant studies on the application of ANNs and SVMs in bankruptcy prediction.

Artificial neural networks

Salchenberger et al. (1992) developed a Backpropagation Neural Network (BPNN) model with five variables from the CAMEL rating system to estimate the probability of default of saving and loan associations. The variables were selected from a sample of 29 using stepwise regression. They achieved better performance compared to LA. Another comparison of BPNN and LA was performed by Bell (1997). He predicted bank failure feeding an ANN with only one hidden layer with six neurons and one output neuron with 12 variables from 28 different candidates. He reported that neither the BPNN model nor the LA predictive power stand out. However, BPNN performed better in more complex situations.

Sharda and Wilson (1993) compared a BPNN model with an MDA model using the five-factor model of Altman. The neural network architecture consists of five input neurons, one hidden layer with ten neurons, and two output neurons. They used the Monte Carlo resampling technique to create 180 distinct training and validation data sets to test the performance strengths of both techniques. The BPNN model outperformed the MDA model in all the cases.

Barniv et al. (1997) described the filing of bankruptcy as a three-state outcome: acquired, emerging, and liquidated. They compared the prediction power of BPNN with a multi-state ordered logit and a Non-Parametric Multiple Discriminant Analysis (NPMDA). For the comparison, two models with twelve and five variables were used, respectively. The BPNN technique outperformed the multi-state ordered LA and NPMDA. Additionally, they also outperformed the nine variable logit model used by Ohlson (1980).

Swicegood and Clark (2001) predict banks failure using BPNN, MDA, and human judgment. Human judgment prediction is based on the experiential and intuitional judgment of individuals such as regulators or analysts with expertise in bank institutions. They made two different comparisons. The first one compared the MDA model with the BPNN using the entire data. The MDA classification's ability ranged from 79% to 86% while the BPNN model ranged between 78% to 81%. For the second comparison, 25% of the original sample was used, and all three models were compared. Here the BPNN model achieved higher prediction rates than both other models. In general, BPNN was the better model for identifying a bank's underperformance.

Yang et al. (1999) proposed a Probabilistic Neural Network (PNN) model and PNN without pattern normalization (PNN^*) to predict the insolvency of firms in the oil and gas industry of the U.S. Both PNNs were compared with a BPNN and Fisher Discriminant Analysis (FDA). They also divided the data set into two subsets, one with deflated ratios to account for fluctuations in factors over time that can affect the ratio values and the second data set without deflation. Two of the main advantages of PNN are its computational speed and that PNN uses the entire available data to build the model; thus, a validation data set is not required. The PNN^* and the BPNN models achieved better classification rates using nondeflated ratios while FDA outperformed all other models with the deflated data set.

Baek and Cho (2003) compared a BPNN with an Auto-Associative Neural Network (AANN). They used data of Korean bankrupt and non-bankrupt firms. As we discussed before, bankruptcy prediction data sets suffered from an imbalance problem, causing the false-negative error rate of the neural network to increase. Their approach to overcome this issue was to train the AANN only using the data of sound firms, so the network learns its "unique" features. This would result in a smaller output error for sound firms and a greater for unsound firms. The AANN achieved classification rates of 80% for sound firms and 50% for unsound firms outperforming the BPNN with 79% for sound firms and 24% for unsound firms.

Lee et al. (1996) developed three hybrid BPNN models to predict the financial distress of Korean firms. The proposed hybrid neural network was (1) an MDA-assisted, (2) an ID3-assisted, and (3) a Self-Organizing Feature assisted-Map (SOFM) neural network. For the first two hybrid neural networks, MDA, and ID3 were used for the variable selection process. The third model used SOFM, a type of unsupervised neural network model for variable selection. In this way, they combined an unsupervised learning (SOFM) method with supervised learning (backpropagation). The variables selected were 57 in total and were categorized into six groups: growth, profitability, stability, cash flow, activity, and credibility. They reported higher prediction rates for the SOFM-assisted neural network than for the other two methods.

Hosaka (2019) proposed a bankruptcy forecasting model using a

Table 1
Ranking of statistical and intelligent models

Model	Number of papers
Logit Analysis (LA)	123
Artificial Neural Network (ANN)	56
Linear Discriminant Analysis (LDA)	52
Multivariate Discriminant Analysis (MDA)	33
Support Vector Machine (SVM)	32
Decision Tree (DT)	21
Genetic Algorithm (GA)	20
Rough Set (RS)	13
Case-Based Reasoning (CBR)	6
Other statistical models	32
Other intelligent models	60

Convolutional Neural Network (CNN). A CNN is a type of ANN used mainly to analyze images. He used 2450 sound firms and 153 failed firms from Japan and utilized the average weight approach to increase its size. Then each observation was transformed into a grayscale image. Each financial ratio corresponded to a fixed pixel position, and the brightness of that pixel depended on the value of the financial ratio. Two different methods were tested to decide the position of each financial ratio. The first method determined the position of the financial ratios randomly, and the second method used correlation. Financial ratios with high correlations were placed as close as possible to each other. He compared the CNN model to LA, Classification and Regression Tree (CART), SVM, BPNN and AdaBoost models, and concluded that the CNN model had the highest accuracy rates.

Mai et al. (2019) compared a hybrid neural network model using textual disclosure with a CNN. To transform the textual data into numerical units, natural language processing was used, and a word embedding model was used to reduce the dimensionality of the textual model. The outputs of the word embedding technique were then fed into a neural network with one hidden layer and two output neurons for the prediction. The database consisted of 11,827 public US firms. They concluded that the embedding model outperformed the CNN.

Support vector machines

Min and Lee (2005) compared an SVM with MDA, LA, and BPNN. They used two kernel functions for the SVM: a Radial Basis Function (RBF) kernel and a polynomial kernel. To determine the kernel function's optimal parameter values and to prevent the problem of overfitting, they employed a grid-search technique using 5-fold cross-validation. The prediction accuracy using the training sample of MDA, LA, and BPNN was 78%, 79%, and 85%, and for the validation sample it was 79%, 78%, and 82%, respectively, while the SVM achieved 85% and 83% outperforming the other models.

Yeh et al. (2010) developed a hybrid model using Rough Set (RS) theory as a variable selection method for SVM and BPNN. Additionally, Data Envelopment Analysis (DEA) was integrated into both models to provide measures for the efficiency of a firm. The RS methodology selected 17 financial ratios, and to those 17 variables, the efficiency variable determined with DEA was added. In this way, four different models were created, alternating the intelligent technique (SVM and BPNN) and the number of variables (17 or 18). They reported that the SVM with DEA model outperformed all other models achieving an average accuracy of 86% followed by the SVM without DEA model with 86%. Both SVM models also achieved lower Type II errors than their counterparts with 10% and 18%.

Horak et al. (2020) also compared an SVM with RBF kernel and BPNN for bankruptcy prediction purposes. Both models used 22 input variables and data from 1582 industrial companies in the Czech Republic. They analyzed different BPNN architectures and opted for one hidden layer with nine neurons and two output neurons. The SVM model correctly predicted 99% of all healthy companies and only 8% of unsound companies, making the SVM not applicable in a real-world situation. In contrast, the BPNN model correctly classified almost 92% of all sound firms and 56% of unsound firms. The overall prediction rate for the SVM model was 76% and for the BPNN model 82%. They concluded that the BPNN outperformed the SVM model.

Yang et al. (2011) suggested a model that combines Partial Least Squares (PLS) based feature selection and SVM. They used three kernel functions: linear, RBF, and polynomial kernel, and compared the prediction accuracy to a Learning Vector Quantization Neural Network (LVQNN). PLS was used for variable selection; namely, PLS achieved promising results identifying correlations between financial indicators and describing the complex non-linear relationship between input and output variables. Hence two model types were selected, the LVQNN model using eight variables and the SVM models using seven variables. The classification rate of the SVM with linear, polynomial of order four and RBF kernel functions were 78%, 94%, and 97%, respectively, while the LVQNN achieved 70%. For the validation sample, the classification rates of the SVM models were 82%, 82%, and 82%, respectively, while for the LVQNN it was 78%. The seven variables SVM models outperformed the LVQNN model.

Kim (2011) tested an SVM model with third-degree polynomial kernel functions against BPNN, MDA, and LA models. He predicted hotel bankruptcy using five explanatory variables with a small data set of 33 Korean tourist hotels from 1995 to 2002. They reported for MDA and LA a classification rate of 72% each and an error rate of 27% and 20%, respectively. Both intelligent models outperformed their statistical counterparts. The classification rate was 95% for SVM and 91% for BPNN, and the corresponding overall error rate was 4% and 8%, respectively. He concluded that there was no significant difference between both intelligent models; thus, the BPNN weighted Type I errors more heavily and committed smaller estimated relative cost errors, suggesting that the BPNN may be a better predictor than the SVM.

2.2.4. Discussion

After reviewing some of the most important statistical and intelligent models, it is still unclear which is the best method to predict firms failure. Shi and Li (2019) ranked the frequency of each model that has been implemented using a sample of 321 papers, see Table 1. Following Table 1, we see that LA (38.32%) and ANN (17.45%) are the most commonly employed models.

Alaka et al. (2018) analyzed the overall accuracy of different models in 151 different studies. ANN model's average accuracy was near 84%, followed by SVM models with approximately 83% and DT models with roughly 80%. When comparing the rate of Type I and Type II errors, ANN was the tool that committed fewer Type I and Type II errors, followed by SVM. Further, they concluded that the AI tool with the lowest overall accuracy is a stand-alone Case-Based Reasoning (CBR) model. Kumar and Ravi (2007) argued that CBR's poor performance lies in its incapability to solve non-linear problems. Chuang (2013) added that CBR's low accuracy results from the tool failing to identify key features from less relevant features and assigning correct weights to each key feature a different.

Despite ANN and SVM models having the highest overall average accuracy, the lack of transparency of both models is also a clear downside. Tseng and Hu (2010) pointed out that the weight assignment of both models is illogical and complicated to interpret. Virág and Nyitrai (2014) allege a sort of trade-off between transparency and predictive ability: the higher the model's accuracy, the less transparent the model becomes.

Moreover, decision rules-generating tools (DT, CBR, and RS) have been proven helpful as feature selection methods, and in some cases, they had a good predictive capacity. Decision rules-generating tools are highly transparent and easy to interpret. Still, one drawback of decision rules tools is the generation of non-deterministic rules. Non-deterministic rules cannot be applied to new observations (firms), leading to the observation's non-classification (Kumar and Ravi (2007)). The non-deterministic problem arises when the set of rules, instead of working as a multivariate system, acts as multiple univariate systems (Alaka et al. (2018)).

