

# Deep Learning Enabled Financial Crisis Prediction Model for Small-Medium Sized Industries

Kavitha Muthukumar\* and K. Hariharanath

SSN School of Management, Kalavakkam, Chennai, 603110, India

\*Corresponding Author: Kavitha Muthukumar. Email: mkavitha@ssn.edu.in

Received: 10 December 2021; Accepted: 27 February 2022

**Abstract:** Recently, data science techniques utilize artificial intelligence (AI) techniques who start and run small and medium-sized enterprises (SMEs) to take an influence and grow their businesses. For SMEs, owing to the inexistence of consistent data and other features, evaluating credit risks is difficult and costly. On the other hand, it becomes necessary to design efficient models for predicting business failures or financial crises of SMEs. Various data classification approaches for financial crisis prediction (FCP) have been presented for predicting the financial status of the organization by the use of past data. A major process involved in the design of FCP is the choice of required features for enhanced classifier outcomes. With this motivation, this paper focuses on the design of an optimal deep learning-based financial crisis prediction (ODL-FCP) model for SMEs. The proposed ODL-FCP technique incorporates two phases: Archimedes optimization algorithm based feature selection (AOA-FS) algorithm and optimal deep convolution neural network with long short term memory (CNN-LSTM) based data classification. The ODL-FCP technique involves a sailfish optimization (SFO) algorithm for the hyperparameter optimization of the CNN-LSTM method. The performance validation of the ODL-FCP technique takes place using a benchmark financial dataset and the outcomes are inspected in terms of various metrics. The experimental results highlighted that the proposed ODL-FCP technique has outperformed the other techniques.

**Keywords:** Small medium-sized enterprises; deep learning; FCP; financial sector; prediction; metaheuristics; sailfish optimization

## 1 Introduction

In recent times, there has been an increase in the financial crisis of companies all over the world, they are giving considerable attention to the field of financial crisis prediction (FCP) [1]. For a financial institution/company, it is highly needed to develop a consistent and early predictive method to predict the possible risks of the company status of earlier financial risk. Commonly FCP produces a binary classification method that has been resolved efficiently [2]. The outcomes from the classification method are classified as follows: failure and non-failure status of a company [3]. So far, several classification methods have



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

been designed by a large number of domain knowledge for FCP. In general, the proposed method is separated into artificial intelligence (AI) /statistical methodologies.

The precision of the FCP plays a significant role to define the financial firms' profitability and productivity [4]. For instance, a small positive adjustment from the accuracy levels of a promising customer with default credit would minimize a great future loss of organizations [5]. FCP is considered as a data classification problem, that represents the user as "default" or user is represented as "non-default" once they return the loan. Several studies have been conducted on the FCP classification, starting at the beginning of the year 1960's. In recent times, conventional approaches applied numerical function to forecast financial crisis that differentiates financial institutions from weaker and stronger ones [6]. During 1990's, the focus has moved towards machine learning (ML) and artificial intelligence (AI) based expert systems such as Support Vector Machines (SVM) and Neural Network (NN). Lately, AI methods are adapted to improve the traditional classification methods. But, the existence of various characteristics in the higher-dimension financial data results in various problems such as low interoperability, overfitting, and high computational complexity [7]. The convenient method to resolve this problem is decreasing the available amount of features with feature selection (FS) methods [8].

The FS method is one of the vital and effective pre-processing phases in Data Mining (DM). It is accountable to extract the redundant and unwanted features from original information [9]. Furthermore, it is employed to extract high possible data through minimal feature subset and potential characteristics such as computation time, noise elimination, minimizing of impure feature, and reduced cost that is crucial to implement an approximation method [10]. Moreover, FS is applied to process the feature subset under the applications of fixed value instead of utilizing elected features. The key challenge in this method is detecting optimum features from available features called NP-hard problems [11]. Various methods are employed to identify partial solutions using shorter time intervals. Certain ML techniques such as gray wolf optimizer (GWO), particle swarm optimization (PSO), and ant colony optimization (ACO) are employed in choosing crucial features, however, such methods aren't relevant in the business application, mainly in FCP.

This paper presents an optimal deep learning based FCP (ODL-FCP) model for SMEs. The major aim of the ODL-FCP technique is to determine the financial status of SMEs. The proposed ODL-FCP technique contains the design of the Archimedes optimization algorithm based feature selection (AOA-FS) algorithm to derive optimal feature subset. In addition, the sailfish optimization (SFO) algorithm with a convolution neural network with long short term memory (CNN-LSTM) is utilized for data classification. To showcase the enhanced classification efficiency of the ODL-FCP technique, a wide range of simulations were carried out against benchmark financial datasets and the outcomes are examined concerning various metrics.

## 2 Prior Works on FCP

Uthayakumar et al. [12] presented an ant colony optimization (ACO) based FCP method that integrates 2 phases: ant colony optimization based feature selection (ACO-FS) and ant colony optimization based data classification (ACO-DC) algorithms. The presented method is confirmed by a group of 5 standard datasets including quantitative and qualitative. For the FS model, the presented ACO-FS technique is compared to 3 current FCP techniques algorithms. Yan et al. [13] developed a new method of DL prediction, according to that, create a DL hybrid predictive method for stock markets—complementary ensemble empirical mode decomposition principal component analysis long term short memory (CEEMD-PCA-LSTM). During this method, CEEMD, as an order smoothing and decomposition model, capable of decomposing the trends/fluctuations of distinct scales of time sequence gradually, generates a sequence of intrinsic mode function (IMF) using distinct characteristic scale. Next, a higher-level abstract feature is individually fed into the LSTM network for predicting the final price of the trading for all the components.

Yang [14] presented a method-based DL architecture for predicting the financial indicator value. The presented method applies the LSTM method as a standard predictive method. This architecture estimates the economic condition to define the number of financial failures to the organization and warns of any fluctuation in the forecasted 3 indicator values. Perboli et al. [15] aimed at mid-and long-term bankruptcy prediction (up to sixty months) targeting smaller/medium enterprises. Samitas et al. [16] examined on “Early Warning System” (EWS) by examining potential contagion risk, according to structured financial network. The presented method improves typical crisis predictive model performances. With machine learning algorithms and network analysis, they detect proof of contagion risks on the date where they witness considerable raise in centralities and correlations.

