Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 27 (2024), No. 4, pp. 1233–1246
DOI : 10.47974/JDMSC-1978

Leveraging ensemble learning for enhanced security in credit card transaction fraudulent within smart cities for cybersecurity challenges

Bharat Kumar Padhi [‡]
Sujata Chakravarty [†]
Department of Computer Science & Engineering
Centurion University of Technology & Management
Bhubaneswar 761211
Odisha
India

Bighnaraj Naik [§]
Department of Computer Application
Veer Surendra Sai University of Technology
Burla 768018
Odisha
India

Soumya Ranjan Nayak*
School of Computer Engineering
KIIT Deemed to be University
Bhubaneswar 751024
Odisha
India

Ramesh Chandra Poonia [@]
Department of Computer Science
CHRIST (Deemed to be University)
Delhi-NCR
Ghaziabad 201003
Uttar Pradesh
India

[‡] *E-mail:* 170506102002@cutm.ac.in
[†] *E-mail:* sujata.chakravarty@cutm.ac.in
[§] *E-mail:* bnaik_mca@vssut.ac.in
* *E-mail:* nayak.soumya17@gmail.com (Corresponding Author)
[@] *E-mail:* rameshcpoonia@gmail.com

Abstract

In the age of digital transactions, credit cards have emerged as a prevalent form of payment in smart cities. However, the surge in online transactions has heightened the challenge of accurately discerning legitimate from fraudulent activities. This paper addresses this crucial concern by introducing a pioneering system for detecting fraudulent credit card transactions, particularly within highly imbalanced datasets, in the realm of cybersecurity. This paper proposes a hybrid model to effectively manage imbalanced data and enhance the detection of fraudulent transactions. This paper emphasizes the efficacy of the hybrid approach in proficiently identifying and mitigating fraudulent activities within highly imbalanced datasets, thereby contributing to the reduction of financial losses for both merchants and customers in smart cities. As cybersecurity in smart cities evolves, this paper underscores the significance of ensemble learning and cross-validation techniques in optimizing credit card transaction analysis and fortifying the security of digital payment systems.

Subject Classification: *Primary 94A60, Secondary 68T05.*

Keywords: *Cyber security, Smart cities, Fraud, Fraud detection, Machine learning, Ensemble learning algorithms.*

1. Introduction

The transition from traditional offline to internet commerce has resulted in an increase in cashless transactions, shaping the landscape of smart cities' development. In this rapidly evolving digital environment, cybersecurity emerges as a pivotal aspect in the operation of fraud detection systems [1], which are integral to safeguarding financial transactions in smart cities. The losses from credit card fraud have increased from USD 9.84 billion in 2011 to USD 32.39 billion in 2020, with worrying estimates of reaching USD 40.62 billion by 2027. Scammers take advantage of weaknesses in smart city systems to access personal information, such as birth dates, social security numbers, and private details, which allow them to do unlawful activities and unauthorized buys that could interfere with smart city services. In the realm of credit card fraud detection within smart cities, complexities like shifting fraud patterns, class imbalance, and feature inaccuracies prevail. Advances in cybersecurity, coupled with machine learning and artificial intelligence [2-3], offer promising solutions. Some of the supervised, unsupervised, reinforcement learning, and ensemble methods, play a crucial role in enhancing the accuracy of credit card fraud detection [4] models, thereby contributing to the secure functioning of smart city ecosystems. However, the dynamic nature of smart city transactions necessitates innovative

hybrid models, ensuring secure detection of fraudulent credit card activities.

The primary goal of this paper is to develop a robust hybrid model that overcomes overfitting difficulties in imbalanced datasets, fostering increased generalizability of the model while fine-tuning critical parameters. The proposed approach includes three main contributions, such as:

- The problem of imbalanced datasets has been mitigated with the implementation of expanded cybersecurity-oriented SMOTE techniques, as outlined in algorithm 1.
- The suggested hybrid SMOTE-LightGBM (SLGBM) model, designed with a primary objective of reducing model complexity and computational requirements, has been successfully implemented to achieve optimal parameter selection, as outlined in algorithms 2 and 3.
- The performance of the investigation was evaluated by employing commonly used metrics such as recall, F1-score, accuracy, precision, and ROC curve.
- Finally, proposed approach is contrasted with existing Machine Learning(ML) models, as well as hybrid models, to demonstrate the efficacy of the proposed model.

