Financial Distress Prediction based on Multi-Layer Perceptron with Parameter Optimization

Magdi El Bannany, Ahmed M. Khedr, Meenu Sreedharan and Sakeena Kanakkayil

Abstract—**Financial Distress Prediction has been a critical concern in the field of finance that sparked a slew of academic interests in the area. In our study, we examine the performance of various data mining models for predicting financial distress in companies in the Middle East and North Africa area, followed by model optimization. The main goal of the study is to find the most reliable deep neural network model for financial distress prediction, with optimized parameters. The study is divided into three phases. The output of various single machine learning classifiers and ensemble techniques for predicting financial distress is compared in the first phase. The best classifier found in the first step, the neural network, is then given different number of hidden layers. Furthermore, to achieve better prediction performance than the second stage, the Multi-Layer Perceptron model is optimised by tuning the hyperparameters such as network depth and network width. The prediction performance of the models is evaluated using real-time data sets containing samples of companies from the MENA region. The technique of re-sampling is used, for all the models, in order to get accurate and unbiased results.**

Index Terms—**Financial Distress Prediction (FDP), Middle East and North Africa (MENA), Machine Learning (ML), Multi-Layer Perceptron (MLP)**

I. Introduction

A financial distress period is a crucial period that occurs prior to bankruptcy when a business faces cash flow problems or credit balance deterioration. Financial distress has a severe impact on businesses, creditors, and, as a result, a country's economy [1]. Decision makers can avert distress by implementing an effective financial risk early warning system (EWS). An accurate prediction of a company's financial situation will assist managers in initiating alternate solution to prevent distress prior to its happening. It also assists investors in adjusting their investment strategies and lowering the expected loss on investment. As a result, FDP models are vital in risk management [2], [3].

For over more than four decades, and since the global financial crisis, many studies have endeavored to help address the financial crisis by creating better methods based on traditional statistical techniques and machine learning methods [4], [5], [6], [7], [8], [9]. Traditional statistical techniques such as Logit regression, Probit regression, univariate analysis, Multiple Discriminant Analysis (MDA), developed mostly in late 20th century, although easy to apply and understand, unfortunately, comes with some limitations themselves [10].Several studies employed statistical approaches to solve the problem; but, machine learning techniques were discovered to produce a more accurate solution. According to the majority of scholars, data mining algorithms can forecast economic distress well than other approaches [11], [12]. Moreover, aggregating the models (ensemble techniques), is seen as a better option, to make advancements in decision-making algorithms [13], [14]. The datasets can be distributed across many networks on a number of sites [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26]. A lot of studies on distress prediction with traditional ML classifiers like Decision Tree (DT) [27], Neural Networks (NN) [28], and Support Vector Machines (SVM) [29], exist in the literature; they are, however, restricted to one or two options.

Ever since, most research have focused on the United States and other advanced economies, with hardly any interest for emerging economies [30]. This work fills the literature void by focusing on FDP in the MENA region, where financially distressed businesses pose a major threat to the economy and growth of the country.

Our research seeks to develop a reliable data mining classifier to assist management and stockholders in managing risks and taking timely steps to avert insolvency before it occurs. The study's findings show that the suggested model has significantly greater prediction accuracy than traditional Machine learning algorithms.

The purpose of the research is to recognize and optimise the best FDP classifier. The core phases of the study are :

- Compare: We compare six major data mining techniques including single classifiers - NN, SVM, DT, and ensemble techniques - Majority Voting (MV), Random Forest (RF) and AdaBoost Ensemble. The classifiers are analyzed based on accuracy and F1-score using cross-validation (Phase 1).
- Optimize: In this phase, we improve the performance of the NN model, identified as the best classifier in phase 1. MLP with varying hidden layers is used for simulation and assessment (Phase 2).
- Fine - tune: Finally, we focus on fine-tuning the hyperparameters of the MLP model from the second phase. There are numerous hyperparameters which can be tuned to increase the MLP's classification accuracy. But the performance of the network is not significantly affected by all of them. IIn this phase, we concentrate on the two most important hyperparameters, network width and network depth. A proper selection of these

Magdi El Bannany is a professor in Department of Accounting, College of Business, University of Sharjah, Sharjah 27272, UAE and Department of Accounting and Auditing, Faculty of Business, Ain Shams University, Cairo, Egypt Email: melbannany@sharjah.ac.ae.

Ahmed M. Khedr is a professor in Department of Computer Science, College of Computing & Informatics, University of Sharjah, Sharjah 27272, UAE, Email: akhedr@sharjah.ac.ae.

Meenu Sreedharan is a Research Assistant in Department of Computer Science, University of Sharjah, Sharjah 27272, UAE, Email: meenus088@gmail.com.

Sakeena Kanakkayil is a Research Assistant in Department of Computer Science, University of Sharjah, Sharjah 27272, UAE, Email: sakeena.k@gmail.com.

parameters highly affects the predictive performance of an MLP model (Phase 3).

The remainder of our paper is structured as given: Section II is dedicated to an overview of the related researches in the literature. Data and methodology used for the study is discussed in Section III. Section IV is reserved for the empirical study and outcomes, and the final section, Section V contains the conclusion and future research.

II. Related Research

FDP models attempt to predict whether a business will face financial difficulties in the future. The primary goal of an FDP model is to determine whether an organisation is likely to experience financial distress. In prior researches, statistical and machine learning methods are used to predict financial distress.

Accounting-based models in [31], [32], [33] are examining the benefits of information included in the financial statements to provide reasonable evaluation of company's financial distress risk. These models are based on one ratio, or a group of financial ratios (i.e. Liquidity, Leverage, and Profitability) which are computed and compared to a benchmark, in order to determine if a company is financially distressed or not. These accounting models are classified as binary or dichotomous models because it measures the company's financial distress using a dichotomous variable which classifies a firm as non financially or financially distressed according to a specific cut-off. Financial data used in these models are historical. The simple structure and access to financial information have made these models the most popular tools for predicting a company's distress for decades. There is no agreement on feature selection in these models [34]. Several studies in the financial distress literature have adopted accounting-based models to predict the potential financial distress of companies [35], [36], [37], [38], [39], [40].

Discriminant analysis and the logit model were the first and most widely utilised statistical methods in the study of distress predictions [4], [5], [6], [8]. Altman [31] tried to predict corporate financial distress using discriminant analysis while Ohlson [32] used a logit model for predicting financial distress. Later, in order to improve the discriminatory power, a modified discriminant analysis model was proposed by Falbo [5] to improve discriminatory power through stability financial ratios in multiple years. A comparison between different traditional models including Z-Score, O-Score, Hazard, Probit and D-Score on FDP was studied by [6].

Financial data is often regarded as non-linear, and therefore statistical methods for generating a robust prediction model are less accurate. Because of their advantage in extracting the non-linear relationship between data without prior knowledge of the input, ML algorithms are often used for FDP. With advancements in data processing technologies, many researchers have used ML techniques for prediction and classification. As a result, later studies introduced ML using data mining techniques such as Logistic Regression, SVM, and NNs as alternatives [9], [41], [42], [32].

