Artificial intelligence assisted Internet of Things based financial crisis prediction in FinTech environment

Irina V. Pustokhina[1] · Denis A. Pustokhin[2] · Sachi Nandan Mohanty[3] · Paulo Alonso Gaona García[4] · Vicente García-Díaz[4]

Abstract

Financial Technology (FinTech) is treated as a distinctive taxonomy which majorly examines the financial technology sectors in a broader set of operations for enterprises by the use of Information Technology (IT) applications. Since the Internet of Things (IoT) is increasing tremendously, artificial intelligence (AI) assisted agile IoT is the way forward for sustainable finance. The deepness of the agile IoT has probably transformed the financial market today, and it may rapidly develop as a dominant tool in the future. The integration of AI and IoT techniques will considerably extract valued financial data and avail better services to the customers. One of the important concepts involved in FinTech is financial crisis prediction (FCP), which is a process of determining the financial status of a company. With this motivation, this paper designs a novel artificial intelligence assisted IoT based FCP (AIAIoT-FCP) model in the FinTech environment. The proposed AIAIoT-FCP model encompasses different stages such as data collection, data preprocessing, feature selection, and classification. At the primary stage, the financial data of the enterprises are collected by the use of the IoT devices such as smartphones and laptops. Besides, a chaotic Henry gas solubility optimization based feature selection (CHGSO-FS) technique is applied to select optimum features. In addition, a deep extreme learning machine (DELM) based classifier is used to determine the class

---

✉ Vicente García-Díaz
garciavicente@uniovi.es

Irina V. Pustokhina
ivpustokhina@yandex.ru

Denis A. Pustokhin
dpustokhin@yandex.ru

Sachi Nandan Mohanty
sachinandan09@gmail.com

[1] Department of Operations Management and Logistics, Graduate School of Business, HSE University, 20 Myasnitskaya ulitsa, Moscow, Russian Federation 101000

[2] Department of Logistics and Marketing, Faculty of Economics and Business, Financial University under the Government of the Russian Federation, Leningradskiy Prospekt 55, Moscow, Russian Federation 125993

[3] Department of Computer Science & Engineering, Vardhaman Engineering College (Autonomous), Hyderabad, India

[4] University of Oviedo, Oviedo, Spain

🖄 Springer

labels of the financial data. Finally, the Nesterov-accelerated Adaptive Moment Estimation (NADAM) based hyperparameter optimizer of the DELM model is involved to boost the classification performance of the DELM model. An extensive simulation analysis is carried out on the benchmark financial dataset to highlight the betterment of the AIAIoT-FCP model. The resultant values portrayed the superior performance of the AIAIoT-FCP model over the state of art techniques in a considerable way.

Keywords Fintech · Financial crisis prediction · Internet of Things · Feature selection · Enterprises · Data classification · Artificial intelligence

1 Introduction

"Financial technology" or "FinTech" refers to the utilization of information technologies to derive financial solutions. FinTech nowadays is frequently considered as an exclusively topical combination of financial service and IT. But the inter-linkage of technology and finance has a long history. FinTech is one of the popular business advancements that utilize technological innovations and share economy modules. It also addresses regulatory and privacy problems for providing novel services and products. The development of FinTech innovation has made tough competition with conventional financial services providers. This competition involves several business entities seeking enhancements on their present business modules or novel paths for investment to remain in business. Generally, companies could witness possible risks by repeatedly observing and analyzing financial data. To assist company in avoiding and disperse financial risk in an effective and timely way, the financial crisis prediction (FCP) is mainly significant in company risk management.

1.1 Role of artificial intelligence in FCP

Financial crisis prediction (FCP) is important for several financial organizations that help to lessen the forthcoming loss using estimation probable threat and evades new credit offers if the default risk exceeds the present acceptance level. This method is so called credit default classifier module which indicates client as "non-default" when he repays loan, or the user is represented as "default". Accurateness of FCP plays a significant process to regulate the financial organization success and production. For instance, a minor positive alteration in the accurateness level of probable client via default credit will decrease the future loss of a firm (Metawa et al., 2018, 2021). FCP is treated as a data classification problem (Ala'raj & Abbod, 2016). Recently, AI methods are executed to improve traditional classification methods. Although the presence of various features in the higher dimension monetary information is the intention for many issues such as overfitting, lower interoperability, and higher computation complexity. As a result of the dimension curse, that occurs on the ratio of sample and features counts. The simple technique to solve this issue is reducing the present feature count by FS methods (Li & Li, 2011).

The FS technique aim is to identify appropriate feature subsets and have important inference for challenges like (i) allowing easy access setting and upgrading methods, (ii) reformation resultant method (iii) noise reduction by removing noise feature and (iv) computation time and cost consumption required to develop suitable method. The chosen subsets of features are suitable for signifying classification functions which affect several dimensions of classifications like cost cohesive with feature, learning duration, and accurateness of the

classification technique (Jensen, 2005). The FS method is employed for diverse application areas such as ML, data mining, and pattern recognition, for minimizing the dimension of feature space and raise the prediction accurateness of a classification method (Xue et al., 2016). Depending upon estimation norm, FS methodologies are divided into wrapper, filter, and embedded based approaches (Kohavi & John, 1997). The technique of FS is evaluated as an optimization problem, with a deliberated effectiveness for every feature subsets that denoted predictive classification effectiveness of the resultant method. The goal is to attain the subset of feature space to discover an optimum or almost an optimum technique based efficiency measures. Various methods were provided to determine the suboptimum outcomes with less time consumption. Random methods like GA, ACO scatter search, and simulated annealing are generally employed to FCP for attaining highest accuracy rate.

1.2 Paper contributions

This paper designs a novel artificial intelligence assisted IoT based FCP (AIAIoT-FCP) model in the FinTech environment. The proposed AIAIoT-FCP model enables the collection of financial data through the IoT gadgets like smartphones, smartwatches, and laptops. Followed by, a chaotic Henry gas solubility optimization based feature selection (CHGSO-FS) technique is applied to select optimum features. In addition, a deep extreme learning machine (DELM) based classifier is used to determine the class labels of the financial data. At last, the Nesterov-accelerated Adaptive Moment Estimation (NADAM) based hyperparameter optimizer of the DELM model is involved to boost the classification performance of the DELM model. A comprehensive experimental validation process takes place on the benchmark financial dataset to showcase the improved performance of the AIAIoT-FCP model.

2 Literature Review

In Aydin and Cavdar (Aydin & Cavdar, 2015), a regular dataset covering the period of 1990 and 2014 belonging to the Turkish economy is used for FCP. It intends to improve an early warning scheme for FCP. To understand this purpose, MLFN would be utilized. With the use of regular information of seven key macro-economic and financial indicators of Turkish economy. Junyu (Junyu, 2020) utilized the information on credit defaults with overall instance of 1,000 instances comprising Germany credit default recordings and few fundamental private data. Xgboost, LR and RF have been employed for discovering beneficial data behindhand information. The result shows that the ML technique suited the information comparatively well, and the accurateness of Xgboost has attained around 80%.

