Journal of Intelligent Systems and Internet of Things
Vol. 09, No. 02, PP. 149-161, 2023
DOI: https://doi.org/10.54216/JISIoT.090211

Intelligent Classification for Credit Scoring Based on a Data Mining algorithm

Mohammed G. Fathi Al-Obaidi
Department of Computer Science, University of Mosul, Mosul, Iraq.
Email: mohammedghanim700@gmail.com

*Corresponding Author: mohammedghanim700@gmail.com

Abstract

Credit scoring has grown in importance and has been thoroughly researched by banks and financial institutions. The amount of redundant and irrelevant features present in credit scoring datasets, however, reduces the classification accuracy. As a result, employing effective feature selection methods has become essential. In this study, a hybrid feature selection approach that combines the backpropagation neural network (BPNN) classifier and the pigeon optimization algorithm (POA) is suggested. With hybridization, the POA works to choose characteristic subgroups through the feature selection (FS) process, and the BPNN then assesses the chosen subsets using a fitness function. The experiment findings show that the suggested hybrid technique outperforms other competing approaches in terms of evaluation criteria, according to three benchmark credit scoring datasets.

Received: February 20, 2023 Revised: May 27, 2023 Accepted: September 06, 2023

Keywords: Feature selection, pigeon optimization algorithm, credit scoring, backpropagation neural network, support vector machine.

1. Introduction

Credit scoring has grown in importance and has been thoroughly researched by banks and financial institutions. However, a significant number of pointless and redundant features are present in credit scoring datasets. Banks and financial institutions place a lot of emphasis on credit rating in order to distinguish between poor and excellent customers [1]. Recently, several systems of credit scoring have been successfully implemented to support credit agreement decisions [2]. Generally, credit scoring problems are related to classification by statistical methods. Credit scoring models have been extensively used for credit agreement evaluation and have become one of the major ways for financial institutions to assess credit risk, get better cash flow, minimize possible risks, and produce managerial decisions [3] [4]. The results required for specific credit score applications are provided by researching the best features and using the more developed classifiers to match samples.

The most significant variable that can affect the categorization accuracy is FS. The size of the search space will be huge if the dataset has a lot of features, which will lower the classification accuracy rate. Eliminating noise, irrelevant features, and redundant features can be a robust and active feature selection strategy [5]. According to the mechanism of selection, FS approaches, in general, can be classified into three groups: filter approaches, wrapper approaches, and approaches [6] [8]. One of the most used FS ways is the use of filters, which base their decisions on learning more about each feature's characteristics and a predetermined criterion. These techniques or strategies function independently and are not based on a classification or dependent upon it. On the other hand, the FS technique largely depends on the effectiveness of the classification algorithms for the wrapper approaches to increase classification accuracy. FS process is incorporated into classification algorithms in embedded systems, allowing for simultaneous FS and classification performance [9] [10]. These approaches provide higher computational efficiency compared with the wrapper approaches [7].

To increase the performance of the classification, the hybrid approaches can be utilized. In hybridization, good properties of at least two approaches are combined to enhance the performance of each approach [11] [12].

The pigeon optimization algorithm (POA) is one of the most important parallel heuristic searches, which is inspired by the natural selection process and the main concepts in genetics [13] [14] [15]. The POA has been used as a tool to perform feature selection of credit scoring.

Using the support vector machine (SVM) as a classifier, a novel hybrid feature selection strategy that combines the characteristics of the POA and the backpropagation neural network (BPNN) is presented in this paper in order to reduce the dimension of data features, eliminate redundant features, and ultimately improve the performance of classification tasks for credit scoring. The results, which are based on three benchmark datasets from the UCI machine learning repository, show that the suggested hybrid technique produces superior classification performance than other rival approaches.

1. Methodology

1.1. The problem of feature selection (FS)

FS method is a procedure that decreases or minimizes the number of characteristics and picks selected features as subsets of total features. Prior to classifying data, key characteristics are identified, and unnecessary features are removed using the FS approach. FS is used for a variety of purposes, however obtaining classification accuracy is the most crucial one. It consists of four basic procedures [16]: (1) subset generation operation, (2) subset estimation operation, (3) stopping condition operation, and (4) result confirmation operation.

A subset of features can be evaluated using the subset generation method depending on a predetermined criterion. According to a specific assessment criterion, each candidate subset is assessed and contrasted with the prior best one. The prior subset is eliminated if the new subset is superior [17] [18].

