Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-021-06147-9

RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

Credit Card Fraud Detection by Modelling Behaviour Pattern using Hybrid Ensemble Model

V. S. S. Karthik[1] · Abinash Mishra[2] · U. Srinivasulu Reddy[2]

Received: 7 April 2021 / Accepted: 23 August 2021

Abstract

The fraud detection system in banking organisation relies on data-driven approach to identify the fraudulent transactions. In real time, detection of each and every fraudulent transaction becomes a challenging task as financial institutions need aggressive jobs running on the log data to perform a data mining task. This paper introduces a novel model for credit card fraud detection which combines ensemble learning techniques such as boosting and bagging. Our model incorporates the key characteristics of both the techniques by building a hybrid model of bagging and boosting ensemble classifiers. Experimentation on Brazilian bank data and UCSD-FICO data with our model shows sturdiness over the state-of-the-art ones in detecting the unseen fraudulent transactions because the problem of data imbalance was handled by a hybrid strategy. The proposed method outperformed by a margin of 43.35–68.53, 0.695–11.67, 43.34–68.52, 42.57–67.75, 3.5–13.06, 24.58–34.35%, respectively, in terms of true positive rate, false positive rate, true negative rate, false negative rate, detection rate, accuracy and area under the curve from the state-of-the-art-techniques, with a Matthews correlation co-efficient of 1.00. At the same time, the current approach gives an improvement in the range of 0.6–24.74, 0.8–24.80, 10–17.00% in terms of false positive rate, true negative rate and Matthews correlation co-efficient respectively from the state-of-the-art techniques with detection rate of 0.6650 and accuracy of 99.18%, respectively.

Keywords Data imbalance · Empirical risk · Ensemble learning · Hybrid ensemble · Bagging · Boosting

1 Introduction

In emerging markets like India, the volume of digital transactions has increased due to various reasons, primarily after demonetisation in the year 2016. A significant amount of those transactions are credit card or debit card transactions; as the usage of digital transactions increased, fraudsters also increased on the respective platforms. It is a challenging task for banking organisations, financial institutions or financial technology firms. Hence, this field has gained a huge interest by the machine learning and business intelligence community in recent past. As the transactions are increasing day by day, it leads to increase in fraudulent activity.

According to Nilson report [25], $24.26 billion loss occurred due to fraudulent activity worldwide; also the report showed a loss of $34.66 billion in 2022. During the years credit card fraud detection has gained increasing interest among researchers around the globe. Fraudulent activity can be done in both ways, i.e. online as well as off-line. Most of the E-commerce companies deploy strenuously working data mining jobs on the logs of their servers. These log data encompass both fraudulent and genuine transactions, the latter in greater numbers than the former. Hence, the distribution of the class label is highly skewed because of which the learning algorithm might not learn perfectly from the minority class. Since the number of fraudulent transactions is very low, the classifier might neglect them by treating them as noise. As described by [19], rule-based fraud detection tools have been employed to catch the fraudsters. Rules for such system are designed by fraud experts; in the contemporary era of Artificial Intelligence (AI) we cannot completely rely on humans

✉ U. Srinivasulu Reddy
usreddy@nitt.edu

V. S. S. Karthik
karthik_vss@outlook.com

Abinash Mishra
405117002@nitt.edu

[1] Indian Institute of Information Technology, Tiruchirappalli, India

[2] Machine Learning and Data Analytics Lab, National Institute of Technology, Tiruchirappalli, India

Published online: 05 September 2021

to solve such problems which involve such high empirical risk. Authors in [19] introduced card holder behaviour model which tries to capture user spending behaviour. Although knowing the structure of data and notion to create behaviour model is critical to our model, but without balanced training data performance of any model cannot be improved. [2] elaborated on problem of class imbalance, and it affects on the performance of the rule-based classifiers. Hence, it is essential to select the subset of the feature space or even feed new variables as features. The imbalance ratio depicts the ratio of genuine transaction to the fraudulent transactions in the database. As the imbalance ratio is very high in UCSD-FICO and Brazilian bank datasets, our strategy to counter data imbalance would be key to our model. The learning algorithm ought to perform better in imbalance scenario, since imbalanced classification is an inherent problem in fraud detection.

This article utilises Adaboost for feature engineering the behavioural feature space. Feature engineering is critical because most of the machine learning models are incompetent decision-makers when a fraudster changes the strategy to commit fraud [31]. An adaptive framework is required to counter the new methods of fraud adopted by the defrauders. Scalability is also a major issue for all the financial institutions having all the transactions in memory. Carcillo et al. [4] have demonstrated the usage of robust framework which could be used for real-time fraud detection using popular tools such as Apache Kafka [8], Apache Spark [9] and Apache Cassandra [7].

The proposed approach in this article aims at maximising the performance state-of-the-art techniques from the literature in terms of precision, recall and Matthew correlation coefficient (MCC) measure. The way to achieve the goal is threefold. First, it proposes a hybrid ensemble model for credit card fraud detection. Second, class imbalance is handled in a lucid manner by using hybrid approach as described by [15]. Finally, it reduces the number of false positives thus asserting its efficacy on real-world data.

The present study proposes a hybrid approach towards the development of credit card fraud detection system. The contribution made by the authors is as follows:

- The feature space is generated by adopting the Adaboost-based feature engineering technique
- This article employs the data distribution, outlier detection and elimination of noisy sample as the pre-processing step
- The present study integrates the feature engineering module with the extra trees classifier and random forest to maximise the effectiveness of the learning algorithm
- The effectiveness of the developed feature engineering approach is compared with the standard baseline classifiers such as logistic regression and boosting ensemble
- The supremacy of the current approach is validated using the state-of-the-art credit card fraud detection system, and it could be used in real time for identifying the fraudulent transaction

2 Related Work

In the present time, credit card fraud detection has gained a lot of interest in the machine learning research community. This section presents the state-of-the-art techniques that have been applied for credit card fraud detection, which is categorised as probabilistic approach, individual learning approach and cost-sensitive learning-based fraud detection system. Risk-induced Bayesian inference bagging is proposed by Akila et al. [1] for credit card fraud detection in which a novel bag creation strategy has been applied to re-balance the distribution of classes in terms of minority (fraudulent transaction) and majority (legitimate transaction) sample. Once the bag creation phase was over, the Bayesian classifier was trained and based on the probability score for a transaction being fraud afterward the class label was assigned using threshold value. The disadvantage of the RIBIB model was: it cannot be efficient for handling the concept drift problem.

