Fraudulent Transactions Detection in Credit Card by using Data Mining Methods: A Review

Amir Aziz[1], Dr. Hamid Ghous[2]

[1]*Research Scholar,* [2]*Research Guide*

Department of Computer Sciences, Institute of Southern Punjab, Multan. Pakistan

Abstract - Right now, it becomes a popular trend in online transactions by using Credit cards and mobile wallets without transferring net currency. By adopting this method number of fraudulent transactions is increasing as opposed to being transferred money physically. Sharing data also becomes the cause of fraud. After following the previous behaviour of users, fraud can be observed. Many different strategies are being obtained and adopted by researchers to minimize massive and complex fraudulent approaches. This paper shows the comparative study of techniques to uncover and detect the ways of fraud by following machine learning (ML) methods like Random Forest (RF), Deep Learning (DL), Support Vector Machine (SVM), Hybrid Methods (HM), and Decision Tree (DT). All these techniques are used to discover common usage patterns of consumers as following their past activities. After reviewing ML methods, it has been revealed tremendous discrepancies amongst different studies for future works. With the help of this study, researchers may follow this to explore their research about credit card frauds. By exploring the actual use of credit cards, researchers may step ahead by following this research in future work.

Keywords: Data Mining, Credit Card Fraud Detection (CCFD), ML Methods, DL Methods.

1 INTRODUCTION

A hand-carry device that is used as a portable card in a credit card identifies records of users with their signature, picture, and title of the authorized man or woman on it. This card is used directly to pay the consignment on the use of unique services. The use of this card is increasing every day beneath the enhancement of technological know-how like ATMs, and the online internet banking system. The credit card has a special range it truly is referred to as code whilst the usage of it that is very necessary to keep secure. These days use of credit cards swiftly increases alongside fraud actions therefore, the foremost goal of credit card detection is to make robust purchaser relationships with the organization also to limit loss for both users and the issuing company. A credit card is a brilliant technology to do transactions electronically barring cash. On one aspect it proves helpful in payment techniques with much less consumption. On the different hand, due to a lack of technological know-how credit card transactions are self-controlled via criminals. The motive is defects in administration and management in transactions. In this era, criminals' activities and strategies are going to their worse stage. To analyse the hazard of crook attacks, outstanding efforts are required.

Tracking down fraud is a massive topic that consists of a lot of industrial equipment containing financial departments, banking systems, groups of authorities' sectors, insurance premiums, regulation enforcement, and many more. An attempt of fraud has been detected in rising for many years and has also become extra indispensable than ever before. Hundreds of millions of bucks are misused to overcome the ratio of fraud however all in vain. Many complicated and problematic elements are mixed up in confirming the fraud case. Data mining and facts are used to become aware of and predict the level of fraud right away and take a surprising action to decrease costs. The equipment of Data mine like cooperation rules, machine learning, decision, boosting or classification trees CHAID and RF, neural networks, the fast-approaching pattern can be developed to rely on records like the danger of fraudulent exercise or how to measure fraud. These predicting fashions for fraud losses can be used in emphasizing assets effectively to prevent or regain a fraudulent process [1].

1.1 SIGNIFICANCE

Nowadays, the main issue for the business of electronic commerce is that there are more fraudulent ways in transactions that are more likely to be genuine. Consequently, demographic fraud detection or straightforward pattern matching methods are not well organized in detecting fraud because there are a few examples of fraud. The data sets for credit card transactions are very unclear. Typically, in real circumstances, ninety-one percent of transactions are legalized, while only one percent of them are fraudulent, so the fraud datasets are extremely unbalanced. The development of productive fraud detection systems is seemly important for all credit card issuing banks to avoid losses. Different modern techniques formed on AI, ML, DM, fuzzy logic, genetic algorithm, etc. are being developed to reveal credit card fraudulent transactions. This research aims to present a contemporary overview of various techniques of classification by comparing

performance across a broad scale of challenges in credit card transaction data sets, and in the conclusion, the study highlights applicability to card fraud detection applications.

The following is the sequence of this paper: Section Second gives a detailed study about credit card fraud background. Section third provides a comprehensive description of various famous techniques, which are utilized to reveal fraud in credit card transactions. The fourth section highlights the discussion utilized to capture fraud in credit card transactions. The fifth section concludes the paper's outcome.