Another criterion to take into account is the dispersion of the data. Data used to develop bankruptcy prediction models is usually imbalanced due to the low number of bankrupt firms compared to the non-bankrupt firms in the real world. High imbalances between the number of bankrupt and non-bankrupt firms can severely weaken the model's performance. Du Jardin (2015) pointed out that the distinctive characteristics of the data describing the failed firms will stay hidden by the high amount of data of non-bankrupt firms. Likewise, most AI tools require large samples to perform correctly. Haykin (1994) stated that to train an ANN properly, the size of the data set needs to be ten times the weight in the network. Kumar and Ravi (2007) reported that DT and LA also need a large sample to perform accurately, but CBR, RS, and SVM can do it with a small sample.

We have discussed the importance of variable selection methods for bankruptcy prediction models and the negative effect of multicollinearity and outliers on statistical models. Commonly, AI tools can achieve a good classification ability with almost any variable selection method and are more robust to multicollinearity (Liang et al. (2015)). Liang et al. (2015) analyzed the effect of variable selection on the performance of AI tools, and they concluded that variable selection does not always lead to an improvement of the classification capacity of the model. Outliers, however, can affect the performance of any tool (Tsai and Cheng (2012)). In general, intelligent models are more flexible and adaptable to other methods than statistical models. Iturriaga and Sanz (2015) showed that an effective combination of a feature selection method and a classification tool is better than a standalone classification tool. It is clear that to develop the best bankruptcy prediction model, the method to be selected should be based on the characteristics of the data and the output preferences.

3. Methodology

In this section, we develop various MLPs to predict bankruptcy. In particular, we propose specific predictive models and discuss different setups of optimization algorithms, activation functions, number of neurons, and number of layers. To this aim, we begin by dealing with the architecture of the MLPs in Section 3.1. Section 3.2 is then devoted to the configuration of the artificial neurons. Subsequently, in Section 3.3, necessary aspects of training the neural networks are discussed. In Section 3.4, we present the developed 12 different MLPs for bankruptcy prediction. Finally, we introduce four diagnostic methods that are useful to identify the parameter setup that achieves the best results (Section 3.5).

3.1. Architecture of the multilayer perceptrons

The prediction of bankruptcy is performed by using an MLP. An MLP is a type of neural network with a multiple-layer feedforward architecture. In other words, an MLP possesses one input layer, a set of hidden layers, and one output layer through which the signal flows strictly from input to output units (unidirectional). While hidden and output layers are composed of artificial neurons, the input layer of an MLP is composed of only input neurons. Input neurons unlike artificial neurons are only capable of receiving the data and passing it to the first hidden layer, where the first computations occur. Considering a model using the entire range of introduced variables, the MLP would have a total of 95 input neurons. Considering the relation between sample size and the number of variables in the sample, performing a dimensionality reduction technique is recommended to improve the performance. On the other hand, determining the remaining architecture of an MLP is not straightforward. For instance, finding an appropriate number of layers and a proper number of neurons per layer depends on the quantity and quality of the data as well as the nature and complexity of the problem (Silva et al. (2017)).

In general, there is no unique architecture or topology of an MLP for every modeling problem that will perform best. As we will see, later on, two different MLP architectures can have the same predictive capabilities. In such a case, the simpler model will always be preferred over the more complex ones. One common rule of thumb is that the number of neurons in the middle layer should not exceed 75% of input neurons. The idea behind this rule of thumb is that a network with too many connections can easily memorize the training sample, and consequently, it will be unable to generalize.

To propose an appropriate architecture of our models, we decided to create 12 different models of which six models use an MLP with one hidden layer (models M1–M6) and the other six use two hidden layers (models M12–M62). To determine the proper number of neurons for our models M1–M6, we begin by calculating the accuracy of each model, starting with one neuron and then iteratively increasing the number up to 50. For models M12–M62, we calculate the accuracy for one neuron in each layer and then increase the number of neurons up to 20. In total, we compare 400 different architectures for every MLP with two hidden layers. For each model, the architecture with the highest accuracy is chosen.

3.2. Configuration of the artificial neurons

In addition to defining the proper architecture of an MLP, the configuration of the artificial neurons plays a substantial role in the modeling process. The artificial neuron is the core element that produces the desired outputs given the predefined set of inputs. To process the data and calculate all fundamental operations, the artificial neuron takes the set of *external impulses* $\mathbf{x}$ and computes a linear combination using a corresponding set of *weights* $\mathbf{w}$ (Silva et al. (2017)):

$$\Sigma(\mathbf{x},\mathbf{w}) := \begin{pmatrix} w_0 & \mathbf{w} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} = \sum_{i=1}^{n} w_i x_i + w_0$$

The aggregated net input signal is then passed to the *activation function* g, which limits the amplitude of the output of the neuron in terms of the *activation potential* u and *activation threshold* Θ, i.e., we have

$$g(u) := \max(0, u(\Sigma,\Theta))$$

with

$$u(\Sigma,\Theta) := \Sigma(\mathbf{x},\mathbf{w}) - \Theta, \quad \Theta \in \mathbb{R}.$$

The *output signal* of the neuron is also often referred to as $\hat{y}$. In a neural network with many neurons, the output signal is also used as input for other interconnected neurons (Silva et al. (2017)).

Not all functions are appropriate to be used as activation functions in the setup of an artificial neuron. Activation functions need to (1) be continuous, (2) have a strictly increasing asymptotic behavior, and (3) be fully or partially differentiable. The most common form of activation functions used in ANN modeling is the *Sigmoid function* that has an *S*-shaped graph and a balanced proportion of linear and nonlinear attributes (Haykin (2009)). Examples of Sigmoid functions are given by the *Logistic function and Hyperbolic tangent function*. Another form of activation function is the *Rectified Linear Unit (ReLU) function* (Silva et al. (2017)). These functions are defined as

$$g(u) = \frac{1}{1+e^{-\beta u}} \quad (\text{Logistic function, see Figure 1}), \tag{3.1}$$

$$g(u) = \tanh(\beta u) \quad (\text{Hyperbolic tangent function, see Figure 2}), \text{and} \tag{3.2}$$

$$g(u) = \max(0,u) \quad (\text{ReLU function, see Figure 3}), \tag{3.3}$$

respectively, where β is a *slope parameter*.

Note that all three functions (3.1)–(3.3) are non-linear. The choice of a non-linear function is because it is better at capturing more complex

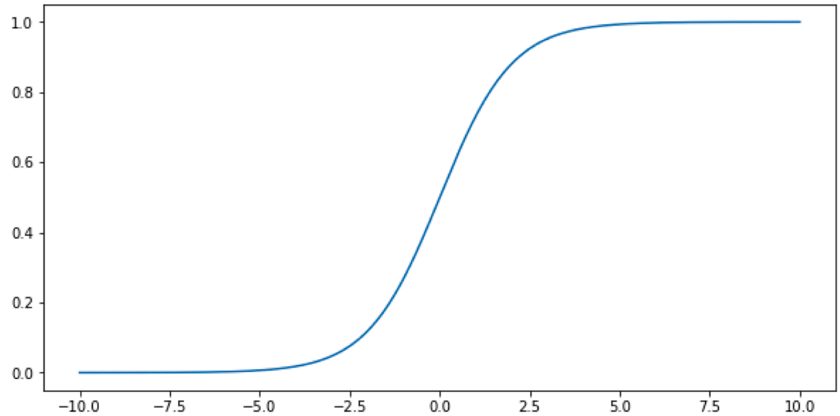

Fig. 1. Logistic function.

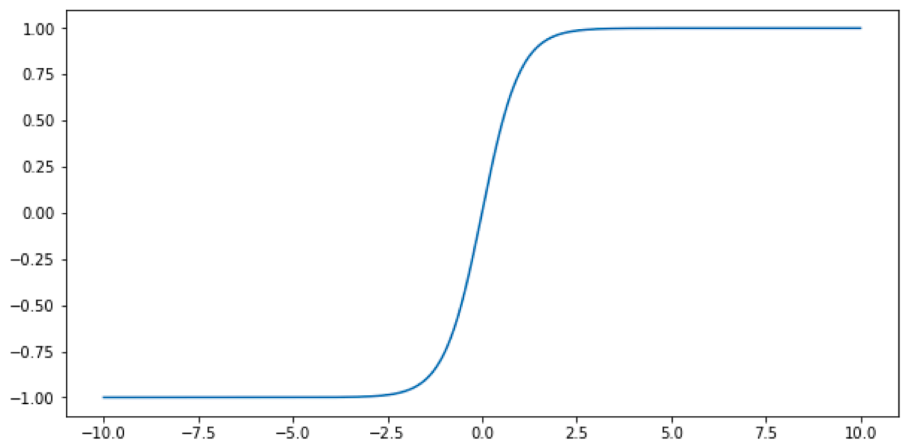

Fig. 2. Hyperbolic tangent function

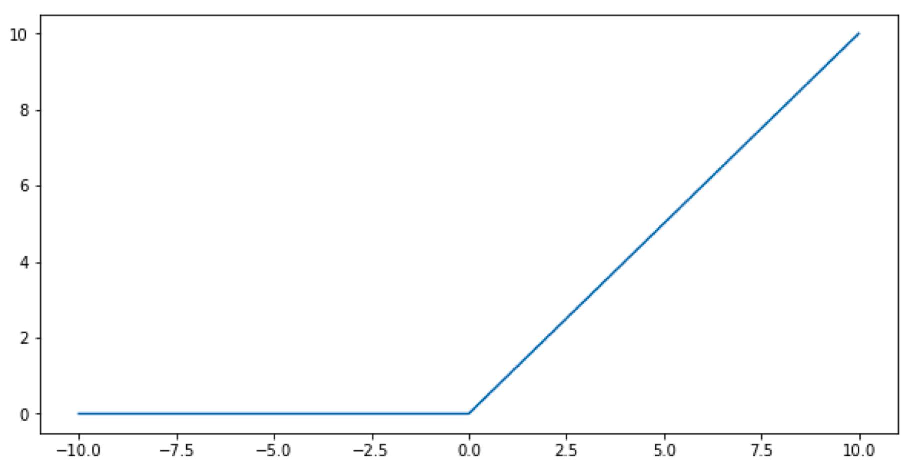

Fig. 3. ReLU function

relations between input and the neuron's output. Linear activation functions can only regress the covariates on the output signal (Denuit (2019)). Thus, the neuron would resemble the multiple regression model. For constructing our 12 different MLPs, we use the above introduced activation functions (3.1)–(3.3).