Kim et al. [18] proposed a champion challenger framework for fraud detection which was based on hybrid ensemble and deep learning. In the champion module, they incorporate the hybrid ensemble, and in the challengers part they have used the deep learning algorithms. Experiments revealed that challengers module beat the champion module; hence, they have chosen the deep learning techniques as the learning algorithm while deploying the fraud detection model in real time.

Liu et al. [23] presented a feature vector selection method that was incorporated with support vector machine (SVM) to maximise the separability between the classes. The proposed work experimented on 26 publicly available data sets and showed the effectiveness of the proposed method in terms of accuracy.

Tao et al. [29] proposed a cost-based learning algorithm in which cost matrix had been applied while fitting the learning algorithm using the training data set in which support vector machine act as learning algorithm. They have performed the experiment on four benchmark data sets available in University of California, Irvine (UCI) repository in terms of G-mean and F-measure.

Zheng et al. [39] have given a generative adversarial network (GAN)-based telecom fraud detection model which can effectively distinguish the fraudulent pattern from the non-fraudulent one. They highlighted the limitation in the performance of the existing model, and it was further improved by applying the loss minimisation technique in the learning approach. They deployed the proposed method in two commercial banks and detected 321 fraud cases in the bank. As a result, customer loss is prevented and improves the reputation of the bank.

Zelenkov [37] proposed an example dependent cost-sensitive adaptive boosting algorithm in order to overcome the problem in base classifiers such as Naive Bayes, logistic regression and decision tree as these algorithms assume the distribution of the class label is balance in nature. They had evaluated the proposed method on three synthetic data sets, and two real data sets also showed that this model outperformed the existing model.

Yu et al. [35] have given a hybrid fraud detection model in which deep belief network (DBN)-based resampling method combined with SVM was proposed. To reduce the misclassification during the learning phase, classifier introduces the revenue matrix. Their proposed work validated in German credit data set and Japanese credit data set. A four-stage approach was used to develop the fraud detection model in which card holders were grouped based on the past transaction; thereby, the group behaviour was similar. In addition to this, the proposed model adopted the window-sliding strategy to aggregate the transactions [17]. The hyper-parameter was selected based on the dynamic weighted entropy (DWE) to overcome the class imbalance problem [20]. A rule-based feature engineering method was proposed to identify the fraudulent and genuine transaction, where both individual and group behaviours were considered. The proposed approach was effective in portraying the individual features into group features; thereby, it improved the model's performance [33]. A set of interpretable features was generated based on the model agnostic nonlinear explanations (MANE). Also, the proposed approach adopted the aggregation to identify the card holders behaviour pattern. Afterward, the cross-features were mined based on the gradient boosting decision tree; thereby, it approximates the local boundary [30]. The legitimate and fraudulent transaction was trained by random forests separately to identify the behaviour pattern. Later, the predictive performance of two random forests was analysed based on the past transaction of card holders, where the base classifier was different [34]. The formalisation of fraud detection model was developed towards the description of operating condition of fraud detection system (FDS) to identify the fraudulent transaction in real time. Also, an effective learning strategy was proposed to overcome the class imbalance, concept drift and verification latency [5]. A weighted extreme learning machine (WELM) was applied to identify the fraudulent transaction from the genuine transaction. The proposed work adopted various intelligent optimisation to optimise the WELM [40]. The behaviour profiles of card holders were extracted from the past transactions, which was used for further classification of incoming transactions. In addition, the proposed approach constructed a logical graph of behaviour profiles (LGBP) based on the attributes of the transactions and gives an information entropy-based diversity co-efficient to analyse the characteristics of transactions [38]. Zhenchuan et al. proposed a new loss function named full centre loss which considers the distance and angle among the feature to classify the fraudulent transaction from the genuine one [21].

3 Proposed Hybrid Model for Credit Card Fraud Detection

A generic credit card fraud detection process which can be employed for real-time fraud detection is shown in Fig. 1. The sources of data are different platforms on which credit card is used, of which point of sale (POS), ATM and online are popular. Suppose a transaction comes to the credit card issuer as request to be verified and accepted or rejected. First, the transaction is input into the streams where outlier detection, pre-processing and feature engineering are performed within the streams. [8] provides such stream pipelines which can be configured to perform our required tasks. In the proposed framework, the pipeline of the present study is shown in Fig. 2. We have employed hybrid ensemble classifier as our primary model because of its sturdiness, described in detail in the following sub-sections. Model decides whether to accept or reject the given request, and accordingly, the data are inserted into the database. Strenuously working monitoring and alerting jobs are employed by transaction monitoring team within the financial institution. In this paper, we mainly focus on the Offline Model Training module. Feature engineering based on Adaboost is used to obtain the feature variables for the hybrid ensemble model which is trained on the historical data added with feature variables. After training, the model is coupled with our fraud detection process for real-time fraud detection.

Several single models were proposed earlier for credit card fraud detection; the need of the hour is a hybrid model which can reduce the risk involved in fraudulent transactions by efficient classification and prediction. Our proposed model for credit card fraud detection constitutes identifying those samples which have a higher probability of being fraudulent. Credit card fraud detection has probably been the most explored problem overall in fraud detection. This problem comes with several challenges such as an unbalanced dataset, dynamic fraudulent behaviour of adversary, noise within the data, size of data and evaluation metrics required as explained by [36]. Outliers and noise within the training set can seriously affect the efficacy of our classifier; pre-processing becomes a much needed step to tackle such issues.

