Hindawi
Applied Computational Intelligence and Soft Computing
Volume 2022, Article ID 1468015, 10 pages
https://doi.org/10.1155/2022/1468015

Research Article

Financial Fraud Detection Approach Based on Firefly Optimization Algorithm and Support Vector Machine

Ajeet Singh,[1] **Anurag Jain**,[1] **and Seblewongel Esseynew Biable**[2]

[1] *University School of Information Communication and Technology, Guru Govind Singh Indraprastha University, Delhi, India*
[2] *Department of Information Systems, Debre Berhan University, Debre Birhan, Ethiopia*

Correspondence should be addressed to Seblewongel Esseynew Biable; sebess2011@gmail.com

Received 12 March 2022; Revised 8 May 2022; Accepted 9 May 2022; Published 17 June 2022

Academic Editor: Manikandan Ramachandran

The usage of credit cards is increasing daily for online transactions to buy and sell goods, and this has also increased the frequency of online credit card fraud. Credit card fraud has become a serious issue for financial institutions over the last decades. Recent research has developed a machine learning (ML)-based credit card fraud transaction system, but due to the high dimensionality of the feature vector and the issue of class imbalance in any credit card dataset, there is a need to adopt optimization techniques. In this paper, a new methodology has been proposed for detecting credit card fraud (financial fraud) that is a hybridization of the firefly bio-inspired optimization algorithm and a support vector machine (called FFSVM), which comprises two sequential levels. In the first level, the firefly algorithm (FFA) and the CfsSubsetEval feature section method have been applied to optimize the subset of features, while in the second level, the support vector machine classifier has been used to build the training model for the detection of credit card fraud cases. Furthermore, a comparative study has been performed between the proposed approach and the existing techniques. The proposed approach has achieved an accuracy of 85.65% and successfully classified 591 transactions, which is far better than the existing techniques. The proposed approach has enhanced classification accuracy, reduced incorrect classification of credit card transactions, and reduced misclassification costs. The evaluation results show that the proposed FFSVM method outperforms other nonoptimization machine learning techniques.

1. Introduction

The ongoing coronavirus disease (COVID-19) pandemic has thrown the global financial system into a loop, hastening the use of digital financial services and posing new hurdles in detecting financial fraud. Academics, industry, and regulatory agencies have all seen the tremendous losses caused by financial fraud. Financial fraud is a significant problem the finance industry faces, and it has affected our daily lives. Financial fraud is a criminal activity involving illegally obtaining goods and money for personal advantage. Financial fraud reduces confidence in the financial industry and affects people's cost of living. Financial frauds (FFs) are classified into five categories: financial statement fraud, insurance fraud, banking fraud, telecommunication fraud, and securities and commodities fraud. Financial fraud detection is the process of identifying and distinguishing between fraudulent and legitimate credit card transactions, insurance claims, and other types of financial transactions [1]. In 2013 [2], fraud was estimated to cost US retailers about $23 billion, but in 2014, the cost of fraud increased to $32 billion. The PwC survey 2016 report [3] showed that approximately 36% of institutions were victims of the economic crime in 2014 and 2015 worldwide. The economic crimes that appear most often in this study are asset misappropriation, cybercrimes, and money laundering. According to Nilson information, credit card losses reached approximately $21.84 billion and a 12% loss increased over 2015 [4]. Credit card fraud is expected to cost more than $35.5 billion globally in 2020, with owners suffering direct or indirect financial losses. Fraudsters in India have stolen the card information of approximately 70 million people and have sold it on the dark web.

This study focused on credit card fraud (CCF). Credit card fraud is a core part of banking fraud. A credit card (CC) is one of the most widely accepted forms of electronic payment worldwide. The development of credit cards has made online transactions easy, comfortable, and convenient for cardholders, but it has also offered new fraud opportunities for cybercriminals and enhanced the fraud rate. Credit card fraud is the illegal use of any system or criminal activity involving a physical card or card information without the cardholder's knowledge. Credit card fraud is one of today's issues. The quantity of fraudulent transactions committed each year has harmed several banks and financial organizations. A credit card is a physical medium that allows cardholders to pay for goods and services online or over the phone using their credit card. The global impact of credit card fraud is alarming. Many organizations and individuals have lost millions of dollars.

Credit card fraud detection relies on the automatic analysis of recorded transactions to detect fraudulent behavior. When a credit card is used, transaction data with various attributes (such as a credit card identifier, transaction date, recipient, and transaction amount) are kept in the service provider database. The most common types of CC fraud are credit application fraud and transaction fraud. Furthermore, credit card fraud is classified as card-not-present (CNP), counterfeit cards, lost or stolen cards, and identity fraud. The illegal use of credit card information for online purchases is called credit card transaction fraud. Credit card transactions are made physically or virtually [5]. Credit card application fraud can happen when someone uses a stolen or fake id. Also, credit card application fraud is called identity crime/fraud when fraudsters use stolen documents such as voter ids, pen cards, and passports, to apply for a new credit card. Cybercriminals use various techniques to obtain credit card information. Some methods are used to prevent application fraud, such as validating physical addresses, phone numbers, and answering questions.

Communal detection and spike detection are the most popular methods to prevent fraud at the application level. Online transaction fraud can be prevented by using AVS, CVV, 3D-Secure (3DS), EMV chip card, one-time password, and encryption and decryption methods [6,7]. These methods reduce card-not-present fraud but do not prevent fraud when caused by loss or stolen cards.

Machine learning techniques have become more powerful and cost-effective for tackling more complicated problems in our society because of the vast amount of data available to organizations and the expansion of hardware capacity. Various machine learning and data mining methods are used to detect CC applications (fraudulent applications) and transaction fraud. These machine learning techniques (MLTs) are the Bayesian network, decision tree, SVM, KNN, neural network, AIS, HMM, and SOM [8]. These techniques cannot gain good detection outcomes by directly applying sparse sample set modeling because these MLTs have some problems: undersampling, overfitting, and local optimal [9].