3.3. Training of the neural network

After determining the hyperparameters of an MLP, the next step is to train the network. Neural networks are powerful predictors because they are capable of learning the relationship between inputs and outputs (Silva et al. (2017)). *Training* in the context of neural networks describes the process of modifying the weights and thresholds of the neurons to produce the desired output. During training, the weights are calibrated using specific modification rules. This process is defined as *learning*. The expected result of training a neural network is to obtain a set of weights w that minimizes the error of the network. Hence, given a training sample $\mathscr{T}$ of size m,

Table 2
Proposed bankruptcy prediction models

Model	Optimization algorithm	Activation function	Neurons in layer 1	Neurons in layer 2
M1	Adam	ReLU	15	0
M2	Adam	Logistic	5	0
M3	Adam	Hyperbolic tangent	12	0
M4	sgd	ReLU	42	0
M5	sgd	Logistic	5	0
M6	sgd	Hyperbolic tangent	3	0
M12	Adam	ReLU	3	4
M22	Adam	Logistic	3	1
M32	Adam	Hyperbolic tangent	1	7
M42	sgd	ReLU	12	3
M52	sgd	Logistic	8	1
M62	sgd	Hyperbolic tangent	19	17

$$\mathscr{T} = \{\mathbf{x}_r, \mathbf{y}_r\}_{r=1}^{m},$$

we are looking to estimate the optimal set of weights $\mathbf{w}$ to solve the following minimization problem (Haykin (2009)):

$$\xi(\mathbf{w}) = \underset{\xi}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} E(y_i(\mathbf{w}), \hat{y}_i(\mathbf{w})) \tag{3.4}$$

In (3.4), ξ denotes a continuously differentiable cost function, E is a penalty function, $\hat{y}_i$ corresponds to the output of the ith neuron and y_i is the ith real observed value. To solve the optimization problem (3.4), we need to find the optimal solution $\mathbf{w}^*$ that satisfies the necessary condition

$$\nabla\xi(\mathbf{w}^*) = 0,$$

where $\nabla\xi(\mathbf{w})$ is the *gradient vector* of the cost function.

Commonly, neural network models predicting insolvency are trained using *supervised learning*. This learning technique requires a training data set consisting of inputs and desired output values (Silva et al. (2017)). The behavior of this training algorithm is often described as a "teacher" showing the network which is the correct response for each set of inputs presented. That is, the weights and thresholds adjustment are repeated until the difference between the network's output and the desired output has an acceptable value, i.e., when the network error reaches the desired minimum.

The most common training algorithm for MLPs is the *back-propagation algorithm*. The back-propagation algorithm is based on an iterative optimization algorithm called *gradient descent*. The training is divided into two phases: the *forward phase* and the *backward phase* (Haykin (2009)). The back-propagation algorithm iterates between both phases applying the correction

$$\Delta w_{ij}(r) := -\eta \frac{\delta\xi(r)}{\delta w_{ij}(r)}$$

to the weight $w_{ij}(r)$ repeatedly until the network error is sufficiently small. The term η is called the *learning rate* parameter of the back-propagation algorithm.

Additionally, the weight adjustment during the back-propagation algorithm is usually performed on an *example-by-example basis*, which is also known as *on-line learning*. In on-line learning, the procedure starts by presenting the first example pair $\{\mathbf{x}(1), \mathbf{y}(1)\}$ in the epoch to the network for which the output is calculated, and then the adjustments of the weights are done by using the gradient descent method. After the first adjustment, we proceed with the next example pair $\{\mathbf{x}(2), \mathbf{y}(2)\}$ and repeat the procedure until the last example $\{\mathbf{x}(m), \mathbf{y}(m)\}$ was accounted for. Modifying the weights on an example-by-example basis gives the search of the optimal weight vector $\mathbf{w}^*$ a stochastic nature that makes the learning process less likely to get trapped in a local minimum.

For our 12 models we implement two optimization algorithms: the *stochastic gradient descent* (sgd) and the *adaptive moment* estimation optimization algorithm (Adam). Adam is an algorithm that makes use of stochastic gradient-based optimization with individual adaptive learning rates for different parameters (Kingma and Ba (2015)). For more information on the Adam algorithm, we refer to Kingma and Ba (2015). Both optimization algorithms use the log-loss function

$$\mathrm{ell}_i := -(y_i \ln(\hat{y}_i) + (1 - y_i \ln(1 - \hat{y}_i)) \text{ with } i \in \mathbb{N}. \tag{3.5}$$

3.4. Proposed bankruptcy prediction models

The MLPs are implemented by using the open source `Python` module `Scikit-learn`. `Scikit-learn` began in 2007 as *Google Summer of Code* project by David Cournapeau and provides a wide number of state-of-the-art machine learning algorithms. In particular, we employ the function `MLPclassifier`, an MLP classifier that uses the log-loss function (3.5). In Table 2, we list all 12 models and their specifications.

To train the different models, we set the batch size to 1, which means that weights adjustment occurs by on-line learning. We also set the parameter `random state` to 42 to randomly generate initial weights values and the parameter `shuffle` to true, which shuffles the sample after each iteration. It is recommended to set the random state parameter to 0 or 42. The initial learning rate of all models is set to 0.001. In addition, for all models using sgd, we set the parameter `learning rate` to adaptive, which maintains the learning rate constant as long as the training loss continues decreasing. To avoid oversampling, the `MLPclassifier` uses early stopping. Therefore, a tolerance value is predefined under a parameter named `tol`. If the value of the log-loss function is not improving by at least `tol` for more than ten consecutive epochs, then convergence is reached, and training is stopped.

3.5. Diagnostic methods

There are several methods to evaluate the prediction performance of classification models. The most common metrics used for bankruptcy prediction are the average accuracy prediction rate as well as Type I and Type II error rates. To examine the results of the proposed models, we employ the average accuracy prediction rate and three other diagnostic methods: specificity, sensitivity and precision. These metrics are calculated using four different values: number of True Positives (TP), number of True Negatives (TN), number of False Positives (FP) and number of False Negatives (FN).

The *Average Accuracy* (AA) prediction rate returns the percentage of all correctly classified samples by the model over a given data set and is defined as

$$\mathrm{AA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}.$$

It is generally agreed upon the costly repercussions of committing Type I errors (FP) for a bankruptcy prediction model. In some cases, Type I errors can be a more critical indicator of the goodness of the model than the average accuracy prediction rate. One way of measuring Type I errors is by calculating the *Specificity* of the model. The specificity is the ability of the model to identify TNs. In other words the specificity returns the probability of identifying a bankrupt firm as bankrupt. A high specificity refers to a lower rate of Type I errors. Specificity or also known as TN Rate (TNR) is defined as

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}.$$

A complementary measure to specificity is the *Sensitivity* of a model. Sensitivity, also known as TP Rate (TPR), is a metric to examine the rate of Type II errors (FN). The sensitivity returns the probability of identifying a non-bankrupt firm as non-bankrupt. A high sensitivity translates in a lower Type II error rate. The sensitivity is defined by

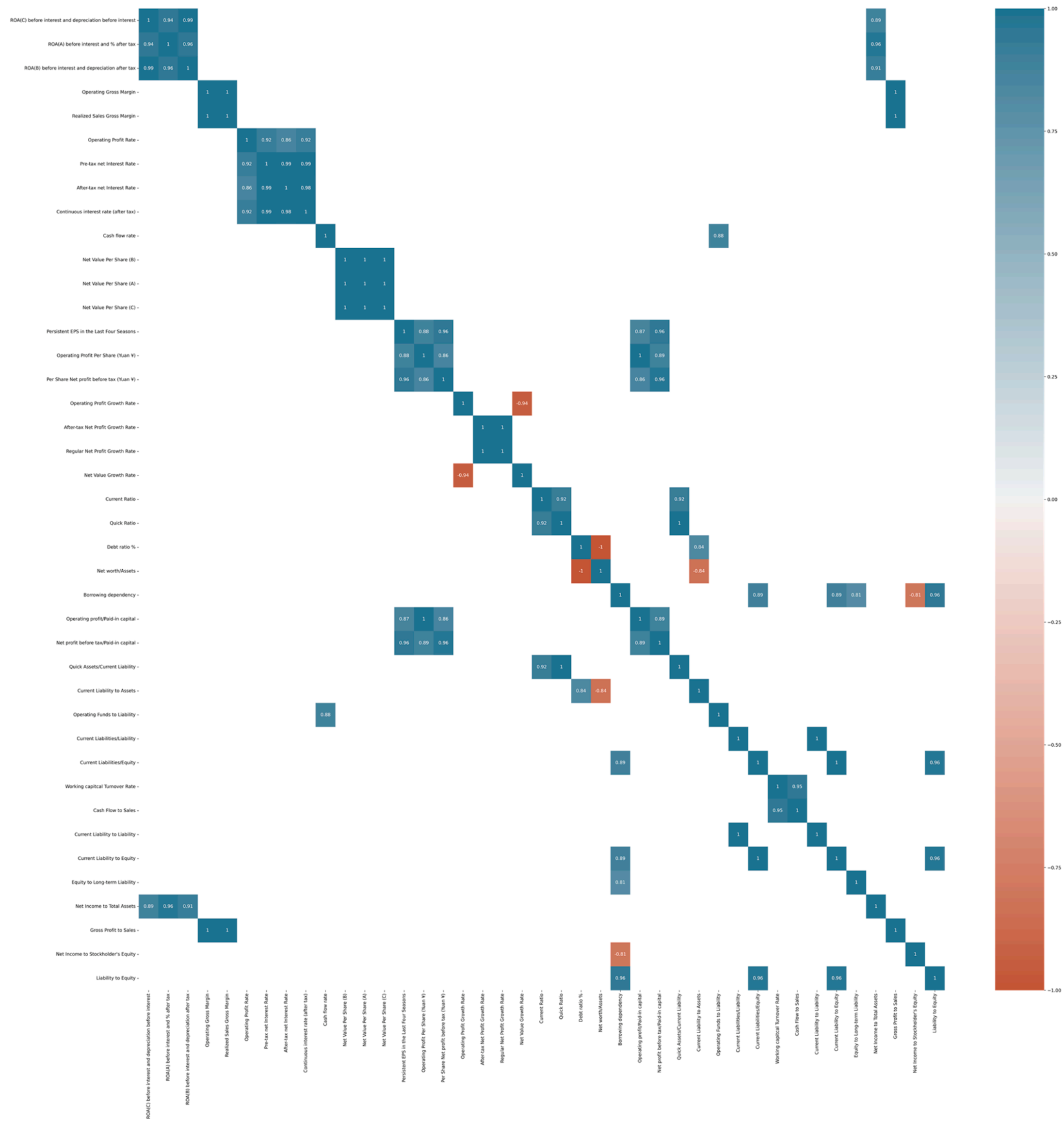

Fig. 4. Correlation heatmap

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}.$$

The fourth performance metric that we use to evaluate the models is the *Precision* rate or also known as Positive Predictive Value (PPV). The Precision rate is defined as

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

and returns the ratio of TP prediction of all positive predictions. In other words, the precision is the percentage of firms classified as non-bankrupt of which are actually non-bankrupt.

4. The data set and data pre-processing

In this section, we present the real data set (Section 4.1) and discuss necessary data pre-processing steps to improve the performance of the proposed MLP models (Section 4.2).