Ohlson [32] was the one to use an ML technique for FDP at the first place - the logistic regression model. The most commonly used data mining approach in the field of distress prediction after logistic regression is NN [41], [42]. Numerous studies have been carried out utilising probabilistic and feed-forward backpropagation neural networks, as cited in [43]. SVM and DT classification methods are two more efficient data mining techniques that have been used to forecast financial crises, according to [9]. [27] describes a comparison of distress predictions using a DT model and a Cox survival analysis model. Many researchers have proposed ensemble methods over base learners to improve the efficiency of their models [44]. The financial distress of firms in Iran was studied using four data mining techniques: artificial neural network, SVM, k-nearest neighbor and naive Bayesian classifier in [45]. Ruibin Geng, Indranil Bose, and Xi Chen assessed the efficacy of ML methods for predicting distress in publicly traded Chinese firms [46]. They analysed the performance of three frequently used data mining classifiers by merging the outputs with the MV ensemble approach. There are additional studies on datasets collected from other nations that use data mining methods[47]. Additionally, Aktham and Basel [48] created a basic hazard model for the FDP of banks in Gulf Cooperation Council nations.

Peat and Jones [49] studied how employing an NN algorithm to incorporate information from company financial statements and stock markets may enhance prediction of business bankruptcy. Utilizing performance metrics relying on the roc curve, the prediction results from the NN outperformed the forecasts produced from a logistic regression method.

Telmoudi et al. [50] suggested a novel hybrid method for dealing with business failure detection. They used financial ratios as inputs and combined rough set theory, Gaussian case based reasoning clustering, real valued genetic algorithm, and support vector machines to forecast whether or not the business entity will collapse. This model is supported by a high degree of accuracy. Anandarajan et al. [51] constructed and evaluated a genetic algorithm NN model, comparing its prediction performance to that of a backpropagation neural network and a model employing multiple discriminant analysis. The findings showed that the genetic algorithm based NN had the lowest misclassification cost among the models.

O'Leary [52] presented a'meta-analysis' on NN to forecast company failure. He examined and analysed fifteen articles to determine "what works and what does not". The formulations of the studies are compared, including the impact of using multiple variations of bankrupt companies, the software used, the input attributes, the type of hidden layer and the number of nodes in them, the output attributes, train and test phases, and empirical evaluation. The outcomes are then examined across a variety of parameters, including the proximity of comparable results, the number of accurate classifications, the influence of hidden layers, and the percentage of bankrupt companies.

Boritz et al. [53] evaluated the effectiveness of ANN in classifying and predicting commercial entities into successful and unsuccessful classes. Backpropagation and Optimal Estimation Theory (OET) are two approaches used to train neural networks to forecast bankruptcy filings. The information is derived from Compustat data tapes covering a diverse range of businesses. The neural network findings are compared with other well known distress prediction approaches such as discriminant analysis, probit, and logit. Even though there are remarkable 'pockets' of higher performance by neural networks, based on the particular combos of percentages of bankrupt samples in the train and test data sets and presumptions about the relative costs of Type I and Type II errors, the neural network methods do not attain the results that the literature in this area frequently claims.

Fanning and Cogger [54] compared the effectiveness of a generalised adaptive neural network algorithm (GANNA) processor to prior techniques - backpropagation ANN and logistic regression techniques. To compare the techniques, they utilised binary classification for distinguishing between losing and non losing companies. The findings demonstrated the potential for saving time and excellent classifying results using a GANNA processor.

The learning time of NN classifiers has been reduced greatly with the development of parallel processing technology. Thus, researchers started building deep neural network-based models for image processing, speech recognition, and prediction [55], [56]. Later, deep learning algorithms were introduced into the field of FDP, because of their outstanding performance in the field of classification [57], [58], [59]. Deep networks were also used for performance prediction. DBN- based prediction model was proposed by Ribeiro and Lopes to predict defaults of French Companies [60]. [61] presents distress prediction using deep learning, which leverages unstructured textual data in statements for prediction.

This paper contributes to the existing literature by analyzing three individual classifiers and three ensemble classifiers to find out the best perfoming one among them for FDP with the help of data collected from MENA companies in various sectors.

III. Data and Methodology

A. Data collection and Preprocessing

The primary aim of this research is to identify the best performing model for predicting distress in MENA firms. The samples are divided into two sets of data, dataset 1 and dataset 2, consisting of 613 and 120 companies respectively. Dataset 1 considers 25 input financial attributes, while dataset 2 takes only 17 financial attibutes. These are selected according to the availability of the data and missing common attributes.

All the selected companies are from various sectors, including manufacturing, technology, energy, telecom, real estate, and insurance. The data are collected from the global company database - Osiris[1], and contains a mix of healthy and non-healthy companies. Dataset 1 is unbalanced, with 407 healthy and 206 unhealthy companies. However, dataset 2 is balanced with 60 healthy and unhealthy companies each. Ratios on 20 different financial variables are calculated and grouped into 6 important financial indicators, as shown in the table given below. The variables collected are two years prior to the financially distressed year of each companies, from 2015 to 2019.

[1]Source:https://www.bvdinfo.com/en-gb/our-products/data/international/Osiris

This paper addresses a binary classification problem, determining whether or not a firm may be characterised as financially distressed. The financial dataset's output or target attribute is divided into two categories: financially troubled firms and financially healthy enterprises. Except for the binary target feature, all of the other features in the dataset are continuous. Initial datasets include partial samples, missing data, and null, which are eliminated during preprocessing. Solvency is considered as the prime indicator to distinguish between healthy and non-healthy firms. All the financial indicators considered are within-firm factors; hence this model may be considered for companies in other countries as well.

1) Financial Indicators : All the financial indicators used in the dataset, discussed below, are taken from financial statements and balance sheets of respective companies, as detailed in the below table.

Solvency ratios: The solvency ratio is a crucial metric used to assess the potential of an organization to meet its debt obligations, and is often used by prospective lenders. The solvency ratio implies if a company's cash flow is adequate to satisfy its short-and long-term liabilities. The lower the solvency ratio for an organization, the greater the probability of it defaulting on its debt obligations. A solvency ratio greater than 20 per cent is considered financially stable as a general rule.

Capital expansion: Capital expansion is any investment that strengthens an existing fixed asset or contributes to the introduction of a new fixed asset. A capital expansion makes a company or other entity's fixed asset base larger.

Profitability ratios: Profitability ratios are a class of financial metrics, that are used to measure the ability of a company to generate profits compared to its sales, operating expenses, balance sheet assets and shareholders equity, using data from a particular time frame. To several profitability ratios, a greater number as compared to a competitor's ratio or the same ratio from the prior quarter indicates that the firm is doing well.

Business development: A business development capacity is the ability to grow a business. This is generally done to identify new income sources and control competitive risks. The ratios of income, assets, and profit of a company, in the current and previous years are taken here.

Operational capabilities: Operational capability is the ability to align essential processes, resources, and technology with the overall driving vision and customer-value propositions combined with the ability to execute these processes efficiently and effectively. The operating ratio indicates how good the management of a business is at holding costs down when generating revenue or profits. The lower the percentage, the more effective the company produces sales over overall expenses.