2 BACKGROUND

After following the above-mentioned brief introduction, this section condensed the background about credit card frauds. Bank America was once the first issuing credit card with all acceptance in 1958 the bank of America and first use in California. Then this card is used all over the world with some guidelines and policies that are strictly followed by the consumer in case of any trouble credit score card services avail them the facility to come or take a look at their card. Also, through 1979, almost 1400 banks give credit score card facilities.

As for real-time credit card concerns, the fast internet speed made it one of the most selling mediums for the retailer. From the ultimate decade due to advancements in technological know-how, the use of the card automatically increases and makes it less complicated for online fraud. The misuse of a deposit card barring informing the proprietor turns into the reason of fraud by getting the complex unlawful source and to reissue deposit card after submitting false records to the bank workforce as well.

In the age of technology, the use of E-commerce with a wide variety of customers has elevated every day. Have a Credit card is one of the digital services that is a very convenient and imperative section of economic lifestyles through presenting a fee device at the spot with these advantages:

2.1 EASY TO PURCHASE

The use of credit cards makes life easy; it allows. Users to buy something on a credit card at any time, place besides carrying cash in hand. With patron credit information: it's easy to preserve information of customers for after use to check Sara's fee and in case of a suspicious transaction. Further, this information is very beneficial for other economic services like loans, launder assessment.

2.2 SAFETY WHILE PURCHASING

The deposit card employer offers extra service to the patron to retail their purchase product in case of theft if they have no longer a unique receipt or lost their record. So, the credit card becomes critical trouble in intra-banking services that supply transactions with a deposit card, it's a violation of public regulation in which fraudsters get a sudden benefit as a result reason sudden damage.

As the range of deposit cards is hastily growing it requires more protection measures for this chip and the pin protection machine plays an essential have an effect on deposit card security. The deposit card offerings should supply education to their user on how to preserve their facts impenetrable from theft. So, the prevention of fraud first desires to recheck and update patron data after a few months and test suspicious transactions by using an environmentally friendly system. So, the prevention of fraud first wishes to recheck and update patron statistics after a few months and test suspicious transactions by using an efficient system.

3 LITERATURE REVIEW

Banks that issue credit cards to their customers need to take effective steps to detect credit card fraudulent transactions. It has been noticed that different techniques could be utilized to reveal credit card fraudulent transactions. Various techniques including RF, HM, DT, SWM, DL methods, and some other approaches are listed and classified as follows for detecting fraudulent credit card transactions:

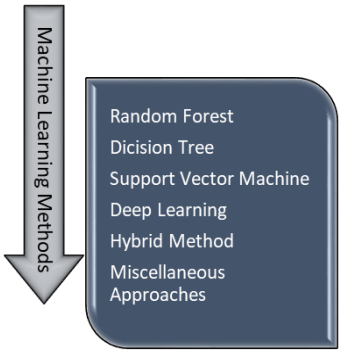

Fig. 1 Review Structure

3.1 RANDOM FOREST

An integrated technique in RF which might be deployed to create approximative models for problems with regression and classification. RF is a method utilized in issues of regression and classification. RF grows many DT, all of which serve as weak learners, but all together become strong learners. RF can easily and efficiently process vast and unbalanced databases with thousands of features [2].

DT, LR, and RF algorithms are operated to measure the operation for credit card fraud identification. On behalf of a rather unbalanced dataset oversampling is required. After oversampling, 60% legal and 40% unlawful transactions are found. R language is used for the implementation of these algorithms. Accuracy of LR is

90.0%, RF 95.5%, and DT 94.3%. Sensitivity, error rates, and specificity are also measured. RF algorithm performs nicely amongst them [3].

The HMM-based method presents automated features engineering so it can increase the effectiveness and also increase in detecting fraud transactions. Feature engineering operates nicely in e-commerce and physical fraud detection: For the f2f transaction, It increases precision-recall AUC of 18.1% and 9.3% for e-commerce [4].

Automatic classification and guide classification are used in fraud identification as properly as one of a kind ML algorithm is compared to pick out the frauds. RF, SVM, and LR are used. They find out about targets to advance a threat scoring model. All the algorithms are tested and RF is performed properly and accomplished with the best possible accuracy. And this algorithm is effortless to practice and works precisely on a massive dataset. The result confirmed that sort of algorithm performs very properly in the real-world [5].