4.1. The data set

The data set is obtained from the UC Irvine Machine Learning Repository (UCIMLR) (https://archive.ics.uci.edu/ml/datasets/Taiwanese+Bankruptcy+Prediction), a center that provides data sets to the machine learning community. Originally the sample was collected by the Taiwan Economic Journal (https://www.finasia.biz/) and includes the financial information of industrial, electronic, shipping, tourism, and retail companies for the years 1999–2009. The sample comprises 96 different financial ratios and 6,819 observations, of which 220 are bankrupt firms and 6,599 are non-bankrupt firms. The definition of bankruptcy is based on the regulations of the Taiwan Stock Exchange.

4.2. Data pre-processing

The aim of pre-processing the data before applying the neural network model is to improve its performance. Therefore, it is essential to identify irregularities in the data. There is an obvious irregularity related to the imbalance between the number of bankrupt and non-bankrupt firms. To address this issue, we retrieved a random subset of which 60% of the firms are non-bankrupt and 40% are bankrupt. In other

Table 3
Table of confusion

Model	TP	FP	FN	TN
M1	98	5	18	44
M2	99	4	19	43
M3	98	5	18	44
M4	94	9	18	44
M5	93	10	21	41
M6	94	9	19	43
M12	97	6	15	47
M22	99	4	19	43
M32	98	5	17	45
M42	94	9	18	44
M52	92	11	20	42
M62	95	8	18	44

Table 4
The performance of MLP prediction models

Model	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)	No. of epochs
M1	86.06	89.80	84.48	95.15	90
M2	86.06	91.49	83.90	96.12	206
M3	86.06	89.80	84.48	95.15	84
M4	83.64	83.02	83.93	91.26	190
M5	81.21	80.39	81.58	90.29	207
M6	83.03	82.69	83.19	91.26	217
M12	87.27	88.68	86.61	94.17	136
M22	86.06	91.49	83.90	96.12	196
M32	86.67	90.00	85.22	95.15	45
M42	83.64	83.02	83.93	91.26	180
M52	81.21	79.25	82.14	89.32	141
M62	84.24	84.62	84.07	92.23	187

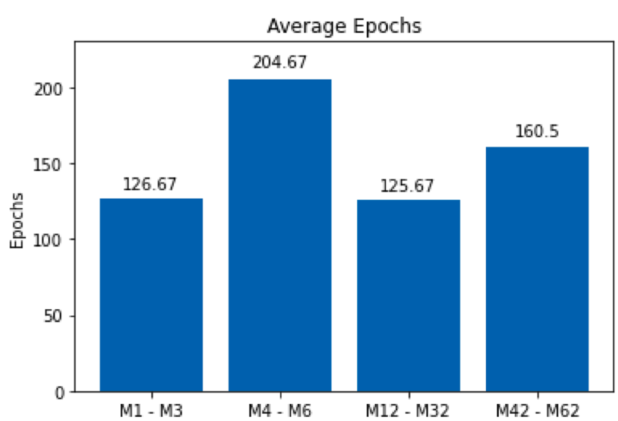

Fig. 5. Average number of epochs of different model groups

words, we randomly selected 330 non-bankrupt firms of the 6,599 available, obtaining a final sample with 550 observations. We also checked for missing values, but there are none.

In Section 2.1, we discussed some important aspects of variable selection techniques. Recapitulating variable selection helps reduce the sample's dimensionality by selecting the most representative variables from the data set. Therefore, we analyzed the correlations between the variables and looked for inconsistencies that permit the removal of irrelevant attributes. In Table 9 (see Appendix), we give a brief description of all the variables. The description includes the total number of observations, the mean, the standard deviation, the 25%, 50% and 75% quantiles, and each variable's minimum and maximum values. Having a closer look at the minimum-maximum values and considering that the values of almost all variables should range between 0 and 1, we identified several outliers among different attributes. We considered a value for a specific variable to be an outlier if the value is far greater than the normal range of the variable.

In general, to ensure the convergence of the activation functions, the values of the observations should always fit the range of the activation functions. For instance, all fully differentiable activation functions introduced in Section 3.4 quickly converge outside a relatively small interval centered around zero (Haykin (2009)). If the input values are not within that convergence interval then the activation function transmits mostly binary signals for many values of the data set. A great amount of information is likely to be lost, because the response of the neurons is to find far in the tails of the activation function, where the derivatives with respect to the input variables are approximately zero. This can have a negative impact on the estimation algorithms that are based on these derivatives, causing the calibration algorithm never to converge (Haykin (2009)).

After examining the outliers in more detail, we can see that 72 variables do not present any outliers while 24 variables have at least one outlier. The values of all outliers are also far greater than 100. Such high values are inconsistent with the nature of financial ratios as financial ratios can rarely be found far outside the range of 0 and 1. We also recognized that the outliers are dependent on the variable and not the observation. There is not a single observation of which all attributes are consistently high. There could be different reasons for such discrepancies in the data, for instance, it could result from measurement errors or false information. It is essential to emphasize the nature and similarities of the outliers along with the sample. Consequently, we removed all variables that have more than 1000 outliers, which results in a total of eight variables. The remaining outliers were substituted by the median of the corresponding variable. In Table 9, we also summarize the number of outliers for every independent variable.

Multicollinearity refers to high correlations between two or more variables and affects regression models. Although neural networks are robust against multicollinearity, it still can affect the convergence of the network. Using correlations to analyze the relations between variables can indicate redundancies and help to reduce the dimensionality of the data set further. As well known, correlation coefficients range from -1 to 1. We considered a correlation as high when the respective correlation coefficient is greater than 0.8 (or lower than -0.8). Figure 4 shows the correlation coefficients of variables with high correlations. Blue color indicates a high positive correlation, while red color indicates a high negative correlation. Following Figure 4, we can identify several groups with high correlation coefficients. Of every group of variables, we selected one and removed the rest. In the end, we were able to remove a total of 27 variables. Table 10 (see Appendix) lists all 61 remaining variables that are used for the development of the model. Finally, we split the data set into a training subset and a test subset. The training sample corresponds to 70% of the final data set and the test subset corresponds to the remaining 30%.

5. Results and comparative analysis

In this section, we firstly examine the performance of the proposed MLPs (Section 5.1). Subsequently, to further evaluate the MLPs developed in this paper, we compare our results with the results presented by Liang et al. (2016) who also used the same data set of Taiwanese firms. To this aim, we briefly give the results of Liang et al. (2016) (Section 5.2) and then we perform extensive comparative analysis (Section 5.3).

5.1. The performance of the proposed multilayer perceptrons

First, we examine the different models presented in Table 2 by implementing four different metrics (see Section 3.5) for the comparison. Table 3 presents the confusion table of all the 12 models M1–M62, and the respective values of TP, FP, FN, TN that are used to compute the different metrics. The results of the metrics for the different MLPs are given in Table 4.

The last column of Table 4 shows the number of epochs needed to

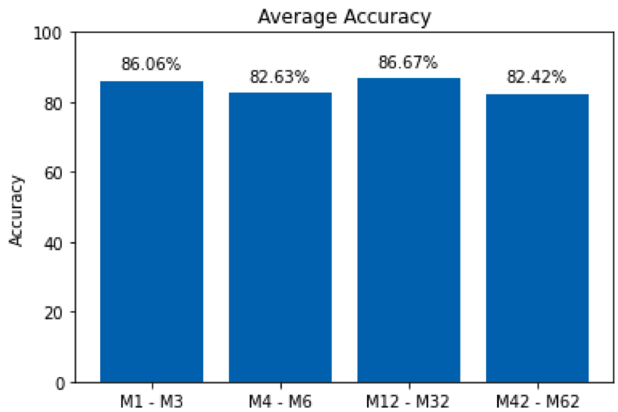

Fig. 6. Average accuracy of different model groups

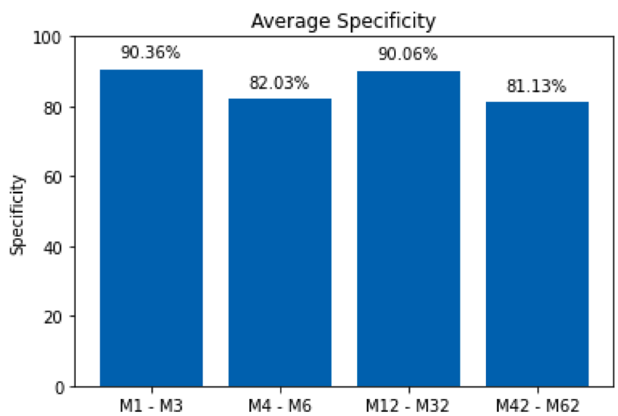

Fig. 7. Average specificity of different model groups

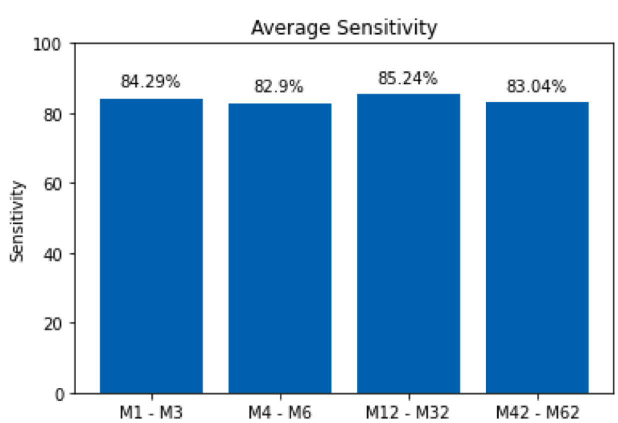

Fig. 8. Average sensitivity of different model groups

train. The number of epochs required to train a neural network may vary depending on the selection of the different hyperparameters. The `MLPclassifier` uses early stopping to avoid over-fitting and ensures the convergence of the MLP. By examining the numbers of epochs for each model, we recognize that models using sgd require more epochs to converge than models using adam. Only the adam model M22 required more epochs, with 196, than its counterpart model M52, with 141 epochs. This difference suggests that adam is a more effective optimization algorithm than sgd. Figure 5 shows the average number of epochs required to train the models divided into four groups. Groups M1–M3 and M12–32 are trained using adam, while groups M4–M6 and M42–M62 are trained using sgd. We see that the training of an MLP

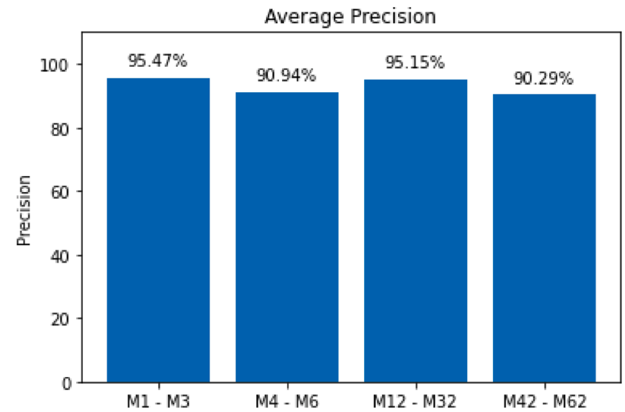

Fig. 9. Average precision of different model groups

using adam averages around 125 epochs despite the number of layers, neurons, and type of activation functions. For those models trained with sgd, the number of connections also affects the average number of epochs needed for the model to converge. Note that models trained with one layer demand more epochs than those with two layers.