Structural soundness: Structure ratios can be defined as financial ratios that measure the company's long-term stability and structure. These ratios provide insight into the funding strategies the company uses, and emphasize on the long term solvency position.

B. Data modeling

Here we describe data modelling, the algorithms utilised, model assessement, and the data which are used for simula-

Indicators	Formula
Solvency	Total liabilities / Total assets Current assets / Current liabilities Current assets-inventory / Current liabilities Total liabilities/total shareholders' equity Current liabilities/total assets Net operating cash flow/current liabilities Earnings before interest and tax /interest expense
Capital expansion	Net profit/number of ordinary shares at the end of year Net assets/number of ordinary shares at the end of year Net increase in cash and cash equivalents/number of ordinary shares at the end of year Capital reserves/number of ordinary shares at the end of year
Profitability	(Sales revenue–sales cost)/sales revenue Net profit/sales revenue Earnings before income tax/average total assets Net profit/average total assets Net profit/average current assets Net profit/average fixed assets Net profit/average shareholders' equity
Business development	Business income of this year/Business income of last year Total assets of this year/total assets of last year Net profit of this year/net profit of last year
Operational capabilities	Main business income/average total assets Sales revenue/average current assets Sales revenue/average fixed assets Main business cost/average inventory Main business income/average balance of accounts receivable Cost of sales/average payable accounts
Structural soundness	Current assets total assets Fixed assets/total assets Shareholders' equity/fixed assets Current liabilities/total liabilities

tion. The part on data modelling is split into three steps. In the first step, we compare single and multiple data mining classifiers for FDP. The objective of this phase is to find the best classifier. NN classifier was identified as the best performer in this phase. In the next phase, we analyze, MLP, a deep NN model, on FDP, followed by fine-tuning of the main two hyperparameters - network depth and network width, in phase 3. To obtain the most unbiased and accurate findings, all of the techniquess in this paper use resampling with cross-validation metrics. Figure 1 depicts a schematic depiction of the process steps in this research.

1) Phase 1: Data mining classifiers evaluation: Three highly preferred data mining classifier models, SVM, DT and NN, are trained and tested for FDP, along with ensemble techniques, MV, RF and AdaBoost ensemble, with the aim of finding out the most accurate predictive models for classification. Python's scikit-learn packages are used for training the models and for generating the results. Input to the model is the financial data collected and the output is binary, indicating whether a company can be labeled financially distressed or not. The output of this phase is the best classifier identified for FDP, which is then used for further analysis in the next phase. The schematic representation of this phase is depicted in Figure 2.

The neural network (NN) is an ML classifier that is based on an artificial model of the human brain. An input layer, an output layer, and one or more intermediate layers

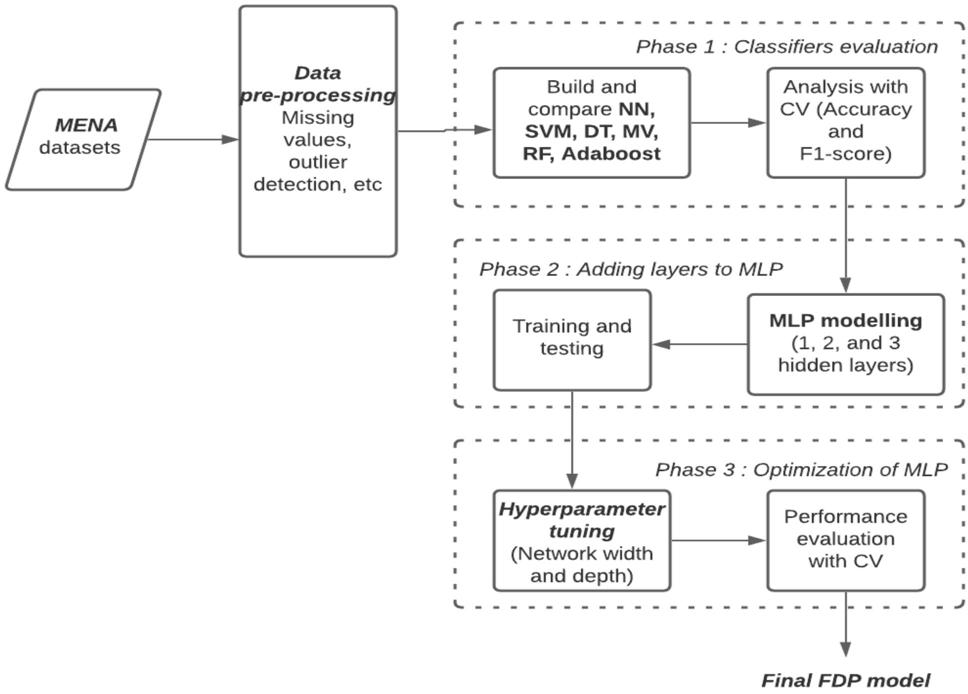

Fig. 1. 3 phases of the proposed study

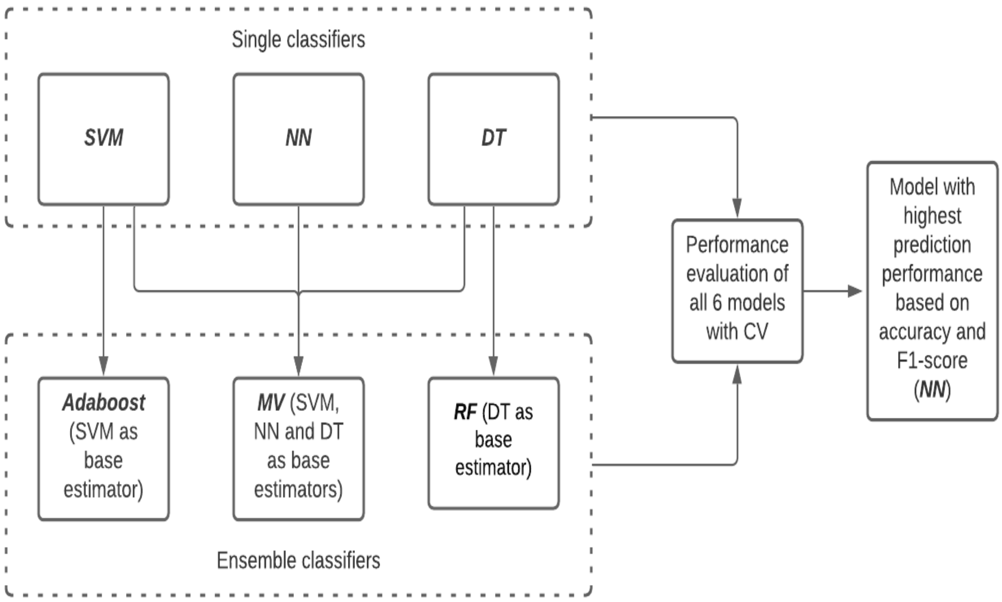

Fig. 2. Schematic diagram of phase 1

comprise the NN architecture. Our NN classifier is built with scikit-learn and has a learning rate of 0.1, as well as back propagation with cross-entropy loss. It contains one hidden layer with ten nodes, and the number of nodes in the input layer is equivalent to the number of financial indicators in the dataset used to train our model. The output comprises of two nodes that indicate whether a data row may be classified as financially distressed or not.

