

## Innovative Computational PSO-EBGWO-CatBoost Approach for Assessing Loan Risks

Suihai Chen<sup>\*</sup>, Bong Chih How, Po Chan Chiu

Computer Science and Information Technology, Universiti Malaysia Sarawak (UNIMAS), Sarawak 943000, Malaysia

Corresponding Author Email: [C51772109@163.com](mailto:C51772109@163.com)

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijse.140429>

### ABSTRACT

**Received:** 4 June 2024

**Revised:** 12 August 2024

**Accepted:** 21 August 2024

**Available online:** 30 August 2024

#### **Keywords:**

*bank, loan risks, prediction, PSO-EBGWO-CB*

Loan risk evaluation is critical for the safety and expansion of financial institutions, but it poses substantial hurdles owing to the intricacy of the data involved. This paper provides an innovative computational approach, the Particle Swarm Optimization-Excited Binary Grey Wolf Optimization-CatBoost (PSO-EBGWO-CatBoost) method, which is intended to improve loan risk forecast accuracy. The proposed framework uses PSO for optimum feature selection, while EBGWO fine-tunes CatBoost's hyperparameters, resulting in better predictive efficiency. Before using the PSO-EBGWO-CatBoost model, the input dataset is preprocessed to remove outliers and missing values. The model's efficiency was verified using a loan dataset, and the findings showed outstanding results in loan risk estimate, with an accuracy of 81.23%, precision of 82.10%, and recall of 80.26%. These findings show that the suggested method greatly outperforms existing strategies, making it an effective instrument for loan risk handling in financial organizations.

## 1. INTRODUCTION

Appropriate loan risk assessment is an essential element of financial stability in the current evolving economic setting because banking institutions play an important role in economic growth. As a consequence, banks, financial institutions, and regulators alike have identified the development of credible loan risk forecasting algorithms as the highest priority. This obligation stems from the inherent ambiguity and complexity incorporated in lending operations, where anticipating and mitigating risks connected with borrower default is critical for protecting both the lenders and borrower's interests [1]. Bank loans are a crucial source of external finance for organizations and consumers facing financial limitations to grow their businesses. Commercial banks benefit greatly from lending to the economy, as loans make up a significant portion of their assets. Increased loan lending poses risks, including default and credit risk, which refers to a borrower's capacity to repay the loan on time and under agreed-upon terms. If the defaulter repays the loan, the creditor receives a profit [2]. Loan risk estimation is a knowledgeable concept, but it is interrelated with the fundamentals of banking and business. At its essence, the determination to anticipate and regulate the inherent risks elaborate with loaning, particularly the risk of borrower failure [3]. Financial institutions attempt to regulate borrowers' creditworthiness and the prospect of payback by applying in-depth knowledge of marketplace influences, borrower customs, and macroeconomic influences [4]. The import of loan risk forecast is far outside individual loaning choices. It has a considerable impact on the global economy. Unreliable risk assessment can trigger financial variability, with communication consequences that replicate across businesses

and continents. In contrast, strong risk prognostication methods increase the lending sector's confidence, permitting enhanced investment and promoting economic development [5]. The growth of diverse lending schemes, along with the growth of big data and artificial intelligence (AI), has resulted in exclusive risk dynamics that are difficult to change and provide creativity in handling risks [6]. Commercial banks, besides being well-capitalized, struggle to influence long-term clients because of inadequate market circumstances. Classifying and regulatory risks connected with these consumers is equally thought-provoking for organizations. Therefore, commercial banks' client borrowing procedures are slowly increasing [7]. Under this setting, the importance of stakeholder engagement and knowledge conversation produces stronger. Considering the interdependence of financial institutions, authorities, and market participants, concentrated efforts are needed to create an environment of collaboration in best practices, and knowledge can be shared and enhanced jointly. Furthermore, aligning regulatory structures with technology changes is critical for fostering innovation while ensuring economic security and consumer safety. However, there are several obstacles in loan risk prediction. Financial institutions face several challenges in the context of fast technological advancement and shifting regulations while transferring money from one person to the other person with security [8]. The goal of this study is to investigate how well loan risk forecasting techniques perform to enhance financial stability. To uncover a critical element that affects loan risk estimation by examining market dynamics, borrower behavior, and macroeconomic information. It aims to enlighten clients about optimal procedures for risk mitigation and sustainability lending methods by investing the influence between technological

advances and regulatory structures in predicting risks.