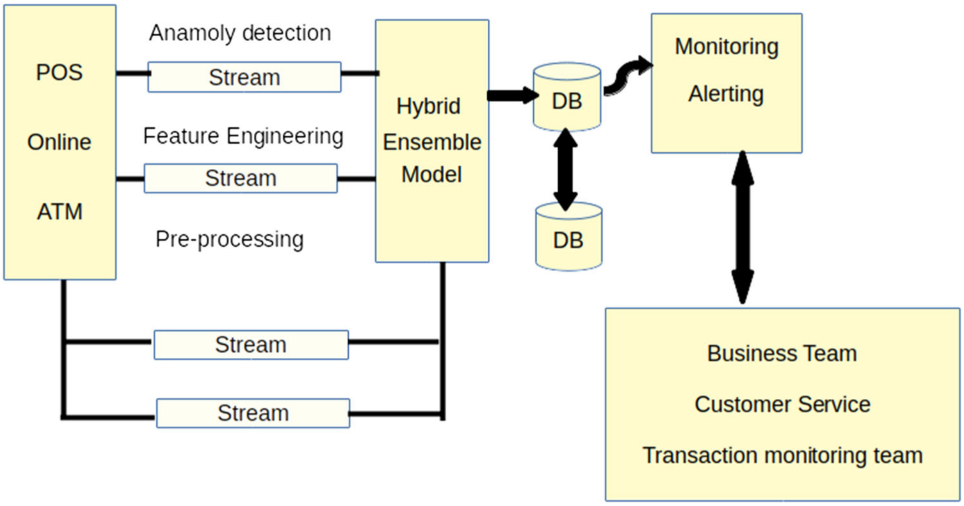


Fig. 1 Credit card fraud detection process

3.1 Pre-Processing

The initial set of features or raw features constitute the information regarding each sample in the dataset. Pre-processing is an essential step as without it our model may produce misleading results and is useful for maintaining the quality of data. Using the raw features, we perform the data distribution, outlier detection and noise elimination task as a part of the pre-processing step.

Outliers are the anomalies; they consist of data patterns that deviate from the normal sample instances. Outliers give us significant insight into the fraudulent uses of a credit card. In our model, we have performed the outlier detection by using isolation forest as described by [22].
For the given dataset X, Y and each sample can be represented as (x_i, y_i), where x_i is a vector of D dimensions and y_i is the class label. Each sample instance is given anomaly score or outlier score $\zeta(x_i, n)$ where $x \in X$ and n denote the number of samples.

$$\zeta(x_i, n) = e^{\frac{-E(h(x))}{c(n)}} \tag{1}$$

$c(n)$ denotes the cost of unsuccessful binary search and score ζ is observed to be monotonous to the function $h(x)$. The constraints to the (1) $0 \leq \zeta \leq 1$ and $0 \leq h(x) \leq n-1$. For a sample, x_i if score s is very close to 1, then x_i outlier. If samples have outlier score $s < 0.5$, then not outliers and considered as normal samples. After detecting the outliers, we experimented by removing them from the samples. Data visualisation was used after performing outlier detection, noise reduction to estimate the performance of hybrid approach.

Algorithm 1: Adaboost + Ensemble learning based Classifier

```
(I) Pre-processing    // Detecting outliers and
    Noise reduction .
(II) Feature Engineering using Adaboost
    Feature Engineering is performed with Adaboost. This step
    varies for the two data sets as the value of $K$ varies .
    (a) $f_1$ = top $K$ features of the data set according to the
        Adaboost algorithm.
    (b) $f_2$ = -ve Log Likelihood probabilities of each sample.
    (c) $f_3 = f_1 + f_2$    // Here the + operation
        indicates the concatenation operation
(III) Split the data into test and train.
(IV) Using the Random Forest model or Extra Trees Classifier
    multi-threading, hyper-parameter tuning and Bootstrap
    sampling are used while training the model.
(V) Classify the test data and generate results.
```

3.2 Feature Engineering

In the current approach, along with misclassification measures, actual financial cost associated with the fraudulent samples and behaviour of fraudster needs to be considered. This insight can only be obtained through feature engineering. In proposed model, we use the classic adaptive boosting (Adaboost) algorithm developed by [10]. [32] explained process of selecting the important features using Adaboost in face detection and facial expression recognition, as it was scalable and could be applied to streaming data. Here, an importance score is assigned to each feature $f_{i,s} \in F$; top features are selected based on this score which indirectly acts as weights for each of the features. The importance score is, in turn, a result of classifier c_i which has to learn only on the feature f_i . The algorithm takes several rounds to converge, and let the number of rounds be T , $0 \leq t \leq T$. The error

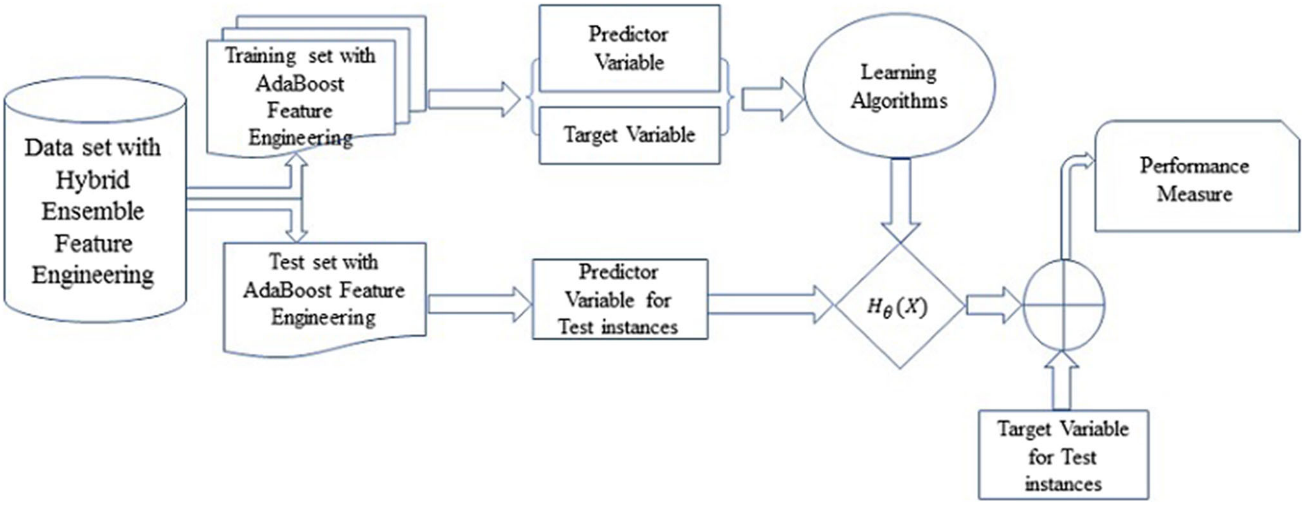

Data set with Hybrid Ensemble Feature Engineering
Training set with AdaBoost Feature Engineering
Test set with AdaBoost Feature Engineering
Predictor Variable
Target Variable
Learning Algorithms
Predictor Variable for Test instances
$H_\theta(X)$
Performance Measure
Target Variable for Test instances