DT is a supervised ML algorithm used in prediction tasks, classification and clustering. An example is Quinlan's ID3 algorithm. Initially, the entire dataset is placed at the root node. The best attribute is placed at the root node. The training dataset is then split into subsets, in such a way that each subset contains data with the same value as in the attribute at the root node. A new node is generated at each branch. This process is repeated until leaf nodes are found at each branch or maximum depth is reached. Information gain is used to decide the best attribute to split on at each step in building the tree. The attribute with the highest information gain is selected as the selection attribute at each node. The leaf nodes of the decision tree indicate the class and decision nodes specify the rules. The test data class is predicted using the decision rules. The DT classifier is trained and evaluated using scikit-learn, with a maximum depth of 5 and entropy as the function to quantify the efficiency of splitting for computing information gain.

SVM is a supervised ML classifier that aims to build a hyperplane that accurately distinguishes between two classes. If the margin of the hyperplane is maximum, then the error is minimized. To achieve this maximum, the quadratic programming optimization is used. The SVM classifier utilised in our work was developed with scikit-learn and a sigmoid kernel function.

Individual base estimator predictions are merged using ensemble approaches to enhance robustness over a single estimator and generate best prediction. In this research, we employed the ensemble methods MV, RF, and AdaBoost.

The MV algorithm integrates the outcomes of the base machine learning models. The model's input, known as base models, can be built using multiple algorithms with the same input dataset or using the same technique with alternative train: test partition ratios or with other approaches. The base estimator methods utilised as input in our work include DT, SVM, and NN. To forecast distress, the test cases are given into each base estimator. The MV model's output for a data set is the class that obtains the best forecast for that sample. The Scikit-learn package is used to train and test MV.

The RF method is a bagging ensemble technique particularly intended for trees. This algorithm's base model is DT, and it tries to decrease the variance of the classifier. We make arbitrary sub-samples of the dataset via replacement and train each subsample using a DT. The mean prediction from every model is computed for the test data. In other words, the ensemble output is the mean prediction of the individual classifiers. W e used scikit-learn to build the model, with the maximum number of features = no. of financial variables in the input dataset and number of trees in forest = 20.

Next, AdaBoost ensemble algorithm is implemented. Scikit-learn is used to create, train, and test the algorithm. For this approach, we utilised SVM as the base model. In each iteration, it assesses the accuracy of predictions by adjusting the weights of classifiers and training the data set.

2) Phase 2: Adding more layers to NN: In this section, we apply deep learning techniques on NN, the most performing model, identified in phase 1, for FDP. The deep network model, MLP is modelled and analyzed for FDP using two sets of data collected from MENA companies. The MLP model is designed with sigmoid activation function and adam optimizer. In this phase, 3 different MLP models are designed and evaluated. The different models contain one, two and three hidden layers respectively. Python's scikit-learn and keras packages using resampling techniques are used for training and testing.

3) Phase 3: Optimization: In phase 2, there was a difference in predicting performance with various layers of MLP. As a result, the third part of our research focuses on further optimising MLP with various architectural combinations. There are several hyperparameters which can be tuned to maximize a deep neural network's prediction performance. We concentrate on fine-tuning the key hyperparameters - network depth and width - that might affect whether the algorithm explodes or converges. The goal of this stage is to study the optimal network depth and network width parameters in order to create an efficient classifier for FDP. The scikit-learn and keras Python libraries are used to train the models and get the outputs. For performance evaluation, we employed the resampling approach of cross-validation, which is further explained in section IV.

IV. Empirical analysis

We built and evaluated the deep learning models using keras, a sophisticated deep learning package in Python. With Python's scikit-learn, the models are assessed using stratified k-fold cross-validation. As a result, we employ the resampling approach to measure model performance.The data is divided into k parts with this approach, and the model is trained using k-1 parts, and the remaining is taken as test data to evaluate the model's performance. We opted to repeat this procedure 10 times, and the mean values of all the created models is then used to estimate the reliable prediction performance.

A. Performance measures

The prediction performance of all classifiers is assessed in terms of accuracy and F1-score, which are popular machine learning assessment metrics. The testing accuracy is used to evaluate performance. Because we employ a k-fold cross-validation score with k=10, the mean and standard deviation for the metrics evaluating accuracy and F1-score are computed over all the ten models. The ratio of correct predictions on the testing data set is referred to as testing accuracy. The F1-score is calculated on the basis of precision and recall, as shown:

$$Precision = \frac{True\ Positive}{(Total\ Predicted\ Positive)} \tag{1}$$

$$Recall = \frac{True\ Positive}{(Total\ Actual\ Positive)} \tag{2}$$

$$F1\ score = \frac{(Precision * Recall)}{(Precision + Recall)} \tag{3}$$

For this research, negative firms are those that are financially troubled, whereas positive companies are those that are financially sound.

B. Analysis

Table I shows the descriptive statistics for the financial attributes in the input data. The accuracy of a machine learning algorithm is heavily reliant on the task at hand, as well as the quality and complexity of the training data set. Table II shows the predicting performance using NN, SVM, and DT models, as well as the results of ensemble techniques MV, RF, and AdaBoost. Figure 3 and Figure 4 show the accuracy and F1-score graphs of the models.

Consistent predictive performance was observed with the NN model throughout all the iterations, although high mean value was shown by MV classifier for both the datasets in terms of accuracy and F1-score. The detailed result of the 10-fold cross validation is given in Table III. Both the

TABLE I
DESCRIPTIVE STATISTICS OF ALL INDIVIDUAL FINANCIAL ATTRIBUTES

Variables	Mean	Min	Max
Total Liabilities	88751.36508	13	1669220
Total Assets	183280.635	5763	3091702
Current Assets	43249.87	138	676520
Current Liabilities	46492.642	10	790074
Accounts Receivable	11041.079	5.63	200071
Accounts Payable	11966.0634	205	374494
Total Shareholders Equity	93131.111	5565	1422482
Net Cash Flow	17297.667	-4600	481539
EBIT	8274	16	267461
Net Profit	4253.9524	-31571	156702
Cash and Cash Equivalent	12907.0158	19	209716
Cost of Goods Sold	22169.1746	345	266764
Sales	1730068.53	93	10301478
Shares	402708	30318.8	3901347
Capital Reserves	5774.65	-615	182827
Fixed Assets	140030.76	599	2415182
Average Total Assets	274315.7778	8815	4608537.5
Average Current Assets	66022.484	210	1048037
Average Fixed Assets	208293.2937	825	3560500
Average Equity	140519.53	8476	2227284
Average Accounts Receivable	13773.69	4069	327807
Average Accounts Payable	16861.05	794	555164.5
Net Increase in Cash	-3622.984	-151247	16129

