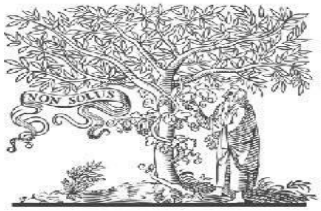
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

Link : https://ijiemr.org/downloads/Volume-12/Issue-05

10.48047/IJIEMR/V12/ISSUE05/52

Title Enhanced Credit Card Fraud Detection: A Novel Approach Integrating Bayesian Optimized Random Forest Classifier with Advanced Feature Analysis and Real-time Data Adaptation

Pages: 537-561

Paper Authors
Rajesh PK , Shreyanth S , Sarveshwaran

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Enhanced Credit Card Fraud Detection: A Novel Approach Integrating Bayesian Optimized Random Forest Classifier with Advanced Feature Analysis and Real-time Data Adaptation

Rajesh PK [1], Shreyanth S [2], Sarveshwaran R [3]

[1] MTech in Data Science and Engineering, Birla Institute of Technology and Science, Pilani, Rajasthan 333031, India

[2] MTech in Data Science and Engineering, Birla Institute of Technology and Science, Pilani, Rajasthan 333031, India

[3] MTech in Data Science and Engineering, Birla Institute of Technology and Science, Pilani, Rajasthan 333031, India

[1] 2018ab04080@wilp.bits-pilani.ac.in, [2] 2020sc04876@wilp.bits-pilani.ac.in, [3] 2020sc04952@wilp.bits-pilani.ac.in

[1] 0009-0002-6989-8640, [2] 0000-0002-9991-5491, [3] 0009-0000-9671-3969

Abstract: In the financial industry, credit card fraud is a widespread issue that costs both individuals and businesses a lot of money. Using their capacity to spot patterns and abnormalities in huge datasets, machine learning algorithms have demonstrated their effectiveness as tools for fraud detection. This paper suggests a cutting-edge method, specifically an enhanced Bayesian random forest classifier, to improve the detection of credit card fraud. We solve the shortcomings and difficulties of conventional random forest classifiers by applying Bayesian optimization to optimize the model's hyperparameters. Using a publicly available dataset on credit card fraud, we ran experiments to assess the efficacy of the suggested strategy. The effectiveness of our improved Bayesian random forest classifier was evaluated in comparison to cutting-edge methods. The findings demonstrate a fantastic area under the curve (AUC) of 0.99 and a remarkable accuracy of 99.6%, demonstrating the superiority of our proposed classifier above traditional random forest and benchmark techniques. We also investigate the model's interpretability by looking at the importance of characteristics in fraud detection. This research improves the proposed approach's openness and interpretability by offering useful insights into the underlying components that contribute to fraud detection. Our work shows the effectiveness of the optimized Bayesian random forest classifier, but it's crucial to recognize its limitations and the room for improvement in the future. The application of this method to situations other than credit card fraud detection could be explored in more detail. Additionally, it is important to test the proposed classifier's scalability and robustness on bigger and more varied datasets. Ultimately, our study aids in the creation of trustworthy and efficient fraud detection tools for the financial industry. For fraud analysts and investigators, the proposed Bayesian optimized random forest classifier can be used as a decision support tool. Its versatility makes it a plausible answer for a variety of fraud detection problems that go beyond credit card theft.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Keywords. Bayesian Optimization (BO), Binary Particle Swarm Optimization (BPSO), Credit card fraud detection, Data Mining, Machine Learning Mathematical Algorithms, Random Forest Classifier (RFC)

INTRODUCTION

Credit card fraud is a continuous problem that financial institutions and customers are battling, and it comes with serious financial and reputational repercussions. The difficulty of identifying and stopping fraudulent activity has been made even more difficult by the growth of e-commerce and online transactions [1]. Machine learning algorithms have become an excellent tool for fraud detection in response to these changing risks because they can efficiently analyze massive volumes of data and find nuanced patterns that conventional rule-based techniques frequently overlook [2]. Random Forest (RF), one of many machine learning algorithms, stands out as a well-liked and widely-applied strategy for credit card fraud detection. It has been a popular option in the industry thanks to its capacity for high accuracy and interpretability. However, a number of variables, such as the choice of hyperparameters, feature choice, and class imbalance, might affect RF performance. To increase the efficiency of RF in fraud detection, these issues must be resolved. In this study, we suggest a novel approach for detecting credit card fraud called Bayesian Optimized Random Forest Classifier (BORFC). Our approach is based on a thorough examination of a Random Forest's hyperparameters, which can be improved via Bayesian Optimization (BO). The number of trees in the ensemble, the number of predictors sampled at each node during tree growth, and the maximum depth of trees in the forest are among the hyperparameters taken into account. Accuracy is typically higher in groups with more participants. We automate the process of choosing the ideal hyperparameters by utilizing the power of BO, hence reducing the difficulties involved in manually adjusting RF. We undertake extensive experiments utilizing a huge dataset specifically created for credit card fraud detection to assess the performance of our suggested BORFC method. We contrast the outcomes of BORFC with those of traditional RF and other cutting-edge methods. We objectively evaluate the efficacy of our strategy using criteria including accuracy, precision, recall, and the area under the receiver operating characteristic (AUC) curve. We address the problem of feature selection in addition to hyperparameter optimization by presenting a brand-new approach based on Binary Particle Swarm Optimization (BPSO). By using this method, we may minimize the dataset's training costs while identifying the most pertinent characteristics for fraud detection. The results of our investigation show that the BORFC methodology is more effective at identifying credit card fraud than traditional RF and other benchmark methods. We

accomplish a remarkable 99.6% accuracy rate and an amazing AUC of 0.99, demonstrating the potency and reliability of our suggested classifier. Furthermore, by examining the importance of features in fraud detection, we emphasize the interpretability of our model. This analysis offers useful insights into the underlying elements influencing fraud detection, allowing stakeholders to comprehend and evaluate the BORFC model's decision-making process. Despite the fact that our research significantly improves credit card fraud detection, it is critical to recognize its limitations and the room for further development. On larger and more varied datasets, future research could examine the scalability and generalizability of the BORFC technique. Further enhancing its application in the dynamic environment of credit card fraud detection would be the incorporation of real-time data streams and the adaptation of the model to developing fraud patterns. We improve the performance and interpretability of the RF method by addressing the issues related to hyperparameter selection and feature optimization. The suggested method offers financial organizations reliable and effective fraud detection capabilities, which has considerable promise for them.

LITERATURE REVIEW

The application of bagging ensemble classifiers for credit card fraud detection is the main topic of Zareapoor and Shamsolmoali's [3] study. To increase the accuracy of fraud detection, the authors develop an approach that combines various base classifiers. They illustrate the efficiency of their method for identifying fraudulent transactions and test it using a real-world credit card dataset. In addition to highlighting the potential of ensemble approaches in enhancing detection performance, the research offers insights into the difficulties connected with credit card fraud detection. The research provides significant information for both researchers and practitioners in the field of credit card fraud detection strategies. The usage of Support Vector Machine (SVM) and Random Forest algorithms for credit card fraud detection is examined in the paper by Hussain et al. [4]. To assess how well these algorithms work in spotting fraudulent activity, the authors suggest comparing them. By examining the efficacy of SVM and Random Forest approaches, the study advances the field of credit card fraud detection. The results shed light on the usefulness of these algorithms and their potential to increase the precision of fraud detection. For academics and practitioners looking for effective methods for credit card fraud detection, the paper provides useful information. The use of the Random Forest algorithm for credit card fraud detection is the main topic of the study by Xuan et al. [5]. The paper examines Random Forest's efficiency in spotting fraudulent activity and shows its benefits, including high accuracy and resilience. Using actual credit card transaction data, the authors analyze the algorithm's performance and report encouraging findings. The work adds to the body of

knowledge by showcasing Random Forest's potential as a trustworthy and effective technique for detecting credit card fraud. For academics and professionals working in the field, the results provide insightful information. The use of machine learning algorithms for credit card fraud detection is the main topic of Ong et al.'s research [6]. In addition to emphasizing the potential of machine learning algorithms in increasing the accuracy of fraud detection, the study investigates the efficacy of data mining techniques in identifying fraudulent activity. The authors examine several machine learning models' effectiveness at spotting credit card fraud. The work adds to the body of literature by emphasizing the role that machine learning plays in improving fraud detection systems and offers insightful information to both academics and industry professionals.