The procedure of evaluation is periodic until a specific stopping factor is checked, and then the most excellent subset needs to be validated by prior information or various tests by means of real datasets [19]. The procedure of subset generation and evaluation is periodic until a given stopping factor is checked, and then we should check the accuracy of most feature subsets by prior information or various tests by means of real datasets [20].

1.2. Pigeon optimization algorithm (POA)

POA is a class of evolutionary algorithms that uses principles of natural evolution and identity of the genetic evolution of organisms, which it introduced by John Holland in 1970 [21]. POA is a heuristic search that modifies the individual functions of coded individuals as real or binary strings by using operators of POA. It finds the optimal solutions from a randomly created population, repeatedly modifies the individual at each stage to be parents and uses the parents to find the offspring for the next generation. The individuals are evaluated using a fitness function that is determined to problem [22].

The POA uses primary operations on the population: selection, crossover (recombination), and mutation to find the optimal solution, and the algorithm is stopped when either a maximum number of generations has been generated or the optimal solutions have been reached by fitness function [23]. Several procedures are important for pigeon optimization algorithms. They are initialization, fitness evaluation, selection, crossover, mutation, and termination.

1.3. Backpropagation Neural Network

A multilayer feedforward neural network (NN) is used in the backpropagation neural network (BPNN) technique to transmit the categorization pattern. The following figures illustrate the general layout of the BPNN, which includes some hidden levels, one input layer, and one output layer [24].

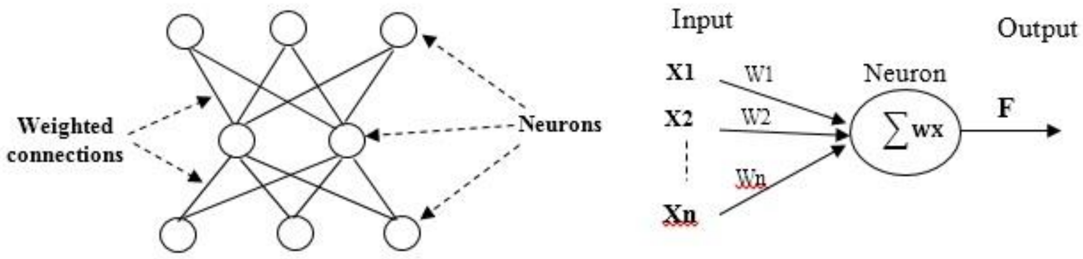

Figure 1: Structure of BPNN

The BPNN structure consists of the input layer, one or more hidden layers, and an output layer. The main steps of the BPNN algorithm are [25] [26]:

1. Create the weights (w) and bias (b) values randomly.
2. Choose the training pair $\{(x,t): x \in X, t \in T\}$ from the training group $[X, T]$, wh X ich represents the input vector, $x \in X$ and T the target (desired) vector.
3. Apply the network forward propagation process to the output account by using the following two equations.

$$o_j = f(net_j) = f(\sum_k w_{ij} x_i), \quad (1)$$

$$o_k = f(net_k) = f(\sum_k w_{jk} x_i), \quad (2)$$

Where Eq. (1) represents the output between the input and hidden layer, while Eq. (2) represents the final output between the hidden layer and the output layer.

4. Compare the final output o_k with the desired output t_k and calculate the error value δ_k of the output as the following.

$$\delta_k = (t_k - o_k) f'(net_k) = (t_k - o_k)(1 - o_k),$$

$$w_{jk}^{new} = w_{jk}^{old} + \eta \delta_k o_k,$$

Where w_{jk} η represents the learning rate, which is supposed to be a small positive real number.

5. Correct the weights of the BP over the network (from the output layer to the hidden layer and to the input layer) by minimizing the error, as follows:

$$\delta_j = o_j (1 - o_j) \sum_{j=1}^{m} w_{jk} \delta_k,$$

Where δ_j represents the errors in hidden layers and w_{jk} represents the weights between the hidden layer and the output layer.

6. Minimize the total error for all inputs used in the training set as follows:

$$w_{ij}^{new} = w_{ij}^{old} + \eta \delta_j o_j,$$

Where w_{ij} represents the weights on the connection from the input layers to the hidden layers.

1.4. Support Vector Machine

Support vector machine (SVM) is a group of supervised machine learning that divides the data into two categories based on statistical learning theory [16]. In the case of linearly separable data, SVM aims to maximize the distance between the training points set and the boundary; however, in the case of nonlinearly separable data, the kernel function maps patterns to a high dimensional space [27].

