ORIGINAL RESEARCH

Deep Learning-Based Model for Financial Distress Prediction

Mohamed Elhoseny[1,2] · Noura Metawa[3,4] · Gabor Sztano[5] ·
Ibrahim M. El-hasnony[1]

Abstract

Predicting bankruptcies and assessing credit risk are two of the most pressing issues in finance. Therefore, financial distress prediction and credit scoring remain hot research topics in the finance sector. Earlier studies have focused on the design of statistical approaches and machine learning models to predict a company's financial distress. In this study, an adaptive whale optimization algorithm with deep learning (AWOA-DL) technique is used to create a new financial distress prediction model. The goal of the AWOA-DL approach is to determine whether a company is experiencing financial distress or not. A deep neural network (DNN) model called multilayer perceptron based predictive and AWOA-based hyperparameter tuning processes are used in the AWOA-DL method. Primarily, the DNN model receives the financial data as input and predicts financial distress. In addition, the AWOA is applied to tune the DNN model's hyperparameters, thereby raising the predictive outcome. The proposed model is applied in three stages: preprocessing, hyperparameter tuning using AWOA, and the prediction phase. A comprehensive simulation took place on four datasets, and the results pointed out the supremacy of the AWOA-DL method over other compared techniques by achieving an average accuracy of 95.8%, where the average accuracy equals 93.8%, 89.6%, 84.5%, and 78.2% for compared models.

✉ Mohamed Elhoseny
   Melhoseny@ieee.org

   Noura Metawa
   Nmetawa@sharjah.ac.ae

   Gabor Sztano
   gabor.sztano@uni-corvinus.hu

   Ibrahim M. El-hasnony
   ibrahimhesin2005@mans.edu.eg

1   Faculty of Computers and Information, Mansoura University, Mansoura, Egypt

2   College of Computing and Informatics, University of Sharjah, Sharjah, United Arab Emirates

3   College of Business Administration, University of Sharjah, Sharjah, United Arab Emirates

4   Faculty of Commerce, Mansoura University, Mansoura, Egypt

5   Corvinus University of Budapest, Budapest, Hungary

🙋 Springer

1 Introduction

Because of the impacts of the COVID-19 outbreak, many businesses are facing financial difficulties. (Abedin et al., 2021) Usually, economic downturns are part of companies' long-term business activity, and problems may be discovered in time and solved adequately. Due to its nature, financial distress is predictable with several types of models. (Uthayakumar et al., 2020a) Generally, businesses could perceive possible threats by analyzing country-, industry, and firm-level information on economic stance. To assist companies in dispersing and avoiding financial threats effectively and in a timely manner, economic crises warning systems have great importance in companies' risk management. Several researchers have developed early warning economic models in various sectors; however, the precision of these methods still needs to be enhanced (Ghisellini & Ulgiati, 2020; Liu & Arunkumar, 2019; Sankhwar et al., 2020).

Now, many related pieces of literature use arithmetical models for building an economic crisis earlier warning system or employ different techniques for assessment (Ashraf et al., 2019; Bluwstein et al., 2020; Shajalal et al., 2021). Also, these studies use a comparative approach and use 2 NN methods for establishing economic crises earlier warning systems (Metawa et al., 2021; Yang & Yang, 2020). This 2 NN model includes RPROP NN and SVM. Such results could recommend the domestic economic sectors and offer a foundation for financial management for public investments. Currently, the ML algorithm has been extensively employed in the fields of company economic distress predictions. But, many algorithms regard economic distress predictions as a more straightforward dichotomy problem and frequently neglect the timeline of the financial distress epidemic in real-time. Also, the possible relations among tags and features may not be deliberated. Worthy economic distress predictions should be efficient and realistic (Chatzis et al., 2018; Huang & Yen, 2019; Petropoulos et al., 2020).

In this study, a new financial distress prediction model uses an adaptive whale optimization algorithm with deep learning (AWOA-DL) technique, including multi-layer perceptron (MLP) and optimization algorithm. Population-based WOA is capable of avoiding local optimums and finding a solution that is optimal globally. With these advantages, WOA is an ideal algorithm for solving a wide range of constrained and unconstrained optimization problems in practical applications without the need for structural reform of the algorithm. The WOA is used for fine-tuning the parameters of the MLP so producing the parameters that maximize the performance of the MLP with better accuracy. It is now easier to fine-tune MLP parameters with the help of the WOA, which allows for more precision.

The goal of the AWOA-DL approach is to determine whether a company is experiencing financial distress or not. The proposed AWOA-DL technique predicts a company's financial distress in four stages, each of which is described in detail below. The data collection phase is responsible for gathering the datasets that will be used to evaluate the performance of the proposed model. Second, the data preprocessing stage improves the quality of the data collected before moving on to the other steps of the model. Preprocessing removes incomplete samples, missing data, and null values. Because all financial indicators are based on company characteristics, this method can be used globally. Third, the AWOA-based parameter optimization phase fine-tunes the hyperparameters of the MLP model in preparation for the final step. We will use sigmoid transfer functions to train the networks to predict distress. The AWOA-DL technique combines deep neural network (DNN) prediction with AWOA hyperparameter tuning. The DNN, including the MLP model, predicts financial distress using financial data.

The last step is to use deep neural networks (DNNs) to predict the company data class labels (DF or NFD).

We examine the performance of real-time and dynamically changing structures in four different datasets in this work. The studies in this work primarily aim to give researchers clear guidance on constructing a reliable deep learning model for predicting future financial difficulties. FD is for the financially distressed firm, while NFD stands for the non-financially distressed corporation. The study's findings indicate that the proposed model has a significantly higher prediction accuracy than conventional machine learning models.

This article is structured as follows. The next section discusses research into financial distress prediction. Section 3 discusses the tools and techniques employed in greater detail. Section 4 introduces and clarifies the model being discussed. Section 5 describes, optimizes, analyzes, and compares the prediction performance of all of the chosen deep models. Section 6 presents a robust model for predicting financial distress using deep neural networks and concludes with recommendations.