Feixiong-Ma et al. (2020) utilized RPROP, SCM, and ANN, with 162 organization for 2 year panel financial indicator details as module instances, and the test instant determined a financial crisis earlier warning module. The concept of complete assessment integrating 2 types of NN approaches is present creatively. Uthayakumar et al. (2018) proposed a cluster based classification module, contains 2 phases: enhanced K means clustering and FSC-GACA based classification module. Initially, an enhanced K-means method is developed for eliminating the incorrectly clustered information. Next, a rule based module is chosen for implementing provided dataset. Finally, FSCGACA is utilized to search optimal parameters of rule-based module. Uthayakumar et al. (2020) proposed a novel ACO based FCP module that includes 2 stages: ACOFS and ACODC methods. The presented ACOFCP module is

authenticated by a group of 5 standard datasets that include both quantitative and qualitative. Sankhwar et al. (2020) present a new prediction architecture for FCP module by incorporating FNC and IGWO. An IGWO method is acquired by combining GWO technique and tumbling effects. The proposed IGWO based FS technique is utilized for discovering an optimum feature and FNC is employed as a classifier model.

Fernández-Arias et al. (2018) proposed a "hybridization method" that integrates ELM and Synthetic Minority Over sampling Method. Because of the imbalanced characteristics of the problems, the researcher employs an oversampling method on the information aims to enhance the classification outcomes on the minority group. Wang and Wu (Wang & Wu, 2017) proposed a novel integration module that integrated BP with PNN based on adaptive boosting method for predicting financial crisis. The BP technique is presented for modifying weight and smoothing variables of PNN. In procedure of creating BP PNN modules, the trained set is separated into research and trained instance for saving computation time. Then the trained module is considered a weaker classifier. And this weaker classifier is combined to establish strong classifiers using adaboost technique. In this Dastkhan (2021), they employed the forward looking condition value at risk (COVaR) as a market based systematic risk metric. A network depiction of asset exposure is introduced according to the value of CoVaR. Based on exposure network, 2 novel network based indices are presented.. The outcomes of the earlier warning scheme for the market return show that the indicator is depending upon both presented techniques has better capability for predicting crises in the market. Alternatively, the indicator distributed a warning signal up to 7 periods earlier the actual crisis in the market.

3 The proposed AIAIoT-FCP model

The workflow involved in the proposed AIAIoT-FCP model is depicted in Fig. 1. From the figure, it is noticeable that the AIAIoT-FCP model initially enables the IoT gadgets like smartphones, laptops, etc. to gather the financial data related to the user which are afterward sent to the cloud server. On the server side, the prediction of financial crisis takes place in three different levels such as data preprocessing, CHGSO-FS technique for selection of features, and NADAM-DELM based classification model. The detailed working of these processes is offered clearly in the subsequent sections.

3.1 Design of CHGSO-FS technique

During the feature selection process, the input financial data gathered by IoT devices in Fintech environment is utilized to elect an optimal set of features. By Henry's law, the highest amount of solute could dissolve in a certain quantity of solvent at particular pressure/temperature named solubility (Mohebbi et al., 2012). Thus, the HGSO was stimulated by the behavior of Henry's law. It is utilized for determining the solubility of low solubility gases in liquids. Furthermore, pressure and temperature are the 2 aspects which influence solubility; at higher temperature, solid becomes more soluble, however, gases are lesser soluble. For pressure, the solubility of gas increased with raising pressure. It can be mathematically defined in the following.

Step 1:   Initiation procedure. The population size N and location of gas is initiated according to succeeding formula:

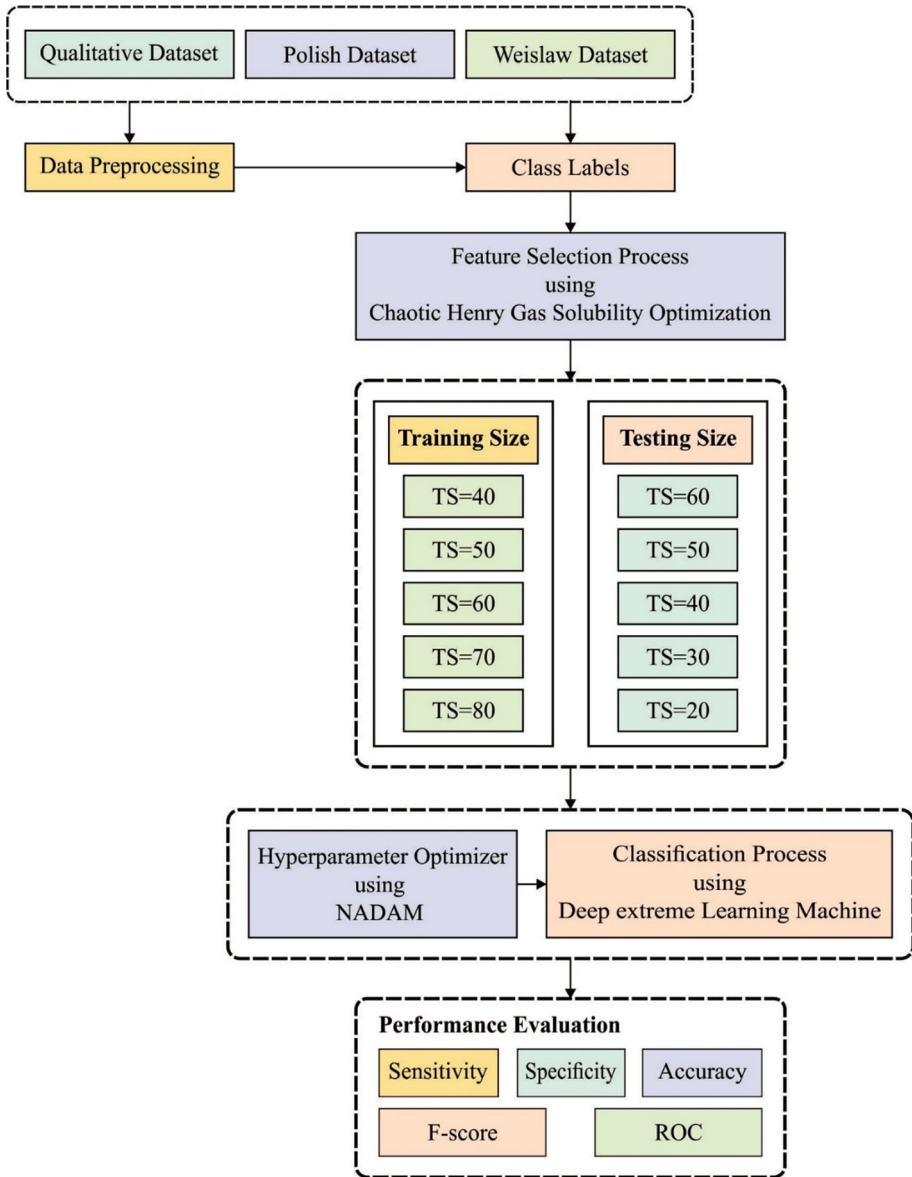$$X_i(t+1) = X_{\text{minm}} + r \times (X_{\text{maxm}} - X_{\text{minm}}) \tag{1}$$

Fig. 1 Working process of AIAIoT-FCP model

where the location of ith gas in population N is represented in $X_{(i)}$, r denotes arbitrary amount among zero and one, and X_{minm}, X_{maxm} indicates boundary values, and t represents iteration time. The amount of gas i, Henry constant of type $j(H_j(t))$, partial pressure $P_{i,j}$ of gas i in the cluster j, and $\nabla_{\mathrm{sol}} E/R$ constant value of type $j(C_i)$ is initiated by the succeeding formula:

$$H_j(t) = l_1 \times rand(0, 1), \ P_{i,j} = l_2 \times rand(0, 1), \ C_j = l_3 \times rand(0, 1) \quad (2)$$

where, l_1, l_2, l_3 are constant values.