Fig. 2 Block diagram of CCFD model

for the round t would be calculated as in (2) where w_t is the weight vector the features in round t. The error for the feature j is ϵ_j which is calculated based on the results of the classifier c_j, the ϵ_j which gives the lowest value to be considered.

$$w_t, \epsilon_{t,j} = \sum_i w_i \left\| c_j(x_i) - y_i \right\| \tag{2}$$

The weights could be modified with (3), l_i is 0 when the sample i is classified as a positive label and 1 is classified as a negative label. β_t is a fraction which is calculated in (4) which incorporates the error ϵ_t for the round t,

$$w_{t+1,i} = w_{t,i}\beta_t^{1-l_i} \tag{3}$$

$$\beta_t = \frac{\epsilon_t}{1-\epsilon_t} \tag{4}$$

The negative log-likelihood generated from the classifier is taken as feature-set f_2 in the feature engineering step. The dimensionality of f_2 is n X $\|C\|$ where $\|C\|$ is number of classes; in the case of credit card fraud detection it is 2, since it is the case of binary classification. For the feature set f_1, we have to tune the parameter K is tuned, where $0 \leq K \leq \|F\|$. We have tuned this parameter by maintaining a threshold value in decrease in accuracy, true positive rate and true negative rate. The importance of feature engineering in the proposed approach is depicted in Fig. 3.

3.3 Bagging based Ensemble Classifier

Bagging stands for bootstrap aggregation, a popular technique used in an ensemble learning-based models that incorporate both classification and regression methods which in turn increase the accuracy and other related metrics. Bagging is based on the grouping weak learners to make a strong learner. We have used decision tree-based bagging classifiers such as extra-trees classifier and random forest-based classifier for our experimentation.

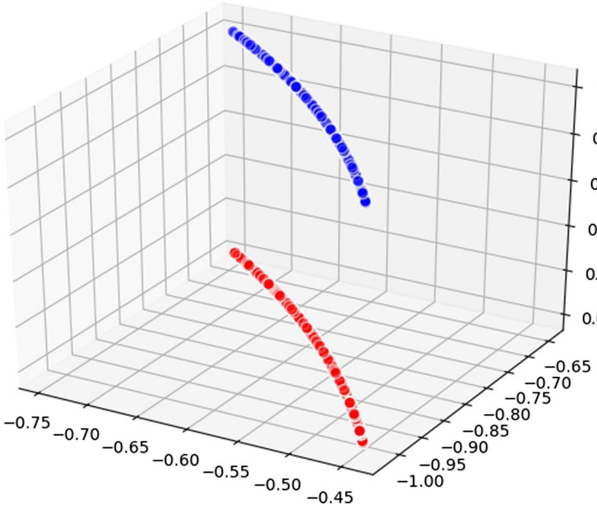


Fig. 3 Advantage of Feature Engineering

Bootstrap sampling uses sampling with replacement which creates a bootstrap sample BS_i such that $\|BS_i\|$ is equal to $\|D\|$, where D is the given data. BS_i is like an independent version of D, and when $\|D\|$ is large enough, the empirical distributions which were assumed are similar to D. So BS_i could be considered as an independent and similar version of D. In bagging, the predictions of the model are averaged to fit the bootstrap sample BS_i in the i^{th} iteration. Ultimately bootstrap sampling tries to kill the tendency of a classifier to overfit.

3.3.1 Random Forest

The idea of random forest was initially introduced by [3].This model is created as a combination of several tree predictors which are independent of each other. Several of such predictors are decision trees, the value of each tree predictor is dependent upon the individual Bootstrap samples on which it learns [24,26]. These tree predictors are then made to vote to choose their best class, based on these votes the predictions of the trees are averaged out. The major difference between this model, and bagging is the additional parameter which denotes how many features to choose from the given set of features in the dataset.

The ensemble consists of several decision trees which are individually weak learners; during training phase when a new sample is given to the classifier, all of the decision trees in ensemble are updated based on the given sample. Although using random forest has several advantages, the main motivation for choosing this method would be its efficiency in handling the imbalance of the data.

3.3.2 Extra Trees Classifier

Extra trees or extremely randomised trees classifier [6] is also an ensemble learning technique which could be considered as a variant of random forest described in 3.3.1 . Conventional decision trees have higher variance which means that they have tendency to overfit the data. This problem was tackled with random forest which has medium variance. Extra trees classifier is better than random forest in the context of overfitting; it has low variance which means that it has less tendency to overfit when compared to random forest.

There are two major differences between random forest and extra trees classifier. The first difference between these two models is that the idea of bootstrap sampling is dropped, and instead, an optimal cut-point is calculated for each of the K randomly chosen features. The second difference is that extra trees classifier splits the features totally or partially at random. Although there is not much difference if we compare the results of both the classifiers in sect. 4, these two modifications make extra trees a better classifier than random forest.

4 Experimental Setting and Fraud Detection Framework

This section describes the accuracy, sturdiness of our proposed model and compares it with the state-of-the-art schemes in this research field. Our primary goal is to improve the fraud detection efficacy of the model. To achieve this, deeper insight into the data is needed.

Table 1 Notation used for metrics definitions

Metric	Symbol
True positive	κ_{11}
False positive	κ_{01}
True negative	κ_{00}
False negative	κ_{10}

Brazillian Bank data and UCSD-FICO dataset are behavioural in nature. Most of the important features we considered were Merchant category code categorical in nature, which denotes the type of business the customer is buying from. Corporate employee program of credit card companies attracts a lot of corporate employees; data from this also become critical. Point of sale machine data are also a critical source; credit card companies also use credit limit and credit score to evaluate the repaying nature of the customer. Location data such as area and state could also play a key role in our decision-making. Multiple transactions of single customer might give us a deeper understanding of the behavioural pattern.

With the above features said, it is important to understand the latent structure of them; it is clear that the features correspond to the behavioural data. From preliminary data analysis, we learned that each feature in the dataset follows the multivariate Gaussian distribution. It might seem that this is a trivial problem to solve, but the curse of data imbalance is the primary hurdle. The empirical risk associated with every transaction is too high; as a result, we cannot afford even a single fraudulent transactions cannot be overlooked.