2 Related work

The logistic regression models (Ohlson, 1980; Sun et al., 2014), and discriminant analysis (Altman, 1968) are examples of early and conventional statistical models that have been used in distress prediction. As seen above, these basic linear processes are impractical since they are very simple when it comes to developing a good model for generating real-time predictions, as seen above. To better predict bank hardship, researchers in the Gulf Cooperation Council created a simple hazard model (Maghyereh & Awartani, 2014). Other methods, such as machine learning utilizing data mining techniques like Logistic Regression (Min & Lee, 2005; Sun et al., 2017) and Support Vector Machines (Chandra et al., 2009; Cleofas-Sánchez et al., 2016; Hall & Holmes, 2003; Moula et al., 2017; Serre et al., 2007) were explored. An evaluation of the performance of machine learning algorithms for predicting financial hardship in publicly traded Chinese enterprises was completed in 2015 by Serre et al. (2007). Yu et al. (2020) developed an advanced complicated network method for simulating an interbank network using general risks contagion, which considered the balance sheet of all banks, where they could find when the economic institution has an adequate capital reserve to avoid risks contagion. Cascade default is also made in the simulations based on the distinct crises triggering (target default) method. They utilize ML methods for identifying the synthetic feature of the networks. Appiahene et al. (2020) presented an integrated DEA using 3 ML methods in calculating bank performance and efficiency with 444 Ghanaian bank branches, DMUs. The outcomes are related to the respective efficacy rating attained from the DEA. Lastly, the predictive accuracy of the 3 ML algorithms is related. The result recommended that the DT and its C5.0 algorithms provide an optimal prediction method. Samitas et al. (2020) studied on EWS model by exploring potential contagion threats according to the structure economic network. Earlier warnings indicators enhance average crisis predictive model performances. The network analyses and ML algorithm found proof of contagion risks on the date in which they observed considerable expansion in centralities and correlations.

Kim et al. (2020) aimed to forecast the directions of US stock price by incorporating time differing ETE and different ML methods. Examining the ETE depending on the three & six months moving window could be considered the explanatory market parameter by examining the associations among the economic crisis and Granger causal relations amongst the stocks. Zeng et al. (2020) proposed a method for FS classification predictions depending on data

source grouping and feature attributes. The present economic distress predictions generally employ the information from financial statements and ignore the timeline of companies. Thus, proposed a company's economic distress predictions, i.e., improved in line using the practice and combines the grouping sparse PCA of financial information, business governance features, and market transaction information using SVM.

Uthayakumar et al. (2020b) introduce a cluster-based classification model comprised of 2 phases: enhanced K-means clustering and FSCGACA based classification methods. Initially, improved K-means algorithms are developed for eliminating the incorrectly clustered information. Later, rule-based models are elected for designing the provided datasets.

Previous research shows that predicting the failure of a business or organization is critical. As the failure of a company is clearly associated with several characteristics and with financial history, modeling the probability of firms' default is a feasible and, nonetheless, essential task. Earlier research focused on developing statistical and machine learning models to predict financial distress. Researchers have identified the economic problems warning model in various sectors, but their precision needs improvement. The ML algorithm has been widely used to predict company financial distress. However, many algorithms treat economic distress prediction as a binary problem, ignoring the epidemic's real-time timeline. Also, tags and features may not be related.

3 Materials and methods

In this chapter, a general description of the DNN and MLP, as a model used in this paper, will be presented and then the Whale Optimization Algorithms will be discussed.

3.1 Deep neural network

A deep neural network (DNN) is an artificial neural network with multiple layers. It is a multilayer network, with the term "deep" denoting the number of layers used to modify the data (Serre et al., 2007)—a subclass of the class of machine learning models. The primary distinction between classical and deep neural networks is the number of hidden layers and the training procedure. DNN can extract higher-order correlations by adding additional hidden layers (Freund & Haussler, 1994; Hinton et al., 2006; Weston et al., 2012). The basic concept behind a DNN is that it learns via repetition from a collection of samples, such as 100 images of various dogs, rather than from a set of artificial criteria, such as "a dog has a black nose and floppy ears." In this manner, a DNN acquires knowledge in the same way the human brain does—via practice and error (Mobahi et al., 2009). This is a graphical depiction of a feedforward DNN architecture in its most basic form. The DNN is given an input vector x of dimension D, which is processed according to a function g and the DNN's parameters by the hidden layers hj (constructed of Nj hidden units) (weights matrices W and bias vectors b) (Erhan et al., 2009). Finally, the output layer O contains the DNN's output for the specified job (for the case of classification, the probability of an input vector to belong to each class C). DNN is a subset of machine learning modeled after how the human brain learns (Bengio, 2009). It has been used for several purposes, some of which you may be familiar with, such as language translation and picture search tools, and others of which you may be unaware, such as medical diagnostics (Zhou et al., 2015). UCLA developed a deep neural network (DNN) to identify cancer cells! We are now utilizing it to charge our new hearing equipment. DNN training is a difficult task to accomplish (Bengio, 2014). The MLP model is used in
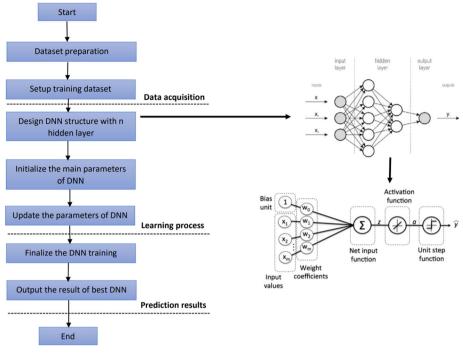
Fig. 1 DNN flowchart

this study to make predictions about the performance of a deep neural network (DNN). DNN architectures include multilayer perceptron (MLP) and convolutional neural network (CNN). In Fig. 1, we can see a DNN design flowchart.

3.1.1 Multilayer perceptron neural network

The processing elements in feedforward neural networks (FFNNs) are referred to as 'neurons.' There are several layers of neurons, and each layer is totally coupled to the one below it. The input layer is the first one, and it acts as a map between the network's variables and the inputs. The output layer is the final and most important layer to consider. Layers between the input and output are referred to as hidden layers. In the field of FFNN, the most commonly used architecture is the multilayer perceptron (MLP). One-way and one-directional connections between neurons are common in MLP. Weights are used to depict connections, and actual integers in the interval [1,1] are used. Using only one hidden layer, Fig. 2 depicts a basic MLP. Each node's output is calculated twice: once for input and once for output. This is done by first calculating the weighted sum of the inputs using Eq. 1, wherein I_i is the input variable I and w_{ij} is the weighted correlation between I_i and the hidden neuron j. Second, a summation function is employed to activate neurons' output via an activation function. MLP could make use of several activation function types. The hidden layer's node j's output can be determined as stated in Eq. 1 using the sigmoid function, which is widely
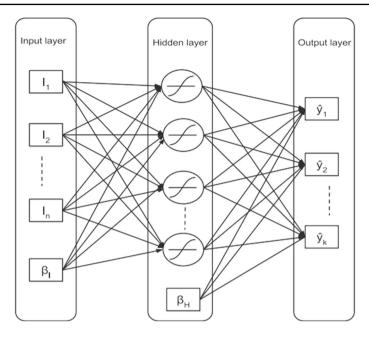
Fig. 2 MLP network

used in the literature (Desai & Shah, 2020).