This paper is designed in the following way: Section 2 gives a summary of the previous research on credit card fraud transactions. Section 3 describes the hybrid method used for identifying credit card fraud. The performance evaluation, using different measures, is shown in Section 4. Finally, Section 5 covers the discussion and conclusion of this research work.

2. Literature review

In this section, we also reviewed twenty articles on approaches to credit card fraud detection (CCFD) systems using machine learning algorithms and hybrid models. The paper [1] proposes a hybrid model that uses ML for an effective CCFD system, while paper [2] shows an improved LightGBM ML approach, boosting fraud detection through Bayesian-based hyperparameter optimization. In [3] explores a hybrid strategy, merging supervised and unsupervised ML techniques within a

CCFD system. In [4], the AdaBoost algorithm is applied with a majority vote approach for credit card fraud detection. Additionally, [5,6] addresses fraud detection challenges, including class imbalance and paradigm drift, aiming to enhance the efficient analysis and detection of credit card fraudulent transactions. In [7], various ML algorithms were tested on an unbalanced dataset to detect fraudulent transactions. [8] utilizes Support Vector Machine (SVM) and Artificial Neural Network (ANN) to address credit card fraud with a focus on time-varying data, achieving optimal accuracy with ANN. [9] proposes an improved Semi-hidden Markov model (SHMM) through an ensemble-based approach, outperforming individual SHMMs in detecting fraudulent transactions. In paper [10] discusses an anomaly detection technique combining oversampling and principal component analysis approaches. In [11], diverse ML algorithms are evaluated for identifying fraudulent transactions in an unbalanced dataset. [12] presents a hybrid approach, combining Neural Networks (NN) and Genetic Algorithms (GA), to detect fraudulent transactions. Meanwhile, [13] emphasizes the efficiency of Neural Networks (NN) in credit card fraud detection, aiming to enhance accuracy in processing credit card transactional data. Moreover, [14] deals with CCFD by using different classification and ensemble methods to handle imbalanced data issues and improve accuracy in identifying whether transactions are legitimate or fraudulent. In [15], the focus is on employing the K-Nearest Neighbors (KNN) algorithm and outlier detection techniques for credit card fraud identification using various machine learning methods. The paper [16, 17] introduces a unique approach to CCFD, incorporating an aggregation strategy and a feedback mechanism to improve the accuracy of fraud detection. In [18], challenges in detecting credit card fraud are reviewed and addressed, while [19] discusses effective approaches for CCFD Systems in 2017, highlighting challenges in implementing artificial intelligence (AI) and machine learning (ML). A novel approach utilizing rule-based filtering, Dempster-Shafer theory, and Bayesian learning for CCFD Systems is proposed in [20]. [21] compares k-nearest neighbor (KNN) with various ML algorithms for fraud detection in credit card transactions. In [22], CCFD Systems based on transactional behavior is explored, with the C4.5 algorithm demonstrating superior results over Random Forest (RF) in detecting fraudulent transactions. The literature suggests ensemble algorithms as potential solutions to improve fraud detection in smart city environments.

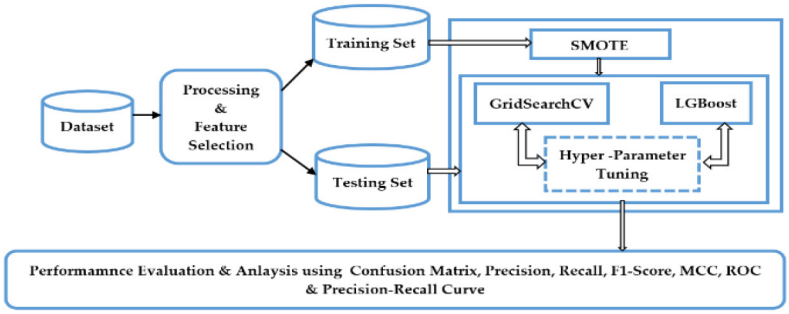

Figure 1

Proposed SLGBM hybrid model

3. Proposed Intelligent System Design

This section shows a graphic illustration of the suggested smart SLGBM hybrid model for identifying credit card fraudulent transactions shown in Figure 1. The model accurately distinguishes between genuine and fraudulent transactions, addressing cybersecurity challenges in the banking and business sectors.