Some ML algorithms are proposed to test the performance of fairly unbalanced data. SVM, RF, DT, and LR are operated to take a look at the potential. These algorithms are examined on pre-processed and uncooked data. The accuracy of these algorithms is SVM 97.5%, RF 98.6%. DT 95.5% and LR 97.7% respectively. The RF performs very well on a large quantity of information however it suffers from speed. If records are more pre-processed then SVM can work properly amongst them [6].

A Hidden Markov Model (HMM) is modelled on each of these sets. Each HMM, given its sequence of previous transactions, associates a likelihood with a transaction. Pre-processing is achieved via Feature Engineering through using HMM Method and with the aid of Splitting Data. In an RF algorithm for fraud detection, certain possibilities are used as greater features. To mannequin the properties of logical order in the dataset with appreciation to the classification work, this more than one standpoint HMM-based strategy lets in for automatic function engineering. Set side by side with the word of the art feature engineering system for the CCFD, this approach makes the AUC precision-recall to increase by 15 percent [7].

The information is used in the research is from a Chinese e-commerce firm. Two types of RF are used to instruct the behavioural characteristics of regular and unusual transactions. To compare and evaluate their overall performance on deposit fraud detection of the two RF algorithms that are distinct in their base classifiers. While RFs operate well on small set data with 98.67% accuracy, some issues such as imbalanced records are still present [8].

Experiments on 6 public datasets show that the RF and every other weighted RF outperform in the system. This study observes the identification of credit card fraud and checks additionally demonstrate that the overall approach is the best. The final approach suggests and examines an improved RF, RWRF, with the RF and TWRF with 97.76% accuracy. Finally, research has been discussed the problem of binary classification and further recommendations are to be followed in multi-classification problems [9].

3.2 LIMITATIONS

In the above literature, some weaknesses could be discussed. Imbalanced datasets are the key problem that should be managed carefully, similar patterns of legal and illegal transactions that should be different, and some classifiers did not work well when detecting face-to-face fraud transactions that may trigger poor precision.

TABLE 1. RANDOM FOREST

References	Dataset	Pre-processing	Methods	Results	Limitations or F/W
[2]	UCSD-FICO Data Mining Contest Dataset 2009. Transaction: 100000. Customers: 73729. Fields: 20	-SMOTE for Oversampling	-Fraud Miner -SVM -NB -RF -KNN	Sensitivity Catching Rate -1 -0.1 -0.9 -0.7 -0.6	Legal and illegal Transaction patterns should be different for a model.
[3]	Credit Card Transaction Dataset from Kaggle. Transactions: 284808	-Random Oversampling -Splitting Data 70%, 30% -Feature Selection	- RF. -Logistic Regression. -Decision Tree	Accuracy -95.5 -90.0 -94.3	NA
[4]	Credit card	Aggregated and	-RF	- Precision-Recall	It would be important

	transaction dataset from the industrial partner	HMM-based Feature Engineering	-LR -AdaBoost	AUC Increase: 18.1% for f2f Transaction. -Increase in For E-Commerce: 9.3% With Feature Engineering	to combine the predictions of an LSTM with the prediction of some improved RF HMM-based features as these classifiers have been shown to not detect the same frauds in face-to-face transactions as a future job.
[5]	The dataset from the fraud analysts a case study company -Training set Observations: 347572 -Testing set observation: 86893	-Missing Value Handling b/w 0 to 1 -Min-Max Standardization -K-fold cross-validation -Random under-sampling	-LR -SVM -RF	RF: -AUC-PR: 0,333 -AUC-ROC: 0,880	Need Comparison of Transforming function and which features to engineer for support guidance. Oversampling is needed for fraud records.
[6]	Credit Card Dataset is taken from ULB Machine Learning Group Transactions: 284786 Features: 30	-K-Fold Cross-Validation -PCA for Feature Selection	-Logistic Regression -SVM -Decision Tree - RF	Accuracy 97.7% 97.5% 95.5% 98.6%	More pre-processing is Required to get better results.
[7]	Credit card transactions dataset from industrial partner	-Feature Engineering by using HMM -Splitting Data in three parts	RF	Precision-recall AUC Increase of 15.1%	NA
[8]	Dataset from e-commerce company of China	-Bootstrap Sampling -Random Under Sampling	RF	Accuracy 98.67%	Some issues, such as imbalanced data, still exist. Our future work will concentrate on addressing these issues. The RF Algorithm should be enhanced.
[9]	Dataset of B2C on credit card transactions and other public datasets are used	-Bootstrap Sampling -Random Under Sampling -Data Splitting ratio of 3:1	RF TWRF RWRF	Accuracy 97.11% 97.11% 97.76%	The authors only include binary classification issues in this article. In the future, we expect to implement it in a multi-classification issue.