Despite advances in loan risk evaluation models, existing approaches generally struggle to effectively estimate risk because of ineffective feature selection and inadequate hyperparameter tuning. This paper fills a gap by presenting a novel strategy that combines Particle Swarm Optimization (PSO) for picking the most important features and Excited Binary Grey Wolf Optimization (EBGWO) for optimizing CatBoost hyperparameters. This novel combination improves the prediction efficiency of loan risk models, providing a more precise and dependable solution to financial organizations.

## 2. RELATED WORKS

Stacking + convolutional neural network (CNN) [9] was a loan risk prediction system that utilizes the CNN and stacking integrating approach. Applying CNN to obtain the local spatial attributes concealed in the stacking method outcomes improved the framework of adaptation capacity, allowing the stacking algorithm's capacity to generalize and the CNN method of feature mining capacity to be fully used. The Machine learning (ML) [10] algorithm was used to examine an enormous quantity of bank loan default information. The objective was to create a loan default prediction framework using the Back Propagation (BP) neural network. It detected the default risks earlier on and enabled aggressive risk reduction using coefficient assessments and significance testing. The concerns of asymmetry, moral hazard, and selection were raised by the significant influence that AI and ML [11] have on credit risk evaluation using data sources. The customer's behavior was examined and the customer's capacity to repay the loans. The outcomes of credit were granted to less fortunate individuals. To enhance that family's finances, get credit cards, advise the financial organization to include bank and credit lending places to increase their investments using the ML technique. An AI [12] model was proposed for credit risk management, specifically for quantifying the risks associated with borrowing credit via peer-to-peer lending platforms. Based on a set of comparable financial criteria, both hazardous and non-risky borrowers may be categorized, which can be utilized to clarify their financial rating and thus predict their subsequent conduct. According to an empirical examination of small and medium-sized businesses that applied for loans, penalized logistical tree regression (PLTR) [13] was a high-performing and easily comprehensible credit assessment approach that employs decision tree information to improve logistic regression's effectiveness. Formally, a penalized logistic regression model used the rules that were taken from several short-depth decision trees and were constructed using the original predictive variables as predictors. Maintaining the fundamental interpretability of the PLTR enabled it to capture non-linear effects that may occur in credit scoring data. Many factors need to make credit-scoring models comprehensible and provide a framework for transparent, auditable, and explicable black-box ML models [14]. Through this structure, they presented an overview of methodologies, demonstrated how to apply for loan scoring, and compared the outputs' accessibility to scorecards. The financial institutions ensure the estimation of the risk connected to the loan defaults. Lenders were more informed about the final approach, which was straightforward to integrate into the present loan processing systems. To revolutionize the way banks, handle their portfolios of loans

by advancing loan risk assessment strategies through ML [15]. To improve the accuracy and efficiency of identifying potential loan defaults. To properly evaluate and identify the prospective risk of default of clients before granting loans, that computes the chance of default of clients, which is the foundation and critical link of modern banking organizations of loan risk management strategies [16]. The statistical examination of historical loan data from banks and other financial organizations was primarily focused on the concept of imbalanced data categorization. Regarding examining loan applications, lending businesses have varied risk preferences. A rating-specific multifaceted ensemble learning structure has been presented. A credit rating-specific method of forecasting was utilized to create other models for different lending demographics with similar default probability. A multi-goal ensemble learning strategy was constructed with a One-Class support vector machine (SVM) [17], a successful unbalanced classification algorithm.

The related research investigated several machine learning approaches for improving loan risk prediction and credit evaluation. For example, using CNN with stacking approaches increased model generalization and feature extraction, resulting in loan default forecasts more flexible. Machine learning techniques such as Back Propagation Neural Networks (BPNN) and Penalized Logistical Tree Regression (PLTR) provided proactive risk reduction and interpretable credit evaluation, correspondingly, whilst AI-powered models concentrated on analyzing borrowers' behavior across various lending situations. These studies highlight the need to use sophisticated algorithms to provide more precise and accessible loan risk projections, especially when dealing with huge and complicated datasets.

Building on these results, the present study proposes a unique strategy that solves constraints noted in prior studies, like the necessity for improved feature selection and hyperparameter tuning. While previous research used different machine learning models to improve prediction accuracy, the PSO-EBGWO-CatBoost approach takes it a step further by integrating Particle Swarm Optimization for feature selection and Excited Binary Grey Wolf Optimization for hyperparameter tweaking. This technique not only increases prediction effectiveness; however, but also gives a more trustworthy framework for assessing loan risk, solving the model adaptability and efficacy issues raised in previous studies.