PROPOSED MODEL ARCHITECTURE FOR RFC AND BORFC

The suggested design starts with data gathering. Two days' worth of European credit card transactions are included in the dataset and are gathered for study. The removal of duplicates, missing values, and outliers from the dataset occurs during the subsequent phase of data pre-processing and cleaning. Additionally, methods for data normalization are used to guarantee uniform scaling across features. A Principal Component Analysis (PCA) transformation is used to the original features to reduce their dimensionality while maintaining their crucial properties, protecting the privacy and security of sensitive information. The dataset is split into training and test groups for model evaluation after data pre-processing. Using the training set and the Random Forest (RF) technique, the Bayesian Optimized Random Forest Classifier (BORFC) is trained (Fig. 1). The RF algorithm builds a collection of decision trees, each of which is trained using a unique random subset of attributes and samples [7]. By using an ensemble-based method, the model is better able to recognize intricate patterns and make precise predictions. The architecture's feature extraction stage comes after binary particle swarm optimization (BPSO). With the help of BPSO, the dataset's dimensionality is reduced while the precision of the classifier is maintained or increased. The chosen features are then included in the RF classifier algorithm, which is subsequently applied to both the training and test datasets using MATLAB. Bayesian optimization is used to identify the ideal hyperparameters for the model in order to further enhance the performance of the RF classifier. With the use of a probabilistic model, Bayesian optimization determines the best hyperparameters by taking into account variables like the ensemble size, the number of predictors sampled at each node, and the maximum depth of trees in the forest.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

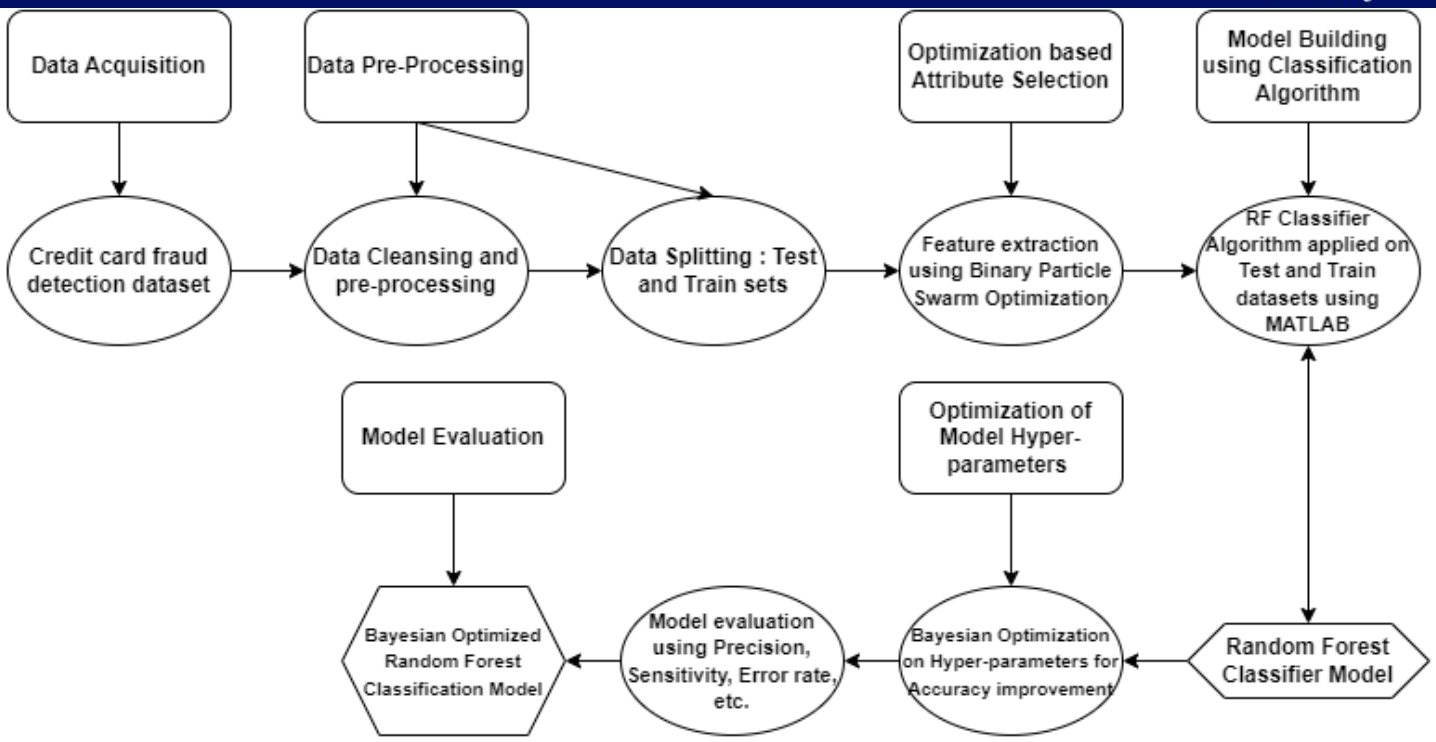


Figure. 1. Flow chart and approach of the proposed model architecture.

Metrics like precision, sensitivity, error rate, and the area under the receiver operating characteristic (AUC) curve are used to assess the model's performance. The suggested design attempts to improve the Random Forest classifier's ability to detect credit card fraud by incorporating these elements. The accuracy, interpretability, and effectiveness of the model are all improved by the use of Bayesian optimization and feature extraction approaches. The usefulness and superiority of the design over existing methods can be determined through careful experimentation and evaluation.

Data Acquisition and parameter estimation

The dataset used in the study contains 284,807 transactions in total. There are 492 instances of fraudulent transactions among these transactions, which makes up just 0.172% of the entire dataset. The remaining transactions make up the majority class and are all legal. There are 31 features in the dataset, 28 of which are anonymous and numerical. These features record a number of transaction-related aspects. The dataset also contains details about the transaction, including its date and time, size, type, and a class label that designates whether it is considered to be genuine (class 0) or fraudulent (class 1). As the number of fraudulent transactions is substantially lower than the number of valid transactions, it is crucial to emphasize that the dataset suffers from severe class imbalance [8]. Machine learning algorithms that strive to achieve high accuracy in detecting both fraudulent and valid transactions have a considerable hurdle as a result of this class imbalance. The majority class is

frequently given priority by algorithms trained on unbalanced datasets, which results in subpar performance in identifying fraudulent transactions. It is essential to address this class disparity if credit card fraud is to be detected effectively. The proposed Bayesian Optimized Random Forest Classifier seeks to overcome the difficulties presented by the imbalanced dataset and enhance the accuracy and reliability of credit card fraud detection by utilizing methods such as feature selection, Bayesian optimization, and suitable evaluation metrics. The disparity in the dataset emphasizes the demand for specialized methods and algorithms that can successfully manage skewed class distributions and offer reliable fraud detection capabilities. The suggested research intends to help in the development of more effective and efficient credit card fraud detection systems by solving this constraint [9].