$$S_j = \sum_{i=1}^{n} w_{ij} I_i + \beta_j$$
$$f_j(x) = \frac{1}{1+e^{-S_j}}$$

(1)

The network's ultimate output is given in Eq. 2 after the hidden layer's neuron outputs have been calculated.

$$\widehat{y}_k = \sum_{i=1}^{m} W_{kj} f_i + \beta_k$$

(2)

3.2 Whale optimization algorithm

There has been a rise in the use of metaheuristic algorithms that replicate the collective behavior of creatures. Particularly concerning optimization, these techniques have had a substantial impact. Metaheuristic algorithms may provide excellent solutions in an acceptable amount of time. A smart way to overcome the restrictions of a lengthy, time-consuming search is to use a more efficient method. On the other hand, many metaheuristic algorithms suffer from a lack of search variety and imbalances between explosive and exploitative performance, making them less effective than other algorithms. One of the new population-based meta-heuristic algorithms is the Whale optimization algorithm (WOA). WOA employs a diminishing encircling mechanism, helical rising, and random learning procedures to keep track of whales. WOA has a lot going for it in terms of its ease of use and accuracy of computation.
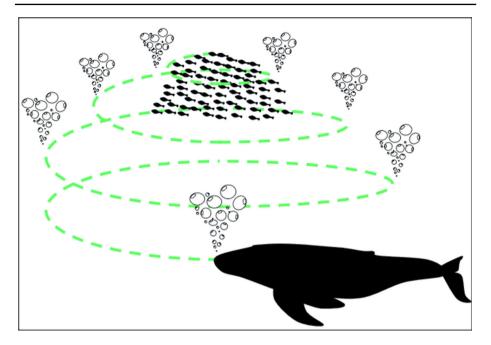
Fig. 3 Bubble-net hunting behavior

The AWOA is developed in the whales' hierarchy. A primary objective of WO concerned the humpback whale and the individual's hunting rule. The present and optimum candidate solutions are set to accomplish objective prey or optimum solutions (Gharehchopogh & Gholizadeh, 2019). The little fish near the surface are the food source for humpback whales (Fig. 3). Mainly, the bubble-net encourage or chase performance is regarded as a positive objective in which humpback whale takes down in unpleasant water from the depth of 10–15 m. Still, another technique is a spiral shape prey surrounded and the air pocket transfer on the surface. The humpback whale notices prey's place and surrounds them. For accomplishing a model place from a search space, present and optimum opponent solution is an objective prey from AWOA (Shankar et al., 2020). Then obtaining the optimum search agents, alternating search agent endeavor for reforming the conditions to effectual search agents,

$$\overrightarrow{U} = \left| S.\overrightarrow{K}^*(t) - \overrightarrow{K}(t) \right| \tag{3}$$

$$\overrightarrow{K}(t+1) = .\overrightarrow{K}^*(t_{best}) - Y^*\overrightarrow{U}. \tag{4}$$

In the beyond purpose, $S = 2.r$, and $Y = 2.I.r - I$, in which "I" is decreased in [2–0] for every iteration, a novel position of search agent was easy from the center of actual place of agent and the place of present optimum agents. A novel solution has been recognized to forecast fitness value and optimum highlight with minimum parameter dependence. According to the functional fitness value, define the coefficient vectors "y", in which the versatile possibilities function has been established:

$$y \Rightarrow Probability = \begin{cases} C_1(f_{\max} - f_x)/(f_{\max} - F_{avg}), f_x \geq F_{ovg} \\ c_3, f_x \leq f_{avg}. \end{cases} \tag{5}$$

At this point, f_{min} and f_{max} represent the maximal and minimal FFs, but C_1 *and* C_3 grades amongst 0 and 1. Therefore, meta-heuristic manners are effective if implemented to an adaptive scheme that outcomes from minimum processing duration to accomplish ideal outcomes like minimum evasion, and effectual convergence. The mathematical illustration refers to the bubble net performance of humpback whales, with 2 enhanced techniques. Primarily, shrink as well as the circle that is decreased to detect the dimensional to coefficient vector. The spiral function was implemented amongst the place of the whale as well as the victim to demonstrating the helix-shaped humpback whale which is shown as:

$$\overrightarrow{K}(t+1) = e^{bt}.\cos(2\pi.y).\overrightarrow{U}' + \overrightarrow{K}^*(t). \tag{6}$$

It is noticeable humpback whales swim through prey in an increasing circle and wind is created to the whole method. For demonstrating the synchronous nature, the maximal possibilities are amongst contract as well as enclose method which refreshes the optimized model. The numerical technique was demonstrated as given under:

$$\overrightarrow{K}(t+1) = \begin{cases} \widehat{K}(t) - Y\overrightarrow{U} & If\ p < 0.5 \\ \overrightarrow{U}'.e^{bs}.\cos(2\pi s) + \overrightarrow{K}(t) & if\ p \geq 0.5, \end{cases} \tag{7}$$

where y refers to the arbitrary score amongst -1 to 1 that implies the synchronous nature which recognizes possibilities amongst the contract encircling model and twisting scheme to the midst of optimization.

The competing method in terms of the arrangement of vectors was implemented to search for prey. Afterward, the whole analyzer request to the administrator was invigorated utilizing an arbitrarily elected search specialist to utilizing a better searching operator

$$\overrightarrow{U} = \left| \overrightarrow{s}.\overline{K}_{rond} - \overrightarrow{K} \right| \tag{8}$$

$$K(t+1) = \overrightarrow{K} - \overrightarrow{Y}.\overrightarrow{U}. \tag{9}$$

So, arbitrary values are implemented frequently amongst $[1, -1]$ to calculate the search operation. Before utilizing an exploitation stage, the search specialist was changed from the exploration phase as demonstrated as without purpose elected search operator if related to an optimum search operator as recognized still now. Therefore, model $\left| \overrightarrow{Y} \right| > 1$ emphasizing search operation and allows the AWOA to precede a global optimal and $\left| \overrightarrow{Y} \right| > 1$ to examine a situation of search agent. This rule is monitored still getting a superior amount of iterations. A novel group of solutions is accomplished and trail on the fundamental of upgrading rules still the end condition is fulfilled.