## 3. METHODOLOGY

In this section, we use a thorough approach to create and verify a novel loan risk forecasting system based on the combination of PSO, EBGWO, and CB. The theoretical basis for using the PSO method lies in its capacity to effectively explore the search space while avoiding local optima, which makes it appropriate for feature selection in high-dimensional data, as supported by Fakhouri et al. [18]. The EBGWO was selected for hyperparameter tuning because it is efficient at balancing exploration and exploitation, which is a vital aspect in improving complicated models [19]. CB was chosen for its capacity to manage categorical data while preventing overfitting, as demonstrated by Luo et al. [20]. The mixture of these techniques is intended to capitalize on their distinct advantages, resulting in a holistic framework that handles both feature selection and model effectiveness, resulting in higher

loan risk prediction accuracy.

The min-max scaler starts by preprocessing the Kaggle dataset and then uses PSO for characteristic determination and EBGWO for hyper-parameter optimization of the CB model. This section describes the data features, preprocessing procedures, and optimization procedures used to progress the accuracy and dependability of loan risk estimates. Figure 1 represents the study flow.

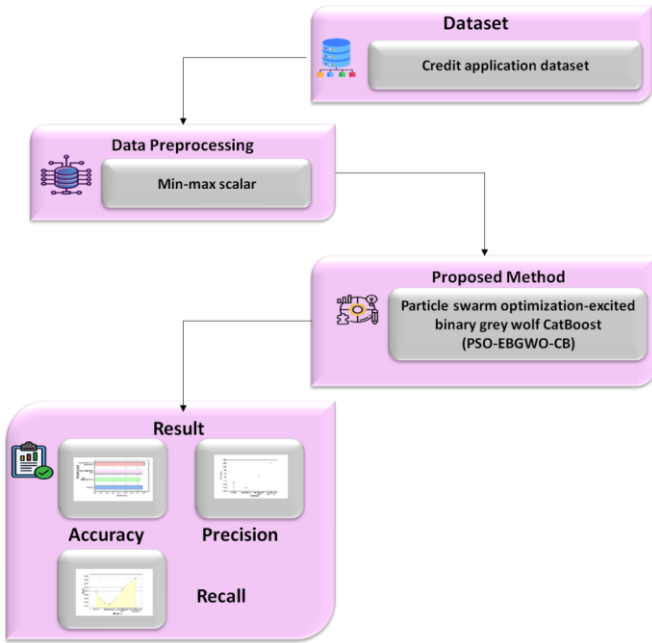


Figure 1. Cavity geometry

### 3.1 Dataset

This dataset (<https://www.kaggle.com/datasets/synthetic-credit-dataset>) is a compilation of credit applications that the data relevant to the Chinese lending sector. Though the data is fictional, it is designed to mirror real-world trends and patterns found in loan application procedures, making it appropriate for loan risk simulation, particularly in the setting of China's banking environment.

### 3.2 Data preprocessing using min-max scalar

The min-max scaler is used to exclude insignificant frequencies that keep the feature vectors inside a specific range of frequencies. The loan data has feature matrices ranging from 0 to 250. Applying the min-max transformation approach, the spectrum is turned through-2 and -3. Eq. (1) defines the min-max scaler procedure.

$$S_{min-max} = \frac{W - W_{min}}{W_{max} - W_{min}} \quad (1)$$

### 3.3 Proposed method

The PSO-EBGWO-CB creates an effective loan risk forecasting system. PSO actively seeks optimal solutions by analyzing and updating particle placements that depend on their fitness scores. EBGWO optimizes to select relevant data that produce more precise risk estimates. CB, the gradient boosting approach used to efficiently mix numerical and categorical data to result in accurate predictions. This

comprehensive method is intended to improve loan risk prediction of reliability and precision for banking institutions.