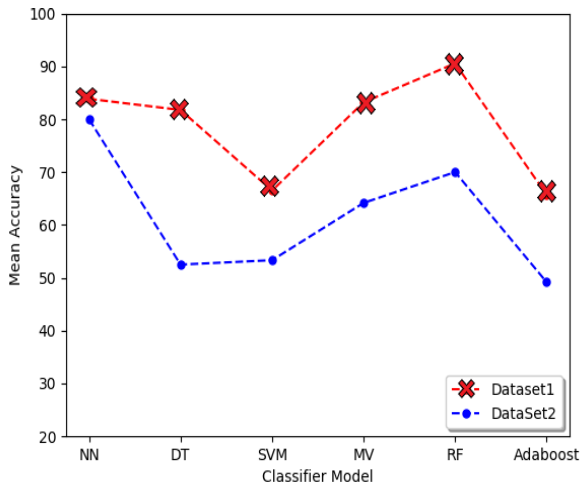

Fig. 3. Mean accuracy of the classifiers in phase 1

MV and the NN classifiers are regarded as the best for data classification with non-linear variable dependencies, however the non-linearity in our data set is better described by the NN classifier than by the MV classifier. As opposed to other ML techniquess, NN is an algorithm that, when tailored for the task, produces the best prediction performance. For this experiment, we have used a hidden layer with ten nodes each and a learning rate of 0.1 to get the best accuracy. The classification of continuous-valued data, like in our data set, is the most accurate with a NN classifier.

The prediction accuracy of DT was high for Dataset1 but comparatively less for Dataset2. DT is a highly preferred categorical data classification model. SVM shows a relatively lower accuracy for both the data sets. The efficiency of

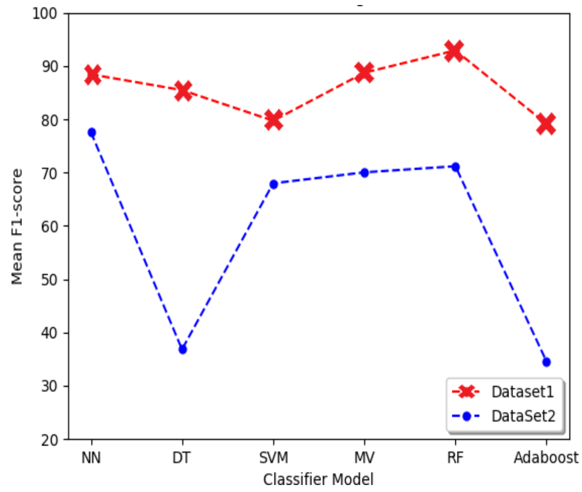

Fig. 4. Mean F1-score of the classifiers in phase 1

SVM is heavily influenced by the parameters selected during training, particularly the kernel function. The parameters that produce the highest accuracy for one task may produce low accuracy for another task. We use the Sigmoid Kernel to train SVM, which is represented by the following equation:

$$k(x, y) = tanh(\alpha x^T y + c) \tag{4}$$

where α, the slope, and c, the intercept constant are adjustable parameters.

AdaBoost is used to improve the efficiency of the SVM classifier, but the results are not as anticipated. For the first dataset, MV has the best average prediction accuracy of the ensemble techniques. We use MV to aggregate the predicted results of all three classifiers, resulting in a stabler classifier than individual ones. While MV aids in the creation of a good accuracy model, it cannot provide consistent output across datasets and surpass the accuracy value given by the single NN model.

Based on the preliminary analysis, NN was found to be the best for FDP, and in the following step of the study, we applied the deep learning technique to the NN classifier. The deep NN model - MLP with sigmoid transfer function and Adam optimizer - was chosen. The results of MLP with one, two and three layers, with 10 nodes in each hidden layer is depicted in figures 5 and 6. The graphs show a clear increase in accuracy and F1-score when deep learning technique is applied. Accuracy of NN in phase 1 was marked 83.86% and 80.00% for Dataset1 and Dataset2, while after applying deep learning technique, the values have increased by 5 to 10%. The results also indicate that the predictive performance of MLP varies based on the number of layers. Hence, we further evaluate the performance of MLP for FDP, with different architectural variations in phase 3.

In phase 3, we further analyze and optimize MLP. In this step, the number of hidden layers and the number of neurons at each layer are adjusted to further optimise the MLP model. Table V and Table VI shows the results of 16 alternative models that were built, trained, and tested using Datasets 1 and 2 respectively. It could be observed that any difference in the number of layers or the number of neurons in each layer have an effect on the model's efficiency. For instance,

TABLE II
MEAN PREDICTION RESULTS OF CLASSIFIERS IN PHASE 1

Classifiers	Dataset1		Dataset2	
	Accuracy	F1-score	Accuracy	F1-score
NN	83.86	88.42	80.00	77.58
DT	81.75	85.46	52.50	36.82
SVM	66.50	79.72	53.33	67.98
MV	83.19	88.72	64.17	70.04
RF	90.57	92.91	70.00	71.18
Adaboost	65.70	79.11	49.17	34.56

TABLE III
DETAILED PREDICTION RESULTS OF CLASSIFIERS IN PHASE 1

Classifiers	Dataset1		Dataset2	
	Accuracy	F1-score	Accuracy	F1-score
NN	74.19	83.67	83.33	83.33
	72.58	80.45	83.33	85.71
	88.52	91.56	75.00	66.66
	77.04	84.44	75.00	72.72
	85.24	88.60	91.66	92.30
	90.16	92.68	91.66	92.30
	83.60	88.37	83.33	80.00
	90.16	92.68	66.66	50.00
	90.16	92.50	75.00	72.72
	86.88	89.18	75.00	80.00
DT	61.29	70.73	58.33	70.58
	66.12	79.20	50.00	00.00
	65.57	77.41	50.00	00.00
	88.52	91.76	58.33	28.57
	93.44	94.87	58.33	28.57
	90.00	90.00	50.00	00.00
	85.24	87.32	58.33	44.44
	73.77	82.22	75.00	80.00
	91.80	93.50	58.33	61.53
	96.72	97.43	41.66	56.82
SVM	67.74	79.16	50.00	66.66
	66.12	79.61	50.00	66.66
	67.21	80.39	50.00	66.66
	67.21	80.39	50.00	66.66
	67.21	80.39	50.00	66.66
	67.21	80.39	50.00	66.66
	65.57	79.20	50.00	62.50
	65.57	79.20	58.33	70.58
	65.57	79.20	75.00	80.00
	65.57	79.20	50.00	66.66

as shown in Table V, MLP with configuration 20-20-20-20-20 has an acceptable percentage accuracy and F1-score of 80.52 and 81.12 respectively, but the network with 50-50-50-50-50 configuration has a poor prediction of 56.14 and 41.63 respectively. An optimized architecture identified for Dataset1 is 10-10-10-10-10 (89.38 for mean accuracy and 91.94 for mean F1-score) and that of Dataset2 is 20-20-20-20 (mean accuracy of 80.00 and mean F1-score of 82.25). The difference in the model performance on the two data sets is because of the change in the number of financial indicators in each data set. A data set with a greater number of features can be trained more effectively and provide a more reliable prediction than a data set with fewer features.