A bio-inspired algorithm with the help of a feature selection method can overcome all these problems. The feature selection (FS) method is necessary to select a good subset of features from a highly skewed dataset. The FS method is greatly needed before credit card fraud classification from a large credit card dataset [10].

The advantages of the FS method are that it makes it easy to understand data, reduces training time, and overcomes the curse of dimensionality issues. Bio-inspired algorithms are generally used for solving hard and combinatorial problems. Various algorithms, such as PSO, GA, and ant colony optimization, have successfully been applied to credit card fraud detection.

The firefly algorithm (FFA) is a famous and efficient bio-inspired optimization algorithm [11]. The FFA is a bio-inspired metaheuristic algorithm proposed by She [12]. FFA is used in this paper because it has a tendency to search both in local and global areas and has both types of search characteristics, i.e., exploration and exploitation.

The research contribution with respect to the research is as follows:

(i) A new approach has been developed based on the firefly bio-inspired algorithm, machine learning techniques such as SVM, and CFS as a feature selection method to detect credit card fraud cases (financial fraud). The firefly algorithm is used to handle the feature optimization process. The developed approach is, namely, FFSVM.

(ii) The developed approach reduces the misclassification cost and enhances classification accuracy. It is also used to maximize correctly categorized transactions and minimize the incorrectly categorized transactions of CC.

(iii) A comparative analysis has been performed between the proposed algorithm and three standard methods using an available real-world dataset.

The rest of the paper is organized as follows: Section 2 covers the related work of credit card fraud and feature selection (FS). Section 3 explains the research methodology for fraud detection. The proposed hybridization of firefly optimization and SVM algorithms is discussed in Section 4. The experimental setup and results of the analysis are evaluated in Section 5. Finally, Section 6 summarizes the conclusion and future work.

2. Related Work

This section presents the current machine learning, bio-inspired, and hybrid techniques used to detect anomalies in financial fraud. The NN, Bayesian network, HMM, DT, SVM, AIS, and KNN algorithms are the most frequently used machine learning techniques to detect financial fraud [13–16].

Sahin and Duman [17] developed a model based on SVM and a decision tree to solve a CCF detection problem. The famous decision tree methods C5.0, C&RT, CHAID, and SVM with four kernels (polynomial, sigmoid, linear, and RBF) are used. The classification accuracy of models lies between 83.02 and 94.76%. Lu and Ju [18] introduced the weight SVM (ICW-SVM) method for CCF detection. The PCA method has been used for feature selection, and the

ICW-SVM method has been used for classification. It has been discovered that the ICW-SVM approach can handle data imbalance. The model has produced a classification accuracy of 91.28%. Furthermore, the result has been compared with BN, C-SVM, and decision tree (C5.0) algorithms. ICW-SVM outperformed the BN, C-SVM, and decision tree algorithms. A compressive survey of ML techniques and nature-inspired techniques has been presented by Aderemi and Andronicus [19] for CCF detection. The various nature-inspired algorithms (AIS and GA), machine learning (SVM and HMM), and hybridized algorithms (ANN + SA, DT + SVM, and KNN + DT + NB) have been discussed in the survey. A novel hybrid fuzzy method has been proposed for the rule-based classification problem. If & then rules have been used in fuzzy logic. Bio-inspired optimization techniques have been used to improve the performance of the fraud detection system. The nine datasets have been used to compute the performance of the proposed approach. The hybrid method's performance and accuracy are very high compared with those of other techniques [20].

In 2008, a hybrid approach called SOM-PSO was introduced by Neill and Brabazon [21]. The PSO algorithm has been used to update the weights of the artificial NN, and the particles represent components of the mapping layer. The classification outcomes increased by using PSO, but the time and space complexity of the SOM-PSO approach also increased because SOM was applied first on the dataset rather than using PSO.

Arora and Kumar [22] have proposed a new hybridization of the SOM and POS methods for detecting credit card fraud. The time and space complexity challenge is solved using their proposed hybridized SOM + POS approach. The proposed SOM + PSO method incorporates both the SOM and PSO algorithms. The PSO method optimizes SOM outputs by updating the weight vector. An improved PSO algorithm has been suggested by Jie [23] to detect electronic transaction fraud. The main goal of the improved PSO algorithm is to detect fraudulent transaction cases. Wang et al. [24] proposed a model based on the whale optimization algorithm and backpropagation neural network (BPNN). They used a whale optimization algorithm to optimize the backpropagation neural network, and the proposed algorithm is called the WOA-BP algorithm. The WOA is first used to get an optimal initial value, and then, the BPNN algorithm is applied to correct the error value and obtain the optimal value. The outcomes of the WOA-BP algorithm have been compared with those of GA-BN and PSO-BN. The WOA-BP algorithm has high detection accuracy and fast convergence speed, which improves the accuracy of CCF detection. The WOA-BP algorithm has the smallest mean square error, the smallest number of interactions, and the fastest convergence rate. Duman and Ozcelik [25] have proposed new techniques that combine genetic algorithms and the scatter search method (GASS) to handle misclassification costs and fraud detection.

West et al. [26] have investigated various categories of financial fraud and fraud detection approaches. This study has also outlined issues and challenges related to existing research techniques and potential future study directions. Kalid et al. [27] have used multiple classifier systems (MCSs) with a cascading decision combination technique to detect fraud. The MCS has been put to the test on the credit card fraud dataset. The output of the first classifier has been used as an input for the second classifier, resulting in the samples being classified multiple times. The C4.5 and Naive Bayes algorithms have been employed for the first and second levels, respectively, and archived to detect fraud at a rate of 0.872. This result is satisfactory, but it could be better. Effective intelligent financial fraud detection methods have been thoroughly explained by Zhu et al. [28]. This study analyzed the new aspects of fraud risk brought on by the pandemic, as well as the evolution of data types employed in fraud detection, from quantitative tabular data to diverse unstructured data. The evolution of financial fraud detection methods has been summarized, focusing on new graph neural network methods in the postpandemic age. Finally, several significant difficulties and promising paths have been presented to motivate future research on intelligent financial fraud detection.