The second column of Table 4 shows the average accuracy rate of all the models. We see that the model with the highest accuracy is model M12 with 87.27%, followed by model M32 with 86.67%. The models with the lowest accuracy are M5 and M52, both with 81.21%. Models M1, M2, M3, and M22 have the third-best average accuracy with 86.06%. In general, there is not a significant difference between the model with the best accuracy rate and the models with the second and third best accuracy rates. The difference is only one percentage point between those six models. However, when we analyze Figure 6, we see that the average accuracy of models using adam is slightly better than those using sgd. This result also aligns with the previous suggestion that adam is a more effective optimization algorithm than sgd.

The third column of Table 4 presents the specificity rates of the models. Note that a high specificity rate indicates a low number of Type I errors. The highest specificity rate is 91.49% and is achieved by models M2 and M22. Model M32 has the second-best specificity rate with 90%. Despite M12 being the most accurate model, the model's specificity is only the fourth-best with 88.68%. The high costs of classifying bankrupt firms as non-bankrupt suggest at first sight that model M2, M22, and M32 are better predictors than model M12. The lowest specificity rate is from model M52 with 79.25%. Figure 7 shows the average specificity of models using adam and models using sgd. Once more, the data suggest that models using adam have, on average, a better performance than their counterparts using sgd. The difference on average in specificity is nearly eight to nine percentage points.

Type II errors may not have severe consequences as Type I errors; however, the opportunity costs of misclassifying non-bankrupt firms as bankrupt should also be considered. The most sensitive model is M12 with 86.61% followed by M32 with 85.22%. The sensitivity of both models differs by only 1.39 percentage points. The model with the lowest sensitivity rate is model M5 with 81.58%. As for the average sensitivity, Figure 8 also suggests that adam performs better than sgd. However, the differences in sensitivity are not as distinct as in specificity. The average sensitivity of adam models is only between 1 and 2 percentage points higher than for sgd models.

Lastly, we calculated the precision rate of all models. The most precise models are model M2 and M22 with 96.12%. The model with the lowest precision rate is model M52 with 89.32%. Figure 9 illustrates the differences in average precision of adam and sgd models. Again, we see the same tendencies of adam models outperforming sgd models.

Table 5 summarizes the average performance of the different activation functions for all 12 models. We observe that the logistic function is the activation function that performed worst. As for the ReLU and

Table 5
Average rates of activation functions

Activation Function	Average Accuracy	Average Specificity	Average Sensitivity	Average Precision	Average no. of Epochs
ReLU	85.15	86.13	84.74	92.96	149.00
Logistic	83.63	85.66	82.88	92.96	187.50
Hyperbolic tangent	85.00	86.78	84.24	93.45	123.25

Table 6
The parameters of the five methods

Method	Parameters
BPNN (MLP)	# hidden layers: 1 # neurons: 64 Learning epochs: 50
SVM	Kernel: linear
KNN	$K = 7$
CART	Tree pruning: cross-validation
NB	Kernel: kernel density estimate

Hyperbolic tangent functions, their performance is nearly the same. The ReLU function outperforms the Hyperbolic tangent function regarding average accuracy, sensitivity, and precision. However, in terms of average specificity and average number of epochs, the Hyperbolic tangent function outperforms the ReLU function. Nonetheless, the difference between both functions is not that distinct to state which of both functions is more suitable for predicting bankruptcy. The performance of all three functions is satisfactory enough to suggest that the choice of the activation functions does not considerably affect the prediction capability of the MLP model.

Essentially, we see that different combinations of activation functions, optimizations algorithms, and architectures can make the MLP perform differently. In our case, we have four models (M2, M12, M22, and M32) that achieved similar results, and identifying which is the best model is not straightforward. The use of different metrics can give an insight into the strengths of each model. For instance, the models that best identified non-bankrupt firms as non-bankrupt are models M12 and M32 (sensitivity). In identifying bankrupt firms as bankrupt (specificity) models M2 and M22 outperform the rest. Further, model M12 has the highest average accuracy overall. Another important observation is that all Adam models outperform all sgd models.

To choose the best model, we use the specificity rate as the primary indicator and the average accuracy as the second one since both metrics have been widely used in the literature. In this sense, we choose model M2 and M22 as the best classifier because it has the best specificity rate and the third-best average accuracy. Models M2 and M22 are almost identical, achieving the same results in every evaluation metric we have determined. The only difference between M2 and M22 is the architecture and the epochs needed to train the MLP. Therefore, we finally chose model M22 over model M2 because it uses fewer weights and requires fewer epochs to be trained. The critical difference to model M32 is that model M22 specificity rate is 1.49% percentage points higher, and the average accuracy is only 0.61% lower. Compared to model M12, the M22 specificity rate is 2.81% higher, but the average accuracy is only 1.21% lower.

5.2. Results from the literature

In this subsection, we briefly review the results presented by Liang et al. (2016). Liang et al. (2016) had the objective to examine the discriminatory power of Corporate Governance Indicators (CGIs) in combination with Financial Ratios (FRs) in the context of bankruptcy prediction. They used five intelligent methods: BPNN, SVM, KNN, CART, and Naïve Bayes (NB) classifier. To reduce the dimensionality of the sample, they compared five different feature selection methods, of which three were filter-based and two wrapper-based. The filter-based methods were Stepwise Discriminant Analysis (SDA), Stepwise Logistic Regression (SLR) and *t*-testing, and two wrapper-based methods, Genetic Algorithm (GA) and Recursive Feature Elimination (RFE). The models were evaluated using four different measurements: average accuracy rate, Type I and Type II error rate, and the Receiver Operating Characteristic (ROC) curve. Several penalty thresholds, misclassification cost and cost ratios were introduced to compare the different ROC curves. Table 6 shows the parameters of each method.

Liang et al. (2016) used the Type I and Type II error rates as metrics to evaluate the prediction models:

$$\text{Type I error} = \frac{\text{FP}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}$$

$$\text{Type II error} = \frac{\text{FN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}$$

To overcome the imbalance of the data set, they used stratified sampling, a method that randomly partitions a sub-sample of the data set. The sub-sample is composed of 239 bankrupt firms and 239 non-bankrupt firms. They also normalized each variable and employed 10-fold cross-validation to divide the sub-sample into ten distinct training and validation subsets. The final prediction results were obtained by individually calculating the average of the ten validation results over the ten testing subsets.

Table 7 presents the results of the most relevant models for our case study. We are primarily focusing on the performance models using solely FRs, but we will take models that use FRs and CGIs also into consideration. The main intelligent methods to be analyzed are (1) SVM because it achieved the highest accuracy rates, and (2) MLP to compare their results with ours. Finally, we compare the performance of three feature selection methods.

Table 7
Part of the results by Liang et al. / (* M22 taken from Table 4).

Model	Method	Variables	Feature selection	Accuracy (%)	Type I error (%)	Type II error (%)
SVM1	SVM	FRs	None	79.1	20.2	21.6
SVM2	SVM	FRs + CGIs	None	81.3	17.8	19.7
SVM3	SVM	FRs + CGIs	SDA	81.5	16.3	20.8
KNN1	KNN	FRs	None	76.5	22.5	24.5
CART1	CART	FRs	None	78.4	23.3	19.9
MLP1	MLP	FRs	None	76.1	24.1	23.8
MLP2	MLP	FRs + CGIs	None	70.4	26.6	32.5
MLP3	MLP	FRs	*t*-test	74.2	22.0	29.5
MLP4	MLP	FRs	SLR	79.8	21.3	19.2
MLP5	MLP	FRs + CGIs	SLR	80.5	18.8	20.3
MLP5	MLP	FRs	SDA	77.7	22.4	22.2
NB1	NB	FRs	None	68.6	26.4	36.5
M22*	MLP	FRs	Multicollinearity	86.06	0.03	10.9

5.3. Comparative analysis

Table 7 shows that the best prediction model of Liang et al. (2016) is SVM3 with an average accuracy rate of 81.5% and 16.3% Type I error rate. The model combines FRs and CGIs and uses SDA to select the most relevant variables. We also notice that models with only FRs and no feature selection do not perform as well as those using one. Only the MLP3 model does not perform better than some models without feature selection, such as SVM1 and KNN1. This result highlights the importance of feature selection methods in classification problems. However, feature selection cannot always ensure better performance.

We also recognize that the average accuracy of the worst model in Table 4 is slightly better than the average accuracy of model SVM3 in Table 7. Identifying the reasons for such discrepancies is not easy due to the lack of transparency such intelligent methods have. Let us take a closer look at the calculation of the three metrics in Table 7. We recall that they used 10-fold cross-validation to create ten distinctive training and testing data sets, and the final prediction is based on the average of the ten testing results. Such an approach tends to reduce the variability of the samples, which, lastly, can affect the performance and minimizes the bias of such models.

The use of cross-validation can prove itself as a helpful tool; nevertheless, it is still necessary to account for the different parameters of each model. Liang et al. (2016) maintained all parameters of the five different models constant. Thus, Table 4 shows that variations of the parameters of an MLP can lead to different prediction results. It is also possible that different parameters obtain the same prediction results, such as M2 and M22. Therefore, when the parameters are held constant, different training samples can increase or reduce the classification capacity of the models. Table 8 illustrates the average accuracy rate and the training epochs when an MLP is trained with different training samples. Twenty training data sets and twenty test data sets were created to train the model M22.

Table 8
Differences in average accuracy and epochs of an MLP for different training sets

Epochs	75	196	119	107	107
Accuracy (%)	82.42	86.06	82.42	80.6	81.21
Epochs	40	62	113	68	101
Accuracy (%)	80	81.21	83.03	80.6	81.81
Epochs	122	143	81	63	107
Accuracy (%)	83.63	81.21	83.63	76.96	83.63
Epochs	122	152	101	145	119
Accuracy (%)	78.78	82.42	80	82.42	79.39

The average accuracy found in Table 8 is 81.69%, and the average number of epochs is 112. The average accuracy in Table 8 also outperforms the best model presented by Liang et al. (2016). Note that training the same MLP with different samples affects its performance. Defining the parameters of an MLP is a complicated task due to the high number of different possibilities. Table 4 shows that variations of the parameters of an MLP can lead to different prediction results. It is also possible that different parameters obtain the same prediction results, such as M2 and M22. Intelligent models, in general, depend on the data's characteristics and output preferences, and their parameters should fit such characteristics. It is not entirely clear if cross-validation alone can account for the different aspects that affect the performance of intelligent models.