Uthayakumar et al. [17] introduced a cluster-based classification method that includes: fitness-scaling chaotic genetic ant colony algorithm (FSCGACA) based classification method and enhanced K-means clustering. Initially, an enhanced K-means method is developed to remove the incorrectly clustered information. Next, a rule-based method is chosen for fitting the provided dataset. Finally, FSCGACA is utilized to search for the best possible parameter of the rule-based method. Tyagi et al. [18] presented a smart Internet of Things (IoT) assisted FCP method with Meta-heuristic models. The presented FCP model contains classification, data acquisition, pre-processing, and FS. Initially, the financial information of the organization is gathered by IoT gadgets like laptops, smartphones, and so on. Then, the quantum artificial butterfly optimization (QABO) method for FS is employed to select optimum sets of features.

### 3 The Proposed ODL-FCP Model

In this paper, an efficient ODL-FCP technique has been presented for the identification of the financial crisis of SMEs. The proposed ODL-FCP technique encompasses major sub-processes namely pre-processing, arithmetic optimization algorithm (AOA) based selection of features, CNN-LSTM based classification, and SFO based hyperparameter tuning. The utilization of AOA for the optimal selection of features and SFO for the hyperparameter optimization process aid to accomplish improved classification performance. Fig. 1 demonstrates the overall block diagram of the proposed ODL-FCP technique.

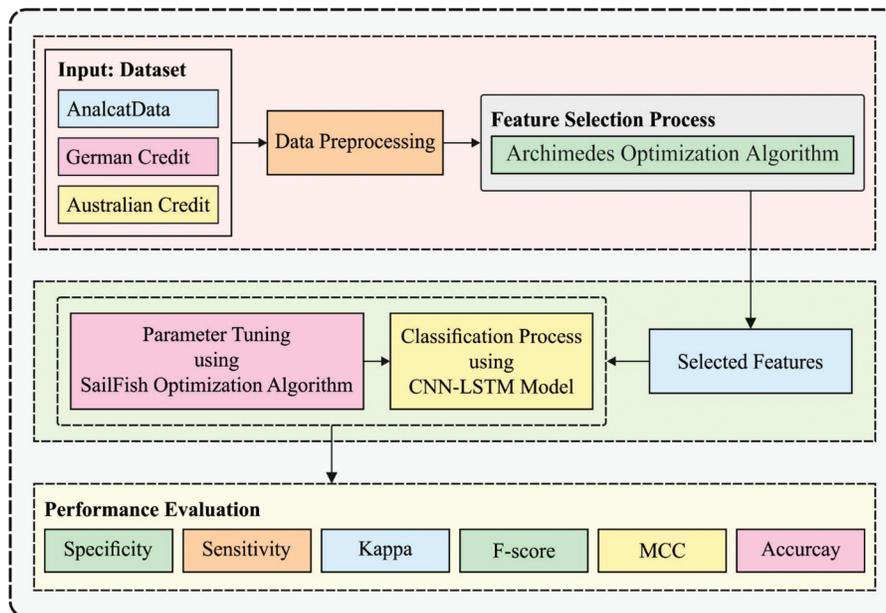


Figure 1: Overall process of proposed framework

### 3.1 Pre-processing

The financial data has extremely difficult and is collected of fundamental signals with many distinct features. But determining the transformation performance from cellular network confidently is help for enhancing networks data forecast techniques. For avoiding load packets with superior numeric values from the network in controlling individuals with lesser numeric values, the data is scaled; it also improves the modeling speed of the technique but continues optimum accuracy. A min-max technique is utilized for transforming the data to value amongst [0,1]; scaling the data is used from increasing the model to forecast network traffics. The two important benefits of scaling are for avoiding samples of higher numeric ranges controlling individuals with minimal numeric ranges and for preventing numerical problems under the forecast. The transformation has realized as follows:

$$z_n = \frac{x - x_{min}}{x_{max} - x_{min}} (New_{max_x} - New_{min_x}) + New_{min_x}, \quad (1)$$

where  $x_{min}$  refers to the minimum of data and  $x_{max}$  signifies the maximal of data.  $New_{min_x}$  defines the minimal number 0, and  $New_{max_x}$  implies the maximal number 1.

### 3.2 Design of AOA-FS Technique

Once the financial data are pre-processed, the appropriate selection of features was carried out using the AOA-FS technique. During this effort, the AOA technique [19] was utilized for solving the presented optimized issue together with NR mathematical technique. The AOA approach is a meta-heuristic technique employed for solving many mathematical optimized issues and is verified their capability for fetching towards a global solution from a short time. The AOA essential condition hinges on Archimedes' rule of buoyancy. The AOA allows several phases determining a near-global solution, and these phases are demonstrated as:

**Phase 1** 'Initialized': In this step, the populations containing the immersed object (solution) are considered by its volume, density, and acceleration. All solutions are initialized with an arbitrary place from the fluid as offered in Eq. (2), afterward the fitness value to all solutions is estimated.

$$O_i = lb_i + rand(0, 1) \times (ub_i - lb_i), \forall i \in \{1, 2, 3, \dots, N\} \quad (2)$$

$$Den_i = rand(0, 1) \quad (3)$$

$$Vol_i = rand(0, 1) \quad (4)$$

$$ACC_i = lb_i + rand(0, 1) \times (ub_i - lb_i), \forall i \in \{1, 2, 3, \dots, N\} \quad (5)$$

where  $O_i$  refer to the  $i^{th}$  solution from the populations and  $N$  implies the population sizes.  $ub_i$  and  $lb_i$  denotes the upper as well as lower bounds of  $i^{th}$  solution correspondingly.  $Den_i$ ,  $Vol_i$ , and  $ACC_i$  signifies the density, volume, and acceleration of  $i^{th}$  solutions.  $rand(0, 1)$  indicates the arbitrary scalar containing a value amongst zero and one.

**Phase 2** 'Upgrade density and volume': During this phase, the density and volumes of all the solutions were upgraded utilizing the subsequent equations:

$$Den_i^{(t+1)} = Den_i^{(t)} + rand(0, 1) \times (Den_{best} - Den_i^{(t)}) \quad (6)$$

$$Vol_i^{(t+1)} = Vol_i^{(t)} + rand(0, 1) \times (Vol_{best} - Vol_i^{(t)}) \quad (7)$$

where  $Den_i^{(t)}$ , and  $Vol_i^{(t)}$  represents the density as well as the volume of  $i^{th}$  solution at  $t^{th}$  iteration.  $Den_{best}$ , and  $Vol_{best}$  stands for the optimum densities as well as volumes of optimum solutions containing the better fitness value.