Nguyen et al. [29] have proposed CNN and LSTM deep learning method-based approach for credit card fraud detection and compared their performance with ANN, RF, and SVM machine learning techniques on European, Small, and Tall cards in three different financial datasets. In this study, sampling techniques have been applied to address the problem of the class imbalance, which improved performance on existing examples but reduced it dramatically on the new unseen data. Experimental results reveal that the proposed deep learning methods outperform traditional machine learning models when it came to detecting credit card fraud, implying that the proposed approaches can be used to detect credit card fraud in real-world situations. When all of the algorithms were compared, the LSTM with 50 blocks came out on top with an F1 score of 84.85%. Zhang et al. [30] proposed a convolutional neural network (deep learning)-based fraud detection model for online transactions that uses an input feature sequencing layer to reorganize raw transaction features into discrete convolutional patterns. According to the experimental result, the model exhibits great fraud detection performance without any derivative features. When compared to the present CNN for fraud detection, its precision and recall can be stabilized at roughly 91% and 94%, respectively, which is an increase of 26% and 2%, respectively.

West and Bhattacharya[31] presented a comprehensive review of data mining-based financial fraud detection research, with a focus on computational intelligence (CI)-based solutions. This study comprised around fifty articles of the scientific literature, most of which have been published between 2004 and 2014. Because no prior review articles examined the link between fraud categories, CI-based detection techniques, and their performance as reported in the literature, a research gap has identified. This study classified and analyzed current fraud detection literature based on key factors such as the detection algorithm used, the fraud type investigated, and the efficacy of detection methodologies for specific financial fraud types. Some of the most important issues and challenges related to existing research procedures have also been outlined.

3. Research Methodology

The research methodology of the financial fraud detection system passes through 4 phases as shown in Figure 1. The first phase discusses the data preprocessing process, which removes redundancy and noise in the data. The second phase is a feature selection phase and is responsible for selecting various features based on the target features A15 (1, 0). The correlation-based feature selection (CFS) method and firefly optimization algorithm are used to select a more relevant subset of features among all the features in a dataset. These features are called "best-selected subset features" or "optimal solutions."

In the third phase, the SVM classifier is applied to the best-selected features for calculating the classification accuracy. The proposed hybrid FFSVM algorithm combines phases 2 and 3. Finally, various performance metrics are used to evaluate the proposed algorithm.

This paper introduces a new method for financial fraud detection called CCF, which combines the firefly algorithm (FFA) with CFS and SVM (FFSVM). The classification technique SVM is used to build the training model and perform classification on the ten cross-validations. The proposed algorithm is housed in the WEKA + Eclipse tool. Furthermore, it is evaluated on the CCF dataset and results are compared with other techniques.

3.1. Proposed Fraud Detection Approach. This section discusses the proposed new hybrid optimization algorithm using the firefly algorithm, the feature selection (CFS) method, and the SVM technique. It has been identified and observed that data mining and machine learning algorithms cannot efficiently handle classification problems and misclassification issues. Furthermore, minimum value transactions must not be taken too lightly, while maximum value transactions have more impact. This research aims to minimize the incidence of these types of problems by using bio-inspired techniques.

3.2. Support Vector Machine. The support vector machine (SVM) is a supervised learning method used for classification, regression, pattern recognition, and outlier detection [20]. In this paper, the SVM performs prediction and classification on the credit card dataset. The SVM algorithm is based on the concept of the hyperplane as a decision maker, which divides credit card transactions into two classes: fraud and genuine transactions. The two essential properties of SVM strength are kernel representation and margin optimization.

3.3. Optimal Hyper Plane. An optimal hyper plane is described as

$$Y_i\left[\left(v^T x_i\right) + b\right] \geq 1, \quad i = 1, 2, 3, \ldots n. \tag{1}$$

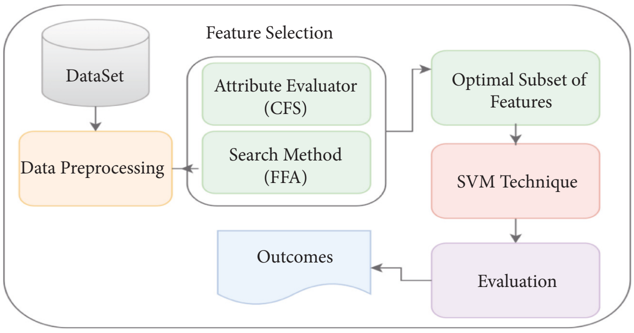

Figure 1: Detailed description of the proposed model for the fraud classification based on a machine learning technique and a bio-inspired algorithm.

Equation (2) presents the objective function:

$$\varphi(vv) = \|(vv)\|, \tag{2}$$

where v is the normal vector to the hyperplane; x_i is CC transactions that want to be classified; and y_i is either 1 or −1, each indicating the type of transaction φ to which the point x_i belongs. The kernel function is described as follows:

$$K(x_1, x_2) = (\varnothing(x_1), \varnothing(x_2)), \tag{3}$$

where $\varnothing$: $X \longrightarrow D$ maps transactions in input space X to higher dimensional space D. The hyperplane is used to separate the transactions after applying KF to the CC dataset which obtains the following form:

$$\{v, x\} + b = 0. \tag{4}$$

The classification for an SVM defined as follows:

$$\sum_i^n (\propto_i, y_i, k(x_i, x) + b) = 0. \tag{5}$$

The kernel function (KF) preference is dependent on the dataset and classification requirements. There are six universally used kernel functions: Gaussian radial basis function (RBF), polynomial function, normalized polynomial kernel, precomputed kernel (matrix kernel), PUK, and string kernel. In this paper, the polynomial kernel function $(k(x, y) \leq x,\ y * \wedge D)$ is used for classification task. In this work, the SVM algorithm is applied to best-selected features to build a classification model. The SVM became popular for solving classification problems and also it has the capability of handling high dimensional Australian datasets.