#### 3.3.1 PSO

PSO is a progressive, global, and chaotic optimization method that uses the social actions of users to intelligently seek the ideal solutions. PSO is suitable for a wide range of issues, including inconsistent, irregular, and multifunctional ones, as it requires no distinct optimization challenges or variations. The process starts with randomly placed particles in the issue area. Initially, every particle is allocated a random speed to predict the risk. Each particle has a fitness value based on its position. Enhancing a particle's position entails minimizing its fitness rating. The method evaluates the particle fitness, updates the velocity, and computes the new position throughout each iteration. A basic PSO technique involves a dimensional searching space and a swarming of  $w_j = (w_{j1}, w_{j2}, \dots, w_{jc})$  particles that are represented as  $\{w_j, \dots, w_n\}$ , where  $w_j = (w_{j1}, w_{j2}, \dots, w_{jc})$ . At iteration  $S$ , particle details are supplied as follows:

Position  $W_j^s = [W_{j1}^s, W_{j2}^s, \dots, W_{jd}^s]^s$  and speed  $U_{j1}^s = [U_{j1}^s, U_{j2}^s, \dots, U_{jc}^s]^s$ , represents the individual highest position  $[O_{j1}^s, O_{j2}^s, \dots, O_{jd}^s]^s$ , and global ideal position  $O_h^s = [O_{h1}^s, O_{h2}^s, \dots, O_{hc}^s]^s$ . At iterations  $O_h^s = [O_{h1}^s, O_{h2}^s, \dots, O_{hc}^s]^s S + 1$ , a particle velocity and location are modified as shown in Eq. (2) and (3):

$$U_{jc}^{s+1} = XU_{jc}^s + D_1 q_1^s (O_{jc}^s - W_{jc}^s) \quad (2)$$

$$W_{jc}^{s+1} = U_{jc}^s + W_{jc}^{s+1} \quad (3)$$

The coefficient  $\omega$  is a measure of inertia that was utilized for balancing exploration and extraction. Social learning parameters ( $D1$  and  $D2$ ) affect particle mobility and global optimal location.

#### 3.3.2 Excited binary grey wolf optimization (EBGWO)

The EBGWO is an innovative optimization technique based on grey wolf hunt behavior. EBGWO can improve loan risk forecasting efficiency by optimizing parameters for models and selecting appropriate data. Financial companies can employ this technique to make more accurate and dependable risk forecasts. One of the often-used strategies for this transition is EBGWO, which makes use of transfer functions. Eq. (4) illustrates the transformation functions that were employed to transform the true values for every solution into a binary.

$$W^i(s+1) = \begin{cases} 1, & \text{if } T(W^i(s+1)) > \rho \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where, the sigmoid functions, as defined by Eq. (5), are denoted by  $T$  and  $\rho[0, 1]$  represent a random boundary.

$$T(w) = \frac{1}{(1 + \exp(-10(w - 0.5)))} \quad (5)$$

A value of  $i^{\text{th}} W^i(s+1) = 1$  indicates that the component  $W^i(s+1)$  is chosen as instructive but a value of  $W^i(s+1) = 0$  indicates an associated  $i^{\text{th}}$  component which is disregarded. In this way, fewer characteristics are used without compromising categorization performance. Since the goal of

the employment is to improve accuracy in classification while using fewer features, Eq. (6) provides the function of the objective *Fit* that is used in this research.

$$Fit = \varepsilon * \frac{|T|}{|M|} - ((1 - \varepsilon) * Acc) \quad (6)$$

where,  $|T|$  is the dimension of the chosen feature subgroup,  $|M|$  is the total amount of characteristics in a dataset, and *Acc* denotes the correctness of a specific classifier. Specifically, the value of  $(1 - \varepsilon)$  represents the dimension of characteristic subgroup and mean reliability, respectively. Algorithm 1 presents the EBGWO pseudo code. It starts with a population of wolves and assesses how fit they are using a loan dataset that has been labeled. It determines a new wolf location and repeatedly modifies the control parameter. Until the maximum number of repetitions is reached, the procedure is repeated. The best fitness rates for loan risk forecasting and the appropriate position of the wolf are finally produced by the algorithm.