The hyperplane is defined by the $w^T x + b = 0$, where the weight $w \in R^N$ $b \in R^N$ is the constant. Giving some training dataset D, as the following:

$$D = \{ (x_i, y_i) : x_i \in R^n, y_i \in \{-1,1\} \}_{i=1}^{m}, \qquad (3)$$

Where x_i is a n-dimensional vector, y_i is the class (either +1 or -1). SVM has two hyperplanes that are defined $w^T x + b = +1$ $w^T x + b = -1$, The two functions can be simplified and combined with one function as follows :

$$y_i (w^T x + b) \geq 1. \qquad (4)$$

SVM finds the optimal separating value $f(x) = w^T x + b$. The classifier is:

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{N} \alpha_i y_i x_i^T x + b), \qquad (5)$$

Where $\mathrm{sgn}(.)$ is the sign function, α_i is Lagrange multiplier, x_i is a training sample, x is a test sample. Let the distance from the data point to the hyperplane be $1/\|w\|$. The training of SVM for the non-separable case is solved using a quadratic optimization problem that is shown in the following equation [28] :

$$\min \phi(w, \xi) = \frac{1}{2}\|w\| + C\sum_{i=1}^{N} \xi_i \quad ; \quad \xi_i \geq 0. \qquad (6)$$

Such that

$$y_i (w^T k(x_i) + b) \geq 1 - \xi_i \quad for \quad 1 \leq i \leq N.$$

Every constraint can be satisfied if ξ_i is sufficiently large and C is a regularization parameter. SVM will transform the data in the non-linear case from lower dimensional space into a higher-dimensional space through special functions called kernel, where the classifier is: [29] [30].

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{N} \alpha_i y_i K(x_i^T, x) + b), \qquad (7)$$

Where $K(.)$ is the kernel function? There exist several types of kernel function, such as radial basis function (RBF) which is defined as:

$$K(x_1, x_2) = \exp(-\|x_1 - x_2\|^2). \qquad (8)$$

2. The proposed algorithm

Our suggested algorithm is introduced in this section. The backpropagation neural network (BPNN) and pigeon optimization algorithm (POA) are the two fundamental components of our hybrid method. The POA works to choose the feature subsets through the FS process, and the BPNN then assesses the feature subsets using the fitness function approach. The capacity of each group of characteristics in POA to discriminate is assessed using the fitness function.

In the global feature selection methods, the number of all possible subsets is calculated from the following equation [3] [31]:

$$n_k = 2^k, \quad (9)$$

Where n_k is the number of feature subsets?

In POA, we select a subset of features randomly by encoding the member to binary representation (0 or 1), where the symbol 1 corresponds to the property selection and 0 is not selected. The diagram of feature selection using POA is represented in Figure 2.

Select subset of features	Yes	No	Yes	No	Yes	……	No

Figure 2 A sample of feature subset solution.

Specific requirements, such as calculating classification accuracy, error, or both, are required by the fitness function in POA. A subset of the chosen characteristics is used to represent each chromosome (person), and the fitness function of each individual is determined by evaluating the BPNN using a training set [32]. The individuals in the current population are evaluated by fitness function based on the error of BPNN.

Eq. 10 was used to calculate the error (Err) between the expected (from BPNN) and observed value. To optimize the weights, learning techniques such as the steepest descent were utilized [33].

$$\text{Err} = \frac{1}{2}\sum_j (out_{t_j} - out_{r_j})^2, \quad (10)$$

Where out_{t_j} represents the target output and out_{r_j} represents the predicted output from BPNN.

Members who are less fit have a better probability of living to see the following generation. The POA lowers the average error rate and chooses the person with the lowest error rate based on fitness value. The POA then chooses the person with the lowest error rate [34].

$$Fit = \frac{E}{n_f} + e^{\frac{-1}{n_f}}, \quad (11)$$

Where E represents BPNN-based classification error and n_f represents the cardinality of the selected features.

BPNNs are employed because they provide strong generalization, even if it may be challenging to identify the ideal network parameters. The input layer with N (number of features) nodes, the hidden layer with 6 nodes, and the output layer with 1 node made up the BPNN architecture. In Table 1, the POA configuration is detailed.

Table 1: Parameter values for POA

Parameter Name	Values
Population Size	25
Initial Population Range	[-100 100]

The important steps in the hybrid BPPOA are shown in Figures 3 and 4.