**Phase 3** ‘Transfer operator and density factor’: During this phase, the collision amongst solutions is still in its equilibrium state. The mathematical process of the collision was demonstrated as:

$$TF = \exp \left\{ \frac{t - t_{max}}{t_{max}} \right\} \quad (8)$$

where  $TF$  refers to the transfer operators able of transmitting the search process under the exploration to exploitation phase.  $t_{max}$  refers to the maximal number of iterations [20]. Also, a reducing density factor (d) is able of supporting the AOA to find near-global solutions.

$$d^{t+1} = \exp \left\{ \frac{t - t_{max}}{t_{max}} \right\} - \left( \frac{t}{t_{max}} \right) \quad (9)$$

**Phase 4** ‘Exploration’: During this phase, the collision amongst solution occurs. Therefore, when  $TF \leq 0.5$ , a random material ( $mr$ ) has chosen in which the acceleration of solution was upgraded as:

$$ACC_i^{(t+1)} = \frac{Den_{mr} + Vol_{mr} \times ACC_{mr}}{Den_i^{(t+1)} \times Vol_i^{(t+1)}} \quad (10)$$

where  $n_{mr}$ ,  $Vol_{mr}$ , and  $ACC_{mr}$  indicate the densities, volumes, and accelerations of random material.

**Phase 5** ‘Exploitation’: During this phase, no collision amongst solution occurs. Therefore, when  $TF \geq 0.5$ , the acceleration of solution was upgraded as:

$$ACC_i^{(t+1)} = \frac{Den_{best} + Vol_{best} \times ACC_{best}}{Den_i^{(t+1)} \times Vol_i^{(t+1)}} \quad (11)$$

where  $ACC_{best}$  refers to the acceleration of solution containing an optimum fitness.

**Phase 6** ‘Normalize acceleration’: The acceleration was normalization for assessing the percentage of the alteration as follows:

$$ACC_{i-norm}^{(t+1)} = g \times \frac{ACC_i^{(t+1)} - \min \{ACC\}}{\max \{ACC\} - \min \{ACC\}} + z \quad (12)$$

where  $g$ , and  $z$  implies the normalized range.  $ACC_{i-norm}^{(r+1)}$  refers to the utilized to step on the percentage which all the agents are phase.

**Phase 7** ‘Evaluation’: The fitness value of all solutions was estimated under this phase, and optimum solutions are registered, for instance, upgrade the optimum solutions ( $x_{best}$ ),  $Den_{best}$ ,  $Vol_{best}$ , and  $ACC_{best}$ .

Different from the classical AOA in which the solution was upgraded from the exploring space near the continued valued place from the BAOA, the searching space was demonstrated as  $n$  dimensional Boolean lattice. Besides, the solution was upgraded on the corner of the hypercube. Moreover, for solving this issue if electing or not, a given parameter and binary solution vectors were implemented, where 1 appears to parameter being chosen for comprising the new datasets, and 0 refers to else.

During the binary techniques, one utilizes the step vector for evaluating the possibility of altering place, the transfer function significantly influences the balance amongst exploitation as well as exploration. During the FS technique, once the size of the feature vector demonstrates  $N$ , the count of various feature combining tends that  $2^N$ , for instance, an enormous space to complete search. The presented hybrid technique was

utilized to resolve to search the feature space vigorously also create the correct group of features. The FS falls in multi-objective tasks as it requires fulfilling many determinations to take better solutions that minimize the subset of FS and at the same time, maximize the accuracy of outcome for providing classifications.

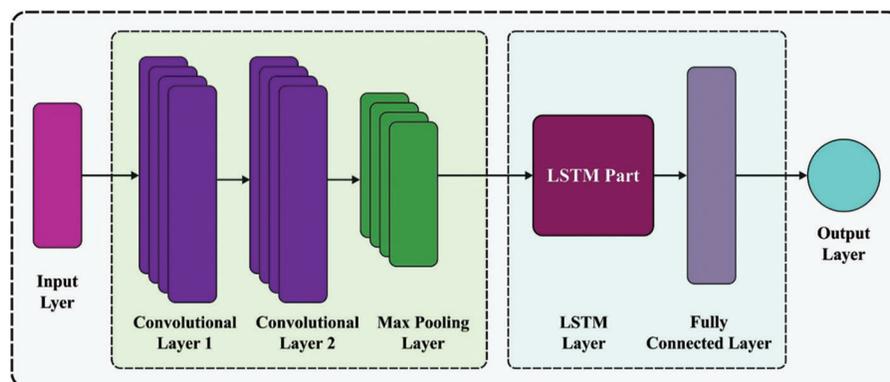
According to the aforementioned, the FF for determining solution under this state produced for attaining a balance amongst the 2 objectives as:

$$fitness = \alpha \Delta_R(D) + \beta \frac{|Y|}{|T|} \quad (13)$$

$\Delta_R(D)$  refers to the classifier error rate.  $|Y|$  represents the size of subsets which the approach elects and  $|T|$  total amount of features comprised from the current datasets.  $\alpha$  refers to the parameter  $\in [0, 1]$  linked as the weight of error rate of classifiers correspondingly also  $\beta = 1 - \alpha$  stands for the significance of decreased features. The classifier efficiency was allowed a vital weight rather than the amount of chosen feature. If the evaluation function is only regarded as classifier accuracy, the effect is neglected of solution that can comprise the same accuracy but has minimum chosen feature that serves as a fundamental factor to decrease the dimensional issue.

### 3.3 Design of CNN-LSTM Classifier

During the classification process, the CNN-LSTM model gets executed to identify the financial status of the SMEs. The CNN layer extracts the data patterns automatically. The order of features can be learned one more time in the LSTM layer. The presented method constantly tunes hyperparameters based on the outcomes from learning LSTM and CNN. CNN layer extracts parameters that are significant for classification. Especially, this is validated by the class activation maps. Also, the pooling layer reduces the spatial size of feature vectors, reduces the number of variables and computation difficulty of the NN. They could automatically alter hyper parameters like several filters, filter size, and several layers. Eq. (14) signifies the process of  $l$  convolution for feature extraction in an SQL query. The kernel  $K_{i,j}^l$  represent distinct weights in every region for extracting significant regions of the feature map. Moreover, the relation among nearby features is removed using the product operations. They utilize the  $B_i^l$  bias matrix to alter the weight in NN operation. They implement product operations on the amount of feature maps  $m_1^{l-1}$  and passes  $y_i^l$  to the following convolution layers. Fig. 2 illustrates the framework of the CNN-LSTM technique [21].