With networks comprising more than 5 layers, accuracy is reduced, and hence a reliable model cannot be generated. Highest classification accuracy is attained with a 4 and 5 layer design, as proven with Datasets 1 and 2. Furthermore, the accuracy score began to decline when the number of neurons at each level approached double the number of features in the input data set. For both data sets, a reliable model was not produced once the number of nodes was increased over 40. As a result, this study shows that when the number of neurons at each level is fewer than twice the number of input features in the data set, prediction performance improves.

TABLE IV
TABLE III (CONTINUED)

Classifiers	Dataset1		Dataset2	
	Accuracy	F1-score	Accuracy	F1-score
MV	74.19	83.67	50.00	57.14
	79.30	85.71	58.33	66.66
	80.32	86.95	83.33	85.71
	75.40	84.21	58.33	66.66
	83.60	88.63	50.00	66.66
	88.52	92.13	75.00	76.92
	85.24	89.88	75.00	72.72
	86.88	90.00	58.33	70.58
	85.24	89.88	75.00	66.66
	93.44	95.23	58.33	70.58
RF	79.03	85.71	75.00	66.66
	75.80	83.14	58.33	44.44
	81.96	87.05	91.66	90.00
	96.72	97.50	83.33	83.33
	98.36	98.76	50.00	50.00
	90.00	90.00	75.00	72.72
	98.36	98.76	83.33	83.33
	90.00	90.00	41.66	46.15
	98.36	98.73	75.00	72.72
	77.04	79.41	50.00	57.14
Adaboost	58.06	72.34	50.00	66.67
	66.12	79.61	50.00	66.66
	67.21	80.39	50.00	66.66
	67.21	80.39	50.00	66.66
	67.21	80.39	50.00	00.00
	67.21	80.39	50.00	66.66
	65.57	79.20	58.33	28.57
	65.57	79.20	58.33	70.58
	65.57	79.20	83.33	80.00
	65.57	79.20	41.66	53.33

TABLE V
PERFORMANCE OF MLP ON DATASET1

Structure	Test Accuracy		Test F1-score	
	Mean	Standard deviation	Mean	Standard deviation
05-05	80.58	9.08	85.88	6.11
05-05-05	86.29	6.36	89.62	4.87
10-10-10	85.29	3.82	88.56	3.18
05-05-05-05	88.57	3.72	91.34	2.94
10-10-10-10	84.96	4.52	88.07	4.46
20-20-20-20	73.19	14.68	76.62	25.80
20-10-20-10	81.57	8.97	84.33	10.93
10-20-10-20	85.80	4.97	88.72	4.17
05-05-05-05-05	86.44	4.96	90.28	3.09
10-10-10-10-10	89.38	4.08	91.94	3.00
20-20-20-20-20	80.52	16.93	81.12	27.27
30-30-30-30-30	72.74	14.19	75.69	21.27
10-20-10-20-10	84.66	7.00	89.11	4.39
20-10-20-10-20	75.13	18.04	75.02	28.76
50-50-50-50-50	56.14	22.75	41.63	40.93
100-50-100-50-100	49.17	18.88	31.61	38.80

TABLE VI
PERFORMANCE OF MLP ON DATASET2

Structure	Test Accuracy		Test F1-score	
	Mean	Standard deviation	Mean	Standard deviation
05-05	65.83	14.65	56.45	26.35
05-05-05	70.83	11.33	72.47	9.91
10-10-10	76.67	11.67	79.57	7.92
05-05-05-05	70.83	11.93	74.70	6.13
10-10-10-10	78.33	15.90	78.46	15.11
20-20-20-20	80.00	13.54	82.25	10.69
20-10-20-10	69.17	14.93	69.42	14.30
10-20-10-20	77.50	15.83	71.22	28.16
05-05-05-05-05	73.33	17	63.59	34.17
10-10-10-10-10	79.17	14.55	79.71	13.14
20-20-20-20-20	72.00	14.43	76.50	11.24
30-30-30-30-30	72.50	12.39	69.01	20.25
10-20-10-20-10	73.33	17.80	68.23	27.51
20-10-20-10-20	80.83	12.94	80.13	14.58
50-50-50-50-50	60.00	9.72	63.95	9.31
100-50-100-50-100	56.67	11.67	62.62	13.77

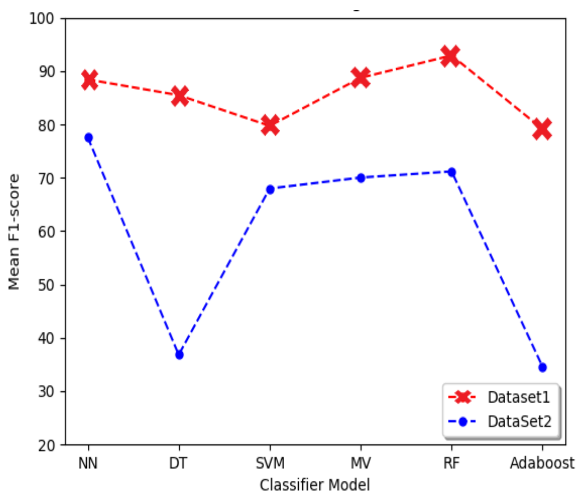

Fig. 5. Mean accuracy of the MLP in phase 2

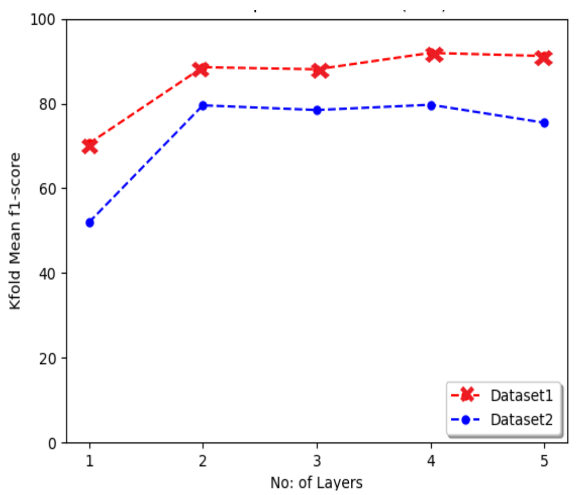

Fig. 6. Mean F1-score of the MLP in phase 2

An architecture with 10 or 20 neuron units in hidden layers achieves the best prediction performance.

Finally, we evaluated the performance of our improved MLP with SVM and DT classifiers. Table VII displays the prediction values in terms of accuracy. The statistical findings showed that the suggested optimised model's prediction accuracy was considerably greater than that of the base ML models utilising both data sets.

V. CONCLUSION AND FUTURE RESEARCH

Finding the best performing FDP model has always been a concern and several FDP models have been created since then. In this paper, we analysed and compared the individual machine learning classifiers and the ensemble techniques used in FDP using two separate datasets from the MENA region. As individual classifiers, we used SVM, DT, and NN, and as ensemble techniques, we used MV, RF, and AdaBoost. Despite the fact that all of the classifier models generated reasonable accuracy rates, the simulation results show that the NN classifier is the most accurate model for FDP. In the final phase of this research, the selected NN classifier is further optimised by incorporating varying numbers of hidden layers (MLP) for improved performance. It was discovered that constructing the model with three or four hidden layers, with the neuron count at each level not surpassing twice the number of input features in the dataset, yielded an optimum predicted performance rate. The experimental findings also demonstrate that the proposed approach outperforms standard machine learning methods like SVM and DT.