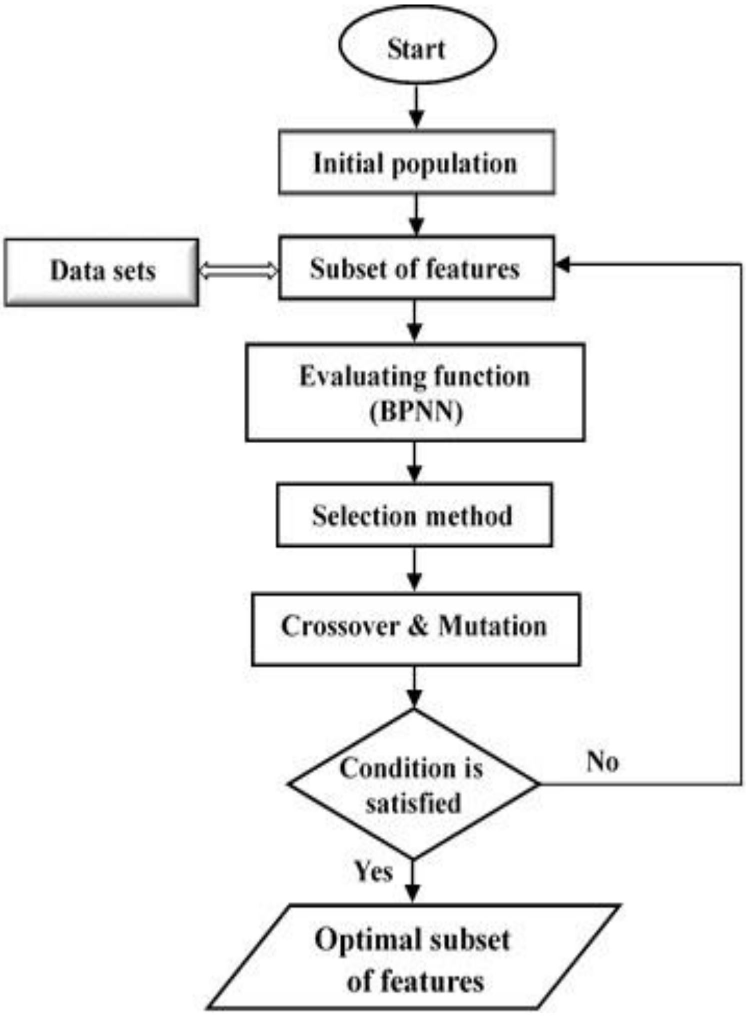

Figure 3: A flow chart of hybrid POA and NN in feature selection.

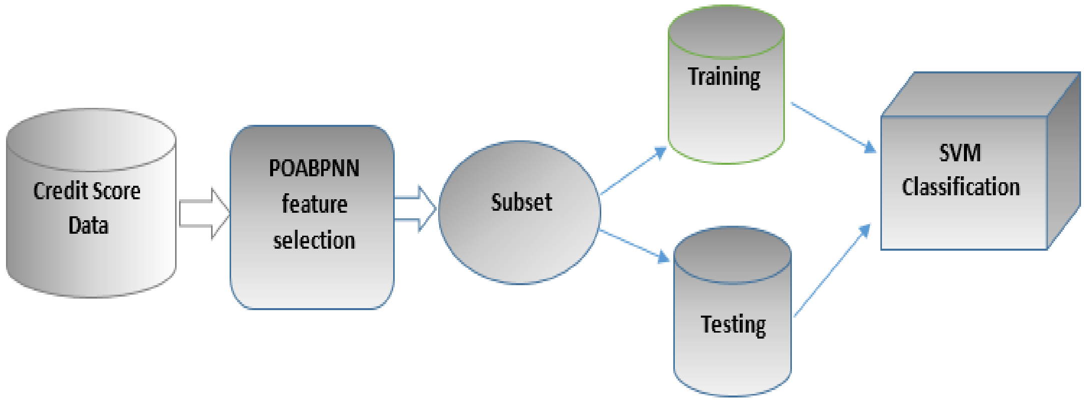

Figure 4: The block diagram of the proposed model.