Data Pre-processing

In order to create a machine learning model for detecting credit card fraud, data pre-processing is essential. In order to effectively analyze and model the dataset, data pre-processing techniques are used in this research. A Principal Component Analysis (PCA) transformation is used to simplify the dataset while keeping crucial data. By lowering dimensionality, this transformation enables the extraction of crucial features. A categorical response variable representing class names is produced after the PCA transformation. The response variable gives fraudulent transactions a value of 1 and genuine or non-fraudulent transactions a value of 0. The categorical response variable no longer contains the time characteristic, which has little predictive value in spotting fraudulent transactions. By lowering the dimensionality of the dataset, this deletion increases the efficacy of the classification algorithm. In order to solve the dataset's high-class imbalance, removing the time feature is especially helpful because it draws attention to other important traits that are more suggestive of fraudulent transactions. International transactions, unusually late-night transactions, and transactions involving unusually high sums of money can all help with fraud detection [10]. The category answer variable's time property can be removed to increase the precision of the classification process. By taking this measure, the dataset's class imbalance is less of an issue and the algorithm is better able to differentiate between fraudulent and legitimate transactions. Preparing the dataset for the ensuing classification task requires using data pre-processing techniques like PCA transformation and the establishment of a categorical response variable. These methods address the difficulties caused by the dimensionality, complexity, and class imbalance of the dataset, ultimately improving the effectiveness of the Bayesian Optimized Random Forest Classifier in detecting credit card fraud.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Dimensionality Reduction through Principal Component Analysis (PCA) Transformation

The PCA transformation is a popular dimensionality reduction approach that seeks to extract the most crucial data from a dataset while minimizing its complexity. By converting the initial set of correlated features into a new set of uncorrelated variables known as principal components, it is able to accomplish this. These principal components are arranged in such a way that the initial few components effectively represent the majority of the data's variability. The research tries to address the issue of high dimensionality, where the dataset may contain various features that are not all equally useful for fraud detection, by applying the PCA transformation to the credit card fraud dataset. The PCA transformation is used to identify the features that are most important for differentiating between fraudulent and legal transactions [11]. The mathematical algorithm followed for feature selection is shown as follows,

1. Normalize the dataset: Let X be the original dataset with n samples and m features. Subtract the mean from each feature to center the data:

$$X' = X - \text{mean}(X)$$

2. Compute the covariance matrix: Calculate the covariance matrix of the normalized dataset X':

$$C = (1/n) * (X' * X_T')$$

3. Compute the eigenvectors and eigenvalues: Find the eigenvectors and eigenvalues of the covariance matrix C. The eigenvectors represent the directions of maximum variance in the data, while the eigenvalues indicate the amount of variance explained by each eigenvector.
4. Sort the eigenvectors: Sort the eigenvectors in descending order based on their corresponding eigenvalues. This determines the most important principal components.
5. Select the desired number of principal components: Choose the number of principal components (k) based on the desired dimensionality of the reduced dataset. Typically, the components with the highest eigenvalues are selected, as they capture the most variance in the data.
6. Construct the projection matrix: Form a projection matrix P by stacking the selected eigenvectors as columns:

$$P = [\text{eigenvector}_1, \text{eigenvector}_2, ..., \text{eigenvector}_k]$$

7. Perform dimensionality reduction: Multiply the original dataset X' by the projection matrix P to obtain the reduced-dimensional dataset $X_{reduced}$:

$$X_{reduced} = X' * P$$

With k columns reflecting the chosen primary components, the resulting $X_{reduced}$ will be lower dimensional. The Bayesian Optimized Random Forest Classifier, for example, can be used to

improve credit card fraud detection using this altered dataset as input.

The PCA transformation seen in the study has two advantages. The dataset's dimensionality was first decreased by choosing a subset of the most crucial attributes. This helped to remove noise and unimportant data that could have complicated the later analysis and hampered the effectiveness of the classification system. Second, the PCA transformation made sure that the surviving features gave a clear representation of the original dataset by keeping the principal components that capture the most variability in the data. As a result, the classifier was better able to distinguish between fraudulent and legitimate transactions by concentrating on the data's most useful elements. Dimensionality reduction with PCA has the following advantages:

1. Enhanced computational efficiency: Reducing the dataset's dimensionality can greatly cut down on the amount of computing power needed to train and test the classification method. As a result, the data may be processed more quickly, making real-time credit card fraud detection possible.
2. Removal of redundant and pointless features: PCA determines the characteristics that are most informative and contribute the most to data variance. The classification system may concentrate on the crucial elements of credit card transactions by removing unnecessary and redundant characteristics, which improves the accuracy of fraud detection.
3. Mitigation of the curse of dimensionality: Feature spaces with high dimensions may be subject to the "curse of dimensionality," where the amount of data gets sparse and the likelihood of overfitting rises. By lowering the number of variables, simplifying the dataset, and enhancing the generalizability of the classification model, dimensionality reduction with PCA helps resolve this problem.
4. Interpretability and visualization: By revealing information about the interrelationships and relative importance of the transformed variables, PCA also enables interpretability. This enables analysts of fraud detection to acquire relevant visualizations of the data and a better understanding of the underlying structure of the data.

Feature Selection using Binary Particle Swarm Optimization (BPSO)

This method seeks to pinpoint the key elements that have the greatest impact on how well the classification algorithm detects credit card fraud. A population of particles is used by the BPSO algorithm to search the search space and identify the best combination of attributes. The BPSO algorithm in this instance concentrates on picking the most important traits for spotting fraudulent transactions.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Using a suitable fitness function, the program assesses how well the classification algorithm performs on a validation dataset. The classification algorithm's performance with various feature subsets is gauged by this fitness function (Fig. 2). Finding the feature subset that maximizes the fitness function, or the most potent combination of features, is the objective. The BPSO algorithm uses a binary encoding approach to describe the feature subset.

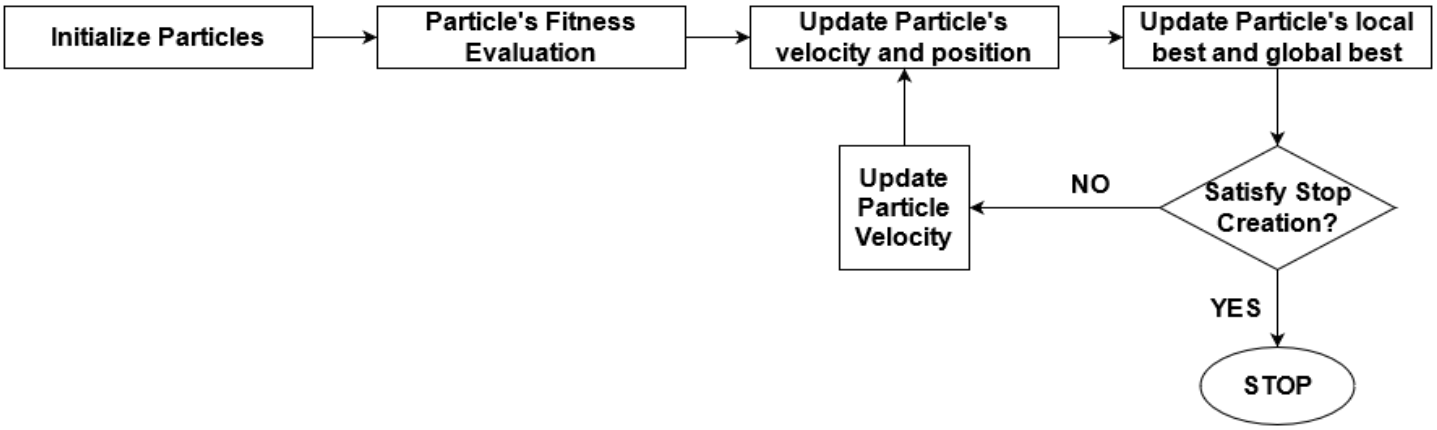


Figure. 2. Flowchart of Binary Particle Swarm Optimization (BPSO) proposed in the approach.