4 The proposed model

The proposed AWOA-DL technique predicts the financial distress of a company in four stages (as shown in Fig. 4). The data collection phase collects the datasets required to evaluate the proposed model's performance. Second, the data preprocessing phase enhances the data collected before moving to the model's other steps. Third, the AWOA-based parameter optimization phase adjusts the MLP model's hyperparameters for the final step. Finally, DNN-based prediction of company data class label (DF or NFD) based on previous phases. A detailed discussion of these modules is offered in the succeeding sections.
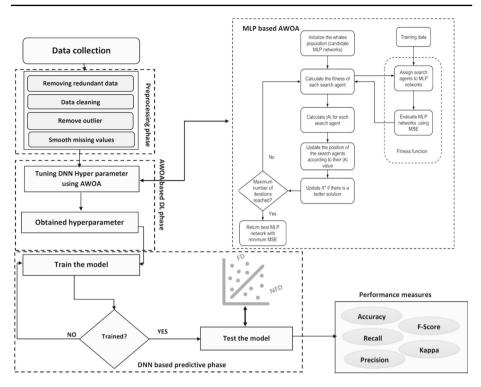
Fig. 4 Workflow of AWOA-DL model

4.1 Data collection

In order to find the most accurate model for predicting financial distress, several datasets have been used. The samples are obtained from four sets of data, Australian (Dua & Graff, 2017), Analecta Dataset (Data archieves, https://pages.stern.nyu.edu/~adamodar/New_Home_Page/dataarchived.html), Polish companies, and Taiwanese datasets (Zikeba et al., 2016).

Australian dataset: considers 690 samples with 14 input financial attributes. The dataset is associated for binary classification tasks. In this file, you'll find information on credit card application processes. To ensure data secrecy, all attribute names and values have been replaced with meaningless symbols…. Because of the good mix of features in this dataset—continuous, nominal with small values, and nominal with large values—it is worth looking into. A few values are omitted as well. The Australian dataset contains 307 non-bankrupt samples and 383 bankrupt samples.

- *Analects dataset* considers 50 samples with about 5 attributes. The dataset is distributed into two or binary classes for bankrupt and non-bankrupt samples. The Analecta dataset contains 25 samples for non-bankrupt and the same for the bankrupt category.
- *Polish company dataset* considers 10,503 samples with 64 input financial attributes. The dataset is associated for binary classification tasks. EMIS, a database of information about emerging markets around the world, was used to gather the information. During the years 2000–2012, the insolvent enterprises were examined, while the still operational firms

Table 1 Dataset description

Dataset	Source	instances	attributes	# of class	Bankrupt/Non-Bankrupt
Australian	UCI	690	14	2	383/307
Analecta Dataset	stern	50	5	2	25/25
Polish companies	UCI	10,503	64	2	495/10008
Taiwanese dataset	UCI	6819	96	2	220/6599

were assessed between the years 2007 and 2013. The Polish dataset contains 10,008 non-bankrupt and 495 bankrupt samples.

- *Taiwanese dataset* considers 6819 samples with 96 input financial attributes. The dataset is associated for binary classification tasks. For the years 1999 to 2009, the information was gathered from the Taiwan Economic Journal. The Taiwan Stock Exchange's business regulations were used to define corporate bankruptcy. The Taiwanese dataset contains 6599 non-bankrupt samples and 220 bankrupt samples.

This article aims to determine whether or not a company may be classified as being in a financial crisis. For preprocessing, incomplete samples are removed as well as missing data and null values. Firm health can be determined by looking at a company's level of solvency. Because all financial indicators are based on characteristics within the company, this approach can be used for businesses in other nations as well. The samples come under two class labels, namely bankrupt (FD) and non-bankrupt (NFD). Table 1 shows the details of the dataset description.

4.2 Data preprocessing phase

The paper addresses a binary classification problem: whether a business should be classified as financially distressed (FD) or not (NFD). The financial dataset's output/target attribute is organized into economically distressed companies and financially healthy companies. Except for the binary target attribute, all other features in the dataset may be of any type. Additionally, the initial datasets include incomplete samples or have missing data or null values, which are removed during preprocessing. Smoothing the missing values is required to handle the lost values of attributes. Normalization of different scales of features is necessary to make all attributes on a homogeneous scale. There may be redundant rows in the datasets so, reducing and removing redundant rows is required. All these tasks are required before modeling the problem into the workflow of the proposed model. All mentioned tasks are performed in the preprocessing phase (Hanbal & Metawa, 2019).

4.3 AWOA based DL phase

For tuning the hyperparameters of the DNN model, the WOA is applied to it. Different MLP layers had varying success forecasting performance in this phase. Our study in this area is focused on finding new ways to improve MLP by combining different architectural styles. The prediction performance of a deep neural network can be improved by adjusting a number of hyperparameters. We focus on fine-tuning the algorithm's critical hyperparameters, such as network width and depth, which might cause it to explode or converge. We want to find out

the best network depth and width to construct a financial distress classifier that works well. The models are trained using Python libraries such as scikit-learn and Keras, and the results are obtained using these same tools. Each search agent in the proposed model is encoded as a one-dimensional vector to represent a potential neural system. It is possible to divide vectors into three parts: an input layer, a hidden-layer weighted layer, and a hidden-to-output layer biases. It is possible to calculate the length of each vector by applying Eq. 10, where n is the number of input variables and m is the number of hidden layer neurons.

$$\text{Individual length} = (n \times m) + (2 \times m) + 1 \qquad (10)$$

To quantify the fitness value of the generated WOA agents, we use the mean square error (MSE) fitness function, which is calculated as the difference between the actual and predicted values by the created agents (MLPs) for all training samples. The MSE is depicted in Eq. 11, where y denotes the actual value, $\widehat{y}$ denotes the predicted value, and n denotes the number of instances in the training dataset.

Figure 5 represent the process of assigning search agent vector to MLP by WOA. The vector represents the input parameters assigned to the WOA for adaptation.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y - \widehat{y})^2 \qquad (11)$$

In other words, the following steps illustrate the workflow of the WOA-based technique used to train the MLP network in this study:

1. A predetermined number of search agents is randomly produced as part of the initialization process. MLP networks are represented by the search agents.
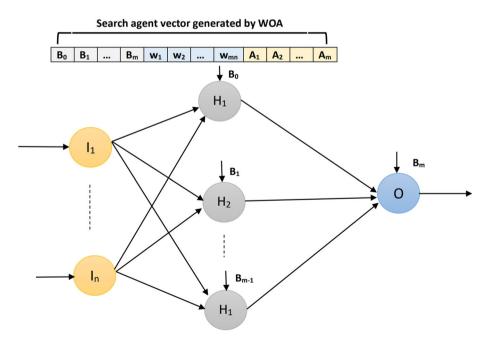