**Algorithm 1: Pseudocode of EBGWO**

Input: A labeled loan information *C*, the whole number of repetitions *MaxIter*, Populace extent *M*, Primary rate of the regulator variable  $b_{initial}$   
Output: Optimum Character’s location  $W_{\alpha}$ , Best fitness rate *Fit* ( $W_{\alpha}$ )  
Arbitrarily set *M* entities’ locations to create a populace utilizing Eq. (3), assess the fitness of all wolves, *Fit* (*W*)  
 $[\sim, Index] = Sort(Fit(W), 'Ascend')$   
 $E\alpha = Fit(W)_{Index(1)}$   
 $E\beta = Fit(W)_{Index(2)}$   
 $E\delta = Fit(W)_{Index(3)}$   
 $EX = Fit(W)_{Index(M)}$   
 $8W\alpha = W(Index(1))$   
 $9W\beta = W(index(2))$   
 $W\delta = W(index(3))$   
For *s* = 1 To *MaxIter*  
For *i* = 1 To *M*  
Regulate  $b_{j,s}$   
Calculate  $W_{vec1}, W_{vec2}$  and  $W_{vec3}$   
Produce  $W_{vec1}^{new}, W_{vec2}^{new}$  and  $W_{vec3}^{new}$   
 $EW_{vec1}$ ,  $EW_{vec2}$ ,  
 $EW_{vec3}$  of the binary vectors  $W_{vec1}^{new}$ ,  
 $W_{vec2}^{new}$  and  $W_{vec3}^{new}$  respectively,  
get the smallest (fittest) of all three examined fitness standards, as well as their index.  
If ( $fittest < Fit(W)_{Index(j)}$ ) then  
 $Fit(W)_{Index(j)} = fittest$   
Upgrade( $W)_{Index(j)}$   
EndIf  
Next *j*  
Repetition steps 3 to 10  
Next *s*

3.3.3 CatBoost (CB)

Binary decision trees serve as essential predictions in CatBoost, a type of gradient-boosting application. Gradient boosting is a powerful ML approach that successfully addresses the obstacles presented by different characteristics, noisy information, and complicated connections. Particularly, the CatBoost technique succeeds at handling both numeric and

category information. It uses an uncontrived decision tree architecture in which all non-terminal nodes from a similar position in the tree have a similar splitting criterion. This assures that the distance between the root node and every leaf is the same as the tree depth. Figure 2 provides a representation of CB.

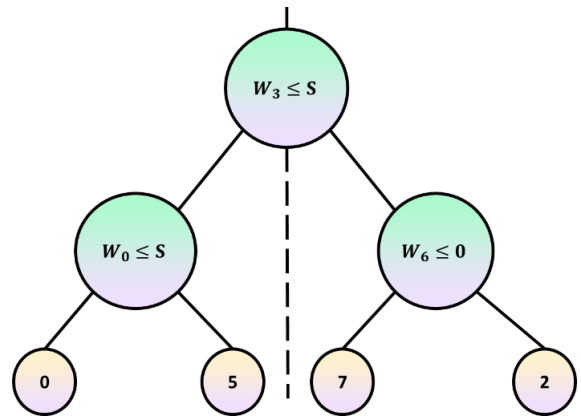


Figure 2. Representation of CB

CB uses sorted target parameters that handle category features. This method involves determining the goal value for every group by computing the matching output value using Eq. (7).

$$F(Z_j | w_j = w_{j,l}) \quad (7)$$

The number of  $j^{th}$  category variables in the  $l^{th}$  training instance from the data provided explodes by  $Z_j$ , resulting in the standard deviation of the output  $Z_j$  in Eq. (8).

$$\hat{w}_{j,l} = \frac{\sum_{i=1}^m [w_{j,i} = w_{j,l}] \cdot z_i + bO}{\sum_{i=1}^m [w_{j,i} = w_{j,l}] + b} \quad (8)$$

If the dot  $[\cdot]$  belongs to the inverting tags, the value  $[w_{j,i} = w_{j,l}]$  represents 1, and the value of  $w_j$  is equivalent to  $w_{j,l}$ , while the value is equivalent to 0. Variable *O* represents the chance of determining the standard category utilizing variable  $b > 0$ . Finally, we predict the risk by using our proposed method PSO-EBGWO-CB.

4. RESULT AND DISCUSSIONS

In this section, the effectiveness of our proposed loan risk assessment approach using PSO-EBGWO-CB was evaluated. We compare it to numerous well-established valuation approaches, such as particle swarm optimization with eXtreme gradient boost (PSO-XGBoost) [21], eXtreme gradient boost (XGBoost) [21], and random forest (RF) [21]. This study emphasizes three vital metrics: accuracy, precision, and recall. These indicators give a full estimation of predictive accuracy and validity for determining loan risk. Through this comparison research, we desire to validate the efficacy of our method in enlightening loan risk prediction accuracy and reliability, thereby showing its potential for use in financial organizations. Table 1 shows the obtained significant numerical outcomes of our proposed PSO-EBGWO-CB method.