The credit card dataset, sourced from European cardholders' transactions in September 2013, consists of 31 columns, with numeric input variables [23]. To preserve customer privacy, 28 variables (V1, V2, …, V28) were converted to PCA out of the 31 input variables. The 'class' input feature designates fraud as 1 and non-fraud as 0. The dataset contains 492 fraudulent transactions, making it strongly unbalanced with a fraud rate of 1.72% and a non-fraud rate of 98.28% [22], as depicted in **Figure 2**. The dataset imbalance reveals a bias against non-fraud transactions. To address this, the training dataset is balanced using **Algorithm 1**, to mitigate overfitting, as depicted in **Figure 3**. Additionally, variables with very low correlations were removed using the correlation factor and Kernel Density Estimation (KDE) method to identify the most crucial features for further processing, as depicted in Figure 4.

Utilizing a correlation plot, Ultimately, features V1-V7, V9-V12, V14, V16-V19, and V21, excluding 'Time' and 'Amount,' were selected.

The modified training dataset, initially (227845, 19) dimensions, becomes (454902, 19) after applying **Algorithm 1**, addressing issues associated with imbalanced datasets to mitigate overfitting effects.

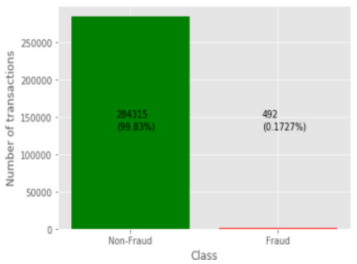

Figure 2

Non-Fraud Vs Fraud Transactions before Sampling

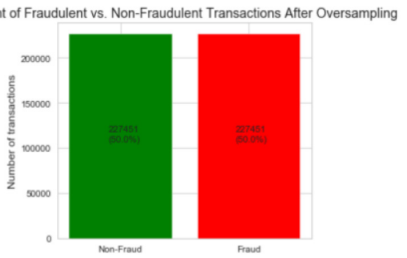

Figure 3

Non-Fraud Vs Fraud Transactions after Sampling

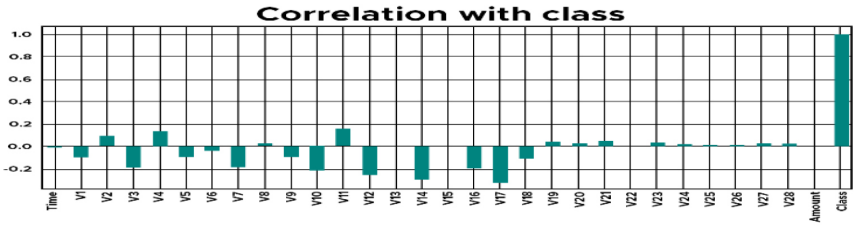

Figure 4

Feature correlations plot using the correlation factor and KDE

Algorithm 1: Implementation of Smote Technique

Input: The minority dataset **M** .

Output: Synthetic Minority dataset **M′**

1. **M** is identified from the imbalanced dataset. Such that class **M** make the smaller proportion in the dataset.
2. For each element $\mathbf{e} \in \mathbf{M}$, the Euclidian distance is calculated to every other element of **M** to get the k-nearest neighbors of **e**.
3. The sampling rate **S** is set such that it corresponds to the imbalance ratio i,e $\mathbf{S_M}$ / $\mathbf{S_N}$, Where $\mathbf{S_M}$ is the size of the minority class and $\mathbf{S_N}$ is the size of the majority class.
4. **S** random values are selected from the class **M** and named **M′**.
5. The new minority values can be generated using the following formula:

$e_{new} = e + |e - e_n| * random()$

For each element $\mathbf{e_n} \in \mathbf{M'}$. Where random() generates a random positive float value less than 1.

6. RETURN **M′**

In addition, **Algorithm 2** is used to optimize the Light Gradient booting algorithm's parameters. We have split the entire dataset in cross-validation as k = 5, and each test fold has 20% of the whole dataset.

Algorithm 2: Grid Search Cross Validation Algorithm

Input: Dataset D; Hyper-parameters values of Light Gradient boosting Machine: X={ X_1 , X_2, X_3, ……., X_n }, Where i =1 to n

Output: Optimized X_i values.