Let $P_k = \{x_i \in R^N \ , \ i = 1,2,...n_k\}$ represents the set of genomes in the k^{th} generation, where k represents the number of generations and $n_k = 1,2,...k$. A pseudo-code for the BPPOA can be written as follows:

START

Create an initial population $p_1 = \{x_i \in R^N \ , \ i = 1,2,3,....n_1\}$

WHILE iteration number < Max number of iteration

  FOR each individual (chromosome), Evaluate_Fitness (p_i):

    Create BP_Neural Network

      BPNN_Train

      BPNN_Validate

      BPNN_Test

      Classification Accuracy (Fitness Function)

  END_FOR

  Genetic Operations

  Selection, Crossover, Mutation

END_ WHILE

Output _Fittest Chromosome (p_k)

END

3.Experimental results

The proposed algorithm, BPPOA, in this section, is evaluated by comparing it with POA-KNN and SVM methods.

3.1. Datasets

Our proposed approach has been tested using three publicly available benchmark credit-scoring datasets that were picked from the UCI machine learning repository (https://archive.ics.uci.edu/ml/index.php). The goal variable in the used datasets is a binary variable that represents the customer's credit status with good=1 and bad=0. The credit score datasets for Australia, Germany, and Japan are described in Table 2. The training and testing data sets were separated for each dataset. The evaluation of the classification utilized a training dataset (80% of all samples), while the evaluation of the external classification of the methods used a test dataset (20%).

Table 2: The description of the used datasets

Dataset	# Samples	# Features	Target class
Australian	690	14	2
German	1000	24	2
Japanese	690	15	2

4.2. Evaluation criteria

For the performance evaluation of the used methods, several criteria were calculated. All of these criteria are based on the confusion matrix. They are defined as:

(1) Classification accuracy (CA)

$$\text{Classification accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}}, \qquad (12)$$

(2) Type I error (T(I))

$$\text{Type I error} = \frac{\text{FP}}{\text{FP+TN}}, \qquad (13)$$

(3) Type II error (T(II))

$$\text{Type II error} = \frac{\text{FN}}{\text{TP+FN}}, \qquad (14)$$

(4) G-mean

$$\text{G-mean} = \sqrt{\frac{\text{FP} \times \text{FN}}{(\text{FP+TN})(\text{TP+FN})}}, \qquad (15)$$

where FP is the number of false positives, FN is the number of false negatives, TP is the number of true positives, and TN is the number of true negatives. Type I error displays the rate of classifying the bad credit status of the customer incorrectly into good credit status. Type II error demonstrates the frequency with which customers' good credit status is mistakenly classified as negative credit status. When the datasets are imbalanced, the G-mean criteria are employed to demonstrate the joint performance of sensitivity and specificity. The best classifiers are those with higher classification accuracy, g-mean, and lower values for both Type I error and Type II error, in accordance with these criteria.

4.3. Classification results

The outcomes of the BPPOA, POA-KNN, and SVM evaluation criteria are averaged over 20 times. The calculated findings for the German dataset are shown in Table 4, those for the Japanese dataset are given in Table 5, and those for the Australian dataset are given in Table 3. Table 6 also includes the corresponding outcomes from the test dataset. The average CA for the BPPOA of the Australian, German, and Japanese credit datasets is 92.326%, 94.885%, and 95.959%, respectively. The average CA for the POA-KNN is 87.909%, 79.197%, and 89.076%, and the average CA for the SVM is 80.372%, 67.093%, and 88.291%. Due to this, when compared to POA-KNN and SVM, our suggested hybrid algorithm achieves the best classification accuracy .

The overall number of characteristics is greatly decreased by employing BPPOA, going from 14 for the Australian credit dataset to 5, from 24 for the German credit dataset to 15, and from 15 for the Japanese credit dataset to 7. Additionally, compared to POA-KNN and SVM, the number of selected features needed to achieve BPPOA's classification accuracy is noticeably smaller. Using the three credit-scoring datasets, our suggested hybrid algorithm chose at least 30.834% fewer characteristics than the other two approaches .

Tables 3 through 5 show that, when compared to POA-KNN and SVM, the average Type I error and Type II error for the BPPOA are significantly reduced for all datasets studied. This suggests that BPPOA has the ability to discriminate favorably between undesirable and desirable applicants. The BPPOA, for example, obtains the best discriminating between the poor and good applicants of both Type I error and Type II error by 93.314% and 93.316% of POA-KNN and by 93.459% and 95.538% of SVM with regard to the German credit dataset. On the other hand, the suggested hybrid algorithm beats POA-KNN and SVM in terms of the average G-mean, which results in a good balance between sensitivity and specificity. This demonstrates that BPPOA is quite good at differentiating between good and poor clients .