Step 2: Clustering. The population agent is separated into equivalent clusters equal to the amount of gas types. Every cluster have same gas and Henry's constant values (H_j).

Step 3: Calculation. Every cluster j is calculated for identifying an optimum gas (Hashim et al., 2019). Later, the gas is rated for obtaining an optimum gas in the whole swarm.

Step 4: Upgrade the Henry $coefficient$. It is upgraded based on succeeding formula:

$$H_j(t + 1) = H_j(t) \times \exp\left(-C_j \times \left(1/T(t) - 1/T^{\theta}\right)\right), T(t) = \exp(-t/iter) \quad (3)$$

where H_j represents Henry's coefficient for cluster j, T indicates temperature, T^{θ} denotes constant and equivalent to 298.15, and iter indicates overall amount of rounds.

Step 5: Upgrade solubility. It is upgraded based on succeeding formula:

$$S_{i,j}(t) = K \times H_j(t + 1) \times P_{i,j}(t) \quad (4)$$

where $S_{i,j}$ denotes solubility of gas i in the cluster j and $P_{i,j}$ indicates partial pressure on gas i in cluster j and K represents constant.

Step 6: Upgrade position and it can be upgraded as follows:

$$X_{i,j}(t + 1) = X_{i,j}(t) + F \times r \times \gamma \times \left(X_{i,best}(t) - X_{i,j}(t)\right)$$
$$+ F \times r \times \alpha \times \left(S_{i,j}(t) \times X_{best}(t) - X_{i,j}(t)\right)$$

$$\gamma = \beta \times \exp\left(-\frac{F_{best}(t) + \varepsilon}{F_{i,j}(t) + \varepsilon}\right), \varepsilon = 0.05 \quad (5)$$

where the gas location i in cluster j is represented by $X_{(i,j)}$, and r and t indicate arbitrary constant and round, correspondingly. $X_{(i,best)}$ represents optimum gas i in cluster j, where X_{best} indicates optimum gas in the swarm. In addition, γ indicates capacity of gas j in cluster i to interrelate with gas in its cluster, α represents effect of other gas on gas i in cluster j and equivalent to one and β denotes constant. $F_{(i,j)}$ indicates the fitness of gas i in cluster j, in contrast F_{best} represents fitness of an optimum gas in the whole scheme. F denotes flag which modifies the course of searching agent and gives diversity $= \pm$.$X_{(i,best)}$ and X_{best} denotes 2 variables in charge to balance exploitation and exploration capacities. Particularly, $X_{(i,best)}$ indicates best gas i in cluster j, and X_{best} represents optimum gas in the swarm.

Step 7: Escape from local optimal. It is utilized for escaping from the local optimal. Select and rank amount of worst agent (N) by the succeeding formula:

$$N_w = N \times (rand(c_2 - c_1) + c_1), c_1 = 0.1 \text{ and } c_2 = 0.2 \quad (6)$$

where N denotes amount of search agent.

Step 8: Upgrade the location of worst agent.

$$G_{(i,j)} = G_{\min(i,j)} + r \times \left(G_{\max(i,j)} - G_{\min(i,j)}\right) \quad (7)$$

where, $G_{(i,j)}$ indicates location of gas i in cluster j, r denotes arbitrary amount and $G_{\min}$, $G_{\max}$ represents bound of the problem.

Figure 2 shows the steps in optimization based FS. From a conceptual viewpoint, HGSO contains exploitation and exploration stages, hence it is considered a global optimization method. Additionally, the amount of operators to be modified in HGSO was reduced to create the technique easier to understand and implement. Consider computation complexity of the presented technique $O(tnd)$ whereas t illustrates highest amount of iteration, n denotes
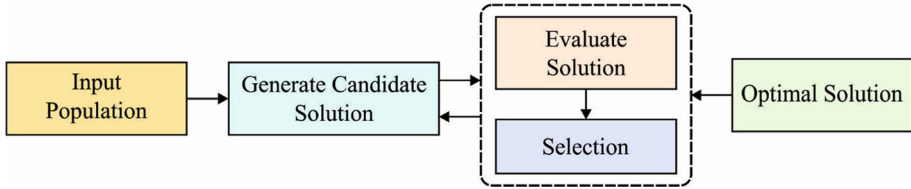
Fig. 2 Steps in optimization based features selection

amount of solution, and d represents amount of parameter. Thus, the whole complexity comprising objective function (obj) determined by Eq. (5) estimated by $O(tnd) * O(obj)$. In order to improve the convergence rate of the HGSO algorithm, the chaotic theory is integrated and derived from the CHGSO algorithm.

Chaos is an unstable form or dynamic performance that is most sensitive to primary situations. The chaos is implemented in several optimization techniques for avoiding traps in the local optimum and enhancing the number of solutions. Every metaheuristic technique depends upon 2 rules: exploitation and exploration. With exploitation, the search is developed to optimal solutions, but exploration permits the search for solutions most efficient (Qasim et al., 2020). The chaos is introduced as metaheuristic models for striking a tradeoff among exploration and exploitation and so attaining a better solution effectually.

The chaos is extra for obtaining enhanced assets in exploration and exploitation in all searching spaces, so enhancing the efficiency of proposed model to determine the optimum global solution. The chaotic map is utilized in determining location x_i^k, where the parameter θ is replaced with value attained with chaotic map is defined using Eq. (8):

$$x_i^{k+1} = x_i^k + C_{map} \times \left(x_{BH} - x_i^k\right), i = 1, 2, \ldots, N_\nabla \qquad (8)$$

where x_i^k and x_i^{k+1} are the locations of the ith star at iterations k and $k+1$, correspondingly. x_{BH} refers the location of BH from space, C_{map} represents chaotic maps, and N_s implies the number of stars..Afterward, chaotic maps are utilized for manipulating the values of arbitrary parameters from HGSO technique. The presented CHSGO technique is utilized to FS by electing the optimum features utilizing the wrapper technique. The optimum feature subsets are individuals that minimizing particular features and maximizing classification accuracy, depends on FF. The FS process take place in two stages namely initialization and update solutions.