1. For each X_i in the list X parameters:
   - 1.1 Create k-folds from the D for k-fold cross-validation {D_1, D_2, … D_k}
   - 1.2 Train Light Gradient boosting algorithm with D_j and X_i, LGBM (D_j, X_i), j=1,2,…k
     - 1.2.1 Find the aggregate predictive error (Err(X_i)) for X_i from all folds
     - 1.2.2 if (Err(X_i)< Err(X_i-1)), then X*=X_i. Else X*=X_i-1
2. Return X*

The proposed hybrid method adopts the gradient boosting technique, operating on three key principles: i) weak learners (decision trees), ii) gradient optimization, and iii) boosting method. The final values, obtained after **Algorithm 2**, include boosting_type = 'gbdt',colsample_bytree = 0.57, learning_rate = 0.01, estimators = 10000, num_leaves = 126, max_depth = -1, importance_type = 'split', min_child_samples = 30, min_child_weight = 0.005, min_split_gain = 0.01, n_jobs = -1, reg_alpha = 0.4, reg_lambda = 0.6, max_bin = 5000, min_data_in_leaf = 150, objective = 'binary',random_state = 0, subsample = 0.4, silent = -1, verbose = -1, subsample_for_bin = 200000, and subsample_freq = 1. To avoid fitting the data too closely, these parameters need to be adjusted by taking into account factors such as choosing small max_bin and num_leaves values, specifying min_data_in_leaf, controlling min_sum_hessian_in_leaf, using bagging with bagging fraction and bagging_freq, applying feature sub-sampling with feature_fraction, enlarging training data, changing lambda_l1 and lambda_l2, setting min_gain_to_split for regularization, and restricting max_depth to prevent deep tree growth.

$$F_p(x) = \arg\arg\min_k \sum_{i=1}^{n} L(y,k) \tag{1}$$

The model is initially determined by **Eq.(1),** where **F** is the function for the iteration **P** to determine the anticipated values. x represents the independent variables (V1-V7, V9-V12, V14, V16-V19, and V21), y

Algorithm 3: Proposed SLGBM hybrid model algorithm

Input: Balanced Training dataset :

$\overline{tr} = \{(x1, y1), (x2, y2), \ldots, (xi, yi)\}$, …….. (Algorithm 1 output)

Where, y=dependent variable: {Non-Fraud (0), Fraud (1)}

x=independent variable, d = $\overline{tr}$

i=0,1,2,….,n

Output: Boosted score after every transaction

1. Sampling ratio: a= large gradient data; b= small gradient data.

$$topN = a \times len(d); randN = b \times len(d)$$

2. Loss function:

$$L(y, F(x))$$

where F(x)= the predicted data.

3. Set the optimized hyper parameter values obtained by Algorithm 2.
4. Number of iterations: P
   - 4.1. Initialize model using Eq. (1)
   - 4.2. For p= 1 to P:
   - 4.3. Compute pseudo- residual values using Eq. (2).
   - 4.4. Apply GOSS for re-sorting the dataset.
   - 4.5. $\text{Sorted} = \text{GetSortedIndices}(\text{abs}(g));$
   - 4.6. $\text{topSet}(A) = \text{sorted}\ [1 : \text{top}N];$
   - 4.7. $\text{randSet}\ (B) = \text{RandomPick}(\text{sorted}[\text{top}N : \text{len}(d)],\ \text{rand}N);$
   - 4.8. $\text{newdataset}(d') = A + B$
   - 4.9. Estimate information gains by Eq. (3)
   - 4.10. new decision tree on the d'
   - 4.11. Append $$F_p(x) = F_{p-1}(x) + F_p(x)$$
   - 4.12. For Loop end
5. Return update $F_p(x)$

represents the dependent variable (class variable), n represents the number of instances, and **min** represents the minimal value of the parameter **k**.

$$g_{ip} = \left| \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right|_{F(x) = F_{p-1(x)}} \quad (2)$$

In **Eq.(2),** g_{ip} is the calculated pseudo-residual values of each parameter, where for i=1,…., N.

$$G_j(dn) = \frac{1}{n}\left(\frac{\left(\sum_{x_i \in A_l} gip + \frac{1-a}{b}\sum_{x_i \in B_l} gip\right)^2}{n_l^j(dn)} + \frac{\left(\sum_{x_i \in A_r} gip + \frac{1-a}{b}\sum_{x_i \in B_r} gip\right)^2}{n_r^j(dn)} \right) \quad (3)$$