3.4. Feature Selection Process. Feature selection (FS) is an essential step executed before the classification procedure. The FS is a technique used for eliminating useless, redundant features and selecting the most important subset of features from a credit card dataset [10]. The primary purposes of FS methods are to improve classification performance, such as accuracy and comprehensibility, by building simpler and understandable data. Feature selection methods are of two types: the wrapper and the filter method. Features are scored based on evaluation criteria, and the lowest-scoring features are removed from a dataset. The wrapper method is more expensive because the search space could be enormous. Due to this, the filter method has been used in this study. Filter methods measure the importance of features by their correlation with a dependent feature. Filter methods use a statistical measure to assign a score to each feature and evaluate a subset of features. The features are ranked by the score and are either selected to be removed or not from the dataset. In the proposed research, the CFS subset feature evaluator method and the greedy stepwise, best first, genetic algorithm, and proposed algorithm (FFSVM) as a search method have been used to select the best features amongst existing features. Finally, an SVM classifier is applied to the best-selected subset of features to calculate the proposed algorithm's performance (Table 1).

3.4.1. Correlation-Based Feature Selection (CFS). CFS is an algorithm used to rank the feature subsets in the search space according to the correlation-based heuristic function in equation (6). CFS evaluates the subset of features that are highly correlated with the target feature and not correlated with each other [32].

CFS uses the interdependency or predictability of one feature with another feature to create the optimal subset of features to improve classification performance and reduce the feature dimension.

$$f_s = \frac{K^* \, r_{\mathrm{cf}}^-}{\sqrt{K + K\,(K-1) r_{\mathrm{ff}}^-}}, \tag{6}$$

where f_s is the heuristic "merit" of feature subset S holding k features, r_{cf}^- is the mean feature-class correlation ($f\epsilon S$), and r_{ff}^- is the average feature-feature intercorrelation [3].

3.5. Firefly Algorithm. The firefly algorithm (FFA) is inspired by firefly flashing patterns and behavior. The FFA used the three idealized rules proposed by [11,12]. The following function is used to calculate the firefly's attractiveness:

$$\beta(r) = \beta o\, e^{-\gamma r2}, \tag{7}$$

where βo is attractiveness between fireflies, r is the distance between 2 fireflies when $r = 0$, and γ is parameter characteristics ("light absorption coefficient). The distance between 2 fireflies i and j at x_i and x_j is calculated by the following formula:

$$r_{i,j} = \left\| x_i - x_j \right\| = \sqrt{\sum_{k=1}^{d} \left(x_i, k - x_j, k \right)^2}, \tag{8}$$

where x_ik is the kth component (transaction) of ith firefly and x_jk is the kth component (transaction) of jth firefly. d is the dimension of the firefly. The movement of the firefly is entirely dependent on its attractiveness. The firefly i can be a step toward the firefly j if and only if the attractiveness of the firefly j is greater than that of the firefly i. The following formula depicts the movement of a firefly i (new position):

$$x_i^{t+1} = x_i^t + \beta(r)^* \left(x_j^t - x_i^t \right) + \alpha \left(\mathrm{rand} - \frac{1}{2} \right), \tag{9}$$

where t represents the number of iterations and $\beta(r)$ is the attractiveness function with $m = 2$. rand is a random number. α is called randomization parameter range [0, 1]. The coefficient α represents a random number controlling the size of the random walk.

3.6. Hybridization of the FFA and SVM Technique. This section presents the working flow of the proposed hybridization algorithm. Cybercriminals are applying a sophisticated technique to get credit card details and steal the physical card. There is an urgent need for an adaptive and dynamic technique to have the ability to learn fraudulent patterns and minimize misclassification costs. In this paper, each firefly means credit card transactions. The proposed hybridization (FFSVM) algorithm is summarized based on 3 rules of the firefly algorithm, correlation base feature selection, and the SVM technique. Algorithm 1 presents the hybridization (FFSVM) algorithm.

Algorithm 1: Hybridization of the support vector machine and firefly optimization algorithm (FFSVM) to credit card fraud classification using feature optimization.

The proposed hybrid optimization approach outperformed the benchmark credit card dataset in terms of detecting credit card fraud and resolving misclassification issues.

4. Experimental Results and Discussion

A study has been performed to investigate the significance of our approach in detecting credit card fraud. An experiment conducted using cutting-edge machine learning techniques is reported in the literature for comparison of results. The experiments are run on a Windows (10) machine with an Intel (*R*) Core (TM) i-3-2310M CPU @ 2.10 GHz and 4 GB of RAM. The WEKA data mining framework + Eclipse has been used to perform all the experiments with default parameter values of Weka. The 10-fold cross-validation method has been used to validate and estimate the classification accuracy of the proposed algorithm. The test was performed on a credit card dataset. The firefly algorithm parameters have been set to alpha (α) = 0.6, beta (β) = 1, and gamma (γ) = 0.1. Firefly and genetic swarm algorithms use 20 iterations and 20 population sizes to get accurate fitness values and precise convergence rates for the experiment.

4.1. Dataset Used. In this study, the Australian credit data have been collected from the UCI-ML Repository [33]. The dataset contains 690 credit applications and 15 features. There are six numerical features, eight categorical features, and 1 class feature. The dataset is unbalanced because it contains 44.5% good (legitimate) and 55.5% bad (fraud) applications (see Table 2). Table 3 shows the best-selected subset of features by the proposed algorithm and another algorithm from the Australian dataset. Both algorithms

Table 1: Description of the feature selection method and classification technique used for fraud detection.