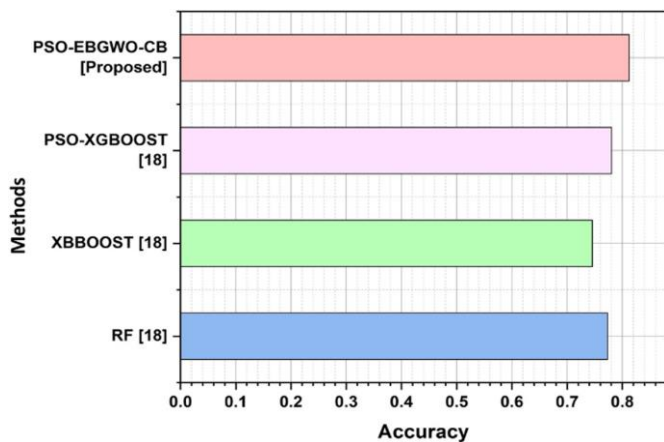
**Table 1.** Numerical outcomes of performance evaluation

| Method    | RF [21] | XGBOOST [21] | PSO-XGBOOST [21] | PSO-EBGWO-CB [Proposed] |
|-----------|---------|--------------|------------------|-------------------------|
| Recall    | 0.7676  | 0.7353       | 0.7745           | 0.8026                  |
| Precision | 0.7645  | 0.7498       | 0.7827           | 0.8210                  |
| Accuracy  | 0.7733  | 0.7458       | 0.7805           | 0.8123                  |

### 4.1 Accuracy

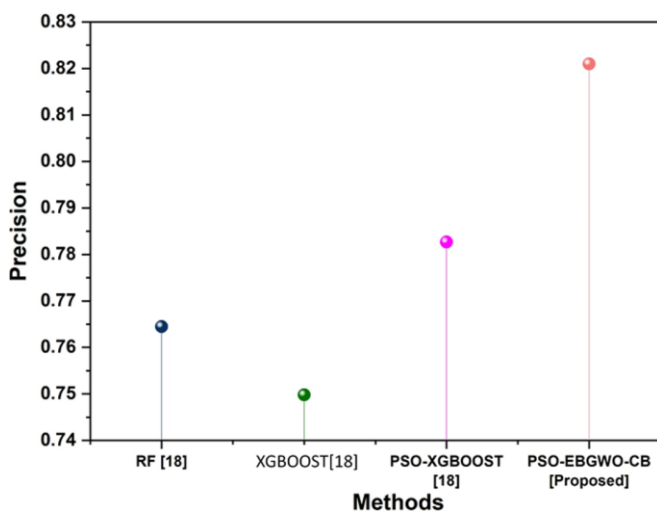
To estimate the accuracy of loan risk prediction, we contrast our proposed method, PSO-EBGWO-CB, with other methods. To determine the usefulness of our method in defining borrower risk, over measuring the accuracy aspect across several models of forecast, such as PSO-XGBoost, XGBoost, and RF. This comparative study establishes the PSO-EBGWO-CB that provides superior prediction accuracy (0.8123) in highlighting its capacity to provide more dependable and accurate loan risk valuations. Figure 3 depicts the graphical representation of accuracy outcomes.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$



**Figure 3.** Graphical representation of accuracy

### 4.2 Precision



**Figure 4.** Graphical representation of precision

The precision rate for our suggested method, PSO-EBGWO-CB was attained at 0.8210. This precision indicator

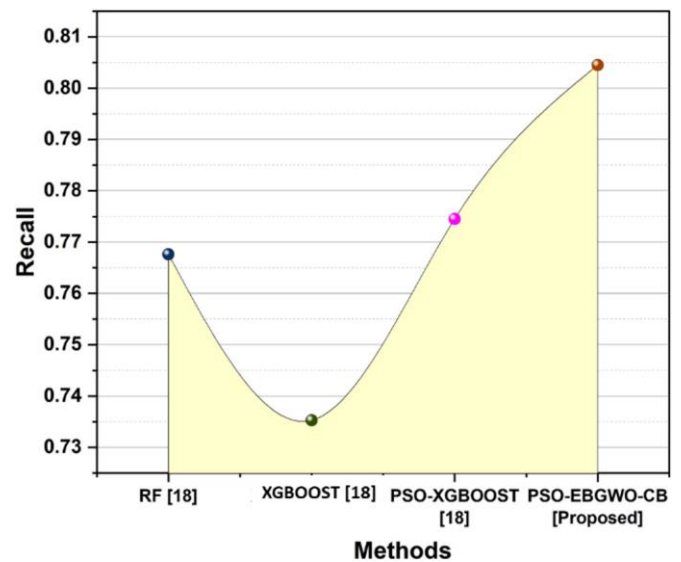
signifies the proportion of properly predictable positive cases, representing the model’s ability to effectually classify at-risk borrowers. To establish PSO-EBGWO-CB’s higher precision performance in knowing loan risks by contrasting this precision standard with the existing methods such as PSO-XGBoost, XGBoost, and RF. Figure 4 describes the graphical representation of precision outcomes.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