During the initialization phase, the HGSO creates early population of N candidate solution, whereas every individual represents a subset of features to be chosen for estimation. This phase performs a major part in the convergence and quality of optimum solutions. The population X^0 is created arbitrarily by Eq. (1). In this work, the lower and upper bounds lb_i and ub_i of every candidate solution, i should in the range of zero and one. For selecting a subset of features, an intermediate stage termed binary alteration is required earlier the fitness procedure. Hence, eve solutions x_i^0 should be transformed to binary x_i^{bin} by Eq. (9):

$$x_i^{bin} = \begin{cases} 1 \text{ if } x_i^0 > 0.5 \\ 0 \text{ otherwise.} \end{cases} \qquad (9)$$

To explain the procedure of transformation, assume the solution x_i that contain 6 components as $x_i^0 = [0.6, 0.1, 0.7, 0.43, 0.2, 0.81]$. The process of transformation is employed by Eq. (10) for generating a binary vector: $x_i^{bin} = [1, 0, 1, 0, 0, 1]$, whereas 1 indicates feature

to be chosen; or else consider unchosen when zero. This implies that initial, third, and final features in actual datasets are appropriate ones and should be chosen, when another is inappropriate then it should be removed. Once defining the subset of chosen features, the FF is processed for every solution x_i^{bin} to define the quality of this feature. The objective value for ith solution is determined by Eq. (10).

$$Fit_i = w_1 \times Err_i + w_2 \times \frac{d_i}{D} \tag{10}$$

where as $w_1 = 0.99$ and $w_2 = 1 - w_1$. The weight w_1 denotes equalization factor which is utilized to maintain balance among the classification error rate Err_i, indicates number of chosen features d_i. In Eq. (10), D represents overall size of features in actual dataset. In this work, the DELM classification is utilized as expert system or evaluator in FS procedure. Err_i denotes error rate of test set processed using DELM model.

Now, the procedure of upgrading solutions contains employing initially the clustering step that goals at separating the population into various clusters, Henry constant in HGSO technique. Afterward, they employ steps 7 to 14 as defined in Eq. 1. Later, they estimate fitness of novel population for determining an optimum solution x_{best}^0. Then, this procedure is repeated till the end criteria are encountered. In this work, the termination condition relates to the higher number of iterations that enables assessing quality of the HGSO method. When completing HGSO procedure, they return an optimum solution x_{best}^0. In actual data, they preserve the features with their values equivalent to one in x_{best}. They utilized hold out approach for classification that indicates to separate arbitrarily the dataset into 2 portions: 20% for test set and 80% for trained set (Neggaz et al., 2020). The DELM module estimates the accurateness via the test set.

3.2 Data classification model

Once the feature subsets are chosen by the CHGSO-FS technique, they are fed into the DELM model to carry out the classification process. The DELM concept was initially presented by Huang (2011); It is superior to the conventional ANN and is one of the significant methods among predictions. The ANN needs huge data instances, minimal learning time and the module of learning turns into over fitted. Figure 3 displays module framework of ELM that contain 3 input layers, multiple hidden layers (HLs) represent hl, and output layer. Initially, the trained instance $[X, Y] = [x_y, y_k]$, $(i = 1, 2, \ldots Z)$, and input instance $X = [X_{k1} X_{k2} X_{k3} \ldots X_{kz}]$ and a target matrix $Y = [y_{11} y_{12} y_{13} \ldots y_{1z}]$ are depicted in the form of matrices X and Y, as given by Eq. (11) and (12). Besides, the weights are adapted using ELM together with input and HL; thus, the input is vk1 of a kth and hidden node of lth displayed in (13).

$$X = \begin{bmatrix} x11 & x12 & \ldots & x1n \\ x21 & x22 & \ldots & x2n \\ x31 & x32 & \ldots & x3n \\ \vdots & \vdots & & \vdots \\ x_{p1} & x_{p2} & & x_{pn} \end{bmatrix} \tag{11}$$
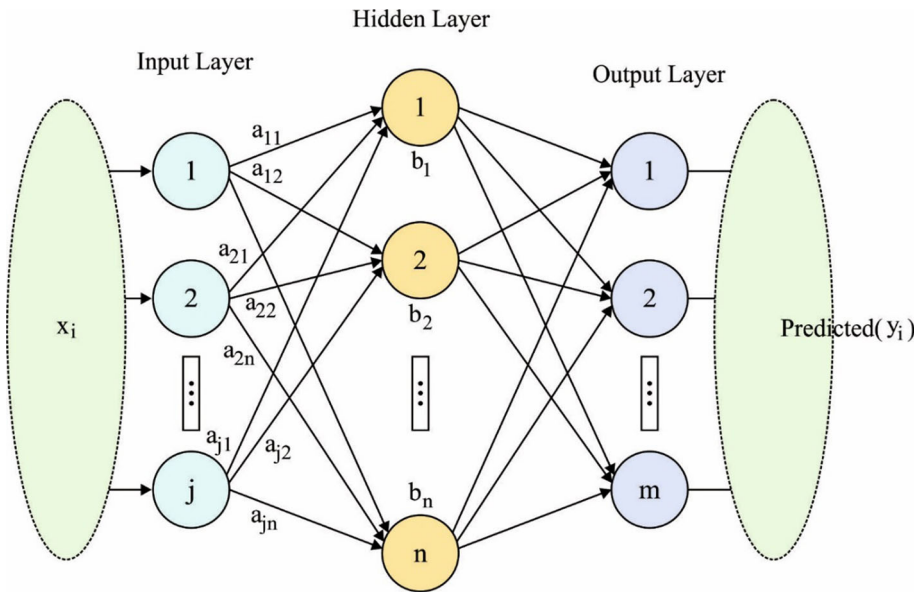
Fig. 3 Structure of ELM

$$Y = \begin{bmatrix} y11 & y12 & \ldots & y1n \\ y21 & y22 & \ldots & y2n \\ y31 & y32 & \ldots & y3n \\ \vdots & \vdots & & \vdots \\ y_{p1} & y_{p2} & & y_{pn} \end{bmatrix} \tag{12}$$

$$V = \begin{bmatrix} v11 & v12 & \ldots & v1n \\ v21 & v22 & \ldots & v2n \\ v31 & v32 & \ldots & v3n \\ \vdots & \vdots & & \vdots \\ v_{p1} & v_{p2} & & v_{pn} \end{bmatrix} \tag{13}$$

$$\mu = \begin{bmatrix} \mu11 & \mu12 & \ldots & \mu1n \\ \mu21 & \mu22 & \ldots & \mu2n \\ \mu31 & \mu32 & \ldots & \mu3n \\ \vdots & \vdots & & \vdots \\ \mu_{p1} & \mu_{p2} & & \mu_{pn} \end{bmatrix} \tag{14}$$

Then, the ELM was arbitrarily chosen by the bias of the HL nodes, using Eq. (15). The resultant matrix as in Eq. (16). Equation (17) denotes column vector of the resultant matrix T.

$$Y = [y1, y2, y3 -------- yp]T \tag{15}$$

$$H = [h1, h2, h3 --------- hz]rxz \tag{16}$$

The result of HL is M and translocation of N as $N^{\cdot}$ and values of weight matrix Y is evaluated by Eq. (19) with least square technique (Siddiqui et al., 2020).

$$h_j = \begin{bmatrix} h_{1j} \\ h_{2j} \\ h_{3j} \\ \vdots \\ h_{rj} \end{bmatrix} = \begin{bmatrix} \sum_{l=1}^{x} N_{k1}g(v_k x_l + y_k) \\ \sum_{l=1}^{x} N_{k2}g(v_k x_l + y_k) \\ \sum_{l=1}^{x} N_{k3}g(v_k x_l + y_k) \\ \vdots \\ \sum_{l=1}^{x} N_{kr}g(v_k x_l + y_k) \end{bmatrix} (l = 1, 2, 3, \cdots .., Q), (j = 1, 2, \cdots z) \quad (17)$$

$$M\mu = N' \quad (18)$$

$$\mu = M^+ N' \quad (19)$$

The μ regularization term is utilized for maximizing the overall network stability. The error and test methods are used to select a number of nodes from HL. The test and error techniques are utilized to specify neurons of HL. The 2nd HL output is attainable by;

$$M_1 = N\mu^+ \quad (20)$$

where μ^+ represents the common inverse of matrix μ. Therefore, HL values are simply achieved by Eq. (19).