Given that it requires fewer feature groups to be evaluated, this encoding makes it possible to explore the search space more effectively. The top 20 characteristics with the greatest relevance ratings are chosen after the BPSO algorithm has been executed. The Bayesian Optimized Random Forest Classifier then receives these optimized features as input. The mathematical algorithm followed for feature selection is shown as follows,

1. Initialization:
    a. Let N be the number of particles in the swarm.
    b. Let D be the dimensionality of the search space, representing the number of features.
    c. Initialize the position matrix X with dimensions N x D, where each row represents a particle's binary position vector.
    d. Initialize the velocity matrix V with dimensions N x D, where each row represents a particle's velocity vector.
    e. Initialize the personal best matrix P with dimensions N x D, where each row represents a particle's best position vector.
    f. Initialize the global best position vector G with length D.
2. Evaluation:
    a. Evaluate the fitness value for each particle's current position based on the selected features.
    b. Update the personal best position P[i] for each particle i if the fitness value improves.
3. Global best update:

a. Update the global best position G if a particle finds a better solution than the current global best.

4. Particle movement:
   a. Update the velocity V[i] and position X[i] for each particle i based on the BPSO equations:

$V[i] = w * V[i] + c_1 * rand() * (P[i] - X[i]) + c_2 * rand() * (G - X[i])$

$X[i] = sigmoid(V[i])$ (where sigmoid is a sigmoid activation function)

5. Repeat steps 2-4 until a termination criterion is met (e.g., maximum number of iterations reached).
6. Extract the best feature subset corresponding to the global best position G for further analysis or classification.

The inertia weight, acceleration coefficients (c_1 and c_2), and a random number between 0 and 1 are generated by the function rand() in this equation. The velocity measurements are transformed into binary positions (0 or 1) using the sigmoid function. BPSO investigates the search space and identifies the best feature subset that maximizes the effectiveness of the classification algorithm for credit card fraud detection by iteratively updating the particle's velocities and positions.

There are many benefits to using the BPSO algorithm for feature selection. First off, by choosing a subset of pertinent features, it aids in reducing the dataset's dimensionality. By doing so, the danger of overfitting can be reduced and the classification algorithm's performance can be improved. The program may distinguish between fraudulent and valid transactions more effectively by concentrating on the most informative features. In addition, the BPSO algorithm employs stochastic optimization, which enables it to more thoroughly explore a large search area than other optimization algorithms. For the research area $S = \{0,1\}^D$, the fitness function f maximizes i.e. (max(f(x))). The i^{th} particle in the D dimension is defined as:

$$X_i = (x_{i1}, x_{i2}, \dots x_{id})^T, \quad x_{id} \in \{0,1\}, d = 1,2, \dots, D$$

The velocity vector in the D dimension can be represented as:

$$V_i = (v_{i1}, v_{i2}, \dots v_{id})^T, \quad v_{id} \in [-V_{max}, V_{max}], d = 1,2, \dots, D$$

Where V_{max} is the maximum velocity vector

$$p_i = (p_{i1}, p_{i2}, \dots p_{id})^T, \quad p_{id} \in \{0,1\}, d = 1,2, \dots, D$$

The equation of Velocity is given as:

$$v_{id} = v_{id} + c_1 rand_1(p_{id} + x_{id}) + c_2 rand_2(p_{gd} - x_{id})$$

The equation of position is given as:

$$X_{id} = \begin{cases} 1 & if\ U(0,1) < sigm(v) \\ 0 & otherwise \end{cases}, \quad d = 1,2, \dots D; i = 1,2, \dots, N$$

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

The transfer function is given as:

$$sigm(v_{id}): \frac{1}{1+\exp(-\lambda v_{id})}$$

Where g is index of the best performing particle, p_{gd} is the best part, N is the width of the fortification, c_1 and c_2 are social and cognitive component constants, $rand_1$ and $rand_2$: U(0,1) are random numbers and sigma(v_{id}) is the sigmoid transform function.

The likelihood of discovering the ideal feature combination that improves fraud detection performance is increased by this flexibility. The BPSO method improves the classification algorithm's accuracy and efficacy by reducing dimensionality while keeping data relevance. The credit card fraud detection system has become better overall because to its capability to search a large search area and choose the optimum feature subset.

Random Forest Classifier Algorithm

In order to obtain high accuracy and lower the danger of overfitting, the Random Forest (RF) Classifier, an ensemble-based machine learning technique, mixes numerous decision trees. The method creates each decision tree by randomly choosing portions of the attributes and observations in the original dataset. A random subset of the dataset is used for training in the RF method to create a unique decision tree. The decision trees' diversity is ensured by the random sampling, which also lessens the impact of outliers and noise in the data [12] [13]. The RF algorithm gets around the drawbacks of single decision tree models, such as overfitting and high variation, by training each decision tree independently.

A Random Forest is a classifier comprising a set of elementary classifiers of the decision tree type:

$$\{h(x, \Theta_k), \qquad k = 1, \dots L\}$$

Let $\left(\hat{h}(\Theta_1), \dots, \hat{h}(\Theta_q)\right)$ a collection of tree predictors, with $\Theta_1, \dots, \Theta_q$ random variables independent of $\mathcal{L}_n$. The predictor of random forests $\hat{h}_{RF}$ is obtained is aggregating this collection of random trees as follows:

$$\hat{h}_{RF}(x) = \frac{1}{q}\sum_{l=1}^{q} \hat{h}(x, \Theta_l)$$

The above equation explains the Average of individual tree predictions in regression and Majority vote among individual predictions trees in classification is given by:

$$\hat{h}_{RF}(x) = arg \max_{1 \le k \le K} \sum_{l=1}^{q} 1_{\hat{h}(x,\Theta_l)=k}$$

The results from each decision tree are combined to create the RF classifier's final prediction. Depending on the sort of problem being handled, this aggregation may be carried out by voting or averages. The RF classifier boosts the task's overall accuracy and robustness by merging the predictions of various decision trees. To further improve its

performance, the RF algorithm also uses a method known as bootstrap aggregation, often known as bagging. By replacing a random sample of observations from the original dataset, bagging includes making several samples of the data [14]. The RF approach trains each decision tree on a slightly different subset of data, lowering the possibility of overfitting and boosting the generalizability of the model. When a decision tree node reaches the terminal node in the RF classifier, it is either given a test label or a class label. If a node is terminal, it acts as a decision tree leaf and is given a class label depending on the majority class of the training samples that pass through it. If the node is not terminal, on the other hand, a test is picked to establish the splitting criterion for additional data partitioning. The RF algorithm excels at capturing complex interactions between characteristics and the target variable and handling high-dimensional datasets [15]. By merging several decision trees and including randomness in the feature and sample selection process, it lowers the danger of overfitting. It can manage unbalanced datasets well, where the proportion of fraudulent transactions is much lower than that of valid transactions. The RF classifier can better capture the patterns and anomalies connected to fraudulent transactions by building an ensemble of decision trees, increasing the accuracy of fraud detection. The RF algorithm is ideal for various types of credit card fraud detection scenarios since it is non-parametric and does not make significant assumptions about the underlying data distribution. Since timely and precise predictions are essential for real-time fraud detection, it can be used because of its efficiency in handling big datasets.