**Figure 2:** Structure of CNN-LSTM model

To generate a non-linear decision boundary,  $f(z)$  in Eq. (15) represents an activation function like ReLU utilized in layer  $l$ . Also, it can be multiplied with coefficients  $g_i$ . They implement feature extracting to database access controls via different layers of convolutional operation.

$$x_i^l = B_i^l + \sum_{j=1}^{m_1^{l-1}} K_{ij}^l * X_j^{l-1} \quad (14)$$

$$Y_i^l = g_i f(y_i^{l-1}), f(z) = \begin{cases} z & \text{if } z \geq 0 \\ 0 & \text{if } z < 0 \end{cases} \quad (15)$$

The pooling layer is utilized for improving the classifier accuracy of the access control scheme and minimalizing the computational costs. Eq. (16) denotes the pooling layer operation. The pooling layer enables to minimize over-fitting and effectively extracts features.  $T$  denotes a stride which represents how close the pooling region will be moving to define the coverage of the pooling region.  $R$  indicates the size of the pooling region and it should be small when compared to the output  $y$  of CNN layers. When the size is large than  $y$ , the SQL query feature data might be lost.

$$p_{ij}^l = \max_{r \in R} Y_{i \times T + r, j}^{l-1} \quad (16)$$

LSTM learns temporal data according to the feature extracted from CNN. Eq. (17) indicates the 3 gates state which completes the LSTM process that controls the sequential data of a database query as a constant value among zero and one. Every cell contains forget input and output gates. Eq. (17) is denoted as the resultant value of  $i, f$ , and  $o$  for all the gates. Additionally, for storing long-term data, the hidden state  $h_r$  of the LSTM cell. Eq. (19) is the hidden state of LSTM. Lastly, Eq. (19) displays the cell state to transfer states under the present cell to the following cell in the LSTM. They utilize activation functions  $\sigma$  like hyperbolic and sigmoid tangents for generating non-linear decision boundaries.

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \text{sigmoid} \\ \text{sigmoid} \\ \text{sigmoid} \\ \text{tanh} \end{pmatrix} w^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix} + \begin{pmatrix} b_i \\ b_f \\ b_o \\ b_c \end{pmatrix} \quad (17)$$

$$c_r = f_t^o c_{t-1} + i_t^o g \quad (18)$$

$$h_t = o_t^o \sigma(c_t) \quad (19)$$

Eq. (20) illustrates the FC layer operation. The output of the FC layer is categorized as zero/one by softmax. Eq. (21) evaluates the role classification possibility.  $C$  indicates the role class,  $L$  represents the final layer index, and  $N_c$  denotes the overall number of roles. The softmax layer categorizes TPCE transactions into 11 classes.

$$d_i^l = \sum_j \sigma(W_{ji}^{l-1} (h_i^{l-1}) + b_i^{l-1}) \quad (20)$$

$$P(c|d) = \operatorname{argmax}_{c \in C} \frac{\exp(d^{L-1} w^L)}{\sum_{k=1}^{N_c} \exp(d^{L-1} w_k)} \quad (21)$$

### 3.4 Hyper Parameter Tuning Using SFO Algorithm

For resolving the limitation of the CNN-LSTM model of trapping into local optima problem at the time of learning and training processes, the SFO algorithm is utilized for optimizing and adjusting the parameters involved in the CNN-LSTM Model and determining the optimum initial weight of the network.

The SFO is a new nature simulated meta-heuristic technique that is demonstrated then a set of hunting sailfish. It depicts competitive efficiency related to famous meta-heuristic techniques. During the SFO technique, it can be considered that sailfish is a candidate solution and which places of sailfish under the search space signify the variables of issue. The place of  $i^{th}$  sailfish from  $k^{th}$  search round was represented as  $SF_{i,k}$ , and their equivalent fitness was estimated by  $f(SF_{i,k})$ . The sardines are other important participants from the SFO technique. The place of  $i^{th}$  sardine has been demonstrated as  $S_i$ , and their equivalent fitness was estimated by  $f(S_i)$ . During the SFO technique, the sailfish that take an optimum place was chosen as elite sailfishes that affect the maneuverability and acceleration of sardines under attack. Also, the place of an injured sardine from all rounds is chosen as a better place to collaborative hunt with sailfish. This process purposes for preventing earlier discarded solutions from being chosen again. An elite sailfish as well as injured sardines were represented as  $Y_{elitesf}^i$  and  $Y_{injureds}^i$  correspondingly, under the  $i^{th}$  iteration [22]. During the hunting, sailfish's attack alternation approach was frequently utilized for improving the success of the hunt. A novel place of sailfish  $Y_{newsf}^i$  was upgraded dependent upon subsequent:

$$Y_{newsf}^i = Y_{elitesf}^i - \lambda_i \times \left( random(0, 1) \times \left( \frac{Y_{elitesf}^i - Y_{injureds}^i}{2} \right) - Y_{currentsf}^i \right), \quad (22)$$

where  $Y_{currentsf}^i$  refers to the present place of sailfish and  $random(0, 1)$  implies the arbitrary number ranging from zero to one.

The variable  $\lambda_i$  signifies the coefficient from  $i^{th}$  iteration and their value was were resulting as:

$$\lambda_i = 2 \times rand(0, 1) \times SD - SD, \quad (23)$$

where  $SD$  stands for the sardine density that implies the number of sardines from all the iterations.

The variable  $SD$  is resultant as:

$$SD = 1 - \left( \frac{N_{sf}}{N_{sf} + N_s} \right), \quad (24)$$

where  $N_{sf}$  and  $N_s$  refer to the number of sailfish and sardines correspondingly.

Initially the hunted, sailfishes are energetic, and sardines aren't tired/injured. The sardines are escape quickly. But, with continued hunting, the control of sailfish attacks is slowly reduced. In the meantime, the sardines are come to be tired, and their awareness of the place of the sailfish is also reduced. Thus, the outcome, the sardines are hunted. According to the algorithmic procedure, a novel place of sardine  $Y_{news}^i$  has upgraded dependent upon the subsequent:

$$Y_{news}^i = random(0, 1) \times \left( Y_{elitesf}^i - Y_{olds}^i + ATP \right), \quad (25)$$

where  $Y_{olds}^i$  refers to the old place of sardine and  $random(0, 1)$  implies the arbitrary number range in zero and one.  $ATP$  stands for the sailfish attack power.