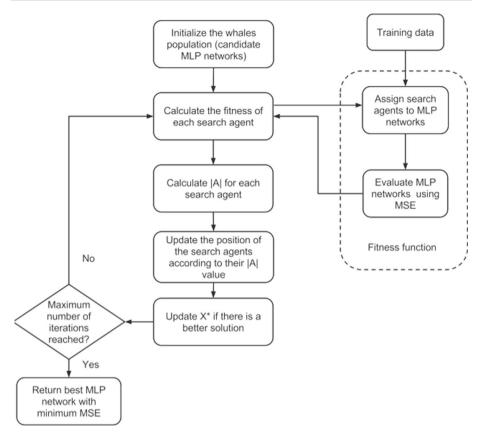Fig. 5 Assigning a vector of a WOA search agent to an MLP

Fig. 6 Flowchart of tuning MLP hyperparameters using WOA

2. The produced MLP networks' quality is assessed using a fitness function. Each MLP network is then assessed to see how well it matches a given set of search agents' weights and biases established during the previous stage. The MSE is used as a fitness function in this study because it is extensively used in evolutionary neural networks. Using the dataset as a training set, the purpose of the algorithm is to find the MLP network with the lowest MSE value.
3. Relocate the search agents to their new locations.
4. Step 4 is the repetition of steps 2 to 3 until the desired number of iterations is reached. Finally, the MLP network with the lowest MSE value is evaluated on an unexplored portion of the dataset.

Figure 6 summarizes the above mentioned steps of the workflow of applying WOA for MLP as a DNN model.

4.4 DNN based predictive phase

MLP is a highly preferred deep learning network model for financial distress prediction at this phase. MLP is developed, trained, and tested. We use Python's scikit-learn libraries to

train the model and generate the results. There are two inputs to the model: financial data gathered by the firm and binary results showing whether the firm is in financial crisis or not. With scikit-learn, we created a classifier that uses back propagation along with cross-entropy loss and has a learning rate of 0.1. We utilized a dataset with an equal number of financial variables to train our model, thus there are an equal number of nodes in the input layer. If a data row can be classed as financially distressed, the output will include two nodes. With the sigmoid activation function and the Whale Optimization Algorithm, the MLP model is created. A total of three MLP models are developed and assessed during this stage. There are one, two, and three hidden layers in each of the models. Resampling techniques from Python's scikit-learn and keras packages are utilised for training and testing.

5 Experimental results analysis

Keras, a powerful Python deep learning tool, built and assessed the deep learning models. Scikit-learn, a Python library, uses stratified k-fold cross-validation to test the models. This is why we use resampling to check the accuracy of our models. As a result of this strategy, the data is divided into k parts, the model is trained on k-1 parts with k = 10, and the remaining data is used as test data to assess how well the model performed. We chose to repeat this technique ten times, and the mean values of all the models developed were then used to measure the accuracy of the predictions.

5.1 Performance measures

Accuracy, Precision, Recall, Kappa, and F-score are common machine learning assessment metrics used to evaluate prediction performance. Performance is assessed based on the correctness of the tests. In order to obtain the mean and standard deviation of the metrics over all models, we utilize a k-fold cross-validation score. When analyzing a test data set, testing accuracy refers to the percentage of correct predictions that were made. Based on the confusion matrix in Table 2, precision and recall are used to determine the F-score, as demonstrated below (Abdelaziz & Mahmoud, 2021):

$$F \text{ score } = \frac{(\text{ Precision } * \text{ Recall })}{(\text{ Precision } + \text{ Recall })} \qquad (12)$$

$$\text{Recall } = \frac{\text{True Positive}}{(\text{ Total Actual Positive })} \qquad (13)$$

$$\text{Precision } = \frac{\text{True Positive}}{(\text{ Total Predicted Positive })} \qquad (14)$$

Table 2 Confusion matrix

	Predicted values	
	Negative	Positive
Actual values		
Negative	True negative (TN)	False positive (FP)
Positive	False negative (FN)	True positive (TP)

Accuracy and Kappa measure are also computed as follows:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total number of objects}} \tag{15}$$

$$\kappa appa = \frac{\text{accuracy} - \text{random accuracy}}{1 - \text{random accuracy}} \tag{16}$$

In order to determine "random accuracy," The confusion matrix in Table 2 tells us that a randomly selected label from the dataset has a probability of being positive with p_1 and a probability of being negative with $(1-p_1)$, where

$$p_1 = \frac{True\ Positive + False\ Negative}{TP + FP + TN + FN} \tag{17}$$

A positive label is produced with a probability of p_2 and an incorrect label will be generated with probability $(1-p_2)$, according to our classifier, where

$$p_2 = \frac{TP + FP}{TP + FP + TN + FN} \tag{18}$$

To define random accuracy, we must assume that the labels produced by these two processes are independent of one another as:

$$\text{Random accuracy} = p_1 p_2 + (1 - p_1)(1 - p_2) \tag{19}$$

Negative firms are the ones that are in financial difficulty, while financially sound firms are used in this study. Where the confusion matrix defines the true positives, true negatives, false positives, and true negatives:

5.2 Performance analysis

A brief comparative classification outcomes analysis of the AWOA-DL technique with the existing one on the Australian Credit dataset takes place in Table 3 and Figures 7 and 8. The results are inspected in terms of different evaluation parameters. Concerning precision, the LR model has exhibited an ineffectual outcome with a minimal precision of 0.6579, whereas the RBF Network model has obtained a moderate precision of 0.8631. At the same time, the DNN and TLBO-DL models have tried to portray reasonable precision of 0.9620 and 0.9243. However, the AWOA-DL technique has resulted in superior performance with a higher precision of 0.9765. In terms of recall, the RBF Network approach has showcased poor results with the least recall of 0.8153, whereas the LR technique has reached a moderate recall of 0.8523. Concurrently, the DNN and TLBO-DL manners have tried to depict good recall

Table 3 Result analysis of AWOA-DL with existing techniques on Australian dataset

Methods	Precision	Recall	Accuracy	F-score	Kappa
AWOA-DL	0.9765	0.9482	0.9689	0.9521	0.9364
TLBO-DL	0.9620	0.9029	0.9405	0.9315	0.8791
DNN	0.9243	0.8762	0.9134	0.8996	0.8230
LR	0.6579	0.8523	0.7971	0.7426	0.7024
RBF Network	0.8631	0.8153	0.8521	0.8386	0.5797

AUSTRALIAN DATASET

■ AWOA-DL ■ TLBO-DL ■ DNN ■ LR ■ RBF Network

Fig. 7 Comparative analysis of AWOA-DL Technique-I on Australian Dataset

AUSTRALIAN DATASET

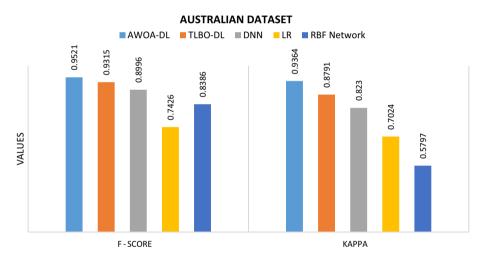■ AWOA-DL ■ TLBO-DL ■ DNN ■ LR ■ RBF Network

Fig. 8 Comparative analysis of AWOA-DL technique-II on Australian Dataset

of 0.8762 and 0.9029. But, the AWOA-DL methodology has resulted in higher performance with a maximum recall of 0.9482. Relating to accuracy, the LR approach has demonstrated worse outcomes with the lower accuracy of 0.7971, whereas the RBF Network technique has attained a moderate accuracy of 0.8521. Besides, the DNN and TLBO-DL techniques have tried to portray reasonable accuracy of 0.9134 and 0.9405. However, the AWOA-DL algorithm has the best performance with maximal accuracy of 0.9689.