$$g(V_1 M + Y_1) = M_1 \quad (21)$$

$$V_{ME} = g^{-1}(M_1)M_E^+ \quad (22)$$

M_E^+ indicates inverse of M_E and evaluate Eq. (13), the active function $g(x)$ is utilized. Therefore, the essential result of the 2nd HL is studied via representing the accurate $g(x)$ activation function:

$$M_2 = g(V_{ME} M_E) \quad \text{where } V_{ME} M_E = Neth_2 \quad (23)$$

$$M_2 = g(Neth_2) \quad (24)$$

Upgrading the weight matrix μ between 2nd and 3rd layers by Eq. (25). M_2^+ denotes inverse of M_2. They evaluated 3rd layer outcomes are demonstrated by Eq. (26).

$$\mu_{new} = M_2^+ N \quad (25)$$

$$M_3 = N\mu^+ new \quad (26)$$

$N\mu^+ new$ denotes inverse of weighted matrix μ_{new}. The DELM is later set the matrix $V_{ME1} = [Y_2, V_2]$. Equations (10) and (11) enables the 3rd layer output to be attained.

$$M_3 = g^{-1}(M_2 V_2 + Y_2) = g(Neth_3) \quad (27)$$

$$V_{ME1} = \left(\mu^{-1}(M_3)M_{E1}^+\right) \quad (28)$$

The M_2 denoted the essential result in Eq. (27), the weight is denoted between the 3rd HL using V_2, the y_2 contain 3 neurons of the HL. M_{E1}^+ represents inverse of $E1$, and $g - 1 (x)$

denotes inverse of $g(x)$ $E1$ In Eq. (29). The logistical Sigmoid is considered. Equation (30) was estimated by 3rd hidden level output.

$$g(x) = \frac{1}{1 + e^{-x}} \tag{29}$$

$$M_3 = g(V_{ME1} M_{E1}) \text{ where } V_{ME1} M_{E1} = Neth_3$$

$$M_3 = g(Neth_3) \tag{30}$$

At last, the resultant weight matrix was estimated by Eq. (31) consider the 3rd hidden level and last layer outcome. Equation (32) signifies possible result of 3rd HL.

$$\mu_{new} = M_4^T \left(\frac{1}{\lambda} + M_4^T M_4 \right)^{-1} N \tag{31}$$

$$M_4 = N\mu_{new}^+ \tag{32}$$

$N\mu_{new}^+$ represents transpose of weight matrix μ_{new}. Later the DELM creates matrix $V_{ME2} = [Y_3, V_3]$. Equations (24) and (33) utilized for attaining the output of 4th layer.

$$M_4 = g^{-1}(M_3 V_3 + Y_3) = g(Neth4_1) \tag{33}$$

$$V_{ME2} = \mu^{-1} \left((M_4) M_{E2}^+ \right) \tag{34}$$

The preferred output is denoted by Eq. (11) using M_3, V_3 indicated weight among 3rd and 4th HLs, the bias of neurons on 3rd HL is y_3. M_{E1}^+ denotes inverse of M_{E1}, and $g - 1(x)$ indicates opposite of $g(x)$ activation feature. The sigmoidal logistical function is considered. In Eq. (35), succeeding the measure of third and fourth HL values;

$$M_4 = g(Neth4_2) \tag{35}$$

Lastly, in Eq. (36). The weighted matrix of output from 4th to the output layer was estimated. Equation (37) denotes calculated result from 5th layer. Equation (38) displays the needed output of DELM scheme.

$$\mu_{new} = M_5^T \left(\frac{1}{\lambda} + M_5^T M_5 \right)^{-1} N \tag{36}$$

$$M_5 = N\mu_{new}^+ \tag{37}$$

$$f(x) = M_5 \beta_{new} (29) \tag{38}$$

concept of cycle is utilized for demonstrating the DELM process. Equations (27)–(31)is re-evaluated for recording the variables of each HL and eventually obtaining the final outcome of the DELM network. When the HL is raised, a similar procedure is reutilized and executed in a similar manner.

$$op = \frac{1}{1 + e^{-Nethj}} \text{ where } j = 1, 2, 3 \cdots, r \tag{39}$$

In order to determine the hyperparameters of the DELM model in an optimally way, NADAM optimizer is used.

The Nadam optimizer explained in Dozat (2016), tries to integrate Nesterov-accelerated adaptive moment estimation as to Adam. The main benefit of this combined model is that utilized adaptive moment estimation uses for performing extremely accurate steps from

gradient direction using upgrades of model parameter through the momentum step previously the calculation of gradient (Hoang, 2021). The upgrade rule of Nadam is expressed as:

$$w_t = w_{t-1} - \alpha \times \frac{\overline{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \tag{40}$$

where

$$\overline{m}_t = \left(1 - \beta_{1,t}\right)\hat{g}_t + \beta_{1,t+1}\hat{m}_t,$$

$$\hat{m}_t = \frac{m_t}{1 - \prod_{i=1}^{t+1} \beta_{1i}} \tag{41}$$

$$\hat{g}_t = \frac{g_t}{1 - \prod_{i=1}^{t+1} \beta_{1i}}$$

4 Performance validation

In this section, the experimental results of the proposed model are assessed and the simulation process take place using Python 3.6.5 tool. The details of the dataset are provided in Table 1. The qualitative dataset includes 250 samples with 6 features. Similarly, the Polish dataset contains a set of 43,405 samples with 6 features. Likewise, the Weislaw dataset encompasses 240 samples with 30 attributes.

The performance of the CHGSO-FS model is examined over the other optimization techniques in Table 2 and Fig. 4. From the table, it is clear that the CHGSO-FS technique has offered optimal FS results on all the applied datasets. For instance, on the applied qualitative bankruptcy dataset, the CHGSO-FS technique exhibited effective performance with the least best cost of 0.06473 whereas the HGSO-FS and GWO-FS techniques have demonstrated ineffectual outcomes with the best cost of 0.08394 and 0.09218 respectively.

Moreover, on the applied polish dataset, the CHGSO-FS technique exhibited effective performance with the least best cost of 0.04351 whereas the HGSO-FS and GWO-FS techniques have demonstrated ineffectual outcomes with the best cost of 0.07260 and 0.09832 respectively. Furthermore, on the applied weislaw dataset, the CHGSO-FS technique exhibited effective performance with the least best cost of 0.04271 whereas the HGSO-FS and GWO-FS techniques have demonstrated ineffectual outcomes with the best cost of 0.08362 and 0.09361 respectively.