The variable  $ATP$  has estimated as:

$$ATP = B \times (1 - (2 \times Itr \times \varepsilon)) \quad (26)$$

where  $B$  and  $\varepsilon$  denote the coefficients which are utilized for reducing the attack power linearly in  $B$  to 0 and  $Itr$  refers to the count of iterations. Once  $ATP$  has maximum, for instance, superior to 0.5, the place of every sardine was upgraded. Conversely, only  $\alpha$  sardines with  $\beta$  variables upgrade its places. The number of sardines that upgrade their place was defined as:

$$\alpha = N_S \times ATP, \quad (27)$$

where  $N_S$  refers to the number of sardines from all iterations. The number of variables of the sardines which upgrade their places are attained as:

$$\beta = d_i \times ATP, \quad (28)$$

where  $d_i$  signifies the number of variables from  $i^{th}$  iteration. If the sardine was hunted, their fitness is superior to sailfish. In these cases, the place of sailfish  $Y_{SF}^i$  has upgraded with the latest place of hunted sardine  $Y_S^i$  for promoting the hunted of novel sardine. The equivalent formula is as follows:

$$Y_{SF}^i = Y_S^i \text{ if } f(S_i) < f(SF_i). \quad (29)$$

#### 4 Experimental Validation

The experimental results analysis of the ODL-FCP technique takes place using three benchmark datasets namely Alancatdata, German Credit, and Australian Credit datasets. The first AlancatData dataset includes 50 samples under two classes. Next, the second dataset comprises 1000 samples with two classes. The final dataset has 690 samples with two classes. [Tab. 1](#) offers the FS outcome of the AOA-FS technique. The results show that the AOA-FS technique has chosen only a minimal number of features. For instance, the AOA-FS technique has selected a total of 3, 10, and 8 features on the test AlancatData, German Credit, and Australian Credit datasets respectively.

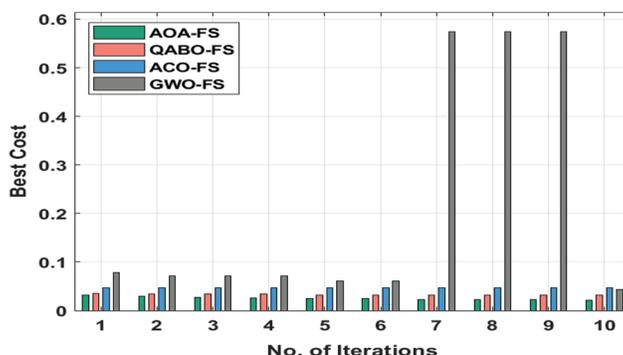
**Table 1:** Selected features of AOA-FS technique on applied dataset

Dataset	Selected features
AlancatData	1,3,5
German Credit	2,3,6,8,11,15,17,19,22,23
Australian Credit	2,3,7,9,10,12,13,14

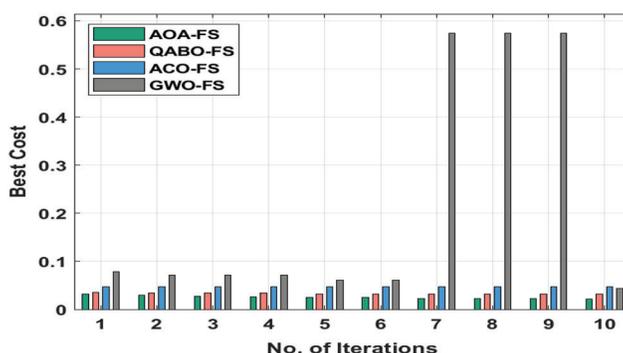
[Fig. 3](#) demonstrates the best cost (BC) analysis of the AOA-FS technique with other FS models on AlancatData Dataset. [Fig. 3](#) shows that the AOA-FS technique has resulted ineffectual FS outcomes with minimal BC. For instance, under 3 iterations, the AOA-FS technique has obtained a lower BC of 0.0269 whereas the QABO-FS, ACO-FS, and grey wolf optimization based feature selection (GWO-FS) techniques have attained higher BC of 0.0347, 0.0466, and 0.0715 respectively. At the same time, under 6 iterations, the AOA-FS approach has reached minimal BC of 0.0246 whereas the QABO-FS, ACO-FS, and GWO-FS methods have obtained superior BC of 0.0316, 0.0466, and 0.0607 correspondingly. Moreover, under 9 iterations, the AOA-FS technique has gained a lower BC of 0.0223 whereas the QABO-FS, ACO-FS, and GWO-FS techniques have attained superior BC of 0.0315, 0.0466, and 0.5743 correspondingly.

[Fig. 4](#) determines the BC analysis of the AOA-FS system with other FS manners on the German Credit Dataset. [Fig. 4](#) outperformed that the AOA-FS technique has resulted in effective FS outcome with the minimal BC. For the sample, under 3 iterations, the AOA-FS technique has obtained least BC of 0.1477 but the QABO-FS, ACO-FS, and GWO-FS techniques have achieved superior BC of 0.1532, 0.1600, and 0.1600 correspondingly. Simultaneously, under 6 iterations, the AOA-FS technique has obtained lower BC of 0.1334 whereas the QABO-FS, ACO-FS, and GWO-FS techniques have attained higher BC of 0.1467, 0.1500, and 0.1600 respectively. Also, under 9 iterations, the AOA-FS technique

has obtained minimum BC of 0.1099 while the QABO-FS, ACO-FS, and GWO-FS techniques have attained a higher BC of 0.1258, 0.1400, and 0.1600 correspondingly.



**Figure 3:** BC analysis of AOA-FS technique on AnalcatData dataset



**Figure 4:** BC analysis of AOA-FS technique on German credit dataset

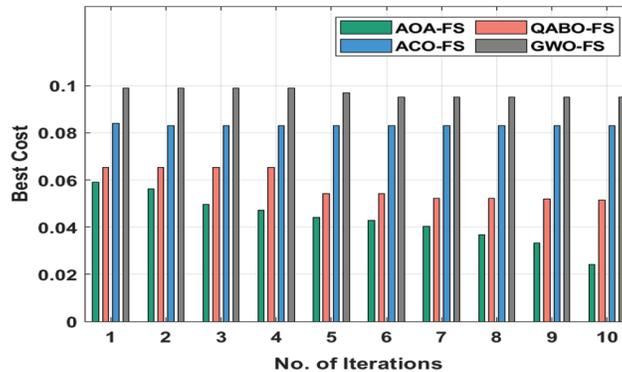
Fig. 5 reveals the BC analysis of the AOA-FS approach with other FS models on the Australian Credit Dataset. Fig. 5 revealed that the AOA-FS system has resulted in effective FS outcomes with minimal BC. For instance, under 3 iterations, the AOA-FS method has reached a minimum BC of 0.0471 whereas the QABO-FS, ACO-FS, and GWO-FS methods have gained superior BC of 0.0652, 0.0830, and 0.0990 respectively. Followed by, under 6 iterations, the AOA-FS technique has reached a lower BC of 0.0429 whereas the QABO-FS, ACO-FS, and GWO-FS manners have attained higher BC of 0.0543, 0.0830, and 0.0950 correspondingly [23]. Likewise, under 9 iterations, the AOA-FS technique has obtained a minimum BC of 0.0333 whereas the QABO-FS, ACO-FS, and GWO-FS methodologies have obtained superior BC of 0.0518, 0.0830, and 0.0950 correspondingly.