4.1 Dataset Description

The experimentation was conducted on a modest Quad-Core,8 GB memory Linux environment, which showed robust results on Brazilian bank data, a real-time transactional data which contain 0.3 million data samples and UCSD-FICO data, an E-commerce data containing two versions of which hard version which contains 0.1 million data samples was used. Brazilian bank data have a data imbalance ratio of 25.7, UCSD-FICO dataset had data imbalance ratio of 45.6. 0.1 million transaction samples belonged to 70124 customers, and there were multiple transactions of a single customer which could lead us to get deeper insight into fraudulent behaviour of the cardholder. Table 1 provides a brief overview of the formal notation adopted for this article.

4.2 Metrics

The performance measure of the learning algorithm showed a skewed behaviour in unbalanced distribution of the classes. Hence, it is required to choose appropriate metrics to evaluate the performance of the classification algorithm. In unbalanced scenario, the learning algorithm exhibits an accuracy paradox; therefore, true positive rate, false positive rate, true negative rate, false negative rate, sensitivity, specificity and Mathews correlation coefficients (MCC) have been chosen as performance evaluation metrics for the proposed work.

True positive rate defined as the number of fraudulent transaction was correctly identified from the total number of transaction which was labelled as fraudulent. The relation for TPR is given in equation 5. It plays a significant role while dealing with a classification task in unbalanced scenario.

$$True\,positive\;\; rate = \frac{\kappa_{11}}{\kappa_{11} + \kappa_{10}} \tag{5}$$

False positive rate (FPR) counts the number of transactions wrongly identified as fraudulent from the total number of legitimate transaction. This is also known as Type-I error. The relation to find the FPR is given in equation 6.

$$False\,positive\;\; rate = \frac{\kappa_{01}}{\kappa_{01} + \kappa_{00}} \tag{6}$$

True negative rate (TNR) defines the number of transactions are correctly identified as legitimate from the total number of legitimate transaction. The relation for TNR is given in terms of true negative (κ_{00}) and false positive (κ_{01}) in equation 7.

$$True\,negative\;\; rate = \frac{\kappa_{00}}{\kappa_{00} + \kappa_{01}} \tag{7}$$

False negative rate (FNR) is also termed as Type-II error. It can be defined as the number of transactions wrongly identified as fraudulent out of the total fraudulent transaction. The mathematical equation is given as 8.

$$False\,negative\;\; rate = \frac{\kappa_{10}}{\kappa_{11} + \kappa_{10}} \tag{8}$$

Accuracy gives the sum of number of correctly identified legitimate transactions and fraudulent transactions over all the test instances. The relation for measure of accuracy given in terms of true positive (κ_{11}) and true negative (κ_{00}) in equation 9.

$$Accuracy = \frac{\kappa_{11} + \kappa_{00}}{\kappa_{11} + \kappa_{00} + \kappa_{01} + \kappa_{10}} \tag{9}$$

Matthews correlation coefficient (MCC): MCC measures the degree of relationship between the actual label and predicted label. It takes **-1** when the actual class label and predicted class label are completely complement to each other. But it takes **1** if and only if actual class label is same as predicted class label.

$$MCC = \frac{(\kappa_{11} * \kappa_{00}) - (\kappa_{01} * \kappa_{10})}{\sqrt{(\kappa_{11} + \kappa_{01})(\kappa_{11} + \kappa_{00})(\kappa_{00} + \kappa_{01})(\kappa_{00} + \kappa_{10})}} \tag{10}$$

Area under precision–recall curve (AUPR): AUPR gives the area under precision and recall for various thresholds. Hence, the measure of precision can be calculated by using Eq.11; however, recall can be calculated by using Eq. 12.

$$Precision = \frac{\kappa_{11}}{\kappa_{11} + \kappa_{10}} \tag{11}$$

$$Recall = \frac{\kappa_{11}}{\kappa_{11} + \kappa_{01}} \tag{12}$$

Detection rate is defined as how accurately the model is predicting the true positive cases in other words how exactly the model is handling the fraudulent cases in credit card fraud detection domain. The mathematical relation is given as 13

$$Detection\ Rate = \frac{\kappa_{11}}{\kappa_{11} + \kappa_{10}} \tag{13}$$

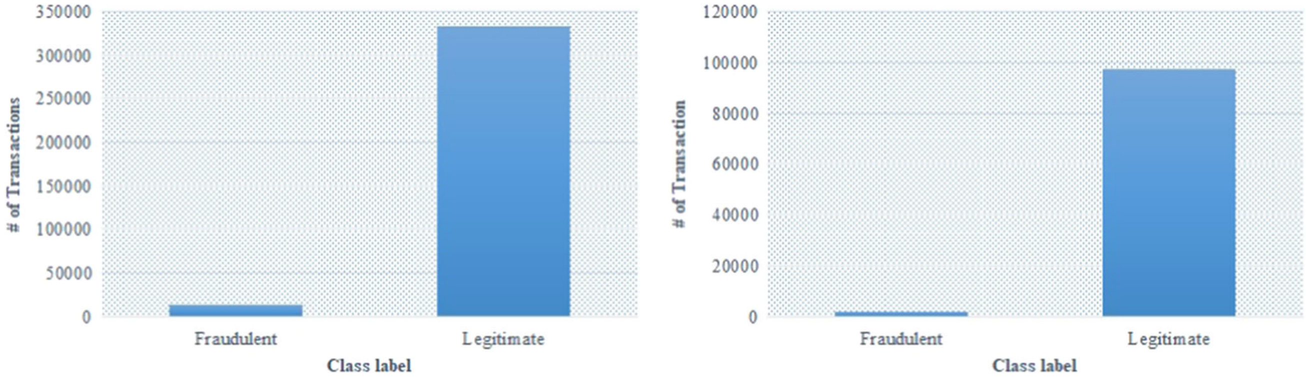
4.3 Data Imbalance

Most of the real-life datasets suffer with data imbalance, in the case of binary classification. In recent times, data imbalance problem has been solved by several people such as [27], [15], [12], etc.