Another critical difference between both papers is how each paper solved the outliers problem. Liang et al. (2016) used normalization. Normalization can be helpful in particular to replace outliers. However, not all variables needed to be normalized into the range from 0 to 1. Most variables were already in that range. Methods such as normalization or standardization are not always suited and do not necessarily lead to a better performance. Therefore, we decided to remove the variables with a high number of outliers and substitute the remaining outliers using the median.

Additionally, we employed multicollinearity to remove redundant attributes. Unfortunately, Liang et al. (2016) did not disclose the removed variables after using the different feature selection methods. Hence, conducting a comparison is not possible. Generally, the dimensionality of all models in Table 7 that do not use a feature selection method is high. They utilize, in some cases, over 100 different variables with a data set composed of only 478 firms. Curiously, feature selection has a relatively small effect on the performance of their models. For example, the average accuracy of model SVM3 is only 0.2 percentage points larger than for model SVM2. For the Type I error rate, the difference between both models is 1.5 percentage points. However, for the models in Table 4, the difference is larger, suggesting a positive effect related to our approach of selecting variables.

As mentioned before, two rules of thumb to properly train a neural network are that the number of observations in a sample should be ten times the number of weights, and the total number of neurons should be no more than 75% of input variables. The total number of weights of model M22 is 187. Therefore, the total number of observations should be near 1,870. In the case of model MLP1 developed by Liang et al. (2016), the total number of weights should reach 6,100. We see a high

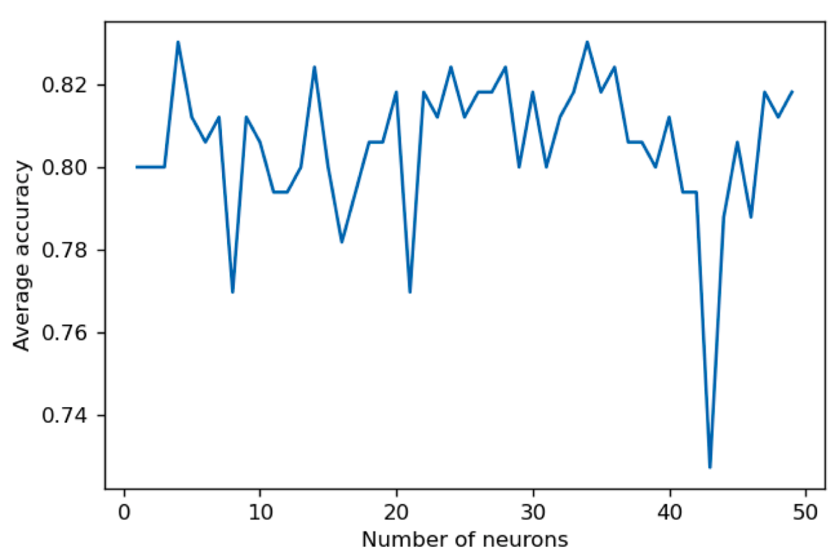

Fig. 10. The effect of increasing the number of neurons on the average accuracy of an MLP with one hidden layer

disproportion between the number of weights and the size of the data set. Model MLP2 uses both FRs and CGIs having even more weights than model MLP1. Except for model M4, the rest in Table 4 do not have such a high number of neurons. Certainly, for models MLP1 and MLP2, the number of neurons does not exceed 75% of the number of variables. Keeping the number of neurons constant and using feature selection methods can easily violate the second rule of thumb. The approach used in this paper to determine the optimal number of neurons for model in Table 4 responds better than maintaining the number of neurons constant.

Changes in the number of layers and neurons alone would generally lead to different prediction results, as shown in Figure 10. Figure 10 shows the effect of changing the number of neurons of an MLP. The highest average accuracy obtained was near 83%, while the lowest was under 74%. It is also remarkable how abrupt the average accuracy can change when one neuron is added to the MLP. As illustrated in Table 8, keeping the number of epochs constant does not assure the convergence of the neuron.

6. Conclusions and outlook

6.1. Conclusions

The objective of this paper was to examine the discriminatory power of an MLP in the context of bankruptcy prediction. The model was developed using a data set of Taiwanese firms composed of 95 financial ratios for the years 1999 to 2009. We compared different setups of four main parameters of MLPs: optimization algorithm, activation function, number of neurons, and number of layers. The goal was to find the parameter setup that achieve the best results in four evaluation metrics that we used: average accuracy, specificity, sensitivity, and precision.

Of all 12 different MLPs, four models made it to the final selection, for which we focused primarily on the specificity rate. Models M2 and M22 had the highest specificity rate. Moreover, both models obtained identical results in all four metrics. However, model M22 is slightly simpler than model M2 because it requires fewer weights. Therefore, model M22 was selected as the best predictor.

Further, the optimization algorithm Adam outperformed the classical sgd optimization. In the case study, we did not find enough evidence that would suggest that any activation performs better than the rest. The architecture of most neural networks is relatively simple and does not exceed 75% of the number of input variables. The data set suffered from a high imbalance between the number of bankrupt and non-bankrupt firms and had a significant number of outliers. Therefore, variables with a lot of noise (outliers) and variables with high correlations were removed. Finally, of the 6,599 non-bankrupt firms, a subsample of 330 firms was randomly selected to create a balanced data set of 550 firms.

To further validate our results, we compared them to the models presented by Liang et al. (2016). Liang et al. (2016) based their study on the same sample of Taiwanese firms and developed five different machine learning models. Additionally to the 95 financial ratios we employed, they also used CGIs. Their approach differed in several aspects from ours. To account for the variability of the sample, which can affect the model's performance, and to reduce the bias, they used 10-fold cross-validation to create ten different training samples and ten test samples. Then an average of all ten testing results was computed as the final outcome. They also used five different feature selection methods. The model with the highest performance in Liang et al. (2016) (SVM3 in Table 7) was outperformed by the worst MLP developed in this paper.

We examined differences between both methodologies to explain the resulted discrepancies. The fact that we did not have access to all the calculations conducted by Liang et al. (2016) does not allow us to confirm our hypothesis. First, we considered that the cross-validation had probably a negative effect on the outcome. Second, we showed the influence that hyperparameters have on the performance of an MLP. Adapting different hyperparameters can help to improve the performance of the intelligent method. However, Liang et al. (2016) maintained all parameters of the five models constant, which can also harm the final performance. Third, different data pre-processing approaches between Liang et al. (2016) and the approach in this paper may also affect the performance. Liang et al. (2016) models that did not use feature selection suffered from high dimensionality. The difference between those models with high dimensionality and those that used feature selection techniques is not significantly big. The impact of feature selection in Liang et al. (2016) is low. Finally, to account for the variability, we created 20 random training and testing data sets and used them with model M22. The average accuracy was 81.69%, still outperforming model SVM3.

The average accuracy of neural networks in the literature is approximately 83%. Therefore, our results align with those published in the literature of bankruptcy prediction models. MLP can be a powerful tool, yet determining the parameters can be challenging. Several aspects should be considered, such as variability of the sample, dimensionality of the data set, and proper pre-processing of the data to fit the needs of the MLP.

6.2. Outlook

This paper highlights the main issues that bankruptcy prediction models face. In general, machine learning techniques have surpassed the capacities of classical statistical models. Machine learning techniques are more flexible and can be quickly adapted to new problems. However, it would be a promising approach to create hybrid models to further improve the performance. A proper combination of a feature selection technique and a classification tool is better than a standalone classification tool (Iturriaga and Sanz (2015)). The objective is to complement the strengths and weaknesses of different methods and finally build a more powerful classifier. For example, decision rules-generating tools are more appropriate for feature selection tools and methods such as neural networks or SVMs for classification. Thus, the tendency to implement hybrid models should be followed further (Alaka et al. (2018)).

We have stressed the influence of the data set on the development of the model. Main factors are the quantity and the quality of the data and the nature and the complexity of the problem. Before committing to a modeling technique, the data set should be analyzed closely. The conditions of the data must match the conditions of the tool. Commonly the data sets are small, and there is a high imbalance between the number of bankrupt and non-bankrupt firms. A better approach could also be to separate data sets by sectors (Shi and Li (2019)). In the future, data recollection techniques should be improved to develop better models.

We have seen the tremendous amount of different variables that have been implemented to create bankruptcy prediction models. Mostly quantitative variables have been used. Still, researchers are continuously looking for new variables to improve the models. An alternative could be the embedding of qualitative variables into the models. However, problems of high dimensionality could arise and, in combination with small and poor data sets, could lead to bad results. Variables should be analyzed from different perspectives to ensure a positive contribution to the model. Newer variables that can predict bankruptcy two or even three years prior are also highly needed. Therefore, implementing new variables with similar abilities to existing ones is counterproductive.

One focus of future researches could also be the improvement of existing models (Bellovary et al. (2007)). As new technologies reveal themselves, it is expected that they will be implemented in new research. However, older methodologies have also achieved good results and can be improved further. There are many new insights we can gain from refining previous models. For instance, one main issue with ANN and SVM models is their lack of transparency. Recent research could make those models easier to interpret (Alaka et al. (2018)). On the other hand, new methodologies should not fall into the mistake of reproducing previous results.

CRediT authorship contribution statement

Raffael Förch Brenes: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Arne Johannssen:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – review & editing, Supervision, Project administration. **Nataliya Chukhrova:** Conceptualization, Formal analysis, Writing – review & editing, Visualization, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank both anonymous reviewers for their valuable feedback and suggestions, which were very important and helpful to significantly improve the paper.