Additionally, the BPPOA produces classification accuracy that is comparable to POA-KNN and SVM for the testing dataset. With a classification accuracy of 91.031% for the Australian credit dataset, BPPOA outperformed POA-KNN and SVM, which had classification accuracy of 87.917% and 79.254% respectively. In contrast, BPPOA identified the consumers for the German credit dataset with a CA of 92.581%, outperforming POA-KNN and SVM. Regarding the Japanese credit dataset, the BPPOA achieved a higher CA of 93.108% as opposed to POA-KNN's and SVM's, which were 87.105% and 86.806%, respectively .

Overall, it appears that the BPPOA hybrid technique we've suggested is suitable for categorizing credit datasets with good classification performance and few features .

Table 3: Comparison of the average evaluation criteria (%) of the used methods over the Australian training dataset. The number in parenthesis is the standard error.

	BPPOA	POA-KNN	SVM
CA	94.621 (0.102)	90.858 (0.112)	80.372 (0.213)
T(I)	8.325 (0.111)	9.969 (0.117)	19.141 (0.197)
T(II)	10.754 (0.108)	18.764 (0.116)	17.972 (0.199)
G-mean	95.101 (0.101)	89.789 (0.113)	79.873 (0.211)
No. selected features	7 (0.002)	11 (0.004)	All

Table 4: Comparison of the average evaluation criteria (%) of the used methods over the German training dataset. The number in parenthesis is the standard error.

	BPPOA	POA-KNN	SVM
CA	97.697 (0.132)	84.767 (0.142)	70.993 (0.334)
T(I)	5.475 (0.137)	40.098 (0.136)	43.945 (0.207)
T(II)	4.348 (0.131)	30.987 (0.135)	45.432 (0.211)
G-mean	95.043 (0.133)	86.999 (0.142)	86.987 (0.341)
No. selected features	19 (0.006)	21 (0.008)	All

Table 5: Comparison of the average evaluation criteria (%) of the used methods over the Japanese training dataset. The number in parenthesis is the standard error.

	BPPOA	POA-KNN	SVM
CA	98.212 (0.102)	92.879 (0.122)	91.691 (0.115)
T(I)	8.657 (0.107)	8.961 (0.127)	10.869 (0.117)
T(II)	9.647 (0.107)	19.207 (0.126)	18.132 (0.121)
G-mean	97.236 (0.101)	98.674 (0.124)	95.675 (0.114)
No. selected features	9 (0.005)	17 (0.005)	All

Table 6: Comparison of the average classification accuracy (%) of the testing dataset over three used datasets. The number in parenthesis is the standard error.

	BPPOA	POA-KNN	SVM
Australian	96.963 (0.053)	93.031 (0.062)	84.769 (0.071)
German	97.089 (0.057)	88.753 (0.061)	79.524 (0.071)
Japanese	98.247 (0.051)	93.690 (0.069)	90.864 (0.072)

4.4. Statistical Test

A non-parametric Friedman test was used to confirm our hybrid algorithm's capacity to choose the key features with high classification success. The training datasets' area under the curve (AUC) criterion was used to perform this test. Table (7) provides a summary of the statistical test findings for the post hoc Bonferroni test under three distinct critical values (0.01, 0.05, and 0.1) when the null hypothesis is rejected. Using the Friedman test statistic and the acquired results, the null hypothesis is rejected at a significance level of 0.05. According to the AUC criterion, this shows that there is statistical significance among the three employed techniques over the three utilized credit scoring datasets. Additionally, when compared to POA-KNN and SVM, the suggested algorithm, BPPOA, has the lowest average rank at 2.371. According to the findings of the Bonferroni test, it is obvious that POA-KNN and SVM have average ranks that are greater than $\alpha_{0.05}$ $\alpha_{0.01}$, and $\alpha_{0.10}$, and. According to these findings, POA-KNN and SVM perform noticeably worse than our suggested method on credit scoring datasets from Australia, Germany, and Japan.