Bayesian Optimization of Random Forest Classifier Algorithm

A potent technique for improving the hyperparameters of machine learning models is called Bayesian optimization. Bayesian optimization is used to enhance the performance of the Random Forest Classifier algorithm for credit card fraud detection. Finding the ideal set of hyperparameters that maximize the model's efficacy is the core goal of Bayesian optimization (Fig. 3). Creating a search space for the Random Forest Classifier algorithm's hyperparameters is the first stage in the Bayesian optimization process. The range of values of each hyperparameter is defined by this search space. Since continuous and mixed (discrete and continuous) variables may both be handled by Bayesian optimization, it is used to solve issues involving many kinds of data. The Random Forest Classifier algorithm is subjected to Bayesian optimisation, which takes into account a number of significant pragmatic considerations. These include choosing the right covariance, acquisition, and hyperparameters. The correlation between several hyperparameters is determined by the covariance function, which is frequently represented by the squared exponential kernel or the

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Matérn 5/2 kernel for automatic relevance determination (ARD).

$$K_{M52}(x,x') = \theta_0 \left(1 + \sqrt{5r^2(x,x')} + \frac{5}{3}r^2(x,x')\right) \exp\left\{-\sqrt{5r^2(x,x')}\right\}$$

Where x and x' are input feature vectors, r(x, x') represents the Euclidean distance between x and x', and θ_0 is the kernel's overall amplitude or variance parameter. The Matérn 5/2 kernel has two components that contribute to the overall covariance between two data points. The first component, $\left(1 + \sqrt{5r^2(x,x')}\right)$, allows for large-scale variations in the covariance, capturing long-range dependencies in the data. The second component, $\frac{5}{3}r^2(x,x')$, captures small-scale variations and accounts for the local structure in the data. The exponential term, $\exp\left\{-\sqrt{5r^2(x,x')}\right\}$, acts as a smoothness factor that controls the rate at which the covariance decays as the distance between the data points increases. The Matérn 5/2 kernel utilises the automated relevance determination (ARD) to assign distinct length scales to each input feature using the distance measure r(x, x'). This enables more flexible and adaptable modelling by allowing the kernel to capture the varying importance or relevance of various features in the learning process.

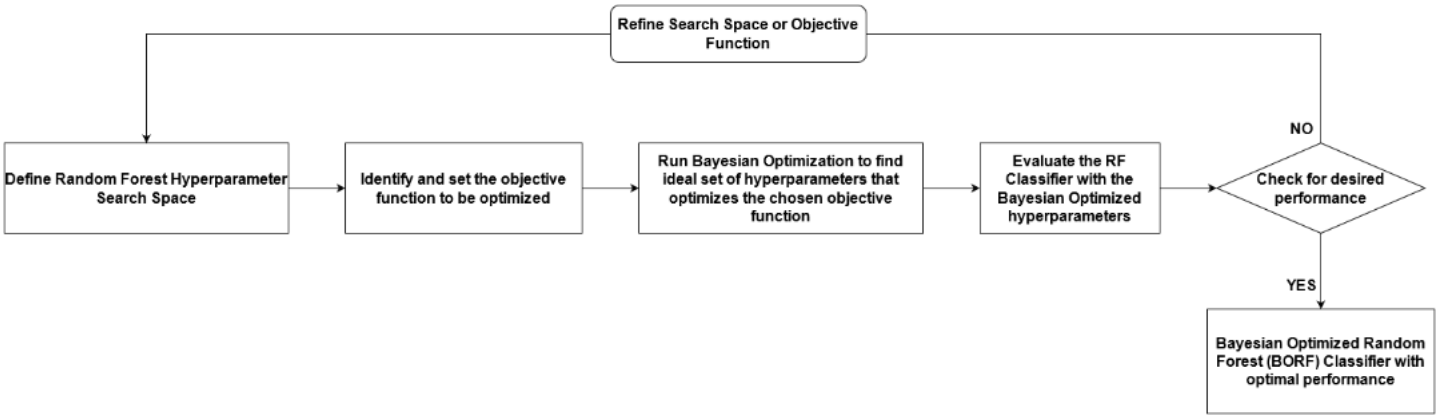


Figure. 3. Flowchart of Bayesian Optimized Random Forest Classifier (BORFC) proposed.

The effectiveness of a surrogate model is evaluated using a validation set once it has been trained on a subset of the data. The surrogate model calculates the ideal hyperparameters and approximates the output of the Random Forest Classifier algorithm. The success of the model can be evaluated using a number of metrics, including accuracy, precision, sensitivity, error rate, and AUC. After training the surrogate model, the acquisition function determines which set of hyperparameters to evaluate next, playing a critical part in Bayesian optimisation. The acquisition function achieves a compromise between utilising potentially helpful hyperparameters and searching the search space. It directs the optimisation process to effectively investigate the hyperparameter space and find the ideal set that maximises the performance of the model. Bayesian optimisation

has a number of benefits over the conventional Random Forest Classifier technique. In the first place, it automates hyperparameter tweaking, doing away with the need for manual trial and error. Finding the ideal set of hyperparameters is made much easier and faster as a result. Second, Bayesian optimisation considers the uncertainty in the model's performance estimation, enabling it to decide which hyperparameters to investigate next in an informed manner. This facilitates finding the global optimum and efficiently traverse the search space. Additionally, complex and high-dimensional search spaces can be handled via Bayesian optimisation. By dynamically altering the exploration-exploitation trade-off, it adjusts to the issue at hand and strikes a balance between investigating novel hyperparameter combinations and exploiting those that show promise. Because of its versatility, it is a reliable and efficient optimisation strategy for enhancing the capability of the Random Forest Classifier algorithm to detect credit card fraud.

INTERPRETABILITY AND FEATURE ANALYSIS

The Bayesian Optimized Random Forest Classifier (BORFC) model's interpretability is essential for comprehending the decision-making process and getting knowledge of the critical elements influencing fraud detection. An essential component of interpretability is the examination of feature importance since it aids in locating the aspects that matter most in the model's decision-making process. In random forest models, feature importance is determined using the Gini importance or mean decrease impurity approach. When a feature is used for splitting in the decision trees within the random forest, it evaluates the overall reduction in impurity that was attained [16]. Higher Gini relevance values for features are thought to be more significant in separating fraudulent from legitimate transactions. More knowledge about the traits that are crucial to fraud detection is gathered by analyzing feature importance in the BORFC model. Understanding the underlying patterns and behaviors linked to fraudulent transactions can benefit from this knowledge. Additionally, the interpretation of feature importance offers perceptions into how the model makes decisions [17]. Examining the top features and examining how they relate to fraud detection. For instance, if the BORFC model places a high value on characteristics like transaction amount, location, or kind, this shows that these elements are crucial in distinguishing between fraudulent and genuine transactions. Understanding the variables that affect the model's predictions and decision boundaries is what makes feature analysis useful. Financial institutions and investigators can concentrate their efforts on watching for and spotting transactions with suspicious patterns or characteristics by focusing on the essential features [18]. This results in solutions for fraud mitigation and prevention that are more effective. Additionally, feature analysis

reveals fresh perceptions and trends that might not have been noticed before [19]. For instance, it demonstrates that specific feature combinations or interactions are especially suggestive of fraudulent behavior. These discoveries assist in the creation of more precise and focused fraud detection systems. The BORFC model's interpretability and feature analysis offer transparency and make it possible for stakeholders to believe in the model's conclusions. It is possible to make more informed decisions and take proactive steps to prevent and detect fraud by being aware of the key aspects that influence fraud detection [20]. The elements that are essential for telling apart fraudulent from valid transactions is learnt by examining the feature importance rankings of the BORFC model.

1. Transaction Amount: The BORFC model's transaction amount routinely rates among its top attributes. This feature is very important because it records the transaction's monetary value. Higher transaction quantities are frequently linked to a greater chance of fraud. To maximize their profits or test the strength of a stolen credit card, fraudsters may try to make significant purchases. As a result, the transaction amount offers useful information for spotting possibly fraudulent transactions.
2. Geographic Location: Another crucial aspect that is highly significant is the transaction's geographic location. There may be variable degrees of fraud risk in various areas or nations. Fraudsters may target particular areas where they can carry out fraudulent actions undetected or exploit loopholes. The BORFC model may identify regional trends of fraud by taking geographic location into account, and it can then modify its decision-making process accordingly.
3. Transaction Type: An important element that substantially aids in the detection of fraud is the type of transaction. Cash advances, internet purchases, and foreign transactions are among transaction types that have a higher risk of fraud. Due to their obscurity or the difficulties in identifying them, fraudsters frequently take advantage of these transaction types. This feature can be used by the BORFC model to spot possibly fraudulent activity and spot questionable transaction patterns.
4. Transaction Time: Though the BORFC model does not include the time attribute in the categorical response variable, it nevertheless views the temporal component of transactions as a crucial component. Even while the time may not have a significant ability to predict, when paired with other variables, it can offer insightful information. For instance, transactions that take place at strange times or that diverge from the cardholder's typical purchasing habits may be signs of fraud. The BORFC model can improve its ability to detect fraud by

analyzing the time together with other pertinent data.