Appendix

Table 9
Descriptive statistics on the variables of the data set

No.	Variable name	Count	Mean	Std	min	25%	50%	75%	max	No. Outliers
0	Bankrupt	6819	0.032	0.177	0	0	0	0	1	0
1	ROA(C) before interest and depreciation before interest	6819	0.505	0.061	0	0.477	0.503	0.536	1	0
2	ROA(A) before interest and % after tax	6819	0.559	0.066	0	0.536	0.560	0.589	1	0
3	ROA(B) before interest and depreciation after tax	6819	0.554	0.062	0	0.527	0.552	0.584	1	0
4	Operating Gross Margin	6819	0.608	0.017	0	0.600	0.606	0.614	1	0
5	Realized Sales Gross Margin	6819	0.608	0.017	0	0.600	0.606	0.614	1	0
6	Operating Profit Rate	6819	0.999	0.013	0	0.999	0.999	0.999	1	0
7	Pre-tax net Interest Rate	6819	0.797	0.013	0	0.797	0.797	0.798	1	0
8	After-tax net Interest Rate	6819	0.809	0.014	0	0.809	0.809	0.809	1	0
9	Non-industry income and expenditure/revenue	6819	0.304	0.011	0	0.303	0.304	0.304	1	0
10	Continuous interest rate (after tax)	6819	0.781	0.013	0	0.782	0.782	0.782	1	0
11	Operating Expense Rate	6819	2.00E+09	3.24E+09	0	1.57E-04	2.78E-04	4.15E+09	9.99E+09	2282
12	Research and development expense rate	6819	1.95E+09	2.60E+09	0	1.28E-04	5.09E+08	3.45E+09	9.98E+09	4033
13	Cash flow rate	6819	0.467	0.017	0	0.462	0.465	0.471	1	0
14	Interest-bearing debt interest rate	6819	1.64E+07	1.08E+08	0	0.000	3.21E-04	0.001	9.90E+08	221
15	Tax rate (A)	6819	0.115	0.139	0	0.000	0.073	0.206	1	0
16	Net Value Per Share (B)	6819	0.191	0.033	0	0.174	0.184	0.200	1	0
17	Net Value Per Share (A)	6819	0.191	0.033	0	0.174	0.184	0.200	1	0
18	Net Value Per Share (C)	6819	0.191	0.033	0	0.174	0.184	0.200	1	0
19	Persistent EPS in the Last Four Seasons	6819	0.229	0.033	0	0.215	0.225	0.239	1	0
20	Cash Flow Per Share	6819	0.323	0.018	0	0.318	0.322	0.329	1	0
21	Operating Profit Growth Rate	6819	0.848	0.011	0	0.848	0.848	0.848	1	0
22	After-tax Net Profit Growth Rate	6819	0.689	0.014	0	0.689	0.689	0.690	1	0
23	Regular Net Profit Growth Rate	6819	0.689	0.014	0	0.689	0.689	0.690	1	0
24	Continuous Net Profit Growth Rate	6819	0.218	0.010	0	0.218	0.218	0.218	1	0
25	Total Asset Growth Rate	6819	5.51E+09	2.90E+09	0	4.86E+09	6.40E+09	7.39E+09	9.99E+09	6017
26	Net Value Growth Rate	6819	1.57E+06	1.14E+08	0	0.000	0.000	0.000	9.33E+09	2
27	Total Asset Return Growth Rate Ratio	6819	0.264	0.010	0	0.264	0.264	0.264	1	0
28	Cash Reinvestment %	6819	0.380	0.021	0	0.375	0.380	0.387	1	0
29	Current Ratio	6819	4.03E+05	3.33E+07	0	0.008	0.011	0.016	2.75E+09	1
30	Quick Ratio	6819	8.38E+06	2.45E+08	0	0.005	0.007	0.012	9.23E+09	9
31	Interest Expense Ratio	6819	0.631	0.011	0	0.631	0.631	0.631	1	0
32	Total debt/Total net worth	6819	4.42E+06	1.68E+08	0	0.003	0.006	0.009	9.94E+09	8
33	Debt ratio %	6819	0.113	0.054	0	0.073	0.111	0.149	1	0
34	Net worth/Assets	6819	0.887	0.054	0	0.851	0.889	0.927	1	0
35	Long-term fund suitability ratio (A)	6819	0.009	0.028	0	0.005	0.006	0.007	1	0
36	Borrowing dependency	6819	0.375	0.016	0	0.370	0.373	0.376	1	0
37	Contingent liabilities/Net worth	6819	0.006	0.012	0	0.005	0.005	0.006	1	0
38	Operating profit/Paid-in capital	6819	0.109	0.028	0	0.096	0.104	0.116	1	0
39	Net profit before tax/Paid-in capital	6819	0.183	0.031	0	0.169	0.178	0.192	1	0
40	Inventory and accounts receivable/Net value	6819	0.402	0.013	0	0.397	0.400	0.405	1	0
41	Total Asset Turnover	6819	0.142	0.101	0	0.076	0.118	0.177	1	0
42	Accounts Receivable Turnover	6819	1.28E+07	2.78E+08	0	0.001	0.001	0.001	9.74E+09	22
43	Average Collection Days	6819	9.83E+06	2.56E+08	0	0.004	0.007	0.009	9.73E+09	18
44	Inventory Turnover Rate (times)	6819	2.15E+09	3.25E+09	0	1.73E-04	0.001	4.62E+09	9.99E+09	2946
45	Fixed Assets Turnover Frequency	6819	1.01E+09	2.48E+09	0	2.33E-04	0.001	0.004	9.99E+09	1228
46	Revenue Per Share (Yuan)	6819	1.33E+06	5.17E+07	0	0.016	0.027	0.046	3.02E+09	5
47	Operating Profit Per Share (Yuan)	6819	0.109	0.028	0	0.096	0.104	0.116	1	0
48	Per Share Net profit before tax (Yuan)	6819	0.184	0.033	0	0.170	0.180	0.193	1	0
49	Realized Sales Gross Profit Growth Rate	6819	0.022	0.012	0	0.022	0.022	0.022	1	0
50	Net Worth Turnover Rate (times)	6819	0.039	0.037	0	0.022	0.030	0.043	1	0
51	Revenue per person	6819	2.33E+06	1.37E+08	0	0.010	0.019	0.036	8.81E+09	2

(*continued on next page*)

Table 9 (*continued*)

No.	Variable name	Count	Mean	Std	min	25%	50%	75%	max	No. Outliers
52	Operating profit per person	6819	0.401	0.033	0	0.392	0.396	0.402	1	0
53	Allocation rate per person	6819	1.13E+07	2.95E+08	0	0.004	0.008	0.015	9.57E+09	12
54	Working Capital to Total Assets	6819	0.814	0.059	0	0.774	0.810	0.850	1	0
55	Quick Assets/Total Assets	6819	0.400	0.202	0	0.242	0.386	0.541	1	0
56	Current Assets/Total Assets	6819	0.522	0.218	0	0.353	0.515	0.689	1	0
57	Cash/Total Assets	6819	0.124	0.139	0	0.034	0.075	0.161	1	0
58	Quick Assets/Current Liability	6819	3.59E+06	1.72E+08	0	0.005	0.008	0.013	8.82E+09	3
59	Cash/Current Liability	6819	3.72E+07	5.10E+08	0	0.002	0.005	0.013	9.65E+09	46
60	Current Liability to Assets	6819	0.091	0.050	0	0.053	0.083	0.120	1	0
61	Operating Funds to Liability	6819	0.354	0.035	0	0.341	0.349	0.361	1	0
62	Inventory/Working Capital	6819	0.277	0.010	0	0.277	0.277	0.277	1	0
63	Inventory/Current Liability	6819	5.58E+07	5.82E+08	0	0.003	0.006	0.011	9.91E+09	99
64	Current Liabilities/Liability	6819	0.762	0.207	0	0.627	0.807	0.942	1	0
65	Working Capital/Equity	6819	0.736	0.012	0	0.734	0.736	0.739	1	0
66	Current Liabilities/Equity	6819	0.331	0.013	0	0.328	0.330	0.332	1	0
67	Long-term Liability to Current Assets	6819	5.42E+07	5.70E+08	0	0.000	0.002	0.009	9.54E+09	109
68	Retained Earnings to Total Assets	6819	0.935	0.026	0	0.931	0.938	0.945	1	0
69	Total income/Total expense	6819	0.003	0.012	0	0.002	0.002	0.002	1	0
70	Total expense/Assets	6819	0.029	0.027	0	0.015	0.023	0.036	1	0
71	Current Asset Turnover Rate	6819	1.20E+09	2.82E+09	0	1.46E-04	1.99E-04	4.53E-04	1.00E+10	1234
72	Quick Asset Turnover Rate	6819	2.16E+09	3.37E+09	0	1.42E-04	2.25E-04	4.90E+09	1.00E+10	2383
73	Working capital Turnover Rate	6819	0.594	0.009	0	0.594	0.594	0.594	1	0
74	Cash Turnover Rate	6819	2.47E+09	2.94E+09	0.00E+00	2.74E-04	1.08E+09	4.51E+09	1.00E+10	4239
75	Cash Flow to Sales	6819	0.672	0.009	0	0.672	0.672	0.672	1	0
76	Fixed Assets to Assets	6819	1.22E+06	1.01E+08	0	0.085	0.197	0.372	8.32E+09	1
77	Current Liability to Liability	6819	0.762	0.207	0	0.627	0.807	0.942	1	0
78	Current Liability to Equity	6819	0.331	0.013	0	0.328	0.330	0.332	1	0
79	Equity to Long-term Liability	6819	0.116	0.020	0	0.111	0.112	0.117	1	0
80	Cash Flow to Total Assets	6819	0.650	0.047	0	0.633	0.645	0.663	1	0
81	Cash Flow to Liability	6819	0.462	0.030	0	0.457	0.460	0.464	1	0
82	CFO to Assets	6819	0.593	0.059	0	0.566	0.593	0.625	1	0
83	Cash Flow to Equity	6819	0.316	0.013	0	0.313	0.315	0.318	1	0
84	Current Liability to Current Assets	6819	0.032	0.031	0	0.018	0.028	0.038	1	0
85	Liability-Assets Flag	6819	0.001	0.034	0	0.000	0	0	1	0
86	Net Income to Total Assets	6819	0.808	0.040	0	0.797	0.811	0.826	1	0
87	Total assets to GNP price	6819	1.86E+07	3.76E+08	0	0.001	0.002	0.005	9.82E+09	20
88	No-credit Interval	6819	0.624	0.012	0	0.624	0.624	0.624	1	0
89	Gross Profit to Sales	6819	0.608	0.017	0	0.600	0.606	0.614	1	0
90	Net Income to Stockholder's Equity	6819	0.840	0.015	0	0.840	0.841	0.842	1	0
91	Liability to Equity	6819	0.280	0.014	0	0.277	0.279	0.281	1	0
92	Degree of Financial Leverage (DFL)	6819	0.028	0.016	0	0.027	0.027	0.027	1	0
93	Interest Coverage Ratio (Interest expense to EBIT)	6819	0.565	0.013	0	0.565	0.565	0.566	1	0
94	Net Income Flag	6819	1.000	0.000	1	1.000	1.000	1.000	1	0
95	Equity to Liability	6819	0.048	0.050	0	0.024	0.034	0.053	1	0

Table 10
Final set of variables.