3.3 DECISION TREE

The C4.5 selection tree is used for extended characteristic resolution and wrapper approach. The DT forest framework is used for ensemble classification via the usage of a cost-sensitive selection tree. A fraud detection dataset is regionally gathered for this proposed method. Evolution metrics are measured by way of F-Measure, Recall, and accuracy and in contrast with different classification methods consisting of ID3, Naïve Bayes, J48, NB, and Bayesian Network. Improvement in the proposed technique yields 1.8% to 2.4%. 27 decision bushes and the best possible rating tree are chosen that generate maximum precision and F-Measure. The proposed approach of precision is 99.6 primarily based on F-Measure [10].

A detection model is proposed via using several classifiers however J48 and Random Tree supply the highest accuracy concerning 93.50% and 94.32%. WEKA is used for model introduction after pre-processing and inspecting the data. Data is not elaborated in detail because of the agreement between the bank and the supporter. The device will ship SMS to these clients who are detected with fraud transactions alternatively of sending SMS to all. Random Tree provides perfect accuracy over J48 [11].

To construct a very powerful fraud detector technique, this lookup brought a DT algorithm supplemented with regression analysis. Data pre-processing is done by way of feature selection. In detection and reporting term of credit card fraud, the scheme protects all regions. The simulation result here exhibits that this method is 81.6 percentage correct with 18.4 percentage misclassification error, and all the inserted incursions used for testing have been checked efficiently with the aid of the system [12].

NB, C4.5 DT, and bagging ensemble ML methods are used to predict the effect of daily transactions and frauds. Algorithm effectivity is calculated by accuracy, recall, PRC area rates. Machine getting to know algorithm output PRC values linking 0,999 and 1,000 recommend that these algorithms in the dataset are very nice at determining binary classification zero. Bagging with a C4.5 DT as a beginner with a cost of 0,825 is among all the first-class performing PRC type 1 algorithms. With the C4.5 DT algorithm, the prediction of fraud transactions with 92.74 percentage success is once effectively predicted [13].

3.4 LIMITATIONS

However, some limitations should be noted. The key concern with some models is that if acquired consumers are told of a fraudulent transaction on time that could not be done and imbalanced data is also an issue in credit card datasets.

TABLE 2. DECISION TREE

Reference	Dataset	Pre-processing	Methods	Results	Limitations or F/W
[10]	Credit Card Dataset has taken form Amirkabir University	- Extended wrapper -based feature selection -Feature Ranking using Chi-Squared Filter, gain ratio, and Relief	Ensemble classification performed in a decision forest model by using cost-sensitive DT's	-Precision: 99.96% -improvement: 2.4%	The majority Voting method can be utilized for feature selection in the future.
[11]	Dataset is taken from the Fraud Cases and transaction log file -Records: 9992	-Data sanitation -Normalization -Binnig	-Random Tree -J48	Accuracy -94.32% -93.50%	The model will be Implement in the financial sector. The benefit of this is to send SMS to those customers who have a fraudulent transaction.
[12]	Credit card transaction dataset	-Normalization B/W 0 to 1 -Regression Analysis	-Decision Tree	Accuracy 81.6% Misclassification Error 18.4%	NA
[13]	Transaction dataset	N/A	-Naïve Bayes -C4.5 DT -Bagging	-PRC Area of Naïve Bayes: Class1= 0,080 Class0=1,000 PRCArea of C4.5 Decision tree Class1=0,745 Class0=0,999 PRCArea of Bagging Class1=0,825 Class0=1,000	NA

3.5 SUPPORT VECTOR MACHINE

Input data is cut up into fraud and normal transactions based on test and training sets. This entire work is accomplished with the help of the SVM classifier. The proposed mannequin carried out nicely in contrast to prevailing techniques. The accuracy of the proposed technique is 90%. It's a lot tricky to pick parameters of kernel characteristic so, its drawbacks of the proposed method [14].