A detailed classification results analysis of the AIAIoT-FCP model on the qualitative bankruptcy dataset under varying training sizes is given in Table 3. On the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the sensitivity of 99.89% whereas the DELM, KELM, and ELM models have attained slightly reduced outcomes with the sensitivity of 98.64, 98.62, and 98.60% respectively. At the same time, on

Table 1 Dataset description

Dataset	Source	Instances	Attributes	Classes	FC/Non-FC
Qualitative	UCI	250	6	2	107/143
Polish	UCI	43,405	64	2	2091/41314
Weislaw	pietruszkiewicz	240	30	2	112/128

Table 2 Selected features and its best cost of proposed CHGSO-FS method with existing methods on applied dataset

Methods	Best cost	Selected features
Qualitative bankruptcy dataset		
CHGSO-FS	0.06473	2,3,5
HGSO-FS	0.08391	1,2,3,5
GWO-FS	0.09218	1,2,4,5,6
Polish dataset		
CHGSO-FS	0.04351	4,5,6,9,10,12,13,14,17,19,22,24,26,28,29,32,24,27,35,38,40,46,48,50,54
HGSO-FS	0.07260	1,2,6,11,13,14,17,20,22,23,24,27,28,35,40,42,43,45,48,52,54,56
GWO-FS	0.09832	1,2,3,6,7,8,9,11,12,13,14,15,17,18,19,20,22,34,35,38,40,42,44,47,53,55,58
Weislaw dataset		
CHGSO-FS	0.04271	2,3,4,5,8,9,11,13,15,18,22,25
HGSO-FS	0.08362	1,2,3,6,7,8,9,11,12,15,17,18,19,22,23,25,27
GWO-FS	0.09361	1,2,3,6,11,12,14,15,16,17,19,20,21,22,23,24,27,28,29
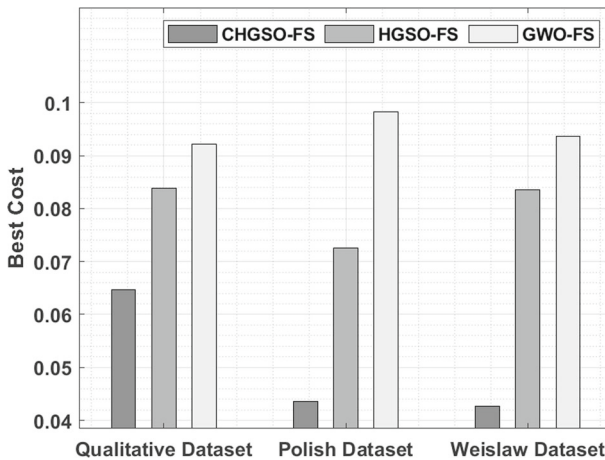
Fig. 4 Best cost analysis of CHGSO-FS Method

the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the specificity of 99.94% whereas the DELM, KELM, and ELM models have attained slightly reduced outcome with the specificity of 98.31, 98.27, and 98.25% respectively. Besides, on the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the accuracy of 99.84% whereas the DELM, KELM, and ELM models have attained slightly reduced outcome with the accuracy of 98, 97.99, and 97.95% respectively. Additionally, on the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the F-score of 99.48% whereas the DELM, KELM, and ELM models have attained slightly reduced outcome with the F-score of 98.62, 98.58, and 98.57% respectively.

A detailed comparative results analysis of the AIAIoT-FCP model with other existing techniques takes place on the qualitative bankruptcy dataset in Table 4 and Fig. 5. From the

Table 3 Result analysis of qualitative bankruptcy dataset various training size on proposed AIAIoT-FCP model with other models in terms of different measures

Training (%)	AIAIoT-FCP	DELM	KELM	ELM
Sensitivity (%)				
TS = 40	99.89	98.64	98.62	98.60
TS = 50	99.61	99.27	97.24	97.22
TS = 60	99.94	98.23	98.20	97.18
TS = 70	99.74	98.93	98.92	98.92
TS = 80	99.81	98.79	97.75	97.75
Average	99.79	98.77	98.14	97.93
Specificity (%)				
TS = 40	99.94	98.31	98.27	98.25
TS = 50	99.86	98.15	98.91	97.08
TS = 60	99.96	99.45	97.33	98.39
TS = 70	99.89	98.58	99.54	98.53
TS = 80	99.91	99.19	98.76	98.13
Average	99.91	98.73	98.56	98.07
Accuracy (%)				
TS = 40	99.84	98.00	97.99	97.95
TS = 50	99.88	97.87	98.87	97.85
TS = 60	99.92	98.48	98.47	98.43
TS = 70	99.96	98.44	98.44	98.40
TS = 80	99.96	98.92	98.91	98.91
Average	99.91	98.34	98.53	98.30
F-score (%)				
TS = 40	99.48	98.62	98.58	98.57
TS = 50	98.87	98.43	98.43	98.39
TS = 60	98.28	97.27	98.25	97.24
TS = 70	99.76	98.56	97.54	97.54
TS = 80	99.85	97.84	97.42	97.80
Average	99.24	98.14	98.04	97.90

Table 4 Result analysis of various classifiers on qualitative bankruptcy dataset

Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)	F-score (%)
AIAIoT-FCP	0.9979	0.9991	0.9991	0.9924
ACO-DC	0.9976	0.9989	0.9979	0.9914
IKM-FSCGACA	0.9636	0.9924	0.9793	0.9769
FSCGACA	0.9074	0.9328	0.9214	0.9116
GACA	0.8909	0.9318	0.9132	0.9032
ACA	0.7946	0.8615	0.8305	0.8127
OlexGA	0.6666	0.7578	0.7148	0.6877
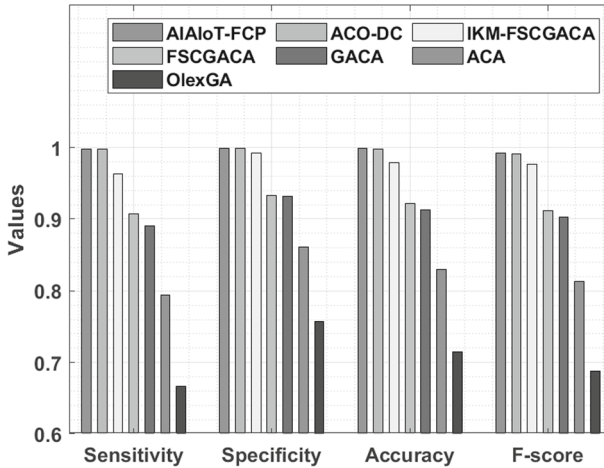
Fig. 5 Result analysis of AIAIoT-FCP model on qualitative bankruptcy dataset

accomplished results, it is noticeable that the Olex-GA model has showcased least classification performance over the other methods. Followed by, the ACA model has tried to exhibit slightly improved outcomes over the Olex-GA model. Then, the GACA and FSCGACA models have showcased moderately closer results. Eventually, the IKM-FSCGACA model has demonstrated reasonable outcome over the other methods. Meanwhile, a near optimal classification performance has been offered by the ACO-DC model. However, the AIAIoT-FCO model has portrayed effectual outcome over the other methods with the sensitivity, specificity, accuracy, and F-score of 0.9979, 0.9991, 0.9991, and 0.9924 respectively.

A detailed classification results analysis of the AIAIoT-FCP model on the polish dataset under varying training sizes is given in Table 5. On the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the sensitivity of 98.86% whereas the DELM, KELM, and ELM models have attained slightly reduced outcomes with the sensitivity of 98.55, 98.21, and 97.86% respectively.