The two datasets used in this paper could be considered as benchmarks in this field of research. Most of the financial institutions such as bank are generally not willing to provide data to the researchers, taking into account the privacy of data of its customers. In our case, data imbalance reflects the scarcity in the count of the fraudulent transactions, as most of the credit card users are not fraudulent. The distribution is shown in Fig. 4. More specifically, the class distribution for legitimate and fraudulent transaction in Brazilian bank dataset and UCSD-FICO dataset is shown in Fig. 4a, b, respectively.

The major concerns for the imbalanced learning problem are the marginalised representation of data, and the data skew is explained by [15]. The detailed description on tackling the data imbalance problem was given by [12]. They introduced three strategies to address this problem; they are data level approaches, algorithm level approaches and hybrid approaches. We have used a hybrid-level approach to tackle the problem of class imbalance. Hybrid-level approach combines both data-level and algorithm-level approaches.

The data-level approach uses re-sampling which is again divided into two sub-approaches under sampling and oversampling. Each approach has its' own snag, for example, under sampling causes loss of information and oversampling may lead to overfitting. Algorithm-level approach tweaks the design of the algorithm and makes suitable changes to



(a) Brazilian bank dataset

(b) UCSD-FICO dataset

Fig. 4 Distribution of fraudulent and legitimate transactions

it to address the class imbalance problem. This level generally employs either a cost-sensitive learning approach or ensemble-based learning approach. We employed hybrid approach by performing oversampling and using an ensemble learning-based classifier as discussed in sect. 3.

4.4 Experimentation Results

The major issues pertaining to credit-card fraud detection problems were addressed by using an Ensemble model, because this problem could be seen as a predictive behavioural modelling problem. The tendency to commit fraud lies well within few customers and to model a deeper insight into historical transactions must be considered. Ensemble models perform on par with the state-of-art models, yet they don't improve the false positive rate and false negative rate significantly due to non-availability of any good variables within the original UCSD-FICO and Brazilian bank datasets. To capture the decision-making pattern of the cardholder and provide us with a clean decision-making variables ensemble models become a good choice. Hence, Adaboost method was adopted to generate an optimal feature space.

These variables were added and tested with ensemble models such as bagging, boosting and predictive models such as logistic regression. On comparing with the results of Table 2 and Table 3, an improvement in specificity and MCC of both the classifiers on Brazilian Bank data was observed. The proposed approach makes use of five-, tenfold cross-validation technique while model building; however, the validation is made with the remaining partition which was kept untouched, i.e. 0.2-0.4 fraction of the original dataset. The results shown in Table 4 and Table 5 signify predictive behaviour of proposed approach on UCSD-FICO dataset, where specificity improved and MCC did not deviate much from the original value. The approach of formulating this problem as a predictive behavioural problem has shown improvement in detecting the false positives; they cause huge damage to financial institutions as they are bad loans. Improvement in specificity solidifies better detection of fraudulent transactions.

The predictive behaviour of the proposed model was analysed with respect to the area under precision–recall (AUPR) curve. The results depicted in Fig. 5a, b give the visualisation of AUPR curve for LR, boosting, Adaboost + random forest and Adaboost + extra trees. The proposed method gives a marginal improvement in the range of 58.03-69.97%, 54.66-69.40%, respectively, on Brazilian bank dataset and UCSD-FICO dataset as reported in Table 6. In other words, the predominant behaviour of the proposed method is presented in Table 6, which signifies that the proposed approach handles the minority class instances in an efficient way by adopting the ensemble feature engineering. In addition, the predictive performance of proposed models outperformed over LR and boosting classifiers shown in Fig. 5. From Table 6, it is evidently inferred that the proposed models were able to capture all the fraudulent transactions.

Performance evaluation of the proposed model described in algorithm 1 has shown sturdy results on both the datasets which are described in Tables 7 and 8, respectively. In addition to this, the predictive performance of the proposed model was validated by taking a partition ratio in the range 0.2-0.4 with an interval of 0.05, while five-, tenfold cross-validation technique was adopted in the remaining partition, i.e. 0.6-0.8, to make the training procedure more effective.

For Brazilian bank dataset, the proposed model has the best results over state-of-the-art models. An improved false positive rate is by 99% and false negatives by 99.9% from the previous best state-of-the-art model which is the testimony of the efficacy of our model. The improvement detection rate, accuracy and AUC were observed as 75%, 3.6% and 32.7%, respectively. Fraudulent transactions in Brazilian bank data contribute to about 3.76 % of the total transactional data. Our model consists of two variants of tree-based ensemble learning classifier with boosting-based ensemble learning

Table 2 Performance measure before applying hybrid ensemble on Brazilian bank data

Learning algorithms	Accuracy	Error rate	Sensitivity	Specificity	F1-score	MCC
LR	0.9625	0.0375	0.9951	0.1162	0.9808	0.2219
Boosting	0.9728	0.0272	0.9764	0.7785	0.9860	0.5307

Table 3 Performance measure after applying feature engineering on Brazilian bank data

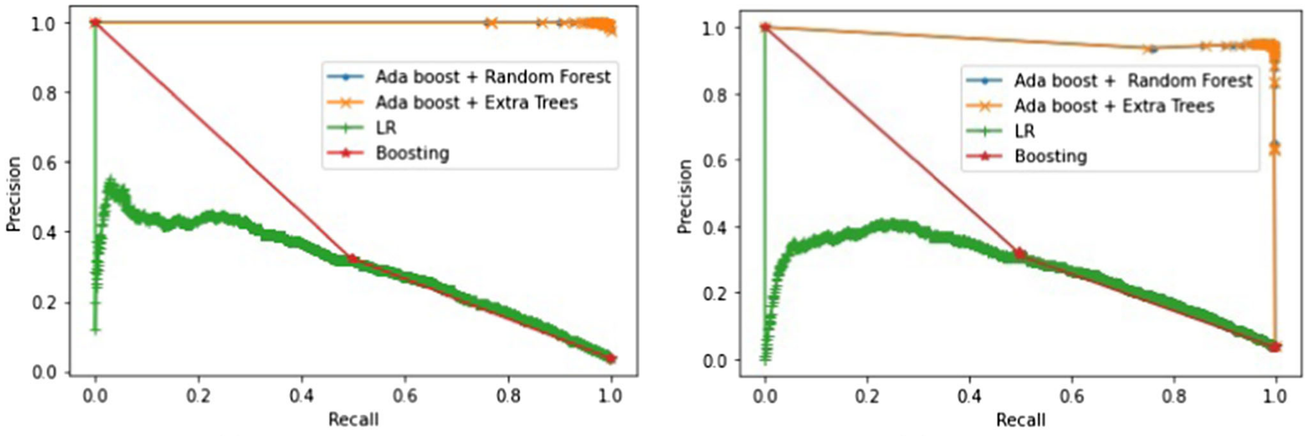
Learning Algorithms	Accuracy	Error rate	Sensitivity	Specificity	F1-Score	MCC
LR	0.9996	0.0004	0.9998	0.9953	0.9998	0.9949
Boosting	0.9743	0.0257	0.9765	0.8428	0.9868	0.5557