### 4.3 Recall

To assess the recall of loan risk prediction, we compare our proposed method, PSO-EBGWO-CB, to other approaches. We show the success of our approach by examining the recall parameter, which measures the model capacity that appropriately detects actual loan risks. Our proposed system has a recall value of 0.8026, showing that it was more effective at distinguishing true positives than PSO-XGBoost, XGBoost, and RF. This robust recall score proves PSO-EBGWO-CB’s ability to give accurate loan risk estimations while reducing the hazard of missing nonpayment clients. Figure 5 labels the graphical representation of recall outcomes.

$$\text{Recall} = \frac{FN}{FN + TP} \quad (11)$$



**Figure 5.** Graphical representation of recall

Through these assessment procedures, our suggested approach, PSO-EBGWO-CB, outperformed every other strategy in the field of risk prediction in lending behavior.

The performance measures show that the PSO-EBGWO-CatBoost model outperformed standard models in terms of accuracy (0.8123), precision (0.8210), and recall (0.8026). This enhancement can be credited to PSO's effective feature selection, which lowered the dataset's dimensionality while retaining crucial data, and EBGWO's hyperparameter

optimization, which fine-tuned the CatBoost model for the highest efficiency. CatBoost's capability to manage categorical attributes without considerable preprocessing further improved the model's prediction power.

When compared to other loan risk evaluation approaches, like the CNN-based stacking methodology, which mainly depends on spatial attributes, the suggested PSO-EBGWO-CatBoost model is more accurate. This can be due to its effective feature selection and parameter optimization procedures. Furthermore, while models like PLTR are interpretable, they lack the thorough optimization given by the incorporation of PSO and EBGWO, making our model better suited to complicated datasets. However, one disadvantage of this strategy is its computational burden, which might be tackled in further study by looking into more effective optimization algorithms. The findings indicate that incorporating this model into previous loan risk evaluation frameworks could greatly improve financial institutions' decision-making procedures.

## 5. CONCLUSIONS

This paper presents the PSO-EBGWO-CB technique, a unique computational technique for loan risk assessment. The fundamental novelty is the combination of PSO for optimum feature selection and EBGWO for hyperparameter tweaking of the CB model, which results in increased prediction accuracy and effectiveness. By tackling the complexities of loan risk evaluation, the proposed framework considerably outperforms previous techniques like PSO-XGBoost and Random Forest, achieving superior accuracy (0.8123), precision (0.8210), and recall (0.8026) when evaluated on a well-preprocessed Kaggle dataset.

The study emphasizes the need to use sophisticated optimization approaches to improve the predictive potential of machine learning models, especially in the field of finance, where precise risk evaluation is important. The results of this study highlight the PSO-EBGWO-CB technique's ability to transform loan risk administration processes, giving financial organizations a more dependable and flexible tool for creating informed lending choices.

Future studies will look into incorporating other optimization approaches to modify and vary the PSO-EBGWO-CB model, to improve its performance in a variety of loan risk prediction situations.