In the Eq.(3), G_j is the calculated largest gain , where **n**= number of variables, The variance gain of **j** or the dividing measure at the point ***dn*** for the node, i=1,..,N, $A_l = \{x_i \in A : x_{ij} \le dn\}$, $A_r = \{x_i \in A : x_{ij} \le dn\}$, $B_l = \{x_i \in B : x_{ij} \le dn\}$, $B_r = \{x_i \in B : x_{ij} \le dn\}$, and $\frac{1-a}{b}$ is the coefficient to normalize the sum of the gradients over B back to the size of A^k. Finally, it is observed that the selection of the right value for parameters has an impact on both the performance and complexity of the model and the output result of **Algorithm 2**, applied to **Algorithm 3**, which is particularly effective at detecting credit card fraud.

4. Result Analysis

This section talks about the proposed hybrid method for CCFD Systems underwent examination using an Intel Core i5 processor with 16 GB of Dynamic Memory. The experiment utilized a Python script and a Jupyter 6.0.1 notebook to assess its performance against other machine learning techniques.

The SLGBM hybrid model was evaluated using a credit card dataset and compared to state-of-the-art techniques like RF, LR, MLP, SVM, LGBoost, and CatBoost Classifier. Additionally, its performance was assessed against other hybrid models [1], such as Adaboost + LGBM, Adaboost + LR, Adaboost + RF, Adaboost + XGBoost, Adaboost + SVM, and OlightGBM models [2].

Table 1 indicates that the proposed SLGBM hybrid method achieved a recall value of 88.9%, signifying its capability to accurately identify fraudulent transactions with a low false-negative rate. Furthermore, the method demonstrated a precision value of 88.0%, accurately identifying only fraudulent transactions. The F1-score of 87.0% shows the high quality of the method that we suggest. Additionally, the method attained an AUC value of 99.0%, showcasing its proficiency in distinguishing between fraudulent and non-fraudulent transactions.

Table 1

Performance comparisons between the proposed hybrid model and other existing single machine learning models.

Model	Performance Metrics						
	Recall	Precision	F1-score	Accuracy	MCC	ROC Score	Precision Recall Score
RF	0.8472	0.8741	0.8601	0.9995	0.8600	0.9500	0.7600
LR	0.8881	0.1023	0.1822	0.9863	0.2984	0.9600	0.0900
MLP	0.8881	0.2154	0.3472	0.9942	0.4363	0.9800	0.1700
SVM	0.1430	0.0020	0.0040	0.8654	0.0010	0.5600	0.0100
LGBoost	0.8272	0.6483	0.7271	0.9989	0.7311	0.9400	0.5400
CART	0.7963	0.3962	0.5291	0.9976	0.5601	0.9000	0.3200
CatBoost	0.8372	0.8201	0.8282	0.9994	0.8281	0.9800	0.6900
Proposed_ Model	0.8890	0.8780	0.8700	0.9996	0.8730	0.9900	0.7400

Figure 5-6 represent the ROC and precision-recall curve for the current machine learning model, while **Figures 7-8** showcase the ROC and precision-recall curve for the proposed model, offering visual insights into their performance in accurately classifying fraudulent and non-fraudulent instances.

The SLGBM hybrid method demonstrated a remarkable recall value of 0.8890, signifying its ability to accurately detect over 88.9% of fraudulent transactions with a small false-negative rate. Moreover, the average F1_

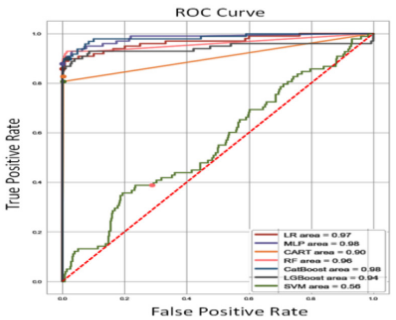

Figure 5

ROC curves of the other machine learning models

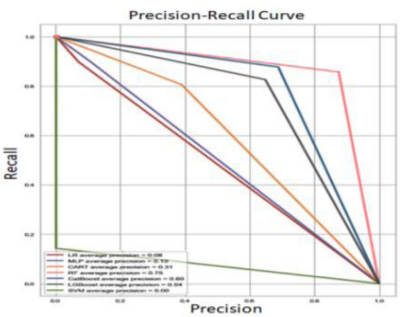

Figure 6

Precision-recall (PR) curves of the other machine learning models.