Table 4 Performance measure before applying hybrid ensemble on UCSD-FICO data

Learning Algorithms	Accuracy	Error rate	Sensitivity	Specificity	F1-Score	MCC
LR	0.9871	0.0129	0.9991	0.2883	0.9935	0.4885
Boosting	0.9906	0.0094	0.990.	0.7886	0.9952	0.6668

Table 5 Performance measure after applying hybrid ensemble on UCSD FICO data

Learning Algorithms	Accuracy	Error rate	Sensitivity	Specificity	F1-Score	MCC
LR	0.9887	0.0013	0.9983	0.4264	0.9943	0.5835
Boosting	0.9904	0.0096	0.9926	0.7917	0.9951	0.6558



(a) Brazilian bank dataset **(b)** UCSD-FICO dataset

Fig. 5 Analysis of AUPR for the proposed approach

Table 6 Measure of AUPR on benchmark datasets

Classifier name	Brazilian bank dataset Measure of AUPR	UCSD-FICO dataset Measure of AUPR
LR	0.300234	0.268294
Boosting	0.419651	0.415617
Adaboost + random forest	0.999920	0.961786
Adaboost + extra trees	0.999928	0.962217

Table 7 Result analysis Brazilian bank data

Name	TPR	FPR	TNR	FNR	DR	Acc	MCC	AUC
RIBIB [1]	0.566154	0.119147	0.880853	0.433846	0.566154	0.869064	–	0.723500
AFDM [14]	0.518400	0.018100	0.981900	0.481500	0.518000	0.964600	–	0.750100
CSNN [13]	0.314300	0.009300	**0.990600**	0.685600	0.314300	0.964100	–	0.652400
AIRS [11]	0.420200	0.021200	0.978700	0.579700	0.420200	0.958500	–	0.699400
Adaboost+ Extra Trees (Proposed method)	**0.999690**	**0.002350**	**0.997000**	**0.000304**	**0.991800**	**0.999600**	**1.000000**	**0.995900**
Adaboost + RandomForest (Proposed method)	**0.999600**	**0.002350**	**0.997000**	**0.000304**	**0.991800**	**0.999600**	**1.000000**	**0.995900**

Table 8 Result analysis on UCSD-FICO dataset

Name	TPR	FPR	TNR	FNR	DR	Acc	MCC	AUC	Recall
RIBIB [1]	**1.00000**	0.00800	**0.99000**	**0.00000**	–	–	0.85000	**0.99000**	**0.9900**
Fraud Miner [28]	0.89000	0.25000	0.75000	0.02500	–	–	0.83000	0.90000	0.89000
Enhanced Fraud Miner [16]	1.00000	0.25000	0.75000	0.00000	–	–	0.90000	0.90000	1.00000
Adaboost+ Extra Trees (Proposed method)	**0.99405**	**0.00256**	**0.99800**	0.00595	0.64900	0.99150	**1.00000**	0.92100	**0.9800**
Adaboost + RandomForest (Proposed method)	**0.99430**	**0.002511**	**0.998050**	0.005670	0.66505	0.99180	**1.00000**	0.92400	**0.98000**

technique Adaboost. Experimentation of our models on the Brazilian bank dataset has shown sturdy results reported in Table 7, and this can be proved when the results of our model are compared to the state-of-the-art fraud detection models such as RIBIB by [1], Fraud miner by [28] and Enhanced fraud miner by [16]. A comparison with the state-of-the-art models of Brazilian bank data is shown in Table 7. For UCSD-FICO dataset, the proposed model has improved false positive rate, MCC by 68% and 10%, respectively, reported in Table 8. These results prove that hybrid ensemble model is better than the cost-based models in detecting fraudulent transactions, thus reducing the loss incurred by the financial institutions.

5 Conclusion and Future Work

This article proposed a hybrid ensemble-based credit card fraud detection. In the first phase of the work, the ensemble feature engineering techniques were used for mapping the original feature space to a best feature space and in the next part using the generated feature space; then, the tree based learning algorithm is applied for classification and prediction process. The proposed model shows a measure of false positive rate of 0.00235, false negative rate of 0.0003048, detection rate of 0.9918, accuracy of 0.9996, MCC of 1 and area under the receiver characteristics measure of 0.9959 which beats the state-of-the techniques like RIBIB. The proposed work shows certain kind of limitation; for example, the imbalance nature of the data set can be handled in efficient way introducing cost-based extreme learning approach, where the cost assigned for the misclassified instances in certain ratio depends upon the imbalance ratio of the dataset. Also, the concept of drift was introduced for analyzing how the behaviour of the customer changes periodically.

Acknowledgements We would like to thank the Machine Learning and Data Analytics Lab and Big Data Analytics Lab for the infrastructure support.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Akila, S.; Reddy, U.S.: Cost-sensitive risk induced bayesian inference bagging (ribib) for credit card fraud detection. J. Comput. Sci. **27**, 247–254 (2018)
2. Batista, G.E.; Prati, R.C.; Monard, M.C.: A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsl **6**(1), 20–29 (2004)
3. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
4. Carcillo, F.; Dal Pozzolo, A.; Le Borgne, Y.A.; Caelen, O.; Mazzer, Y.; Bontempi, G.: Scarff: a scalable framework for streaming credit

card fraud detection with spark. Information Fusion **41**, 182–194 (2018)