Feature evaluator method	Search method	Classification technique
CfsSubsetEval	FFA	SVM
CfsSubsetEval	Genetic algorithm	SVM
CfsSubsetEval	Greedy stepwise (forward)	SVM
CfsSubsetEval	Best first search	SVM

Table 2: Characteristics of the Australian credit dataset.

No.of instances	Legitimate	Fraud	Features	Class type
690	307	383	15	1 and 2 (+& −)

Table 3: Detailed description of a best-selected subset of features that are used for credit card fraud detection.

CFS technique	Features	Description
FFA	A4	{1, 2, 3}
	A5	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14}
	A7	Numeric
	A8	{1, 0}
	A9	{1, 0}
	A13	Numeric
	A14	Numeric
BS, GS, and GA	A4	{1, 2, 3}
	A5	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14}
	A7	Numeric
	A8	{1, 0}
	A10	Numeric
	A13	Numeric
	A14	Numeric

select the seven best subsets of features, but only one feature is different from the other algorithm. Finally, we perform an SVM classifier on seven features to determine classification accuracy.

4.2. Evaluation Criteria. The performance evaluation measures are discussed in this section to validate the proposed approach. The accuracy rate is not sufficient to measure the performance of the proposed hybridization algorithm. The performance of the proposed algorithm is validated in terms of PPV, EER, TPR, accuracy rate, ROC, MCC, and *f*-measures based on Table 4 [5,10]. The confusion matrix presents the predicted class versus the actual class outcomes in Table 4:

(i) Positive predictive value (PPV)/ precision = TP/ (TP + FP)

(ii) Recall/TPR/sensitivity = TP/ (TP + FN)

(iii) *F*-measure = 2 ∗ (recall ∗ precision)/ (recall + precision)

(iv) Accuracy rate = (TP + TN)/ (TP + TN + FP + FN)

(v) Specificity/TNR = TN/ (TN + FP)

(vi) Error rate (ERR)/misclassification rate

Table 4: Demonstration of the confusion matrix, which is used to detect credit card fraud.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN
	Total	*P* = TP + FN	*N*=FN + TN

$$\text{ERR} = \frac{(\text{FP} + \text{FN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \tag{10}$$

(vii) Relative operating characteristic (ROC) curve: it is a comparison of TPR and FPR as the criterion changes

4.3. Results and Discussion. This section analyzes the empirical outcomes obtained by the proposed FFSVM algorithm and three other algorithms: BSVM, GSVM, and GASVM. The proposed FFSVM algorithm has been applied to the Australian dataset. The performance of the FFSVM algorithm has been evaluated using a variety of performance measures. Tables 5 and 6 illustrate the outcomes of the four algorithms (FFSVM, BSVM, GSVM, and GASVM) based on PPV, TPR, accuracy, precision, *f*-measure, error rate, and MCC. All experiments have been implemented in WEKA + Eclipse. The classification time of FFSVM is 0.83 seconds less than the other methods. Figures 2–7 show the comparison of the four algorithms based on the positive predictive value (PPV), true positive rate (TPR) or recall, *f*-measure, accuracy, error rate, fraud detection rate, ROC, and MCC performance metric.

The results of GSVM are very close to those of BSVM. The proposed hybrid algorithm has 85.65% accuracy. The PPV, TPR, *f*-measure, error rate value, and accuracy are almost similar for BSVM and GSVM. Also, the PPV value of the proposed algorithm is the lowest among all other techniques, which means that the proposed algorithm is the most accurate in classifying transactions correctly into their respective fraud and nonfraud classes. The TPR value of the proposed algorithm is the highest, which is a direct compute of the accuracy of an algorithm, among all other techniques. The FFSVM algorithm's performance is slightly higher than that of BSVM, GSVM, and GASVM because of the best subset of features selected. The best seven features are selected from the high dimensional dataset. Table 7 shows the abbreviations for their real names used in this paper.

Figure 2 shows the accuracy of four different techniques. It clearly shows that the proposed FFSVM technique

Table 5: Outcomes of the proposed algorithm and the three other algorithms (BSVM, GSVM, and GASVM) based on various parameters such as PPV, TPR, *f*-measure, accuracy, error rate, experimental time, and best-selected feature (FS).

Parameters	FFSVM	BSVM	GSVM	GASVM
PPV	0.795	0.815	0.815	0.798
TPR	0.912	0.873	0.873	0.886
f-measure	0.850	0.843	0.843	0.840
Accuracy	85.65%	85.50%	85.51%	84.93%
Error rate	14.34%	14.50%	14.49%	15.07%
Exp. time	0.83	0.32	0.22	0.34
FS	7	7	7	7

Table 6: Outcomes of the proposed algorithm and the three other algorithms (BSVM, GSVM, and GASVM) based on various parameters such as CCI, ROC, MCC, and instance classification.

Model	CCI	ICI	ROC	MCC
FFSVM	591	99	0.862	0.72
BSVM	590	100	0.857	0.71
GSVM	590	100	0.857	0.71
GASVM	586	104	0.853	0.702

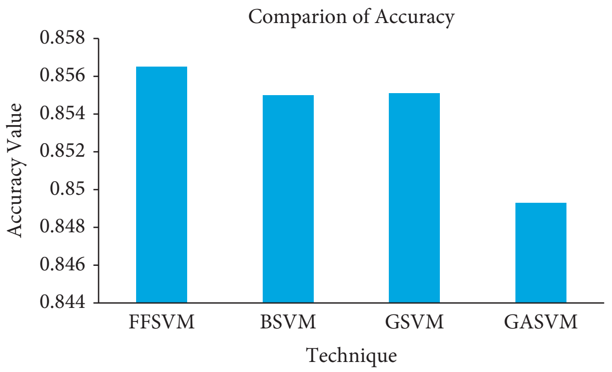

Figure 2: Classification accuracy comparison between the proposed algorithm and three other search algorithms (BSVM, GSVM, and GASVM) for fraud detection which shows the proposed algorithm to be the most accurate.