The variables considered in our study are all within-firm variables. In the future research, macroeconomic and other industrial factors may be included to get a better model. Furthermore, time series values may be included in the analysis so as to propose a more effective classifier; in our

TABLE VII
COMPARISON OF MLP, SVM, AND DT

Classifiers	Dataset 1		Dataset 2	
	Accuracy	F1-score	Accuracy	F1-score
MLP	89.38	91.94	80.00	82.25
DT	81.75	85.46	52.50	36.82
SVM	66.50	79.72	53.33	67.98

paper, we have considered only the values two years prior to the distress.

REFERENCES

[1] Wanke, Peter, Carlos P. Barros, and João R. Faria, "Financial distress drivers in Brazilian banks: A dynamic slacks approach," *European Journal of Operational Research*, vol. 240 no. 1, pp. 258–268, 2015.

[2] Hu, Hui, and Milind Sathye, "Predicting Financial Distress in the Hong Kong Growth Enterprises Market from the Perspective of Financial Sustainability," *Sustainability*, vol. 7, no. 02, pp. 1186–1200, 2015.

[3] Valášková, katarína, Tomas Kliestik, Lucia Svabova, and Peter Adamko, "Financial Risk Measurement and Prediction Modelling for Sustainable Development of Business Entities Using Regression Analysis," *Sustainability*, vol. 10, no. 06, pp. 21–44, 2018.

[4] Sun, Jie, Hui Li, Qing-Hua Huang, and Kai-Yu He, "Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches," *Knowledge-Based Systems* vol. 57, pp. 41–56, 2014.

[5] Falbo, P., "Credit-scoring by enlarged discriminant models," *Omega* vol. 19, no. 4, pp. 275 – 289, 1991.

[6] Ashraf, Sumaira, Elisabete G. S. Félix, and Zélia Serrasqueiro, "Do Traditional Financial Distress Prediction Models Predict the Early Warning Signs of Financial Distress?," *Journal of Risk and Financial Management*, vol. 12, no. 2, pp. 55-59, 2019.

[7] Khedr, A, Arif, I, P V, PR, El-Bannany, M, Alhashmi, S, S, M., "Cryptocurrency price prediction using traditional statistical and machine learning techniques: A survey," *Intell Sys Acc Fin Mgmt.* vol.28, pp. 3—34. https://doi.org/10.1002/isaf.1488, 2021.

[8] Charalambakis, Evangelos C., and Ian Garrett, "On corporate financial distress prediction: What can we learn from private firms in a developing economy? Evidence from Greece," *Review of Quantitative Finance and Accounting*, vol. 52, no. 2, pp. 467–491, 2018.

[9] jae Kim, Kyoung, Kichun Lee, and Hyunchul Ahn. "Predicting Corporate Financial Sustainability Using Novel Business Analytics," *Sustainability*, vol. 11, no. 1, pp. 64–71, 2018.

[10] Khoja, Layla, Maxwell Chipulu, and Ranadeva Jayasekera, "Analysis of financial distress cross countries: Using macroeconomic, industrial indicators and accounting data," *International Review of Financial Analysis*, vol. 66, pp. 101–379, 2019.

[11] Vellido, A, P.J.G Lisboa, and J Vaughan, "Neural networks in business: a survey of applications (1992–1998)," *Expert Systems with Applications*, vol. 17, no. 1, pp. 51–70, 1999.

[12] Lee, Kidong, David Booth, and Pervaiz Alam, "A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms," *Expert Systems with Applications*, vol. 29, no. 1, pp. 1–16, 2005.

[13] Tax, David M.J., Martijn [van Breukelen], Robert P.W. Duin, and [Josef Kittler], "Combining multiple classifiers by averaging or by multiplying?," *Pattern Recognition*, vol. 33, no. 9, pp. 1475–1485, 2000.

[14] Sreedharan, Meenu, Ahmed M Khedr, and Magdi El Bannany, "A Multi-Layer Perceptron Approach to Financial Distress Prediction with Genetic Algorithm," *Automatic Control and Computer Sciences*, vol. 54, no. 6, pp. 475–482, 2020.

[15] Khedr, Ahmed M, "Learning k-nearest neighbors classifier from distributed data," *Computing and Informatics*, vol. 27 no. 3, pp.355–376, 2008.

[16] Khedr, Ahmed M, and Raj K Bhatnagar, "New algorithm for clustering distributed data using K-means," *Computing and Informatics*, vol. 33, no. 4, pp. 943–964, 2014.

[17] Khedr, Ahmed M. "Nearest neighbor clustering over partitioned data," *Computing and Informatics*, vol. 30, no.5, pp. 1011–1036, 2011.

[18] Khedr, AM—Salim, and Ahmed Salim, "A Decomposable Algorithms for Finding the Nearest Pair," *J. Parallel Distrib. Comput*, vol. 68, pp. 902–912, 2008.

[19] Walid Osamy, Khedr, Ahmed M., Ahmed Aziz, and Ahmed El-Sawy, "Cluster-tree routing based entropy scheme for data gathering in WSNs," *IEEE Access*, vol. 6, pp. 77372-77387, 2018.

[20] Khedr, Ahmed M., Walid Osamy, Ahmed Salim, Sohail Abbas, "A Novel Association Rule-Based Data Mining Approach for IoT Based WSNs," *IEEE Access*, vol. 8, pp. 151574–151588, 2020.

[21] Omar, Dina M, Khedr, Ahmed M, D. P Agrawal, "Optimized clustering protocol for balancing energy in WSNs," *IJCNIS*, vol. 9, no. 3, pp. 367–375, 2017.

[22] Khedr, Ahmed M., and Raj Bhatnagar, "Agents for integrating distributed data for complex computations," *Computing and Informatics*, vol. 26, no. 2, pp. 149–170, 2007.

[23] Khedr, Ahmed M, and Rania Mahmoud, "Agents for integrating distributed data for function computations," *Computing and Informatics*, vol. 31, no. 5, pp. 1101–1125, 2012.

[24] Khedr, Ahmed M, "Decomposable algorithm for computing k-nearest neighbours across partitioned data," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 31, no. 4, pp. 334–353, 2016.

[25] Khedr, Ahmed M, Walid Osamy, Ahmed Salim, and Abdel-Aziz Salem, "Privacy preserving data mining approach for IoT based WSN in smart city," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 8, pp. 555–563, 2019.

[26] Khedr, Ahmed M, Mohamed H Ibrahim, and Amal Al Ali, "LPB: A New Decoding Algorithm for Improving the Performance of an HMM in Gene Finding Application," *IAENG International Journal of Computer Science* vol. 47, no. 4, pp. 723–729, 2020.

[27] Gepp, Adrian, and Kuldeep Kumar, "Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques," *Procedia Computer Science*, vol. 54, pp. 396–404, 2015.

[28] Wu, Desheng(Dash), Liang Liang, and Zijiang Yang, "Analyzing the financial distress of Chinese public companies using probabilistic neural networks and multivariate discriminate analysis," *Socio-Economic Planning Sciences*, vol. 42, no. 3, pp. 206 – 220, 2008.