#### 4.1 Comparative Results Analysis on AnalcatData Dataset

Tab. 2 offers a brief comparative study of the ODL-FCP with other techniques on the AnalcatData dataset.

Fig. 6 offers the  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  analysis of the ODL-FCP with recent techniques on the AnalcatData dataset. The results show that the AdaBoost manner has shown poor performance with the  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  of 0.6500, 0.6700, and 0.6583 respectively. In addition, the multilayer perception (MLP) and SVM systems have gained slightly improved outcomes [24]. Followed by, the QABOLSTM, LSTM-RNN, and ACO models have resulted in reasonable values of  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$ . However,

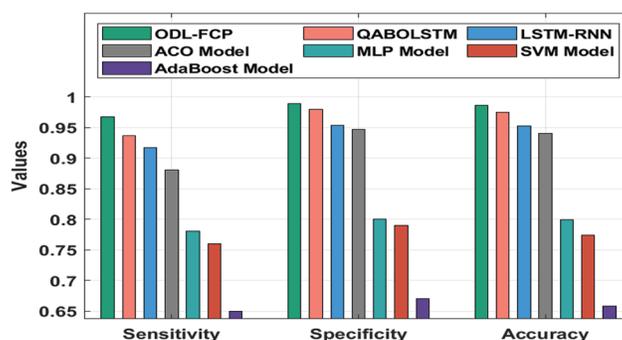
the presented ODL-FCP technique has accomplished superior performance with the maximum  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  of 0.9675, 0.9886, and 0.9862 correspondingly.



**Figure 5:** BC analysis of AOA-FS technique on Australian credit dataset

**Table 2:** Result analysis of proposed methods with existing methods on AnalcatData dataset

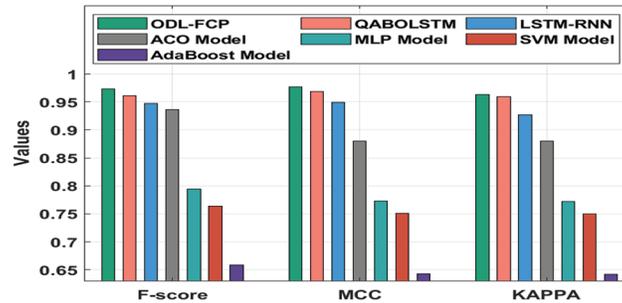
Classifiers	Sensitivity	Specificity	Accuracy	F-score	MCC	KAPPA
ODL-FCP	0.9675	0.9886	0.982	0.9735	0.9767	0.9634
QABOLSTM	0.9368	0.9760	0.9751	0.9609	0.9687	0.9593
LSTM-RNN	0.9167	0.9534	0.9523	0.9472	0.9489	0.9274
ACO Model	0.8800	0.9465	0.9400	0.9361	0.8800	0.8800
MLP Model	0.7800	0.8000	0.7987	0.7945	0.7733	0.7721
SVM Model	0.7600	0.7900	0.7743	0.7641	0.7512	0.7502
AdaBoost Model	0.6500	0.6700	0.6583	0.6582	0.6431	0.6424



**Figure 6:**  $SEN_y$ ,  $SEP_y$ , and  $ACC_y$  analysis of ODL-FCP technique on AnalcatData dataset

Fig. 7 provides the  $F_{score}$ ,  $MCC$ , and  $Kappa$  analysis of the ODL-FCP with recent approaches on the AnalcatData dataset. The outcomes exhibited that the AdaBoost manner has outperformed worse effectiveness with the  $F_{score}$ ,  $MCC$ , and  $Kappa$  of 0.6582, 0.6431, and 0.6424 respectively. Besides, the MLP and SVM techniques have obtained somewhat higher outcomes. Then, the QABOLSTM, LSTM-RNN, and ACO models have resulted in reasonable values of  $F_{score}$ ,  $MCC$ , and  $Kappa$ . But, the

presented ODL-FCP method has accomplished superior performance with the maximum  $F_{score}$ ,  $MCC$ , and  $Kappa$  by 0.9735, 0.9767, and 0.9634 correspondingly.



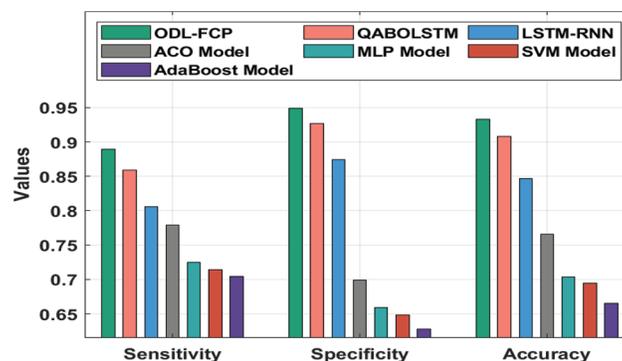
**Figure 7:**  $F_{score}$ ,  $MCC$ , and  $kappa$  analysis of ODL-FCP technique on AnalcatData dataset

#### 4.2 Comparative Results Analysis on German Credit Dataset

Tab. 3 provides a brief comparative analysis of the ODL-FCP with other approaches on the German Credit dataset. Fig. 8 suggests the  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  analysis of the ODL-FCP with existing approaches on the German Credit dataset. The outcomes depicted that the AdaBoost manner has exhibited the least efficiency with the  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  of 0.7045, 0.6274, and 0.6654 respectively. Besides, the MLP and SVM models have reached somewhat increased outcomes. Followed by, the QABOLSTM, LSTM-RNN, and ACO models have resulted in reasonable values of  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$ . Finally, the presented ODL-FCP technique has accomplished superior performance with the maximal  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  of 0.8895, 0.9487, and 0.9328 correspondingly.