5. Dal Pozzolo, A.; Boracchi, G.; Caelen, O.; Alippi, C.; Bontempi, G.: Credit card fraud detection: a realistic modeling and a novel learning strategy. IEEE Trans. Neural Netw. Learn. Syst. **29**(8), 3784–3797 (2017)
6. Désir, C.; Petitjean, C.; Heutte, L.; Salaun, M.; Thiberville, L.: Classification of endomicroscopic images of the lung based on random subwindows and extra-trees. IEEE Trans. Biomed. Eng. **59**(9), 2677–2683 (2012)
7. Foundation, A.S.: Apache cassandra (2016). http://cassandra.apache.org
8. Foundation, A.S.: Apache kafka (2017). https://kafka.apache.org
9. Foundation, A.S.: Apache spark (2018). https://spark.apache.org/
10. Freund, Y.; Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**(1), 119–139 (1997)
11. Gadi, M.F.A.; Wang, X.; do Lago, A.P. : Credit card fraud detection with artificial immune system. In: International Conference on Artificial Immune Systems, pp. 119–131. Springer (2008)
12. Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F.: A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans. Syst., Man, Cybernet., Part C (Applications and Reviews) **42**(4), 463–484 (2011)
13. Ghobadi, F.; Rohani, M.: Cost sensitive modeling of credit card fraud using neural network strategy. In: 2016 2nd international conference of signal processing and intelligent systems (ICSPIS), pp. 1–5. IEEE (2016)
14. Halvaiee, N.S.; Akbari, M.K.: A novel model for credit card fraud detection using artificial immune systems. Appl. Soft Comput. **24**, 40–49 (2014)
15. He, H.; Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowledge Data Eng. **9**, 1263–1284 (2008)
16. Hegazy, M., Madian, A., Ragaie, M.: Enhanced fraud miner: credit card fraud detection using clustering data mining techniques. Egyptian Computer Science Journal (ISSN: 1110–2586) **40**(03) (2016)
17. Jiang, C.; Song, J.; Liu, G.; Zheng, L.; Luan, W.: Credit card fraud detection: A novel approach using aggregation strategy and feedback mechanism. IEEE Internet Things J. **5**(5), 3637–3647 (2018)
18. Kim, E.; Lee, J.; Shin, H.; Yang, H., Cho, S.; Nam, S.k., Song, Y., Yoon, J.a., Kim, J.i. , : Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning. Expert Syst. Appl. **128**, 214–224 (2019)
19. Kültür, Y.; Çağlayan, M.U.: A novel cardholder behavior model for detecting credit card fraud. Intelligent Automation & Soft Computing 1–11 (2017)
20. Li, Z.; Huang, M.; Liu, G.; Jiang, C.: A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. Expert Syst. Appl. **175**, 114750 (2021)
21. Li, Z.; Liu, G.; Jiang, C.: Deep representation learning with full center loss for credit card fraud detection. IEEE Trans. Comput. Soc. Syst. **7**(2), 569–579 (2020)
22. Liu, F.T.; Ting, K.M.; Zhou, Z.H.: Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422. IEEE (2008)
23. Liu, J.; Zio, E.: Integration of feature vector selection and support vector machine for classification of imbalanced data. Appl. Soft Comput. **75**, 702–711 (2019)
24. Livingston, F.: Implementation of breiman's random forest machine learning algorithm. ECE591Q Machine Learning Journal Paper 1–13 (2005)
25. Nilson: (2019). https://www.nilsonreport.com
26. Pal, M.: Random forest classifier for remote sensing classification. Int. J. Remote Sens. **26**(1), 217–222 (2005)
27. Raghuwanshi, B.S.; Shukla, S.: Underbagging based reduced kernelized weighted extreme learning machine for class imbalance learning. Eng. Appl. Artif. Intell. **74**, 252–270 (2018)
28. Seeja, K.; Zareapoor, M.: Fraudminer: A novel credit card fraud detection model based on frequent itemset mining. The Scientific World Journal **2014**,(2014)
29. Tao, X.; Li, Q.; Guo, W.; Ren, C.; Li, C.; Liu, R.; Zou, J.: Self-adaptive cost weights-based support vector machine cost-sensitive ensemble for imbalanced data classification. Inf. Sci. **487**, 31–56 (2019)
30. Tian, Y., Liu, G.: Mane, : Mane: Model-agnostic non-linear explanations for deep learning model. In: 2020 IEEE World Congress on Services (SERVICES), pp. 33–36. IEEE (2020)
31. Van Vlasselaer, V.; Bravo, C.; Caelen, O.; Eliassi-Rad, T.; Akoglu, L.; Snoeck, M.; Baesens, B.: Apate: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decis. Support Syst. **75**, 38–48 (2015)
32. Wang, R.: Adaboost for feature selection, classification and its relation with svm, a review. Phys. Procedia **25**, 800–807 (2012)
33. Xie, Y.; Liu, G.; Cao, R.; Li, Z.; Yan, C.; Jiang, C.: A feature extraction method for credit card fraud detection. In: 2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS), pp. 70–75. IEEE (2019)
34. Xuan, S.; Liu, G.; Li, Z.; Zheng, L.; Wang, S.; Jiang, C.: Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., Jiang, C.: Random forest for credit card fraud detection. In: 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), pp. 1–6. IEEE (2018)
35. Yu, L.; Zhou, R.; Tang, L.; Chen, R.: A dbn-based resampling svm ensemble learning paradigm for credit classification with imbalanced data. Appl. Soft Comput. **69**, 192–202 (2018)
36. Zareapoor, M.; Shamsolmoali, P.: Application of credit card fraud detection: Based on bagging ensemble classifier. Procedia Comput. Sci. **48**, 679–686 (2015). https://doi.org/10.1016/j.procs.2015.04.201
37. Zelenkov, Y.: Example-dependent cost-sensitive adaptive boosting. Expert Systems with Applications (2019)
38. Zheng, L.; Liu, G.; Yan, C.; Jiang, C.: Transaction fraud detection based on total order relation and behavior diversity. IEEE Trans. Comput. Soc. Syst. **5**(3), 796–806 (2018)
39. Zheng, Y.J.; Zhou, X.H.; Sheng, W.G.; Xue, Y.; Chen, S.Y.: Generative adversarial network based telecom fraud detection at the receiving bank. Neural Netw. **102**, 78–86 (2018)
40. Zhu, H.; Liu, G.; Zhou, M.; Xie, Y.; Abusorrah, A.; Kang, Q.: Optimizing weighted extreme learning machines for imbalanced classification and application to credit card fraud detection. Neurocomputing **407**, 50–62 (2020)