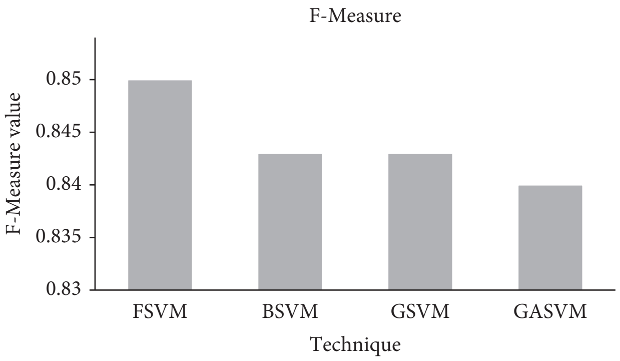

Figure 3: *f*-measure performance of four search algorithms. The *f*-measure is calculated by weighing the harmonic means of precision and recall equally. It allows a model to be evaluated with a single score that accounts for both accuracy and recall, which is valuable for comparing models and summarizing their performance.

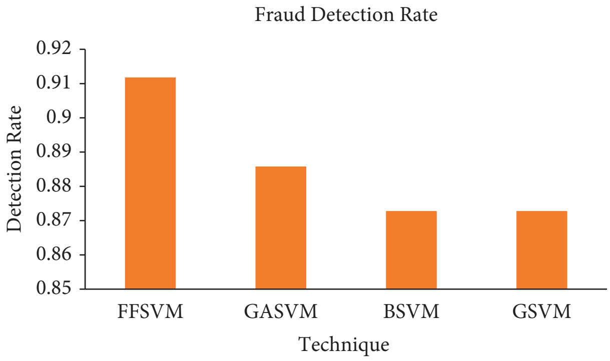

Figure 4: Comparison between the proposed algorithm and three other search algorithms (BSVM, GSVM, and GASVM) in terms of the fraud detection rate.

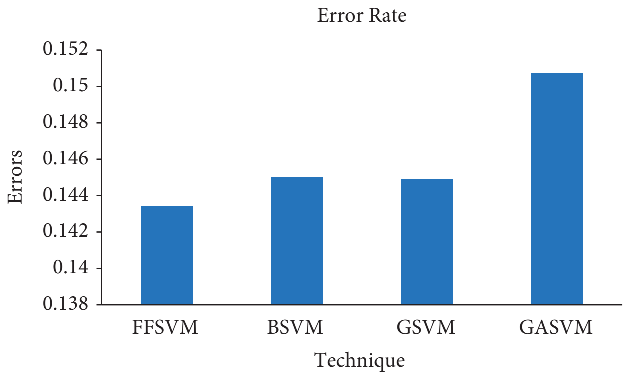

Figure 5: Comparison of error rates between the proposed algorithm and three other search algorithms (BSVM, GSVM, and GASVM). The error rates of the BSVM and GSVM algorithms are about to 14.50%. However, in the case of the GASVM algorithm, which has a high error rate (15.07%), it means the performance of a fraud detection algorithm can be decreased. The proposed algorithm has achieved a much lower error rate (14.34%). It means that the financial fraud detection correctness is high, and the most accurate way to detect fraud is by FFSVM.

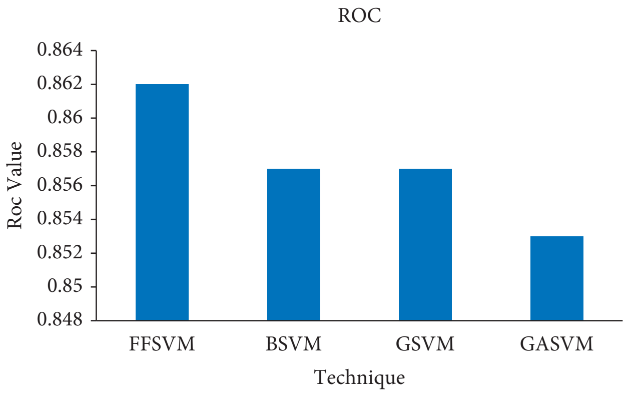

Figure 6: Comparative analysis of ROC of FFSVM, BSVM, GSVM, and GASVM techniques used to determine the quality and effeteness of the prediction techniques.

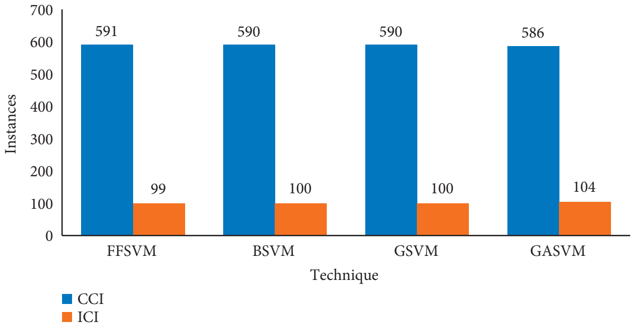

Figure 7: Classification of credit card transactions as correctly classified instances (CCI) and incorrectly classified instances (ICI). The outcomes show that the proposed approach has performed well, with the highest correctly classified instances of 591 and incorrectly classified instances of 99 over 690 CC applications than the three different hybrid algorithms.

Table 7: Abbreviation used in this research paper.

Abbreviation	Corresponding name
CCI	Correctly classified instances
ICI	Incorrectly classified instances
FS	Feature selected
BSVM	Best first search with SVM
GSVM	Greedy stepwise with SVM
GASVM	Genetic algorithm with SVM
FFSVM	Proposed firefly algorithm with SVM
CC	Credit card
FF	Financial fraud
FFA	Firefly algorithm
SVM	Support vector machine
CFS	Correlation-based feature selection

Table 8: Comparative analysis of the proposed method and the existing method.