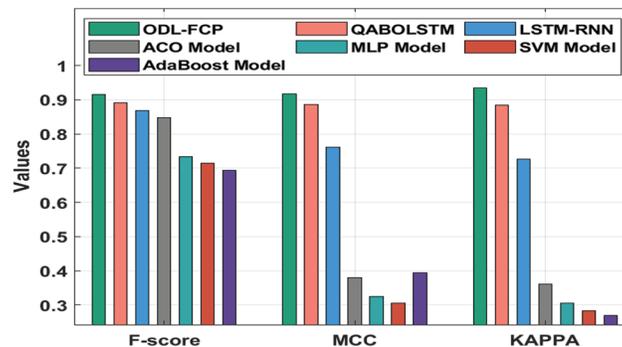
**Table 3:** Result analysis of proposed techniques with existing methods on German credit dataset

Classifiers	Sensitivity	Specificity	Accuracy	F-score	MCC	KAPPA
ODL-FCP	0.8895	0.9487	0.9328	0.9156	0.9163	0.9351
QABOLSTM	0.8589	0.9265	0.9078	0.8908	0.8863	0.8842
LSTM-RNN	0.8060	0.8743	0.8467	0.8693	0.7621	0.7261
ACO Model	0.7793	0.6987	0.7660	0.8474	0.3800	0.3613
MLP Model	0.7244	0.6593	0.7031	0.7340	0.3249	0.3054
SVM Model	0.7138	0.6483	0.6942	0.7146	0.3056	0.2840
AdaBoost Model	0.7045	0.6274	0.6654	0.6942	0.3951	0.2698



**Figure 8:**  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  analysis of ODL-FCP technique on German credit dataset

Fig. 9 gives the  $F_{score}$ ,  $MCC$ , and  $Kappa$  analysis of the ODL-FCP with recent algorithms on the German Credit dataset. The outcomes demonstrated that the AdaBoost methodology has exposed poor performance with the  $F_{score}$ ,  $MCC$ , and  $Kappa$  of 0.6942, 0.3951, and 0.2698 respectively. Along with that, the MLP and SVM techniques have obtained slightly improved outcomes [25]. Afterward, the QABOLSTM, LSTM-RNN, and ACO models have resulted in reasonable values of  $F_{score}$ ,  $MCC$ , and  $Kappa$ . Eventually, the presented ODL-FCP technique has accomplished maximum performance with the higher  $F_{score}$ ,  $MCC$ , and  $Kappa$  by 0.9156, 0.9163, and 0.9351 correspondingly.



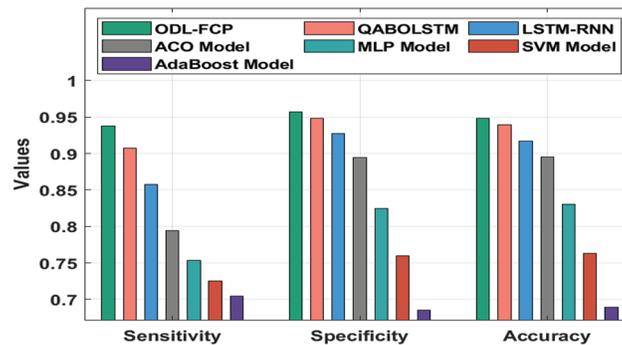
**Figure 9:**  $F_{score}$ ,  $MCC$ , and  $kappa$  analysis of ODL-FCP technique on German credit dataset

#### 4.3 Comparative Results Analysis on Australian Credit Dataset

Tab. 4 depicts a detailed comparative analysis of the ODL-FCP with other techniques on the Australian Credit dataset. Fig. 10 shows the  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  analysis of the ODL-FCP with recent approaches on the Australian Credit dataset. The outcomes displayed that the AdaBoost method has shown poor efficiency with  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  of 0.7044, 0.6847, and 0.6890 respectively [26]. Besides, the MLP and SVM techniques have gained slightly maximal outcomes. Next, the QABOLSTM, LSTM-RNN, and ACO approaches have resulted in reasonable values of  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$ . However, the presented ODL-FCP methodology has accomplished higher performance with the increased  $SEN_y$ ,  $SPE_y$ , and  $ACC_y$  of 0.9375, 0.9567, and 0.9485 correspondingly.

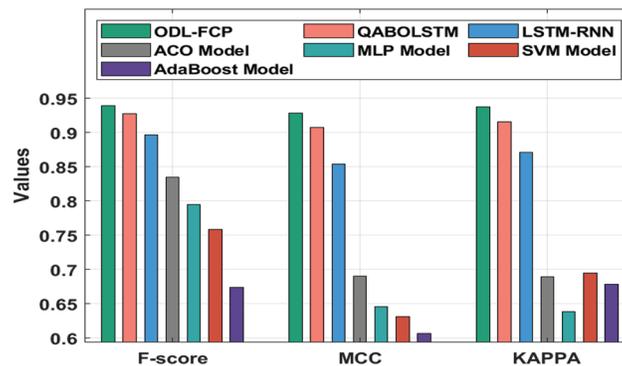
**Table 4:** Result Analysis of Proposed Methods with Existing Methods on Australian Credit Dataset

Classifiers	Sensitivity	Specificity	Accuracy	F-score	MCC	KAPPA
ODL-FCP	0.9375	0.9567	0.9485	0.9399	0.9284	0.9377
QABOLSTM	0.9071	0.9478	0.9390	0.9273	0.9080	0.9163
LSTM-RNN	0.8578	0.9274	0.9166	0.8967	0.8542	0.8709
ACO Model	0.7941	0.8942	0.8949	0.8346	0.6900	0.6893
MLP Model	0.7533	0.8247	0.8304	0.7952	0.6459	0.6387
SVM Model	0.7248	0.7600	0.7628	0.7584	0.6313	0.6948
AdaBoost Model	0.7044	0.6847	0.6890	0.6741	0.6066	0.6786



**Figure 10:**  $SEN_y$ ,  $SEP_y$ , and  $ACC_y$  analysis of ODL-FCP technique on Australian credit dataset

Fig. 11 examines the  $F_{score}$ ,  $MCC$ , and  $Kappa$  analysis of the ODL-FCP with recent techniques on the Australian Credit dataset. The outcomes exhibited that the AdaBoost approach has shown poor effective with the  $F_{score}$ ,  $MCC$ , and  $Kappa$  of 0.6741, 0.6066, and 0.6786 respectively. Similarly, the MLP and SVM systems have attained slightly improved outcomes. Followed by, the QABOLSTM, LSTM-RNN, and ACO manners have resulted in reasonable values of  $F_{score}$ ,  $MCC$ , and  $Kappa$ . Finally, the projected ODL-FCP system has accomplished superior performance with the maximum  $F_{score}$ ,  $MCC$ , and  $Kappa$  of 0.9399, 0.9284, and 0.9377 respectively.



**Figure 11:**  $F_{score}$ ,  $MCC$ , and  $kappa$  analysis of ODL-FCP technique on Australian credit dataset

From the detailed results and discussion, it can be ensured that the ODL-FCP technique can be applied as a proficient method to determine the financial crisis of SMEs.