Step with the aid of-step system is used for detection of fraud as the first step is checking the transaction in the dataset then finding the item set in credit card after that remember the number of transactions and counting is accomplished primarily based on the group. After that how to check many transactions have been made through every user. At the quit classification algorithms are used like apriori and SVM. Apriori is used to refine the dataset. SVM is better for the separation of fraud transactions. Results assessment is once made with present strategies but the proposed method carried out well on the exclusive variety of instances. Accuracy of HMM is 68.2% and SVM get 72.3% accuracy [15].

SVM is an ML algorithm based on input and output matching. This approach analyses how customers use credit cards to detect fraudulent transactions. This approach analyses how customers use credit cards to detect fraudulent transactions. If the new transaction behaves differently than normal, it is considered fraud. The ability to emphasize the customer's cost structure is selected from the dataset. Good results are obtained when a few behavioural functions are used. Therefore, SVMs are used for pattern recognition and classification. Classify templates as rogue and non-illegal. This guarantees efficient operation and high accuracy [16].

SVM is used to identify transactions as valid or fraudulent. The SVM analyses the past transaction habits of the cardholder. When a new transaction happens, by marking it as an unlawful transaction, it deviates from its previous behaviour. On SVM, the highest fraud detection score is 91% [17].

To create SVM models for monitoring credit card fraud transactions on the Internet, the SVM algorithm is used. Each model's test results are then analysed. To find the best SVM, data is pre-processed and then compared with the hybrid ID3 + BP model. SVM performed well against the hybrid model [18].

3.6 LIMITATIONS

The SVM has speed and size limitations that arise during the SVM training and testing process. The parameters of the kernel function are therefore not easy to select, considered as a drawback of the algorithm. The use of unoptimized kernel functionality is also a concern.

TABLE 3. SUPPORT VECTOR MACHINE

Reference	Dataset	Pre-processing	Methods	Results	Limitations or F/W
[14]	-10 Years Historical Data about Credit card	Regression Model is used for pre-processing	SVM Classifier	Accuracy 90%	Restrictions of this research are that it so hard to pick the parameters of the kernel function as well as speed and size
[15]	Credit Card Transaction Dataset. UCI -Instances: 600 -Attributes: 23	-Splitting the Dataset in legal and fraud	Classification Method: -HMM -SVM	Accuracy: - -68.2% -72.3%	NA
[16]	Download Transaction Dataset from UCI Website	Feature Extraction Using PCA	SVM	Accuracy Above 80%	In the future, for lower error rates, a cost-based SVM with an optimized kernel feature would be used to detect fraud.
[17]	Financial Institution Dataset	-SMOTE for Oversampling -CNN and Random Under-Sampling for	-SVM -NB -LR	Accuracy -91% -83% -74%	Hope to focus on developing the levels of prediction to achieve a better

		Under-Sampling	-KNN	-72%	prediction. Future enhancements are also planned to concentrate on location-based fraud.
[18]	Commercial Bank's Dataset	-Remove Dirty Data -Handling Missing Values -Correct Error Data -Data Conversion -Data Selection	-SVM	SVM Performance is Higher than Hybrid Model	NA

3.7 DEEP LEARNING

CNN is a nonparametric method of classification and it performs a vital role in decision making. KNN mannequin is used in previous research that has some problems like very excessive statistical complexity and it only works nicely on a big quantity of datasets. On the other hand, CNN performs very properly on a small number of samples and it undertakes a hundred percent perfection on the education setting. CNN offers a condensed set that is a subset of a true set [19].

A system's mannequin is designed for fraud detection. A supervised anomaly algorithm for detection is applied to notice real and fake transactions. This algorithm is primarily based on a neural community that works like a human brain. A subsystem is modelled that can be used with purposes and software programs to notice fraud in transactions [20].

This is the first task in which exposure to the credit card fraud detection issue has ever been used through 3D ConNet. Compared with different modern-day baseline approaches, strategies obtain successful AUC and precision-recall curves. Also, inspect the analysis of obtained interest amounts in research to find out fraud trends. In a digital payment, post-analysis scheme, the counselled method is completely examined. The result suggests that strategies can notice fraudulent transactions effectively [21].