Approach	Feature selection	Accuracy%
FFSVM (Proposed)	CfsSubsetEval method	85.65
SOM-PSO [34]	No	75.25
ANFIS-PSO [35]	No	79.2
HKFA [36]	No	72.57

achieved better accurate (85.65%) results than the BSVM, GSVM, and GASVM.

The f-measure value of the four different algorithms is presented in Figure 3. The FFSVM algorithm has the highest f-measures rate of 0.850% compared to the other three algorithms for detecting financial fraud.

The proposed algorithm has a higher fraud detection rate (91.20%) than the BSVM, GSVM, and GASVM (see Figure 4). It means that the algorithm is the most accurate in classifying instances correctly.

Figure 6 demonstrates the ROC curve value of four algorithms. The proposed FFSVM approach got an incredible ROC curve value of 0.867%. The performance of FFSVM is better than that of BSVM, GSVM, and GASVM.

4.4. Comparison of Proposed Approach Performance with the Existing Techniques

4.4. Comparison of Proposed Approach Performance with the Existing Techniques. This section compares the proposed hybridization approach with the existing fraud detection method based on the feature selection method. The comparison is shown in Table 8, and it can be seen that the feature selection approach is not employed in any of the existing methods (SOM-PSO, ANFIS-PSO, and HKFA). It is clear that the proposed strategy improves the performance of the fraud detection system. With a true positive rate of 0.912, an error rate of 14.34, and an accuracy of 85.65, it detects all fraud cases. These findings show that the proposed strategy outperforms the existing model regarding classification accuracy and true positive rate, indicating that fraudulent instances are accurately categorized.

As a result, the proposed strategy improves the performance by reducing the error rate, incorrectly classified instances, and improving fraud case detection accuracy.

5. Conclusion

Financial fraud is a form of theft that occurs when someone takes money for personal gain. This study has focused on credit card fraud detection in the banking domain (financial fraud). Detecting credit card fraud is a classification issue in which machine learning methods and bio-inspired algorithms have been used to classify a transaction as fraudulent or legitimate.

The optimization is needed to understand the feature; else, the framework will include all the features (the features that are not more relevant), and the framework's performance will be decreased.

Keeping the same track in view, a novel hybridization FFSVM approach for building a credit card fraud detection system has been proposed that combines the firefly algorithm, feature selection, and the SVM algorithm. The CfsSubsetEval as a feature selection method and the firefly algorithm as a search method have been used to address the newly built-up algorithm. The CfsSubsetEval feature selection method with the firefly algorithm has optimized the best seven subsets of features out of the 15 features in the dataset. This method has optimized the best subsets of features.

The firefly algorithm-based approach has optimized far better features than nonoptimization algorithms. Because this optimization technique can search both local and global areas, it has both exploration and exploitation search characteristics. The firefly optimization algorithm is an efficient technique that can handle multimodality and has two key advantages: automatic and automatic subdivision of swarms.

Finally, an SVM machine learning algorithm has been applied to the best subset of the features for classifying credit transactions as fraudulent or legitimate. A support vector machine (SVM) is a binary classifier. The support vector machine has been applied to distinguish between fraudulent and authentic transactions using the hyperplane and polynomial kernel function.

The proposed hybridization approach has been evaluated on the Australian credit card approval dataset. The performance of the proposed hybridization FFSVM approach has been compared with that of the BSVM, GSVM machine learning algorithm, and GASVM bio-inspired algorithm. This approach has increased correctly classified credit transactions and decreased incorrectly classified credit transactions. This approach also achieved a higher fraud detection rate than other algorithms. It signifies that the proposed approach is the most accurate in correctly classifying fraud and legitimate instances. Experimental results show that the proposed hybridization FFSVM approach outperformed the BSVM, GSVM, and GASVM algorithms regarding detection efficiency, classification accuracy, and correctly classified instances.

This study indicates that bio-inspired techniques are feasible for building credit card fraud detection systems and it is the best optimization technique. The results clearly show that the proposed FFSVM approach outperforms the other three algorithms. In the future, our aim is to improve the classification accuracy and reduce the experimental time of the credit card fraud detection system in real time by using new bio-inspired algorithms and deep learning. The proposed approach can be used to detect fraud incidents in real-time applications.

Data Availability

On reasonable request, the corresponding author will provide the datasets used and/or analyzed during the current study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge University School of Information, Communication Technology, Guru Gobind Singh Indraprastha University, Delhi, India, and the University Grants Commission of India for providing fellowship to research work.

References

[1] J. West and M. Bhattacharya, “An investigation on experimental issues in financial fraud mining,” in *Proceedings of the 11th Conference on Industrial Electronics and Applications*, Hefei, China, 2016.

[2] B. Insider, “Payments companies are trying to fix the massive credit-card fraud problem,” 2019, http://www.businessinsider.com/how-payment-companies-aretrying.

[3] Pwc, “Global economic crime survey 2016,” 2018, https://www.pwc.com/gx/en/economic-crime-survey/.

[4] Nilson report 2016, “Card fraud worldwide,” 2018, https://www.nilsonreport.com.

[5] M. Zareapoor, K. R. Seeja, and M. Afshar Alam, “Analysis on credit card fraud detection techniques: based on certain design criteria,” *International Journal of Computer Application*, vol. 52, no. 3, pp. 35–42, 2012.

[6] V. Mareeswari and G. Gunasekaran, “Prevention of credit card fraud detection based on hsvm,” in *Proceedings of the Information Communication and Embedded Systems (ICICES), 2016 International Conference*, pp. 1–4, Chennai, India, 2016.

[7] A. Singh and A. Jain, “Cost-sensitive metaheuristic technique for credit card fraud detection,” *Journal of Information and Optimization Sciences*, vol. 41, no. 6, pp. 1319–1331, 2020.

[8] S. C. Dubey, K. S. Mundhe, and A. A. Kadam, “Credit card fraud detection using artificial neural network and back-propagation,” in *Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 268–273, Madurai, India, 2020.