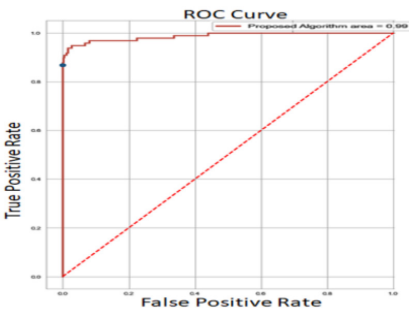

Figure 7

ROC curves of the proposed models.

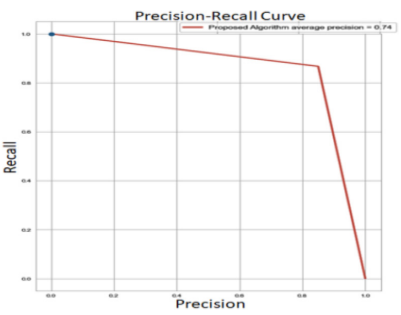

Figure 8

PR curves of the proposed models.

score of 87.0%, precision value of 88.0%, and AUC value of 99.0% showcased outstanding performance, making it one of the top-performing techniques in the proposed hybrid approach. These results were thoroughly analyzed and evaluated in the paper [1],[2] and are summarized in **Table 2**.

As shown in **Table 1**, the proposed hybrid method achieved the highest ROC score of 99.0%. The other existing machine learning models, such as MLP, and CatBoost obtaining the second-highest AUC scores of 98.0% and 98.0%, respectively. The SVM algorithm had the lowest recall, precision, f1-score, MCC value of 14.3 %, 0.2%, 0.4%, and 0.1% respectively. The suggested hybrid approach received the highest precision and recall score of 88% and 89% among hybrid models (as seen in **Table 2**). A higher

Table 2

Performance comparisons between the proposed hybrid model and other hybrid models based on the Recall, Precision, F1_Score, and AUC.

Model	Performance Metrics				Ref.
	Recall	Precision	F1-score	AUC	
Adaboost + LGBM	0.64	0.97	0.77	0.82	[1]
Adaboost+LR	0.36	0.83	0.50	0.68	[1]
Adaboost+RF	0.50	0.97	0.66	0.75	[1]
Adaboost+ XGBoost	0.59	0.94	0.73	0.79	[1]
Adaboost +SVM	0.18	0.91	0.30	0.59	[1]
OlightGBM	0.41	0.97	0.57	0.91	[2]
Proposed_Model	0.89	0.88	0.87	0.99	

ROC score indicates better prediction performance, which the proposed SLGBM hybrid method achieves with a score of 99%. The proposed SLGBM hybrid approach achieves a precision-recall score of 74%, indicating improved prediction performance.

5. Conclusion & Future Research Directions

The conclusion of this paper states the proposed hybrid model stands out from other frameworks by implementing an efficient hybrid approach for identifying fraudulent credit card purchases. It offers a more effective solution to the banking industry and presents a cost-effective option for the banking and financial sectors.

Currently, the hybrid model faces limitations in dealing with high-dimensional datasets, which will be a primary area of investigation for future research. Incorporating meta-heuristic tools aims to overcome these challenges and continuously refine the proposed hybrid approach, enhancing its effectiveness and applicability in smart cities and beyond.

References

[1] Malik, Esraa Faisal, et al. "Credit card fraud detection using a new hybrid machine learning architecture." Mathematics 10.9 : 1480 (2022).

[2] Taha, Altyeb Altaher, and Sharaf Jameel Malebary. "An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine." IEEE Access 8 : 25579-25587 (2020).

[3] Carrillo, Fabrizio, et al. "Combining unsupervised and supervised learning in credit card fraud detection." *Information sciences* 557 : 317-331 (2021).

[4] Randhawa, Kuldeep, et al. "Credit card fraud detection using AdaBoost and majority voting." IEEE access 6 : 14277-14284 (2018).

[5] Dal Pozzolo, Andrea, et al. "Credit card fraud detection: a realistic modeling and a novel learning strategy." IEEE transactions on neural networks and learning systems 29.8 : 3784-3797 (2017).