[29] Min, Jae H., and Young-Chan Lee, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," *Expert Systems with Applications*, vol. 28, no. 4, pp. 603–614, 2005.

[30] Wanke, Peter, Md Abul Kalam Azad, Ali Emrouznejad, and Jorge Antunes, "A dynamic network DEA model for accounting and financial indicators: A case of efficiency in MENA banking," *International Review of Economics & Finance*, vol. 61, pp. 52–68, 2019.

[31] Altman, Edward I, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankrupcy," *The Journal of Finance*, vol. 23, no.4, pp. 589–609, 1968.

[32] Ohlson, James A., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, vol. 18, no. 1, pp. 109–116, 1980.

[33] Zmijewski, Mark E., "Methodological Issues Related to the Estimation of Financial Distress Prediction Models," *Journal of Accounting Research*, vol. 22, pp. 59–82, 1984.

[34] Balcaen, Sofie, and Hubert Ooghe, "35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems," *The British Accounting Review*, vol. 38, no. 1, pp. 63–93, 2006.

[35] Altman, Edward I., Malgorzata Iwanicz-Drozdowska, Erkki K. Laitinen, and Arto Suvas, "Financial and Non-Financial Variables as Long-Horizon Predictors of Bankruptcy," *Available at SSRN 2669668 Electronic Journal*, 2015.

[36] Chiaramonte, Laura, and Barbara Casu, "Capital and liquidity ratios and financial distress. Evidence from the European banking industry," *The British Accounting Review*, vol. 49, no. 2, pp. 138–161, 2017.

[37] Agarwal, Vineet, and Richard Taffler, Wrapquotes, "Comparing the performance of market-based and accounting-based bankruptcy prediction models," *Journal of Banking & Finance*, vol. 32, no. 8, pp. 1541–1551, 2008.

[38] Laitinen, Erkki K, "FINANCIAL RATIOS AND DIFFERENT FAILURE PROCESSES," *Journal of Business Finance & Accounting*, vol. 18, no. 5, pp. 649–673, 1991.

[39] Skogsvik, Kenth, "CURRENT COST ACCOUNTING RATIOS AS PREDICTORS OF BUSINESS FAILURE: THE SWEDISH CASE," *Journal of Business Finance & Accounting*, vol. 17, no. 1, pp. 137–160, 1990.

[40] Casey, Cornelius, and Norman Bartczak, "Using Operating Cash Flow Data to Predict Financial Distress: Some Extensions," *Journal of Accounting Research* vol. 23, no. 1, pp. 384–396, 1985.

[41] Cleofas-Sánchez, L., V. García, A.I. Marqués, and J.S. Sánchez, "Financial distress prediction using the hybrid associative memory with translation," *Applied Soft Computing*, vol. 44, pp. 144–152, 2016.

[42] Ravisankar, P., and V. Ravi, "Financial distress prediction in banks using Group Method of Data Handling neural network, counter propagation neural network and fuzzy ARTMAP," *Knowledge-Based Systems*, vol. 23, no. 8, pp. 823–831, 2010.

[43] Lee, Sangjae, and Wu Sung Choi, "A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis," *Expert Systems with Applications*, vol. 40, no. 8, pp. 2941–2946, 2013.

[44] Sun, Jie, Hamido Fujita, Peng Chen, and Hui Li, "Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble," *Knowledge-Based Systems*, vol. 120, pp. 4–14, 2017.

[45] Salehi, Mahdi, Mahmoud Mousavi Shiri, and Mohammad Bolandraftar Pasikhani, "Predicting corporate financial distress using data mining techniques," *International Journal of Law and Management*, vol. 58, no. 2, pp. 216–230, 2016.

[46] Geng, Ruibin, Indranil Bose, and Xi Chen, "Prediction of financial distress: An empirical study of listed Chinese companies using data mining," *European Journal of Operational Research*, vol. 241, no. 1, pp. 236–247, 2015.

[47] Bae, Jae Kwon, "Predicting financial distress of the South Korean manufacturing industries," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9159–9165, 2012.

[48] Maghyereh, Aktham I., and Basel Awartani, "Bank distress prediction: Empirical evidence from the Gulf Cooperation Council countries," *Research in International Business and Finance*, vol. 30, pp. 126–147, 2014.

[49] Peat, Maurice, and Stewart Jones, "Using neural nets to combine information sets in corporate bankruptcy prediction," *Intelligent Systems in Accounting, Finance and Management*, vol. 19, no. 2, pp. 90–101, 2012.

[50] Telmoudi, Fedya, Mohamed El Ghourabi, and Mohamed Limam, "RST–GCBR-Clustering-Based RGA–SVM Model for Corporate Failure Prediction," *Intelligent Systems in Accounting, Finance and Management*, vol. 18, no. 2-3, pp. 105–120, 2011.

[51] Anandarajan, Murugan, Picheng Lee, and Asokan Anandarajan, "Bankruptcy prediction of financially stressed firms: An examination of the predictive accuracy of artificial neural networks," *Intelligent Systems in Accounting, Finance & Management*, vol. 10, no. 2, pp. 69–81, 2008.

[52] O'leary, Daniel E, "Using neural networks to predict corporate failure," *Intelligent Systems in Accounting, Finance & Management*, vol. 7, no. 3, pp. 187–197, 1998.

[53] Boritz, J Efrim, Duane B Kennedy, and Augusto de Miranda e Albuquerque, "Predicting corporate failure using a neural network approach," *Intelligent Systems in Accounting, Finance and Management*, vol. 4, no. 2, pp. 95–111, 1995.

[54] Fanning, Kurt M, and Kenneth O Cogger, "A comparative analysis of artificial neural networks using financial distress prediction," *Intelligent Systems in Accounting, Finance and Management*, vol. 3, no. 4, pp. 241–252, 1994.

[55] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[56] Shen, Furao, Jing Chao, and Jinxi Zhao, "Forecasting exchange rate using deep belief networks and conjugate gradient method," *Neurocomputing*, vol. 167, pp. 243–253, 2015.

[57] Glorot, Xavier, Antoine Bordes, and Yoshua Bengio, "Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach," *ICML*, pp. 513–520, 2011.

[58] Sreedharan, Meenu, Ahmed M Khedr, and Magdi El Bannany, "A comparative analysis of machine learning classifiers and ensemble techniques in financial distress prediction," *2020 17th International Multi-Conference on Systems, Signals & Devices (SSD)*, IEEE, pp. 653–657, 2020.

[59] El-Bannany, Magdi, Meenu Sreedharan, and Ahmed M Khedr. "A Robust Deep Learning Model for Financial Distress Prediction," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 2, pp. 170-175, 2020.

[60] Ribeiro, Bernardete, and Noel Lopes, "Deep Belief Networks for Financial Prediction," *International Conference on Neural Information Processing*, pp. 766–773, 2011.

[61] Matin, Rastin, Casper Hansen, Christian Hansen, and Pia Mølgaard, "Predicting distresses using deep learning of text segments in annual reports," *Expert Systems with Applications*, vol. 132, pp. 199–208, 2019.