Concerning F-score, the LR manner has showcased an ineffectual result with the reduced F-score of 0.7426, whereas the RBF Network manner has gained a moderate F-score of 0.8386. Followed by, the DNN and TLBO-DL methods have tried to portray a reasonable F-score of 0.8996 and 0.9315. Finally, the AWOA-DL technique has resulted in higher performance with an improved F-score of 0.9521. Finally, concerning kappa, the RBF Network

algorithm has outperformed ineffectual outcomes with the lesser kappa of 0.5797, whereas the LR model has obtained a moderate kappa of 0.7024. At the same time, the DNN and TLBO-DL approaches have tried to reveal reasonable kappa of 0.8230 and 0.8791. Nevertheless, the AWOA-DL methodology has resulted in a maximum performance with an increased kappa of 0.9364.

To see how well the AWOA-DL method compares to our current one on the Analcat dataset, see Table 4 and Figs. 9 and 10. The outcomes are evaluated based on various calculating parameters. LR models have a moderate precision of 0.9200, while RBF Network models provide ineffective results with a minimal precision of 0.8000. Parallel to each other, DNN and TLBO-DL models have attempted to convey a reasonable degree of accuracy, with precision values between 0.96 and 1.0000. The AWOA-DL approach, on the other hand, has shown better results, with a precision of 1.0000.

Regarding recall, the RBF Network technique has shown poor results with a minimum recall of 0.7142, while the LR method has obtained an average recall of 0.8518. Both TLBO-DL and DNN attempted to demonstrate a fair recall of 0.8500 and 0.8518 simultaneously. The AWOA-DL approach, on the other hand, has a high recall of 0.9675.

The RBF network approach has a low accuracy of 0.7400, while the LR method has a middling accuracy of 0.8800. Also, DNN and TLBO-DL tried to portray 0.9000 and 0.9600

Table 4 Result analysis of AWOA-DL with existing techniques on Analcat dataset

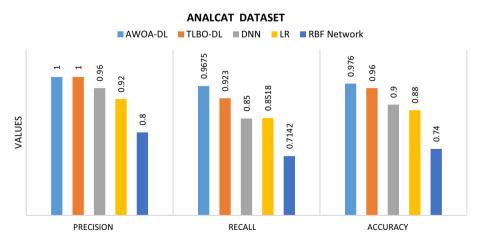| Methods	Precision	Recall	Accuracy	F-score	Kappa
AWOA-DL	1.0000	0.9675	0.9760	0.9763	0.9472
TLBO-DL	1.0000	0.9230	0.9600	0.9600	0.9201
DNN	0.9600	0.8500	0.9000	0.9056	0.8000
LR	0.9200	0.8518	0.8800	0.8846	0.7600
RBF network	0.8000	0.7142	0.7400	0.7547	0.4800

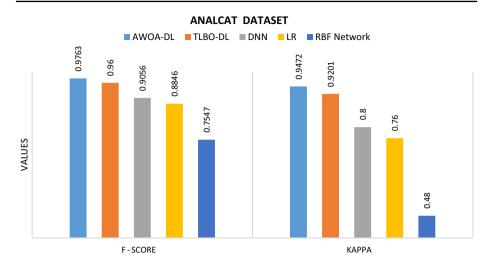**Fig. 9** Comparative analysis of AWOA-DL Technique-I on Analcat Dataset

Fig. 10 Comparative analysis of AWOA-DL technique-II on Analcat Dataset

accuracy. However, the AWOA-DL technique outperforms with 0.9760 accuracies. For example, the LR methods yielded ineffective results with an F-score of 0.7426, whereas the RBF Network methods yielded 0.8386. DNN and TLBO-DL tried to portray an F-score of 0.8996 and 0.9315. Finally, the AWOA-DL approach has a high F-score of 0.9763. Eventually, the RBF Network methods exceeded ineffective outcomes with a kappa of 0.4800 versus 0.7600 for the LR model. At the same time, DNN and TLBO-DL approaches tried to disclose kappa of 0.8000 and 0.9201. The proposed model achieved maximum performance with a kappa of 0.9472.

Table 5 and Figs. 11 and 12 quickly compare the AWOA-DL method's classification results with those from the current one on the Polish dataset. Different computation settings are examined to see how the results change. While the RBF Network models have shown ineffective results, the LR and DNN models have achieved moderate precision of 0.8010 and 0.8632, respectively, with low accuracy of 0.7135. Due to their reliance on deep learning techniques, for the most part, the TLBO-DL models have attempted to show an acceptable level of accuracy (0.95412). The AWOA-DL approach has a higher accuracy of 0.9823 due to the tuning and adaptation phases of hyperparameters, which primarily improve the phase of prediction. The RBF Network approach has shown low recall outcomes, with a minimum recall of 0.7022 compared to the LR method's 0.8010, which was observed in

Table 5 Result analysis of AWOA-DL with existing techniques on Polish credit dataset

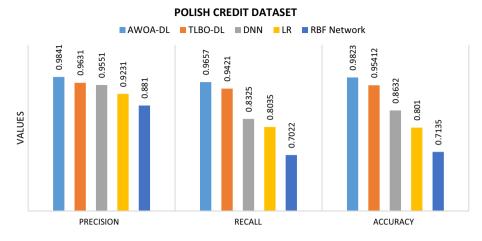| Methods	Precision	Recall	Accuracy	F-score	Kappa
AWOA-DL	0.9841	0.9657	0.9823	0.9887	0.9318
TLBO-DL	0.9631	0.9421	0.95412	0.9553	0.9113
DNN	0.9551	0.8325	0.8632	0.9123	0.8234
LR	0.9231	0.8035	0.8010	0.8744	0.7768
RBF network	0.8810	0.7022	0.7135	0.7643	0.6691

Fig. 11 Comparative analysis of AWOA-DL technique on Polish Credit Dataset
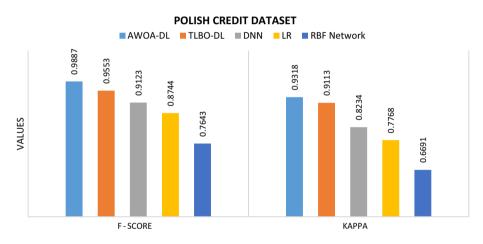

Fig. 12 Comparative analysis of AWOA-DL technique-II on Polish Credit Dataset

several datasets due to the propagation error utilized. The highest recall for the AWOA-DL technique is 0.9657, indicating excellent performance.