At the same time, on the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the specificity of 99.94% whereas the DELM, KELM, and ELM models have attained slightly reduced outcome with the specificity of 98.1, 96.71 and 96.32% respectively. Besides, on the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the accuracy of 97.87% whereas the DELM, KELM, and ELM models have attained slightly reduced outcome with the accuracy of 96.95, 97.58 and 96.18% respectively. Additionally, on the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the F-score of 99.56% whereas the DELM, KELM, and ELM models have attained slightly reduced outcome with the F-score of 98.39, 98.03, and 97.71% respectively.

A detailed comparative results analysis of the AIAIoT-FCP model with other existing techniques takes place on the polish bankruptcy dataset in Table 6 and Fig. 6. From the accomplished results, it is noticeable that the Olex-GA model has showcased least classification performance over the other methods. Followed by, the ACA model has tried to exhibit slightly improved outcomes over the Olex-GA model. Then, the GACA and FSCGACA models have showcased moderately closer results. Eventually, the IKM-FSCGACA model has demonstrated reasonable outcome over the other methods. Meanwhile, a near optimal

Table 5 Result analysis of polish dataset various training size on proposed AIAIoT-FCP model with other models in terms of different measures

Training (%)	AIAIoT-FCP	DELM	KELM	ELM
Sensitivity				
TS = 40	98.86	98.55	98.21	97.86
TS = 50	98.35	97.76	97.43	97.11
TS = 60	97.82	97.50	97.17	96.83
TS = 70	98.49	98.04	97.73	97.34
TS = 80	98.68	98.27	97.92	97.60
Average	98.44	98.02	97.69	97.34
Specificity				
TS = 40	99.94	98.1	96.71	96.32
TS = 50	99.96	98.39	97.07	96.71
TS = 60	99.91	96.92	96.57	96.27
TS = 70	99.88	97.40	97.05	96.74
TS = 80	99.92	98.79	98.13	97.75
Average	99.92	97.92	97.10	96.75
Accuracy				
TS = 40	97.87	96.95	97.58	96.18
TS = 50	98.52	97.54	97.24	96.93
TS = 60	98.41	98.47	98.08	96.78
TS = 70	99.10	98.17	97.78	97.39
TS = 80	99.31	98.53	97.19	97.83
Average	98.64	97.93	97.57	97.02
F-score				
TS = 40	99.56	98.39	98.03	97.71
TS = 50	99.37	97.56	97.17	96.85
TS = 60	98.94	97.50	97.16	96.78
TS = 70	98.23	98.84	97.54	97.21
TS = 80	99.23	98.54	98.15	97.81
Average	99.06	98.16	97.61	97.27

Table 6 Result analysis of various classifiers on polish bankruptcy dataset

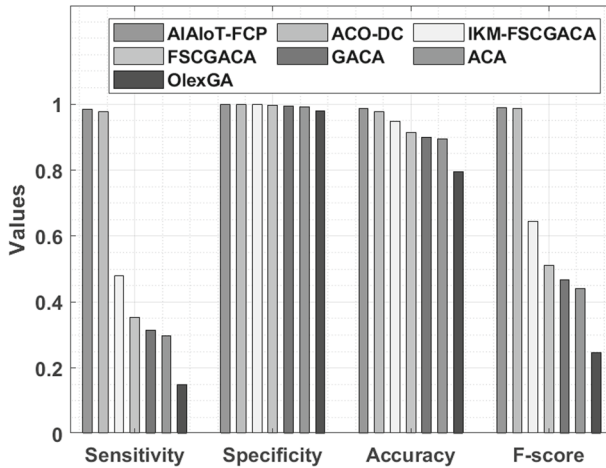| Methods	Sensitivity	Specificity	Accuracy	F-score
AIAIoT-FCP	0.9844	0.9992	0.9864	0.9906
ACO-DC	0.9763	0.9990	0.9769	0.9880
IKM-FSCGACA	0.4791	0.9989	0.9472	0.6439
FSCGACA	0.3529	0.9958	0.9139	0.5112
GACA	0.3155	0.9948	0.8998	0.4684
ACA	0.2965	0.9920	0.8941	0.4409
OlexGA	0.1499	0.9806	0.7946	0.2464

Fig. 6 Result analysis of AIAIoT-FCP model on polish bankruptcy dataset

classification performance has been offered by the ACO-DC model. However, the AIAIoT-FCO model has portrayed effectual outcome over the other methods with the sensitivity, specificity, accuracy, and F-score of 0.9844, 0.9992, 0.9864, and 0.9906 respectively.

A detailed classification results analysis of the AIAIoT-FCP model on the Wieslaw dataset under varying training sizes is given in Table 7. On the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the sensitivity of 99.71% whereas the DELM, KELM, and ELM models have attained slightly reduced outcomes with the sensitivity of 99.55, 99.24 and 98.89% respectively. At the same time, on the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the specificity of 98.23% whereas the DELM, KELM, and ELM models have attained slightly reduced outcome with the specificity of 97.05, 96.65 and 96.34% respectively. Besides, on the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the accuracy of 97.53% whereas the DELM, KELM, and ELM models have attained slightly reduced outcome with the accuracy of 98.16, 97.79 and 97.43% respectively. Additionally, on the training set (TS) of 40%, the AIAIoT-FCP model has obtained maximum performance with the F-score of 98.39% whereas the DELM, KELM, and ELM models have attained slightly reduced outcome with the F-score of 97.46, 97.14 and 96.75% respectively.

A detailed comparative results analysis of the AIAIoT-FCP model with other existing techniques takes place on the weislaw bankruptcy dataset in Table 8 and Fig. 7. From the accomplished results, it is noticeable that the Olex-GA model has showcased least classification performance over the other methods. Followed by, the ACA model has tried to exhibit slightly improved outcomes over the Olex-GA model. Then, the GACA and FSCGACA models have showcased moderately closer results.

Eventually, the IKM-FSCGACA model has demonstrated reasonable outcome over the other methods. Meanwhile, a near optimal classification performance has been offered by the ACO-DC model. However, the AIAIoT-FCO model has portrayed effectual outcome over the other methods with the sensitivity, specificity, accuracy, and F-score of 0.9886, 0.9921, 0.9826, and 0.9857 respectively. From the above mentioned tables and figures, it is apparent that the proposed model can be employed as an effective FCP model in IoT environment. The proposed model has achieved superior results due to the inclusion of CHGSO-FS technique

Table 7 Result analysis of weislaw various training size on proposed aiaiot-fcp model with other models in terms of different measures

Training (%)	AIAIoT-FCP	DELM	KELM	ELM
Sensitivity				
TS = 40	99.71	99.55	99.24	98.89
TS = 50	99.23	98.4	98.10	97.78
TS = 60	99.26	98.56	98.24	97.91
TS = 70	97.72	96.98	96.65	96.31
TS = 80	98.40	97.56	97.24	96.85
Average	98.86	98.21	97.89	97.54
Specificity				
TS = 40	98.23	97.05	96.65	96.34
TS = 50	98.75	98.34	97.95	97.60
TS = 60	99.51	96.92	96.60	96.27
TS = 70	99.83	98.73	97.69	97.33
TS = 80	99.77	98.35	96.99	96.68
Average	99.21	97.87	97.17	96.84
Accuracy				
TS = 40	97.53	98.16	97.79	97.43
TS = 50	98.48	98.12	97.81	96.44
TS = 60	98.73	97.15	97.48	97.45
TS = 70	99.13	98.34	98.00	97.66
TS = 80	97.43	97.22	96.82	96.44
Average	98.26	97.79	97.58	97.08
F-score				
TS = 40	98.39	97.46	97.14	96.75
TS = 50	99.44	98.21	97.90	97.57
TS = 60	97.74	96.83	96.43	96.10
TS = 70	98.32	98.62	97.63	97.23
TS = 80	98.96	97.27	96.89	96.51
Average	98.57	97.67	97.19	96.83