[6] Raj, S. Benson Edwin, and A. Annie Portia. "Analysis on credit card fraud detection methods." 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET). IEEE, (2011).

[7] Makki, Sara, et al. "An experimental study with imbalanced classification approaches for credit card fraud detection." IEEE Access 7 : 93010-93022 (2019).

[8] Chen, Rong-Chang, et al. "Personalized approach based on SVM and ANN for detecting credit card fraud." 2005 international conference on neural networks and brain. Vol. 2. IEEE (2005).

[9] Prakash, A., and C. Chandrasekar. "An ensemble approach for credit card fraud detection." *International Journal of Computer Applications* 59.19 (2012).

[10] Lee, Yuh-Jye, Yi-Ren Yeh, and Yu-Chiang Frank Wang. "Anomaly detection via online oversampling principal component analysis." *IEEE transactions on knowledge and data engineering* 25.7 : 1460-1470 (2012).

[11] Mittal, Sangeeta, and Shivani Tyagi. "Performance evaluation of machine learning algorithms for credit card fraud detection." 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE (2019).

[12] Patidar, Raghavendra, and Lokesh Sharma. "Credit card fraud detection using neural network." *International Journal of Soft Computing and Engineering* (IJSCE) 1.32-38 (2011).

[13] Ghosh, Sushmito, and Douglas L. Reilly. "Credit card fraud detection with a neural-network." System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on. Vol. 3. IEEE (1994).

[14] Mishra, Ankit, and Chaitanya Ghorpade. "Credit card fraud detection on the skewed data using various classification and ensemble techniques." 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE (2018).

[15] Malini, N., and M. Pushpa. "Analysis on credit card fraud identification techniques based on KNN and outlier detection." 2017 third international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB). IEEE (2017).

[16] Dorronsoro, Jose R., et al. "Neural fraud detection in credit card operations." *IEEE transactions on neural networks* 8.4 : 827-834 (1997).

[17] Jiang, Changjun, et al. "Credit card fraud detection: A novel approach using aggregation strategy and feedback mechanism." *IEEE Internet of Things Journal* 5.5 : 3637-3647 (2018).

[18] Kumar, Ankit, Pankaj Dadheech, Vijander Singh, Ramesh C. Poonia, and Linesh Raja. "An improved quantum key distribution protocol for verification." *Journal of Discrete Mathematical Sciences and Cryptography* 22, no. 4 : 491-498 (2019).

[19] Ryman-Tubb, Nick F., Paul Krause, and Wolfgang Garn. "How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark." *Engineering Applications of Artificial Intelligence* 76 : 130-157 (2018).

[20] Panigrahi, Suvasini, et al. "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning." Information Fusion 10.4 : 354-363 (2009).

[21] Awoyemi, John O., Adebayo O. Adetunmbi, and Samuel A. Oluwadare. "Credit card fraud detection using machine learning techniques: A comparative analysis." 2017 international conference on computing networking and informatics (ICCNI). IEEE (2017).

[22] Fernández, Alberto, et al. "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary." Journal of artificial intelligence research 61 : 863-905 (2018).

[23] Padhi, B. K., S. Chakravarty, and B. N. Biswal. "Anonymized credit card transaction using machine learning techniques." Advances in Intelligent Computing and Communication: Proceedings of ICAC 2019. Springer Singapore (2020).

[24] Credit Card Fraud Dataset. Accessed: Sep. 4, 2023. [Online]. Available:https://www.kaggle.com/mlg-ulb/creditcardfraud/data.

[25] Rajesh Kumar, E., et al. "Suicidal ideation prediction in twitter data using machine learning techniques." *Journal of Interdisciplinary Mathematics* 23.1 : 117-125 (2020).

[26] Nayak, Soumya Ranjan, and Jibitesh Mishra. "A modified triangle box-counting with precision in error fit." *Journal of Information and Optimization Sciences* 39.1 : 113-128 (2018).

[27] Nayak, Soumya Ranjan, et al. "A statistical analysis of COVID-19 using Gaussian and probabilistic model." *Journal of Interdisciplinary Mathematics* 24.1 : 19-32 (2021).

[28] Roy, Arpita, et al. "Fuzzy rule based intelligent system for user authentication based on user behaviour." *Journal of Discrete Mathematical Sciences and Cryptography* 23.2 : 409-417 (2020).