[9] K. R. Seeja and M. Zareapoor, “FraudMiner: a novel credit card fraud detection model,” *Science World Journal*, vol. 10, 2014.

[10] A. Singh and A. Jain, “Adaptive credit card fraud detection techniques based on feature selection method,” in *Proceedings of the Springer International Conference on Computer, Communication, and Computational Sciences*, Bangkok, Thailand, 2018.

[11] S. Binitha and S. S. Sathya, “A survey of bio inspired optimization algorithm,” *International Journal of soft computing and engineering (IJSCE)*, vol. 2, 2012.

[12] Y. X. She, “Firefly algorithms for multimodal optimization,” in *Proceedings of the 5th symposium on stochastic algorithms, foundations and applications*, vol. 5792, pp. 169–178, Sapporo, Japan, 2009.

[13] A. Abdallah, M. A. Maarof, and A. Zainal, “Fraud detection system: a survey,” *Journal of Network and Computer Applications*, vol. 68, pp. 90–113, 2016.

[14] A. U. S. Khan, N. Akhtar, and M. N. Qureshi, “Real-time credit-card fraud detection using artificial neural network tuned by simulated annealing algorithm,” in *Proceedings of the International Conference on Recent Trends in Information, Telecommunication and Computing*, pp. 113–121, Chandigarh, India, 2014.

[15] N. S. Alfaiz and S. M. Fati, “Enhanced credit card fraud detection model using machine learning,” *Electronics*, vol. 11, no. 4, p. 662, 2022.

[16] A. Alharbi, M. Alshammari, O. D. Okon et al., “A novel text2IMG mechanism of credit card fraud detection: a deep learning approach,” *Electronics*, vol. 11, no. 5, p. 756, 2022.

[17] Y. Sahin and E. Duman, “Detecting credit card fraud by decision trees and support vector machines,” in *Proceedings of the International MultiConference of engineers and computer scientists (IMECS 2011)*, vol. 1, pp. 1–6, Kowloon, Hong Kong, 2011.

[18] Q. Lu and C. Ju, “Research on credit card fraud detection model based on class weighted support vector machine,” *Journal of Convergence Information Technology*, vol. 6, pp. 62–68, 2011.

[19] O. A. Aderemi and A. Andronicus, “A survey of machine-learning and nature-inspired based credit card fraud detection techniques,” *International Journal of System Assurance Engineering and Management Springer*, vol. 8, 2016.

[20] M. B. Pouyan, R. Yousefi, S. Ostadabbas, and M. Nourani, “A hybrid fuzzy-firefly method for rule-based classification,” in *Proceedings of the Twenty-Seventh International Florida*

Artificial Intelligence Research Society Conference, Richardson, USA, 2014.

[21] M. O'Neill and A. Brabazon, "Self-organising swarm (SOSwarm)," *Soft Computing*, vol. 12, no. 11, pp. 1073–1080, 2008.

[22] S. Arora and D. Kumar, "Hybridization of SOM and PSO for detecting fraud in credit card," *International Journal of Information Systems in the Service Sector*, vol. 9, 2017.

[23] S. J. Jie, "Electronic transaction fraud detection based on improved PSO algorithm," in *Proceedings of the 2nd International Conference on Computer Science and Network Technology*, Changchun, China, 2012.

[24] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai, and S. Pan, "Credit card fraud detection based on whale algorithm optimize BP neural network," in *Proceedings of the 13th International Conference on Computer Science and Education*, Colombo, Sri Lanka, 2018.

[25] E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," *Expert Systems with Applications*, vol. 38, no. 10, Article ID 13057, 2011.

[26] J. West, M. Bhattacharya, and R. Islam, "Intelligent financial fraud detection practices: an investigation," in *Proceedings of the International Conference on Security and Privacy in Communication Networks*, Cham, Australia, 2014.

[27] S. N. Kalid, K.-H. Ng, G.-K. Tong, and K.-C. Khor, "A multiple classifiers system for anomaly detection in credit card data with unbalanced and overlapped classes," *Institute of Electrical and Electronics Engineers Access*, vol. 8, Article ID 28210, 2020.

[28] X. Zhu, X. Ao, Z. Qin et al., "Intelligent financial fraud detection practices in post-pandemic era," *Innovation*, vol. 2, no. 4, Article ID 100176, 2021 Nov.

[29] T. T. Nguyen, H. Tahir, M. Abdelrazek, and A. Babar, "Deep learning methods for credit card fraud detection," 2020 Dec, https://arxiv.org/abs/2012.03754.

[30] Z. Zhang, X. Zhou, X. Zhang, L. Wang, and P. Wang, "A model based on convolutional neural network for online transaction fraud detection," *Security and Communication Networks*, vol. 2018, Article ID 5680264, 2018.

[31] J. West and M. Bhattacharya, "Intelligent financial fraud detection: a comprehensive review," *Computers and Security*, vol. 57, pp. 47–66, 2016.

[32] M. A. Hall, *Correlation-based feature selection for machine learning*, Ph.D. thesis, University of Waikato Hamilton, Hamilton, New Zealand, 1999.

[33] "UC irvine machine learning repository," 2020, https://archive.ics.uci.edu/ml/index.%20php.

[34] S. Arora and D. Kumar, "Hybridization of som and pso for detecting fraud in credit card," *International Journal of Information Systems in the Service Sector*, vol. 9, no. 3, pp. 17–36, 2017.

[35] M. Ghodsi and M. S. Abadeh, "Fraud detection of credit cards using neuro-fuzzy approach based on tlbo and pso algorithms," *Journal of Computer & Robotics*, vol. 10, no. 2, pp. 57–68, 2017.

[36] A. Kaur, S. K. Pal, and A. P. Singh, "Hybridization of k-means and firefly algorithm for intrusion detection system," *International Journal of System Assurance Engineering and Management*, vol. 9, no. 4, pp. 901–910, 2018.