DNN and TLBO-DL try to portray a realistic F-score of 0.9123and 0.9553. Finally, the AWOA-DL approach enhanced performance by 0.9887. For kappa, the AWOA-DL technique resulted in maximum performance of 0.9318 due to the use of the MLP prediction model the proved the superiority over many DNN models in classification results.

Table 6 and Figs. 13 and 14 provide a quick comparison of classification results using the AWOA-DL method and the current one on the Taiwan dataset. RBF Network models provide ineffective results with the minimal precision of 0.8000. The AWOA-DL approach, on the other hand, has shown better results, with a precision of 1.0000. The DNN and TLBO-DL approach attempted to demonstrate a good recall of 0.8500 and 0.8518 simultaneously.

In contrast, the AWOA-DL approach has a maximum recall of 0.9675, which indicates outstanding performance. Finally, an improved F-score of 0.9521 was achieved using the

Table 6 Result analysis of AWOA-DL with existing techniques on Taiwan economic dataset

Methods	Precision	Recall	Accuracy	F-score	Kappa
AWOA-DL	0.9347	0.9033	0.9034	0.9122	0.9224
TLBO-DL	0.9141	0.8924	0.9001	0.9055	0.9125
DNN	0.9025	0.8910	0.9102	0.8944	0.9132
LR	0.8934	0.8933	0.9048	0.8970	0.9090
RBF network	0.8855	0.8316	0.8233	0.8699	0.8514

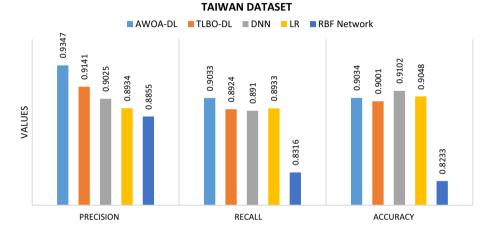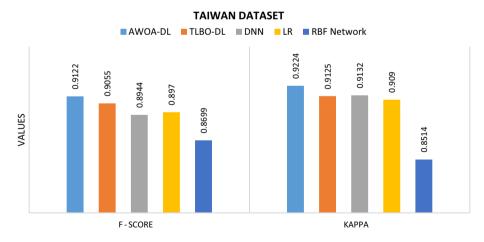**Fig. 13** Comparative analysis of AWOA-DL technique on Taiwan Dataset

Fig. 14 Comparative analysis of AWOA-DL technique-II on Taiwan Dataset

Table 7 Result analysis of AWOA-DL accuracy over four datasets

Methods	Australian dataset	Analcat dataset	Polish credit	Taiwan	Average
AWOA-DL	0.9689	0.9760	0.9823	0.9034	**0.95765**
TLBO-DL	0.9405	0.9600	0.95412	0.9001	0.93868
DNN	0.9134	0.9000	0.8632	0.9102	0.8967
LR	0.7971	0.8800	0.801	0.9048	0.845725
RBF network	0.8521	0.7400	0.7135	0.8233	0.782225

Fig. 15 Radar chart of accuracy of AWOA-DL

AWOA-DL method. According to RBF Network's kappa, it performed better with a smaller value of 0.5797 than an ineffectual result, but LR's middling value of 0.7024 beat RBF's. However, using the AWOA-DL approach yielded the best results, with a kappa of 0.9364.

5.3 Results discussion

As shown in Table 7 and Fig. 15, the average accuracy of AWOA-DL over the four datasets used in this study, when compared to other techniques, demonstrated that the proposed model outperformed the others. Several factors contribute to the proposed model's superiority; for example, using the WOA optimization algorithm for adjusting the classifier's parameter is one of them. WOA is familiar with high-performance advertisements that have the least amount of MSE for the predicted values. Secondly, unlike earlier probability-based models, MLP Neural networks make no assumptions about the underlying probability density functions or the pattern classes under consideration, which is a significant advantage. Additionally, the MLP demonstrated high accuracy when dealing with large datasets and input sizes.

As a detailed evaluation of the results, we conducted the Friedman rank-sum test, shown in Table 8.

As shown in Table 8, the Australian dataset, the test statistic (19.04), and degrees of freedom (4) are reported for the Friedman test. The Friedman test was significant (p = 0.0007718 < 0.001), so the scores' distributions for the whole models being compared are different. This explanation for all the compared datasets.

Table 8 Friedman rank sum test

Friedman rank sum test, data: Sample

Dataset	Friedman chi-squared	Degree of freedom (DF)	p-value
Australian	19.04	4	0.0007718
Analcat	19.192	4	0.0007206
Polish credit	20	4	0.0004994
Taiwan economic	13.28	4	0.009986

6 Conclusion and future work

The ability to estimate or predict the failure of any business or organization is critical for financial institutions because bad decisions impact overall revenue. In this study, a new financial distress prediction model uses an adaptive whale optimization algorithm with deep learning (AWOA-DL) technique. The goal of the AWOA-DL approach is to determine whether a company undergoes financial distress or not. The AWOA-DL method involves a deep neural network (DNN) based predictive process and AWOA based hyperparameter tuning process. Primarily, the DNN model receives the financial data as input and predicts financial distress. In addition, the AWOA is applied to tune the MLP model's hyperparameters, thereby raising the predictive outcome. To validate the improved performance of the AWOA-MLP technique, a comprehensive simulation took place on the Australian Credit dataset, and the results pointed out the supremacy of the AWOA-DL technique. The proposed model outperformed other techniques with an average accuracy of 95.8% versus 93.8, 89.6%, 84.5, and 78% for compared models. WOA has a lot going on in terms of ease of use and accuracy, but its convergence speed is poor. In the future, the predictive outcomes can be further improved by using clustering and feature selection approaches. Moreover, using a hybrid of metaheuristic algorithms for improving the convergence rate.

References

Abdelaziz, A., & Mahmoud, A. N. (2021). Intelligent system for forecasting failure of agile projects. *Journal of Intelligent Systems and Internet of Things, 5*(1), 8–15.