No.	Variable
0	Bankrupt
1	Operating Gross Margin
2	Operating Profit Rate
3	Non-industry income and expenditure/revenue
4	Cash flow rate
5	Interest-bearing debt interest rate
6	Tax rate (A)
7	Net Value Per Share (A)
8	Persistent EPS in the Last Four Seasons
9	Cash Flow Per Share
10	Revenue Per Share (Yuan)
11	Realized Sales Gross Profit Growth Rate
12	After-tax Net Profit Growth Rate
13	Continuous Net Profit Growth Rate
14	Net Value Growth Rate
15	Total Asset Return Growth Rate Ratio
16	Cash Reinvestment %
17	Current Ratio
18	Interest Expense Ratio
19	Total debt/Total net worth
20	Debt ratio %
21	Long-term fund suitability ratio (A)
22	Borrowing dependency
23	Contingent liabilities/Net worth
24	Net profit before tax/Paid-in capital
25	Inventory and accounts receivable/Net value
26	Total Asset Turnover
27	Accounts Receivable Turnover
28	Average Collection Days
29	Net Worth Turnover Rate (times)
30	Revenue per person
31	Operating profit per person
32	Allocation rate per person
33	Working Capital to Total Assets
34	Quick Assets/Total Assets
35	Current Assets/Total Assets
36	Cash/Total Assets
37	Cash/Current Liability
38	Inventory/Working Capital
39	Inventory/Current Liability
40	Current Liabilities/Liability
41	Working Capital/Equity
42	Long-term Liability to Current Assets
43	Retained Earnings to Total Assets
44	Total income/Total expense
45	Total expense/Assets
46	Cash Flow to Sales

(*continued on next page*)

Table 10 (*continued*)

No.	Variable
47	Fixed Assets to Assets
48	Cash Flow to Total Assets
49	Cash Flow to Liability
50	CFO to Assets
51	Cash Flow to Equity
52	Current Liability to Current Assets
53	Liability-Assets Flag
54	Net Income to Total Assets
55	Total assets to GNP price
56	No-credit Interval
57	Degree of Financial Leverage (DFL)
58	Interest Coverage Ratio (Interest expense to EBIT)
59	Net Income Flag
60	Equity to Liability

References

Adnan Aziz, M., & Dar, H. A. (2006). Predicting corporate bankruptcy: where we stand? *Corporate Governance, 6*(1), 18–33.

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayie, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications, 94*, 164–184.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589–609.

Baek, J., & Cho, S. (2003). *Bankruptcy Prediction for credit risk using an auto associative neural networks in Korean firms*. IEEE International Conference on Computational Intelligence for Financial Engineering, Hong-Kong.

Bardos, M. (1995). Détection précoce des défaillances d'entreprises á partir des documents comptables. *Bulletin de la banque de france* (pp. 57–71). Supplément Études, 3ème trimestre.

Barniv, R., Agarwal, A., Laitinen, E. K., & Leach, R. (1997). Predicting the outcome following bankruptcy filing: a three-state classification using neural networks. *Intelligent Systems in Accounting, Finance and Management, 6*(3), 177–194.

Beaver, W. (1996). Financial ratios as predictors of failure. *Journal of Accounting Research, 5*, 71–111.

Bell, T. B. (1997). Neural nets or the logit model? a comparison of each model's ability to predict commercial bank failures. *International Journal of Intelligent Systems in Accounting, Finance and Management, 6*, 249–264.

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education, 33*, 1–42.

Bradley, D. B., & Cowdery, C. (2004). *Small Business: Causes of Bankruptcy. Small Business Advancement National Center*. University of Central Arkansas, College of Business Administration, Research Paper.

Canbas, S., Cabuk, A., & Kilic, S. B. (2005). Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The turkish case. *European Journal of Operational Research, 166*(2), 528–546.

Chuang, C. L. (2013). Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction. *Information Sciences, 236*, 174–185.

Clement, C. (2006). Machine learning in bankruptcy prediction – a review. *Journal of Public Administration, Finance and Law, 17*, 178–196.

Coakley, J. R., & Brown, C. E. (2000). Artificial neural networks in accounting and finance: modeling issues. *Intelligent Systems in Accounting, Finance and Management, 9*, 119–144.

Denuit, M., et al. (2019). *Effective Statistical Learning Methods for Actuaries III*. Springer.

Dietrich, J. R., & Kaplan, R. S. (1982). Empirical analysis of the loan classification decision. *The Accounting Review, 57*, 18–38.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics, 35*(5–6), 352–359.

Du Jardin, P. (2012). The influence of variable selection methods on the accuracy of bankruptcy prediction models. *Bankers, Markets & Investors, 116*, 20–39.

Du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research, 242*(1), 286–303.

Esen, H., Esen, M., & Ozsolak, O. (2017). Modelling and experimental performance analysis of solar-assisted ground source heat pump system. *Journal of Experimental and Theoretical Artificial Intelligence, 29*(1), 1–17.

Esen, H., Inalli, M., Sengur, A., & Esen, M. (2008). Performance prediction of a ground-coupled heat pump system using artificial neural networks. *Expert Systems with Applications, 35*(4), 1940–1948.

Esen, H., Ozgen, F., Esen, M., & Sengur, A. (2009a). Artificial neural network and wavelet neural network approaches for modelling of a solar air heater. *Expert Systems with Applications, 36*(8), 11240–11248.

Esen, H., Ozgen, F., Esen, M., & Sengur, A. (2009b). Modelling of a new solar air heater through least-squares support vector machines. *Expert Systems with Applications, 36*(7), 10673–10682.

Gupta, J., Barzotto, M., & Khorasgani, A. (2018). Does size matter in predicting SMEs failure? *International Journal of Finance & Economics, 23*(4), 571–605.

Hashi, I. (1997). The economics of bankruptcy, reorganization, and liquidation: Lessons for east european transition economies. *Russian and East European Finance and Trade, 33*(4), 6–34.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. New York: Mc Millan.

Haykin, S. S. (2009). *Neural Networks and Learning Machines*. Pearson, 3rd edition.

Horak, J., Vrbka, J., & Suler, P. (2020). Support vector machine methods and artificial neural networks used for the development of bankruptcy prediction models and their comparison. *Journal of Risk and Financial Management, 60*(13).

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

Iturriaga, F. J. L., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of US commercial banks. *Expert Systems with Applications, 42*(6), 2857–2869.

John, G., Kohavi, R., & Pfleger, K. (1998). *Irrelevant Features and the Subset Selection Problem* (pp. 121–129). New Jersey: Proceedings of the 11th International Conference, New Brunswick.

Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: A mixed logit model. *The Accounting Review, 79*(4), 1011–1038.

Karels, G. V., & Prakash, A. J. (1987). Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance and Accounting, 14*(4), 573–593.

Kim, S. Y. (2011). Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis. *The Service Industries Journal, 31*, 441–468.

Kingma, D. P., & Ba, J. (2015). *Adam: A method for stochastic optimization. 3rd International Conference for Learning Representations*. San Diego.

Kirkos, E. (2015). Assessing methodologies for intelligent bankruptcy prediction. *Artificial Intelligence Review, 43*, 83–123.

Kristóf, T., & Virág, M. (2012). Data reduction and univariate splitting'do they together provide better corporate bankruptcy prediction? *Acta Oeconomica, 62*(2), 205–228.

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.

Lee, K. C., Han, I., & Kwon, Y. (1996). Hybrid neural network models for bankruptcy predictions. *Decision Support Systems, 18*(1), 63–72.

Liang, D., Lu, C.-C., Tsai, C.-F., & Shih, G. A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research, 252*(2), 561–572.

Liang, D., Tsai, C.-F., & Wu, H. T. (2015). The effect of feature selection on financial distress prediction. *Knowledge-Based Systems, 73*(1), 289–297.

Lim Xiu Yun, T., Lim, J., Siwei, G., & Jiang, H. (2015). Bankruptcy prediction: Theoretical framework proposal. *International Journal of Management Sciences and Business Research, 1*(9), 571–605.

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

Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking and Finance, 1*(3), 249–276.

Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications, 28*, 603–614.

Ohlson, J. A. (1980). Financial rations and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109–131.

Omatu, S., Neves, J., Rodriguez, J. M. C., Santana, J. F. P., & Gonzalez, S. R. (2021). *Distributed Computing and Artificial Intelligence*. Special Sessions, 17th International Conference. Springer International Publishing.

Onakoya, A. B., & Olotu, A. E. (2017). Bankruptcy and insolvency: An exploration of relevant theories. *International Journal of Economics and Financial Issues, 7*, 706–712.

Priddy, K., & Keller, P. E. (2005). *Artificial Neural Networks: An Introduction. **68***. SPIE.

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.

Raghupathi, W., Schkade, L. L., & Raju, B. S. (1991). A neural network application for bankruptcy prediction. *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences, 4*, 147–155.

Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). Neural networks: A tool for predicting thrift failures. *Decision Sciences, 23*, 899–916.

Salmi, T., & Martikainen, T. (1994). A review of the theoretical and empirical basis of financial ratio analysis. *Finnish Journal of Business Economics, 4*(94), 426–448.

Sharda, R., & Wilson, R. L. (1993). *Performance comparison issues in neural network experiments for classification problems*. Proceedings of the 26th Hawai International Conference on System Scientists.

Shi, Y., & Li, X. (2019). An overview of bankruptcy prediction models for corporate firms: A systematic literature review. *Intangible Capital, 15*(2), 114–127.

Da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., & Alves, S. F.d. R. (2017). *Artificial Neural Networks*. Springer.

Swicegood, P., & Clark, J. A. (2001). Off-site monitoring for predicting bank under performance: A comparison of neural networks, discriminant analysis and professional human judgment. *Intelligent Systems in Accounting, Finance and Management, 10*(3), 169–186.

Tsai, C.-F., & Cheng, K. C. (2012). Simple instance selection for bankruptcy prediction. *Knowledge-Based Systems, 27*, 333–342.

Tseng, F. M., & Hu, Y. C. (2010). Comparing four bankruptcy prediction models: logit, quadratic interval logit, neural and fuzzy neural networks. *Expert Systems with Applications, 37*(3), 1846–1853.

Virág, M., & Nyitrai, T. (2014). Is there a trade-off between the predictive power and the interpretability of bankruptcy models? the case of the first hungarian bankruptcy prediction model. *Acta Oeconomica, 64*(4), 419–440.

Wilcox, J. W. (1973). Prediction of business failure using accounting data, empirical research in accounting: Selected studies. *Journal of Accounting Research, 11*, 163–179.
Yang, Z., You, W., & Ji, G. (2011). Using partial least squares and support vector machines for bankruptcy prediction. *Expert Systems with Applications, 38*(7), 8336–8342.
Yang, Z. R., Platt, M. B., & Platt, H. D. (1999). Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research, 44*(2), 67–74.
Yeganeh, A., & Shadman, A. (2021). Monitoring linear profiles using artificial neural networks with run rules. *Expert Systems with Applications, 168*, 114237.
Yeganeh, A., Shadman, A., & Abbasi, S. A. (2022a). Enhancing the detection ability of control charts in profile monitoring by adding RBF ensemble model. *Neural Computing and Applications, 34*(12), 9733–9757.
Yeganeh, A., Shadman, A., Abbasi, S. A., Pourpanah, F., Johannssen, A., & Chukhrova, N. (2022b). An ensemble neural network framework for improving the detection ability of a base control chart in non-parametric profile monitoring. *Expert Systems with Applications, 204*, 117572.
Yeh, C.-C., Chi, D.-J., & Hsu, M. F. (2010). A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications, 37*(2), 1535–1541.