## REFERENCES

- [1] Samsir, S., Suparno, S., Giatman, M. (2020). Predicting the loan risk towards new customer applying data mining using nearest neighbor algorithm. *IOP Conference Series: Materials Science and Engineering*, 830(3): 032004. <https://doi.org/10.1088/1757-899X/830/3/032004>
- [2] Ndayisenga, T. (2021). Bank loan approval prediction using machine learning techniques. Doctoral dissertation, University of Rwanda.
- [3] Shingi, G. (2020). A federated learning based approach for loan defaults prediction. In *2020 International Conference on Data Mining Workshops (ICDMW)*, Sorrento, Italy, pp. 362-368. <https://doi.org/10.1109/ICDMW51313.2020.00057>
- [4] Lee, J.W., Lee, W.K., Sohn, S.Y. (2021). Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. *Expert Systems with Applications*, 168: 114411. <https://doi.org/10.1016/j.eswa.2020.114411>
- [5] Liang, K., He, J. (2020). Analyzing credit risk among Chinese P2P-lending businesses by integrating text-related soft information. *Electronic Commerce Research and Applications*, 40: 100947. <https://doi.org/10.1016/j.elerap.2020.100947>
- [6] Rishehchi Fayyaz, M., Rasouli, M.R., Amiri, B. (2021). A data-driven and network-aware approach for credit risk prediction in supply chain finance. *Industrial Management & Data Systems*, 121(4): 785-808. <https://doi.org/10.1108/IMDS-01-2020-0052>
- [7] Hu, B., Zhang, Z., Zhou, J., et al. (2020). Loan default analysis with multiplex graph learning. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Virtual Event, Ireland*, pp. 2525-2532. <https://doi.org/10.1145/3340531.3412724>
- [8] Doko, F., Kalajdziski, S., Mishkovski, I. (2021). Credit risk model based on central bank credit registry data. *Journal of Risk and Financial Management*, 14(3): 138. <https://doi.org/10.3390/jrfm14030138>
- [9] Li, M., Yan, C., Liu, W. (2021). The network loan risk prediction model based on Convolutional neural network and Stacking fusion model. *Applied Soft Computing*, 113: 107961. <https://doi.org/10.1016/j.asoc.2021.107961>
- [10] Gao, B. (2024). Financial loan default risk prediction based on big data analysis. In *2024 IEEE 4th International Conference on Power, Electronics and Computer Applications (ICPECA)*, Shenyang, China, pp. 398-403. <https://doi.org/10.1109/ICPECA60615.2024.10471076>
- [11] Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*, 9(3): 39. <https://doi.org/10.3390/ijfs9030039>
- [12] Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, 57(1): 203-216. <https://doi.org/10.1007/s10614-020-10042-0>
- [13] Dumitrescu, E., Hué, S., Hurlin, C., Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297(3): 1178-1192. <https://doi.org/10.1016/j.ejor.2021.06.053>
- [14] Bücke, M., Szepannek, G., Gosiewska, A., Biecek, P. (2022). Transparency, auditability, and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society*, 73(1): 70-90. <https://doi.org/10.1080/01605682.2021.1922098>
- [15] Robinson, N., Sindhvani, N. (2024). Loan default prediction using machine learning. In *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, pp. 1-5. <https://doi.org/10.1109/ICRITO61523.2024.10522232>
- [16] Gu, X., Chen, Y., Kang, Z., Wang, Y. (2022). Bad loan risk prediction algorithm in financial market. *Frontiers in Economics and Management*, 3(1): 614-621. [https://doi.org/10.6981/FEM.202201\\_3\(1\).0074](https://doi.org/10.6981/FEM.202201_3(1).0074)

- [17] Song, Y., Wang, Y., Ye, X., Zaretzki, R., Liu, C. (2023). Loan default prediction using a credit rating-specific and multi-objective ensemble learning scheme. *Information Sciences*, 629: 599-617. <https://doi.org/10.1016/j.ins.2023.02.014>
- [18] Fakhouri, H.N., Hudaib, A., Sleit, A. (2020). Multivector particle swarm optimization algorithm. *Soft Computing*, 24(15): 11695-11713. <https://doi.org/10.1007/s00500-019-04631-x>
- [19] Segera, D., Mbuthia, M., Nyete, A.M. (2020). An excited binary grey wolf optimizer for feature selection in highly dimensional datasets. In *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2020)*, pp. 125-133. <https://doi.org/10.5220/0009805101250133>
- [20] Luo, M., Wang, Y., Xie, Y., Zhou, L., Qiao, J., Qiu, S., Sun, Y. (2021). Combination of feature selection and catboost for prediction: The first application to the estimation of aboveground biomass. *Forests*, 12(2): 216. <https://doi.org/10.3390/f12020216>
- [21] Rao, C., Liu, Y. Goh, M. (2023). Credit risk assessment mechanism of personal auto loan based on PSO-XGBoost Model. *Complex & Intelligent Systems*, 9(2): 1391-1414. <https://doi.org/10.1007/s40747-022-00854-y>