Table 7: Friedman and Bonferroni test results over the three datasets

Friedman	Friedman test	Bonferroni test

	average rank		
BPPOA	2.371	$\chi^2_{\text{Friedman}} = 15.386$, p-value (0.05)=0.0021	$\alpha_{0.05} = 6.185$, $\alpha_{0.01} = 6.839$, $\alpha_{0.10} = 5.907$
POA-KNN	7.069		
SVM	10.152		

4.5. Comparisons with other proposed methods

To further highlight the performance of our proposed method, comparisons with other proposed methods for credit scoring in the literature are also presented in this paper. Our proposed method, BPPOA, outperformed these existing methods (see Table 8). It is clearly seen that our proposed method, BPPOA, yielded the highest classification accuracy in all datasets except that the proposed method by [38] yielded higher classification accuracy in the Australian dataset. Generally speaking, our proposed method is superior to other methods.

Table 8: Classification performance for several proposed methods in the literature

Year	Method (Reference)	Dataset Australian	German	Japanese
2014	(Oreski and Oreski, 2014)	-	78.90	-
2018	(Jadhav et al., 2018)	90.75	82.80	-
2019	(Tripathi et al., 2019)	93.85	88.42	84.51
2019	(Zhang et al., 2019)	86.16	74.83	86.38
2020	Our proposed method (BPPOA)	92.32	94.88	95.95

5. Conclusion

To accomplish feature selection and enhance the credit scoring classification, a hybrid method is presented in this research. The backpropagation neural network (BPNN) classifier and the pigeon optimization algorithm are combined to achieve this. Through three collections of well-known credit scoring datasets, the suggested hybrid algorithm was examined and contrasted with existing accepted techniques. Four features of the hybrid algorithm's classification criterion are presented: classification accuracy, Type I error, Type II error, and G-mean. The proposed algorithm is nominated as an effective feature selection approach that is helpful for credit scoring classification since it simultaneously meets these four requirements. Additionally, picking a few features could greatly enhance your categorization outcome. Overall, the applicability and utility of the proposed hybrid method in other kinds of classification datasets connected to a different field serve to highlight its superiority.

References

[1] Koutanaei FN,Sajedi H and Khanbabaei M. A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. Journal of Retailing and Consumer Services. 2015; 27: 11-23.

[2] Lunn DJ,Thomas A,Best N and Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and computing. 2000; 10: 325-337.

[3] Huang C-L,Chen M-C and Wang C-J. Credit scoring with a data mining approach based on support vector machines. Expert systems with applications. 2007; 33: 847-856.

[4] Al-Thanoon NA,Qasim OS and Algamal ZY. Tuning parameter estimation in SCAD-support vector machine using firefly algorithm with application in gene selection and cancer classification. Computers in biology and medicine. 2018; 103: 262-268.

[5] Guyon I and Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003; 3: 1157-1182.

[6] Bolón-Canedo V,Sánchez-Maroño N and Alonso-Betanzos A. An ensemble of filters and classifiers for microarray data classification. Pattern Recognition. 2012; 45: 531-539.