Abedin, M. Z., Moon, M. H., Hassan, M. K., & Hajek, P. (2021). Deep learning-based exchange rate prediction during the COVID-19 pandemic. *Annals of Operations Research, 5*, 1–52. https://doi.org/10.1007/s10479-021-04420-6

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

Appiahene, P., Missah, Y. M., & Najim, U. (2020). Predicting bank operational efficiency using machine learning algorithm: comparative study of decision tree, random forest, and neural networks. *Advances in Fuzzy Systems, 2020*, 1–12.

Ashraf, S., Félix, E. G. S., & Serrasqueiro, Z. (2019). Do traditional financial distress prediction models predict the early warning signs of financial distress? *Journal of Risk and Financial Management, 12*(2), 55.

Bengio, Y. (2009). *Learning deep architectures for AI* (pp. 1–127). Now Publishers Inc.

Bengio, Y. (2014). Evolving culture versus local minima. In T. Kowaliw, N. Bredeche, & R. Doursat (Eds.), *Growing adaptive machines* (pp. 109–138). Springer.

Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S., & Simsek, Ö. (2020). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3520659

Chandra, D. K., Ravi, V., & Bose, I. (2009). Failure prediction of dotcom companies using hybrid intelligent techniques. *Expert Systems with Applications, 36*(3), 4830–4837.

Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Systems with Applications, 112*, 353–371.

Cleofas-Sánchez, L., Garcia, V., Marqués, A. I., & Sánchez, J. S. (2016). Financial distress prediction using the hybrid associative memory with translation. *Applied Soft Computing, 44*, 144–152.

Data archieves. Retrieved from https://pages.stern.nyu.edu/~adamodar/New_Home_Page/dataarchived.html

Desai, M., & Shah, M. (2020). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical eHealth, 4*, 1–11.

Dua, D., & Graff, C. (2017). {UCI} machine learning repository.

Erhan, D., Manzagol, P.-A., Bengio, Y., Bengio, S., & Vincent, P. (2009). The difficulty of training deep architectures and the effect of unsupervised pre-training. In *Artificial Intelligence and Statistics*, pp. 153–160.

Freund, Y., & Haussler, D. (1994). Unsupervised learning of distributions of binary vectors using two layer networks.

Gharehchopogh, F. S., & Gholizadeh, H. (2019). A comprehensive survey: Whale Optimization Algorithm and its applications. *Swarm and Evolutionary Computation, 48*, 1–24.

Ghisellini, P., & Ulgiati, S. (2020). Circular economy transition in Italy. Achievements, perspectives and constraints. *Journal of Cleaner Production, 243*, 118360.

Hall, M. A., & Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering, 15*(6), 1437–1447.

Hanbal, H., & Metawa, S. (2019). Study on reasons of failure of small and medium enterprises: Looking into Egypt case. *American Journal of Business and Operations Research*.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*(7), 1527–1554.

Huang, Y.-P., & Yen, M.-F. (2019). A new perspective of performance comparison among machine learning algorithms for financial distress prediction. *Applied Soft Computing, 83*, 105663.

Kim, S., Ku, S., Chang, W., & Song, J. W. (2020). Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques. *IEEE Access, 8*, 111660–111682.

Liu, C., & Arunkumar, N. (2019). Risk prediction and evaluation of transnational transmission of financial crisis based on complex network. *Cluster Computing, 22*(2), 4307–4313.

Maghyereh, A. I., & Awartani, B. (2014). Bank distress prediction: Empirical evidence from the Gulf Cooperation Council countries. *Research in International Business and Finance, 30*, 126–147.

Metawa, N., Pustokhina, I. V., Pustokhin, D. A., Shankar, K., & Elhoseny, M. (2021). Computational intelligence-based financial crisis prediction model using feature subset selection with optimal deep belief network. *Big Data, 9*(2), 100–115.

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*(4), 603–614.

Mobahi, H., Collobert, R., & Weston, J. (2009). Deep learning from temporal coherence in video. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 737–744.

Moula, F. E., Guotai, C., & Abedin, M. Z. (2017). Credit default prediction modeling: An application of support vector machine. *Risk Management, 19*(2), 158–187.

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

Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting, 36*(3), 1092–1113.

Samitas, A., Kampouris, E., & Kenourgios, D. (2020). Machine learning as an early warning system to predict financial crisis. *International Review of Financial Analysis, 71*, 101507.

Sankhwar, S., Gupta, D., Ramya, K. C., Rani, S. S., Shankar, K., & Lakshmanaprabu, S. K. (2020). Improved grey wolf optimization-based feature subset selection with fuzzy neural classifier for financial crisis prediction. *Soft Computing, 24*(1), 101–110.

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Progress in Brain Research, 165*, 33–56.

Shajalal, M., Hajek, P., & Abedin, M. Z. (2021). Product backorder prediction using deep neural network on imbalanced data. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2021.1901153

Shankar, K., Lakshmanaprabu, S. K., Gupta, D., Khanna, A., & de Albuquerque, V. H. C. (2020). Adaptive optimal multi key based encryption for digital image security. *Concurrency and Computation: Practice and Experience, 32*(4), e5122.

Sun, J., Li, H., Huang, Q.-H., & He, K.-Y. (2014). Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based System, 57,* 41–56.

Sun, J., Fujita, H., Chen, P., & Li, H. (2017). Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble. *Knowledge-Based Systems, 120,* 4–14.

Uthayakumar, J., Metawa, N., Shankar, K., & Lakshmanaprabhu, S. K. (2020a). Financial crisis prediction model using ant colony optimization. *International Journal of Information Management, 50,* 538–556.

Uthayakumar, J., Metawa, N., Shankar, K., & Lakshmanaprabu, S. K. (2020b). Intelligent hybrid model for financial crisis prediction using machine learning techniques. *Information Systems and e-Business Management, 18*(4), 617–645.

Weston, J., Ratle, F., Mobahi, H., & Collobert, R. (2012). Deep learning via semi-supervised embedding. In G. Montavon, G. Orr, & K. R. Müller (Eds.), *Neural networks: Tricks of the trade* (pp. 639–655). Springer.

Yang, Y., & Yang, C. (2020). Research on the application of GA improved neural network in the prediction of financial crisis. In *2020 12th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 625–629.

Yu, J., & Zhao, J. (2020). Prediction of systemic risk contagion based on a dynamic complex network model using machine learning algorithm. *Complexity, 2020,* 1–13.

Zeng, S., Li, Y., Yang, W., & Li, Y. (2020). A financial distress prediction model based on sparse algorithm and support vector machine. *Mathematical Problems in Engineering, 2020,* 1–11.

Zhou, J., Li, X., & Mitri, H. S. (2015). Comparative performance of six supervised learning methods for the development of models of hard rock pillar stability prediction. *Natural Hazards, 79*(1), 291–316.

Zikeba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications, 58,* 93–101.