Table 8 Result analysis of various classifiers on weislaw bankruptcy dataset

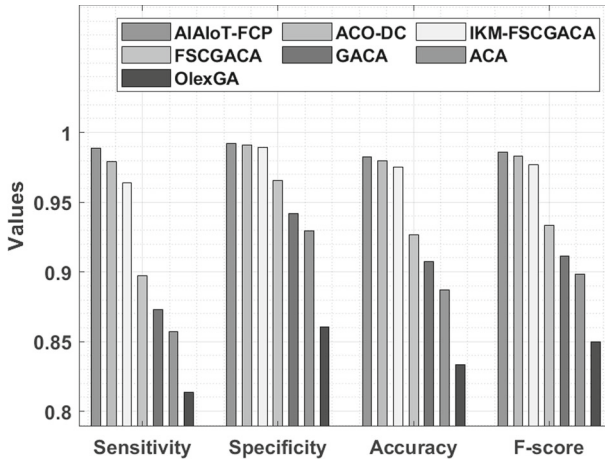| Methods	Sensitivity	Specificity	Accuracy	F-score
AIAIoT-FCP	0.9886	0.9921	0.9826	0.9857
ACO-DC	0.9789	0.9912	0.9798	0.9834
IKM-FSCGACA	0.9639	0.9892	0.9755	0.9772
FSCGACA	0.8974	0.9655	0.9264	0.9333
GACA	0.8729	0.9417	0.9074	0.9115
ACA	0.8571	0.9294	0.8872	0.8986
OlexGA	0.8135	0.8605	0.8333	0.8496

Fig. 7 Result analysis of AIAIoT-FCP model on Weislaw bankruptcy dataset

for feature subset selection and NADAM optimizer in the determination of optimal prameter values. The presented model can be employed in real time financial sectors such as banking, e-commerce, companies, etc.

5 Conclusion

This paper has designed an effective AIAIoT-FCP model to determining the financial crisis of the organization in the IoT environment. The proposed AIAIoT-FCP model initially enables the IoT gadgets like smartphones, laptops, etc. to gather the financial data related to the user which are afterward sent to the cloud sever. On the server sides, the prediction of financial crisis takes place in three different levels such as data preprocessing, CHGSO-FS technique for selection of features, and NADAM-DELM based classification model. The use of CHGSO-FS technique helps to eradicate the unwanted features in such a way that the classification performance of the NADAM-DELM model gets improved to a considerable extent. A comprehensive experimental validation process takes place on the benchmark financial dataset to showcase the improved performance of the AIAIoT-FCP model. The resultant values portrayed the superior performance of the AIAIoT-FCP model over the compared methods in a considerable way. As a part of future scope, the proposed AIAIoT-FCP model can be extended to the inclusion of outlier detection techniques and metaheuristic based hyperparameter optimization.

Data availability Data sharing not applicable to this article as no datasets were generated during the current study.

Declaration

Conflict of interest The authors declare that they have no conflict of interest. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

References

Alaraj, M., & Abbod, M. F. (2016). Classifiers consensus system approach for credit scoring. *Knowledge-Based System, 104*, 89–105.

Aydin, A. D., & Cavdar, S. C. (2015). Prediction of financial crisis with artificial neural network: An empirical analysis on Turkey. *International Journal of Financial Research, 6*(4), 36.

Dastkhan, H. (2021). Network-based early warning system to predict financial crisis. *International Journal of Finance and Economics, 26*(1), 594–616.

Dozat T. (2016) Incorporating Nesterov Momentum intoAdam ICLR Workshop, (1), 2013–2016.

Feixiong-Ma, Yingying-Zhou, Xiaoyan-Mo and Yiwei-Xia, 2020. The establishment of a financial crisis early warning system for domestic listed companies based on two neural network models in the context of COVID-19. Mathematical Problems in Engineering, 2020.

Fernández-Arias, D., López-Martín, M., Montero-Romero, T., Martínez-Estudillo, F., & Fernández-Navarro, F. (2018). Financial soundness prediction using a multi-classification model: Evidence from current financial crisis in OECD banks. *Computational Economics, 52*(1), 275–297.

Hashim, F. A., Houssein, E. H., Mabrouk, M. S., Al-Atabany, W., & Mirjalili, S. (2019). Henry gas solubility optimization: A novel physics-based algorithm. *Future Generation Computer Systems, 101*, 646–667.

Hoang, N.D., 2021. Automatic Impervious Surface Area Detection Using Image Texture Analysis and Neural Computing Models with Advanced Optimizers. *Computational Intelligence and Neuroscience*, *2021*.

Huang, G. B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: A survey. *International Journal of Machine Learning and Cybernatics, 2*(2), 107–122.

Jensen, R. (2005). *Combining rough and fuzzy sets for feature selection."*. University of Edinburgh.

Junyu, H., 2020, August. Prediction of Financial Crisis Based on Machine Learning. In 2020 The 4th International Conference on Business and Information Management (pp. 71–75).

Kohavi, R., & John, G. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*, 273–324.

Li, C., & Li, H. (2011). One dependence value difference metric. *Knowledge-Based System, 24*(5), 589–594.

Metawa, U. J. N., Shankar, K., & Lakshmanaprabu, S. K. (2018). Financial crisis prediction model using ant colony optimization. *International Journal of Information Management, 50*, 538–556.

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.

Mohebbi, V., Naderifar, A., Behbahani, R., & Moshfeghian, M. (2012). Determination of henry's law constant of light hydrocarbon gases at low temperatures. *The Journal of Chemical Thermodynamics, 51*, 8–11.

Neggaz, N., Houssein, E. H., & Hussain, K. (2020). An efficient henry gas solubility optimization for feature selection. *Expert Systems with Applications, 152*, 113364.

Qasim, O. S., Al-Thanoon, N. A., & Algamal, Z. Y. (2020). Feature selection based on chaotic binary black hole algorithm for data classification. *Chemometrics and Intelligent Laboratory Systems, 204*, 104104.

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.

Siddiqui, S.Y., Khan, M.A., Abbas, S. and Khan, F., 2020. Smart occupancy detection for road traffic parking using deep extreme learning machine. *Journal of King Saud University-Computer and Information Sciences*.

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

Uthayakumar, J., Metawa, N., Shankar, K. and Lakshmanaprabu, S.K., 2018. Intelligent hybrid model for financial crisis prediction using machine learning techniques. Information Systems and e-Business Management, pp.1–29.

Wang, L., & Wu, C. (2017). A combination of models for financial crisis prediction: Integrating probabilistic neural network with back-propagation based on adaptive boosting. *International Journal of Computational Intelligence Systems, 10*(1), 507–520.

Xue, B., Zhang, M., Browne, W., & Yao, X. (2016). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation, 20*(4), 606–626.