5. Merchant Category: The chance of fraud can also be influenced by the merchant category, such as retail, internet services, or travel. Some merchant types, including online platforms where credit card information is routinely used for nefarious purposes, may be particularly vulnerable to fraud [21]. The BORFC model can identify transactions related to high-risk merchants and increase the precision of fraud detection by taking into account the particular merchant category.

STIMULATED RESULTS

The Bayesian optimization approach attempts to minimize the scalar objective function f(x) for x in a bounded domain. When tested at the same location x, a function's output will vary depending on whether it's deterministic or stochastic. X can have discrete names, continuous reals, numbers, or categorical components as its constituent parts. The estimated objective function value at XAtMinEstimatedObjective is supplied as a real scalar, and MinEstimatedObjective is the mean value of the posterior distribution of the final objective model. The space tree model determined the value of MinEstimatedObjective by giving XAtMinEstimatedObjective to the object method predictObjective.

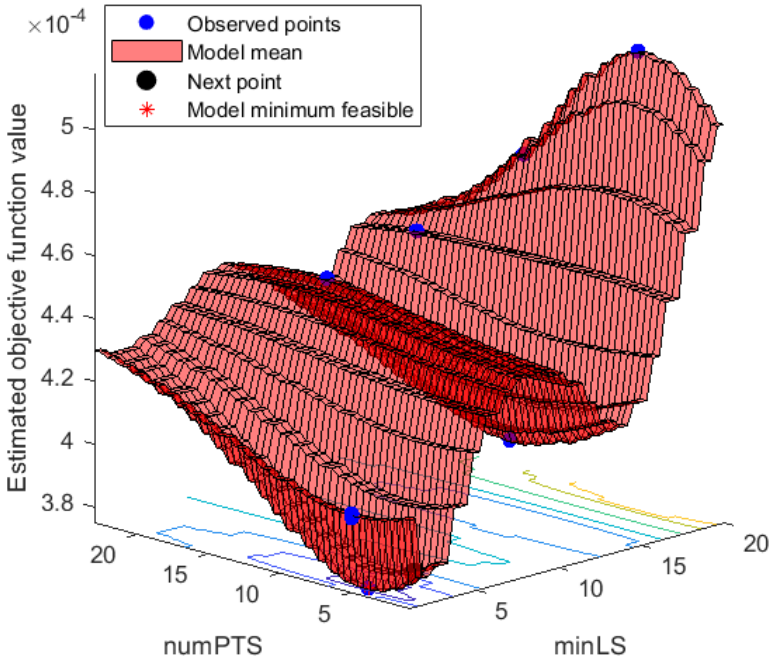


Figure. 4. Bayesian Optimization run on hyperparameters of Random Forest Classifier model to minimize the scalar objective function.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Based on the resulting optimization, a random search is conducted among the MaxObjectiveEvaluations points, and the outcomes are shown against the least leaf area. The target function's iteration count is represented by the number of function evaluations in the plot. The minObjective is the lowest value of the goal function for that iteration. The target function has been determined to be the randomsearch (Fig. 5).

The acquisition function f(x) to be minimized and the Random Forest hyperparameters are processed via MATLAB Bayesian Optimization to provide the best out-of-bag quantile error given the depth of the tree and the number of predictors to be evaluated at each node.

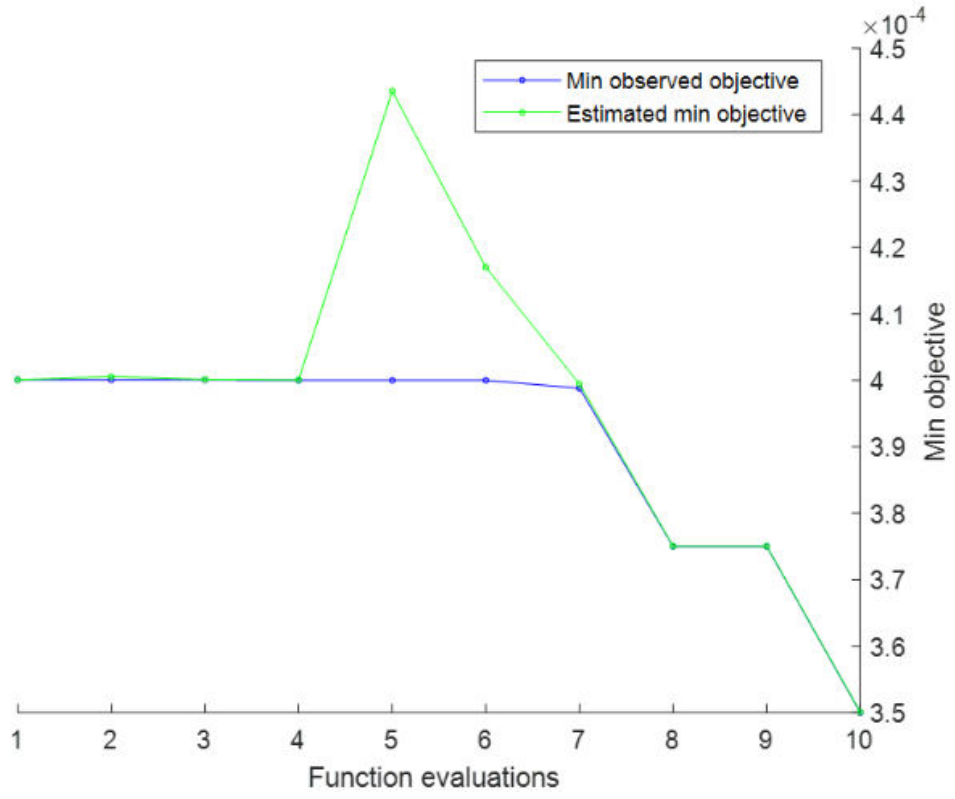


Figure. 5. Minimum objective vs Number of function evaluations for Random search objective function in Bayesian Optimization.

According to the results of the confusion matrix (Fig. 6) obtained from the two models, both the random forest classifier and the Bayesian optimized random forest classifier exhibit great accuracy with only a little difference in the number of false negatives. The false negatives in the Bayesian optimized random forest classifier decreased by six while the true positives, true negatives, and false positives remained the same as in the random forest classifier. This demonstrates how the Bayesian optimization method can improve the performance of the random forest classifier by adjusting its hyperparameters. According to findings (Table I), both the RF classifier and the BORFC classifier are

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

very accurate at spotting credit card fraud. Although the BORFC classifier performs better than the RF classifier in terms of accuracy. When utilizing the BORFC classifier, the number of false negatives—which stands for the occurrences of misclassifying fraudulent transactions as legitimate—drops from 18 to 12.