[7] Ferreira AJ and Figueiredo MAT. Efficient feature selection filters for high-dimensional data. Pattern Recognition Letters. 2012; 33: 1794-1804.
[8] Qasim OS and Algamal ZY. Feature Selection Using Different Transfer Functions for Binary Bat Algorithm. International Journal of Mathematical, Engineering and Management Sciences. 2020; 5: 697-706.
[9] Mai Q and Zou H. The Kolmogorov filter for variable screening in high-dimensional binary classification. Biometrika. 2013; 100: 229-234.
[10] Al-Talib ZS and Al-Azzawi SF. Projective and hybrid projective synchronization of 4-D hyperchaotic system via nonlinear controller strategy. Telkomnika. 2020; 18: 1012-1020.
[11] Kabir MM,Islam MM and Murase K. A new wrapper feature selection approach using neural network. Neurocomputing. 2010; 73: 3273-3283.
[12] Yu Z,Li L,Liu J and Han G. Hybrid adaptive classifier ensemble. IEEE transactions on cybernetics. 2015; 45: 177-190.
[13] Haoyang W,Changchun Z,Bingguo C and Junhua L. Adaptive pigeon optimization algorithm to Improve group premature convergence [J]. Journal of Xi'an Jiaotong University. 1999; 11: 007.
[14] Alhafedh MAA and Qasim OS. Two-Stage Gene Selection in Microarray Dataset Using Fuzzy Mutual Information and Binary Particle Swarm Optimization. Indian Journal of Forensic Medicine & Toxicology. 2019; 13: 1162-1171.
[15] Abed KA. Solving Kuramoto–Sivashinsky equation by the new iterative method and estimate the optimal parameters by using PSO algorithm. Indonesian Journal of Electrical Engineering and Computer Science. 2020; 19: 709-714.
[16] Steinwart I and Christmann A. 2008. Support vector machines: Springer Science & Business Media.
[17] Cristianini N and Shawe-Taylor J. 2000. An introduction to support vector machines and other kernel-based learning methods: Cambridge university press.
[18] Danenas P and Garsva G. Selection of Support Vector Machines based classifiers for credit risk domain. Expert Systems with Applications. 2015; 42: 3194-3204.
[19] Lu H,Chen J,Yan K,Jin Q,Xue Y and Gao Z. A hybrid feature selection algorithm for gene expression data classification. Neurocomputing. 2017; 256: 56-62.
[20] Al-Thanoon NA,Qasim OS and Algamal ZY. A new hybrid firefly algorithm and particle swarm optimization for tuning parameter estimation in penalized support vector machine with application in chemometrics. Chemometrics and Intelligent Laboratory Systems. 2019; 184: 142-152.
[21] Aickelin U and Dowsland K. Exploiting problem structure in a pigeon optimization algorithm approach to a nurse rostering problem. arXiv preprint arXiv:0802.2001. 2008.
[22] Deb K,Pratap A,Agarwal S and Meyarivan T. A fast and elitist multiobjective pigeon optimization algorithm: NSPOA -II. IEEE transactions on evolutionary computation. 2002; 6: 182-197.
[23] Kozeny V. Pigeon optimization algorithms for credit scoring: Alternative fitness function performance comparison. Expert Systems with Applications. 2015; 42: 2998-3004.
[24] Goh A. Back-propagation neural networks for modeling complex systems. Artificial Intelligence in Engineering. 1995; 9: 143-151.
[25] Hagan MT and Menhaj MB. Training feedforward networks with the Marquardt algorithm. IEEE transactions on Neural Networks. 1994; 5: 989-993.
[26] Gaxiola F,Melin P,Valdez F and Castillo O. Interval type-2 fuzzy weight adjustment for backpropagation neural networks with application in time series prediction. Information Sciences. 2014; 260: 1-14.
[27] Gu B and Sheng VS. A robust regularization path algorithm for $\nu $-support vector classification. IEEE Transactions on neural networks and learning systems. 2017; 28: 1241-1248.
[28] Kim K-j. Financial time series forecasting using support vector machines. Neurocomputing. 2003; 55: 307-319.
[29] Trabelsi I and Bouhlel MS. 2017. Feature selection for GUMI kernel-based SVM in speech emotion recognitionArtificial intelligence: concepts, methodologies, tools, and applications. IGI Global. p 941-953.
[30] Maldonado S,Pérez J and Bravo C. Cost-based feature selection for Support Vector Machines: An application in credit scoring. European Journal of Operational Research. 2017; 261: 656-665.
[31] Huang J,Cai Y and Xu X. A hybrid pigeon optimization algorithm for feature selection wrapper based on mutual information. Pattern Recognition Letters. 2007; 28: 1825-1844.
[32] Motieghader H,Najafi A,Sadeghi B and Masoudi-Nejad A. A hybrid gene selection algorithm for microarray cancer classification using pigeon optimization algorithm and learning automata. Informatics in Medicine Unlocked. 2017; 9: 246-254.
[33] Bohte SM,Kok JN and La Poutre H. Error-backpropagation in temporally encoded networks of spiking neurons. Neurocomputing. 2002; 48: 17-37.

[34] Min S-H,Lee J and Han I. Hybrid pigeon optimization algorithms and support vector machines for bankruptcy prediction. Expert Systems with Applications. 2006; 31: 652-660.
[35] https://archive.ics.uci.edu/ml/index.php.
[36] Oreski S and Oreski G. Pigeon optimization algorithm-based heuristic for feature selection in credit risk assessment. Expert Systems with Applications. 2014; 41: 2052-2064.
[37] Jadhav S,He H and Jenkins K. Information gain directed pigeon optimization algorithm wrapper feature selection for credit rating. Applied Soft Computing. 2018; 69: 541-553.
[38] Algamal, Z. Y., Abonazel, M. R., & Lukman, A. F. (2023). Modified Jackknife Ridge Estimator for Beta Regression Model With Application to Chemical Data. International Journal of Mathematics, Statistics, and Computer Science, 1, 15–24. https://doi.org/10.59543/ijmscs.v1i.7713