## 5 Conclusion

In this paper, an efficient ODL-FCP technique has been presented for the identification of the financial crisis of SMEs. The proposed ODL-FCP technique encompasses major subprocesses namely pre-processing, AOA-based selection of features, CNN-LSTM based classification, and SFO based hyperparameter tuning. The utilization of AOA for the optimal selection of features and SFO for the hyperparameter optimization process aid to accomplish improved classification performance. To showcase the enhanced classification performance of the ODL-FCP technique, a wide range of simulations were carried out against benchmark financial datasets and the outcomes are examined concerning various metrics. The experimental results highlighted that the proposed ODL-FCO technique has outperformed the other techniques. As a part of

future extension, the ODL-FCP technique is extended to the design of data clustering approaches in a big data environment.

**Funding Statement:** The authors received no specific funding for this study

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] M. Ala'raj and M. F. Abbod, "Classifier's consensus system approach for credit scoring," *Knowledge-Based System*, vol. 104, pp. 89–105, 2016.
- [2] A. Martin, V. Aswathy, S. Balaji, T. M. Lakshmi and V. P. Venkatesan, "An analysis on qualitative bankruptcy prediction using fuzzy ID3 and ant colony optimization algorithm," in *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, Salem, India, pp. 416–421, 2012.
- [3] Guojun Gan, Chaoqun Ma and Jianhong Wu, "Data clustering: theory, algorithm and application," *ASA-SIAM Series on Statistics and Applied Mathematics, Philadelphia*, vol. 1, pp. 3–17, 2007.
- [4] A. Ekbal and S. Saha, "Joint model for feature selection and parameter optimization coupled with classifier ensemble in chemical mention recognition," *Knowledge Based Systems*, vol. 85, pp. 37–51, 2015.
- [5] A. F. Atiya, "Bankruptcy prediction for credit risk using neural networks: A survey and new results," *IEEE Transactions on Neural Network*, vol. 12, no. 4, pp. 929–935, 2001.
- [6] G. Ou, "Research on early warning of financial risk of real estate enterprises based on factor analysis method," *Social Scientist*, vol. 9, no. 56, pp. 56–63, 2018.
- [7] Y. Yang, "Research on information disclosure of listed companies and construction of financial risk early warning system," *Modern Business*, vol. 1, pp. 151–152, 2018.
- [8] M. Chen, "Predicting corporate financial distress based on integration of decision tree classification and logistic regression," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11261–11272, 2011.
- [9] Y. Lee and H. Teng, "Predicting the financial crisis by mahalalanobis – Taguchi system – Examples of Taiwan's electronic sector," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7469–7478, 2009.
- [10] V. Ravi and C. Pramodh, "Threshold accepting trained principal component neural network and feature subset selection: Application to bankruptcy prediction in banks," *Applied Soft Computing*, vol. 8, no. 4, pp. 1539–1548, 2008.
- [11] W. Y. Lin, Y. H. Hu and C. F. Tsai, "Machine learning in financial crisis prediction: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 42, no. 4, pp. 421–436, 2012.
- [12] J. Uthayakumar, N. Metawa, K. Shankar and S. K. Lakshmanaprabu, "Financial crisis prediction model using ant colony optimization," *International Journal of Information Management*, vol. 50, no. 5, pp. 538–556, 2020.
- [13] B. Yan and M. Aasma, "A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM," *Expert Systems with Applications*, vol. 159, no. 4, pp. 113609, 2020.
- [14] S. Yang, "A novel study on deep learning framework to predict and analyze the financial time series information," *Future Generation Computer Systems*, vol. 125, no. 12, pp. 812–819, 2021.
- [15] G. Perboli and E. Arabnezhad, "A machine learning-based DSS for mid and long-term company crisis prediction," *Expert Systems with Applications*, vol. 174, no. 4, pp. 114758, 2021.
- [16] A. Samitas, E. Kampouris and D. Kenourgios, "Machine learning as an early warning system to predict financial crisis," *International Review of Financial Analysis*, vol. 71, no. 2, pp. 101507, 2020.
- [17] J. Uthayakumar, N. Metawa, K. Shankar and S. K. Lakshmanaprabu, "Intelligent hybrid model for financial crisis prediction using machine learning techniques," *Information Systems and e-Business Management*, vol. 18, no. 4, pp. 617–645, 2020.
- [18] S. K. S. Tyagi and Q. Boyang, "An intelligent internet of things aided financial crisis prediction model in FinTech," *IEEE Internet of Things Journal*, vol. 10, pp. 1, 2021.

- [19] F. A. Hashim, K. Hussain, E. H. Houssein, M. S. Mabrouk and W. Al-Atabany, "Archimedes optimization algorithm: A new metaheuristic algorithm for solving optimization problems," *Applied Intelligence*, vol. 51, pp. 1–21, 2020.
- [20] Z. M. Ali, I. M. Diaaeldin, A. El-Rafei, H. M. Hasanien, S. H. A. Aleem *et al.*, "A novel distributed generation planning algorithm via graphically-based network reconfiguration and soft open points placement using Archimedes optimization algorithm," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1923–1941, 2021.
- [21] I. E. Livieris, E. Pintelas and P. Pintelas, "A CNN-LSTM model for gold price time-series forecasting," *Neural Computing and Applications*, vol. 32, no. 23, pp. 17351–17360, 2020.
- [22] M. Li, Y. Li, Y. Chen and Y. Xu, "Batch recommendation of experts to questions in community-based question-answering with a sailfish optimizer," *Expert Systems with Applications*, vol. 169, no. 9, pp. 114484, 2021.
- [23] L. Wensheng, W. Kuihua, F. Liang, L. Hao, W. Yanshuo *et al.*, "A region-level integrated energy load forecasting method based on CNN-LSTM model with user energy label differentiation," in *5th Int. Conf. on Power and Renewable Energy (ICPRE)*, Shanghai, China, pp. 154–159, 2020.
- [24] T. Zheng, "Scene recognition model in underground mines based on CNN-LSTM and spatial-temporal attention mechanism," in *Int. Sym. on Computer, Consumer and Control (IS3C)*, Taichung City, Taiwan, vol. 20, pp. 513–516, 2020.
- [25] S. J. Bu, H. J. Moon and S. B. Cho, "Adversarial signal augmentation for CNN-LSTM to classify impact noise in automobiles," in *IEEE Int. Conf. on Big Data and Smart Computing (BigComp)*, Jeju Island, Korea (South), vol. 10, pp. 60–64, 2021.
- [26] Y. Wang, X. Wang and X. Chang, "Sentiment analysis of consumer-generated online reviews of physical bookstores using hybrid LSTM-CNN and LDA topic model," in *Int. Conf. on Culture-Oriented Science & Technology (ICCST)*, Beijing, China, vol. 24, pp. 457–462, 2020.