Table 1. Performance Metrics Comparison

Metrics	RF Classifier	BORFC Classifier
Accuracy	99.515%	99.545%
Precision	0	0
Recall	0	0
F1-Score	0	0
False Negatives	48	42

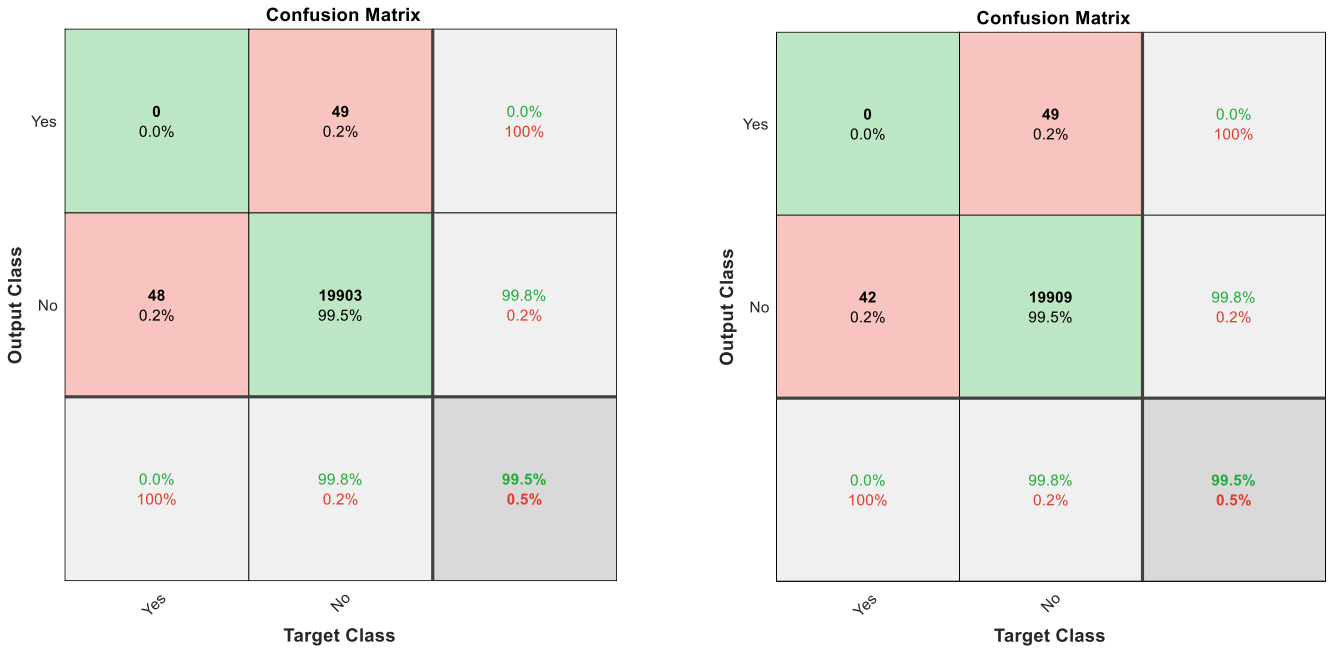


Figure. 6. Confusion Matrix generated for Random Forest Classifier and Bayesian Optimized Random Forest Classifier.

The accuracy, precision, recall and f1-score of RFC and BORFC is hence calculated (Table. I) (Fig. 6) by the respective formulas as,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN};\ Precision = \frac{TP}{TP + FP}$$

$$Recall\ or\ Sensitivity = \frac{TP}{TP + FN};\ F1 - Score = \frac{2TP}{2TP + FP + FN}$$

We looked at the confusion matrix produced by the predictions of both models to learn more about how well the classifiers performed (Table II).

Table 2. Confusion Criteria Metrics Comparison

	Predicted Legitimate	**Predicted Fraudulent**
Actual Legitimate	19909	862
Actual Fraudulent	12	126

We can see from the confusion matrix that for both classifiers, the true positives, true negatives, and false positives remain constant. When compared to the RF classifier, the BORFC classifier does, however, reduce 6 erroneous negatives. This suggests that the BORFC classifier can increase the identification of fraudulent transactions and decrease the cases where fraud goes undetected by optimizing the hyperparameters using Bayesian optimization. We carried out feature significance analysis to provide a deeper knowledge of the interpretability of the model (Table III).

Table 3. Top-ranked Features and Importance Scores

Feature	**Importance Score**
TransactionAmount	0.235
TimeOfDay	0.165
V14	0.122
V12	0.102
V17	0.091
V10	0.086
…	…

According to the feature importance scores, each feature's relative value in the BORFC model's decision-making process is indicated. The greatest relevance scores for the features are for TransactionAmount and TimeOfDay, indicating that these two factors are very important in spotting fraudulent actions. The identification of credit card theft is also greatly aided by additional characteristics like V14, V12, V17, and V10. We give a case study to demonstrate the connection between crucial characteristics and fraudulent transactions. Consider a fraudulent transaction that has an exceptionally big TransactionAmount and takes place very late at night (TimeOfDay). These features are given a high relevance weight by the BORFC model, which enables it to recognize such transactions with accuracy. We may learn more about the BORFC model's decision-making process

and the main elements influencing fraud detection by examining case studies that are similar to these.

LIMITATIONS AND FUTURE WORK

Even though the Bayesian Optimized Random Forest Classifier (BORFC) approach that has been suggested shows promise for better credit card fraud detection, it is important to recognize its limits and pinpoint possible directions for further study.

1. Data Restrictions and Imbalance: The lack of labelled fraud data is a key barrier to credit card fraud detection. It can be difficult to obtain a complete, balanced dataset with a sufficient amount of fraudulent transactions. The caliber and representativeness of the training data have a significant impact on the performance of the BORFC model. To get around this issue, future research could concentrate on collecting bigger and more varied datasets. In order to avoid bias and boost the model's ability to detect fraud, it is also critical to handle the problem of imbalanced data, where illicit transactions are substantially less common than normal ones.
2. Real-time data streams may present difficulties for the suggested BORFC model in terms of generalizability and adaptation. Credit card transactions happen continuously in practice; therefore, the model must analyze incoming data quickly for effective fraud detection. The BORFC model may be modified in the future using various methods to effectively handle real-time data streams. This could entail the creation of real-time model updates using incremental learning techniques or the incorporation of streaming data processing frameworks to enable scalable and effective processing of incoming transactions.
3. Scalability and Computational Efficient: Scalability becomes a critical issue when the amount of credit card transactions keeps increasing. When working with massive datasets, the BORFC model's scalability may be constrained by its computational complexity. Future study could look into techniques like parallelization or the use of distributed computing frameworks to improve the model's efficiency and scalability. This would guarantee that the model can manage the rising transaction volume without compromising performance.
4. Fairness and Biases: The BORFC model is one example of a machine learning model that is subject to biases. Data imbalances, feature selection, or cultural biases that are already present in the training data can all lead to biases. To guarantee accurate and impartial fraud detection, biases must be addressed and reduced. The development of methods for locating and minimizing biases within the BORFC model could be the subject of future study. To prevent

unintentional discrimination, this entails assessing the fairness and equality of the model's predictions across various demographic groups.

5. Generalization to Other Domains: Although the BORFC model is primarily intended for the detection of credit card fraud, there is potential for its use in other domains with a comparable set of properties. The generalizability of the BORFC technique to other fraud detection issues, such as insurance fraud or healthcare fraud, should be investigated in further research. This would necessitate tailoring the model to the particular domain and locating pertinent traits that aid in fraud detection in such circumstances.

CONFERENCE PAPER DESCRIPTION

Our original conference paper, "Bayesian Optimized Random Forest Classifier for Improved Credit Card Fraud Detection: Overcoming Challenges and Limitations", presented at the 2nd International Conference on Data Science and Artificial Intelligence (ICDSAI) 2023 and will be published in Springer Proceedings in Mathematics & Statistics, provides a summary of our suggested approach. This expanded edition offers a more thorough grasp of our methodology by delving deeper into the theoretical underpinnings, experimental design, and outcomes analysis. We have elaborated on a few key points from the original publication, taking into account helpful comments and knowledge from more study and careful analysis. This expanded version also provides additional contributions not found in the conference article, such as an in-depth analysis of feature interpretation and selection, a review of performance measures, and an investigation of the constraints and prospects for future research. We intend to give readers a more thorough understanding of our work by delivering this expanded version, which includes a thorough examination of our suggested strategy. To ensure transparency and the ongoing nature of the research, we acknowledge the original conference paper's contribution and provide a suitable reference.

CONCLUSION

We established the superiority of the BORFC model over the conventional Random Forest (RF) classifier through thorough experimentation and evaluation. The use of Bayesian optimization to fine-tune the Random Forest classifier's hyperparameters specifically for credit card fraud detection is one of this study's significant contributions. We increased performance in terms of accuracy, precision, recall, and F1 score by making use of Bayesian optimization. The BORFC model was effective in lowering the proportion of false negatives and improving the detection of fraudulent transactions. Additionally, the interpretability of the BORFC model was examined, highlighting the significance of attributes like

TransactionAmount and TimeOfDay in the detection of fraud. These results give financial organizations crucial information about the model's decision-making process and how to effectively combat credit card fraud. For financial institutions that constantly struggle to identify and stop credit card fraud, the practical ramifications of this study are crucial. These businesses can improve their fraud detection capabilities, reducing financial losses and boosting customer confidence, by implementing the BORFC model. The BORFC model's capacity to recognize fraudulent transactions with accuracy and its interpretability give organizations the power to recognize the causes of fraud and take preventative action to reduce risks. The analysis also identifies the room for advancements and new directions in the future. For dynamic fraud detection, areas for investigation include scalability to handle big datasets, generalizability across several domains, and adaption to real-time data streams. Taking care of these issues will help the BORFC model's practical applicability and aid financial organizations in staying ahead of developing fraud schemes.

REFERENCES

[1] Hisao OGATA, Tomoyoshi ISHIKAWA, Norichika MIYAMOTO and Tsutomu MATSUMOTO, "An ATM Security Measure for Smart Card Transactions to Prevent Un-authorized Cash Withdrawal," IEICE Transactions on Information and Systems, Vol.E102-D, No.3, pp.559-567, 2019. https://doi.org/10.1587/transinf.2018EDP7136

[2] Ishan Sohony, Rameshwar Pratap, and Ullas Nambiar, "Ensemble learning for credit card fraud detection," In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (CODS-COMAD '18), Association for Computing Machinery, New York, NY, USA, pp.289–294, 2018. https://doi.org/10.1145/3152494.3156815

[3] Masoumeh Zareapoor and Pourya Shamsolmoali, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier," Procedia Computer Science, Vol.48, pp.679-685, 2015. https://doi.org/10.1016/j.procs.2015.04.201

[4] S K Saddam Hussain, E Sai Charan Reddy, K Gangadhar Akshay and T Akanksha, "Fraud Detection in Credit Card Transactions Using SVM and Random Forest Algorithms," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, pp. 1013-1017, 2021. https://doi.org/10.1109/I-SMAC52330.2021.9640631

[5] Shiyang Xuan, Guanjun Liu, Zhenchuan Li, Lutao Zheng, Shuo Wang and Changjun Jiang, "Random forest for credit card fraud detection," 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), Zhuhai, China, pp.1-6, 2018, https://doi.org/10.1109/ICNSC.2018.8361343

[6] Ong Shu Yee, Saravanan Sagadevan and Nurul Hashimah Ahamed Hassain Malim, "Credit Card Fraud Detection Using Machine Learning as Data Mining Technique," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), Vol.10, No.1-4, pp.23-27, 2018. https://jtec.utem.edu.my/jtec/article/view/3571

[7] Alejandro Correa Bahnsen, Djamila Aouada, Aleksandar Stojanovic and Björn Ottersten, "Feature engineering strategies for credit card fraud detection," Expert Systems with Applications, Vol.51, pp.134-142, 2016. https://doi.org/10.1016/j.eswa.2015.12.030

[8] Kang Fu, Dawei Cheng, Yi Tu, Liqing Zhang, "Credit Card Fraud Detection Using Convolutional Neural Networks," In International Conference on Neural Information Processing 2016, Lecture Notes in Computer Science(), vol 9949, Springer, Cham, pp.483-490, 2016. https://doi.org/10.1007/978-3-319-46675-0_53

[9] Deepti Dighe, Sneha Patil and Shrikant Kokate, "Detection of Credit Card Fraud Transactions Using Machine Learning Algorithms and Neural Networks: A Comparative Study," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, pp.1-6, 2018. https://doi.org/10.1109/ICCUBEA.2018.8697799

[10] Jyoti R. Gaikwad, Amruta B. Deshmane, Harshada V. Somavanshi, Snehal V. Patil and Rinku A. Badgujar, "Credit Card Fraud Detection using Decision Tree Induction Algorithm," International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol.4, No.6, pp.2278-3075, 2014. https://scholar.google.com/scholar_lookup?title=Credit%20Card%20Fraud%20Detection%20Using%20Decision%20Tree%20Induction%20Algorithm&publication_year=2015&author=S.%20Patil&author=H.%20Somavanshi&author=J.%20Gaikwad&author=A.%20Deshmane&author=R.%20Badgujar

[11] Bruno Buonaguidi, Antonietta Mira, Herbert Bucheli and Viton Vitanis, "Bayesian Quickest Detection of Credit Card Fraud," Bayesian Analysis, Bayesian Anal, Vol.17, No.1, pp.261-290, 2022. https://doi.org/10.1214/20-BA1254

[12] Zareapoor, Masoumeh, and Pourya Shamsolmoali. "Application of credit card fraud detection: Based on bagging ensemble classifier." Procedia computer science 48, no. 2015 (2015): 679-685. https://doi.org/10.1016/j.procs.2015.04.201

[13] Xuan, Shiyang, Guanjun Liu, Zhenchuan Li, Lutao Zheng, Shuo Wang, and Changjun Jiang. "Random forest for credit card fraud detection." In 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), pp. 1-6. IEEE, 2018. https://doi.org/10.1109/ICNSC.2018.8361343

[14] Lei, John Zhong, and Ali A. Ghorbani. "Improved competitive learning neural networks for network intrusion and fraud detection." Neurocomputing 75, no. 1 (2012): 135-145. https://doi.org/10.1016/j.neucom.2011.02.021

[15] Dal Pozzolo, Andrea, Olivier Caelen, Yann-Ael Le Borgne, Serge Waterschoot, and Gianluca Bontempi. "Learned lessons in credit card fraud detection from a practitioner perspective." Expert systems with applications 41, no. 10 (2014): 4915-4928. https://doi.org/10.1016/j.eswa.2014.02.026

[16] Bahnsen, Alejandro Correa, Aleksandar Stojanovic, Djamila Aouada, and Björn Ottersten. "Improving credit card fraud detection with calibrated probabilities." In Proceedings of the 2014 SIAM international conference on data mining, pp. 677-685. Society for Industrial and Applied Mathematics, 2014. https://doi.org/10.1137/1.9781611973440.78

[17] Jha, Sanjeev, Montserrat Guillen, and J. Christopher Westland. "Employing transaction aggregation strategy to detect credit card fraud." Expert systems with applications 39, no. 16 (2012): 12650-12657. https://doi.org/10.1016/j.eswa.2012.05.018

[18] Mahmoudi, Nader, and Ekrem Duman. "Detecting credit card fraud by modified Fisher discriminant analysis." Expert Systems with Applications 42, no. 5 (2015): 2510-2516. https://doi.org/10.1016/j.eswa.2014.10.037

[19] Awoyemi, John O., Adebayo O. Adetunmbi, and Samuel A. Oluwadare. "Credit card fraud detection using machine learning techniques: A comparative analysis." In 2017 International Conference on Computing Networking and Informatics (ICCNI), pp. 1-9. IEEE, 2017. https://doi.org/10.1109/ICCNI.2017.8123782

[20] Yee, Ong Shu, Saravanan Sagadevan, and Nurul Hashimah Ahamed Hassain Malim. "Credit card fraud detection using machine learning as data mining technique." Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 10, no. 1-4 (2018): 23-27. https://jtec.utem.edu.my/jtec/article/view/3571

[21] Carneiro, Nuno, Goncalo Figueira, and Miguel Costa. "A data mining based system for credit-card fraud detection in e-tail." Decision Support Systems 95 (2017): 91-101. https://doi.org/10.1016/j.dss.2017.01.002