This work aims to predict transactions involving credit card fraud through the use of two MLP and ELM artificial neural community algorithms. The effects exhibit that MLP exceeds ELM the usage of different techniques, like accuracy, recall, false effective frequency matrix, truly high-quality rate, and classification time. That being said, the creator can infer that ELM is very effortless to predict new fraudulent transactions due to its simple architecture compared to MLP by using counting the predicting time for each algorithm [22].

The study is suggested a deep network method for fraud detection. To manipulate the data skew troubles that occur in the dataset, the log transformation is used. For the training of difficult examples, the focal reduction is utilized to the network. The effects show that the different classical models such as SVM and LR are outperformed by using the neural network model [23].

3.8 LIMITATIONS

DL model computational complexity is a big challenge and also these methods perform badly on small datasets. Some models are not tested in a real-time environment. The computational complexity of the DL model is a major challenge and these approaches often work poorly on small datasets. In a real-time environment, certain models are not tested.

TABLE 4. DEEP LEARNING

References	Dataset	Pre-processing	Methods	Results	Future Direction
[19]	Credit card transaction dataset	Sampling	CNN Algorithm	Accuracy 100% on training Set	Computational Complexity still a challenge in the condensed training set
[20]	Credit Card data collected from different sources.	Data Selection	Neural Network -> Anomaly Detection Algorithm	Error Free Model is Proposed	More study is needed for the enhancement and testing of this system
[21]	Benchmark dataset	-Feature Exaction by combining users with multiple cards. (Temporal Slice, Spatial Slice, Features) -The three-sigma law of Cut Off outliers	-Neural Network (STAN) -LSTM-seq -CNN-max -Deep & wide	AUC -0.8865 -0.8290 -0.8267 -0.8108	In time ahead, writers are ready to develop online fraud detection system instead of offline.

		-Down-sampling	-AdaBM -LR	-0.8232 -0.7199	
[22]	European Credit card fraud dataset	SMOTE for Feature selection (Oversampling)	Artificial Neural Network -MLP -ELM	MLP Model Accuracy=97.84% ELM Accuracy= 95.46%	Deep Networks Performs best on large datasets instead of Small
[23]	CIS dataset Download from Kaggle	-Filling missing values with 0 -Standard Normalization for Numeric Values -One Hot Coding is used for Categorical Values	-Naïve Bayes -Logistic Regression -SVM -NN	The accuracy of algorithms is Naïve Bayes=0.875 LG=0.911 SVM=0.932 NN=0.954	NA

3.9 HYBRID METHODS

Input data is allotted in certain portions one is the training section and the difference is the trying out part. In the end, an evaluation made between the voting-based classification model and Hybrid based classification model, the exactness and performance of the proposed model is observed very well. Hybrid Model and this model additionally reduce execution time. The accuracy of the voting-based model is 99.12 and Hybrid based mannequin accuracy determined 99.95 [24].

A massive dataset is used to put in force this model so records cleaning is an essential technique to put off the noise in records and records normalization is additionally performed. Dataset had 1356243 data after performing the cleansing 1000023 archives remained handy. The original dataset is so huge so 7 subsets of dataset are created like A, B, C, D respectively. The overall performance is estimated by way of the usage of Kappa Statistics, F-Measure, and Recall. The hybrid model suggests proportion enhancement is 71.81% and contribution is some initial discoveries [25].

Existing systems have the restraint that are time constraints, scalability, and imbalanced class. HSVM is delivered to reduce these restraints with the way of Spike Identification and Communal detection. In Classification and pattern recognition Hybrid SVM is often operated. Prediction of fraud and crime things are finished by using the machine in the initial stage. HSVM is used to identification of frauds for every attribute and it is weighted for Communal and SD. A scalable and efficient gadget is proposed [26].

A hybrid strategy is proposed for identifying credit card frauds by operating the DT and Rough Set method that can be used in detection. The total work has utilized the usage of the software's WEKA and MATLAB. After the 10-time execution of the proposed and existing method, the proposed technique performs very well with 84.25% [27].

This research observes credit card fraud via computer mastering algorithms. Generic versions are used first. Hybrid strategies that use AdaBoost and techniques of majority voting are then implemented. Then, it analyses a real-world deposit card information series from an economic institution. Moreover, to similarly check the robustness of the methods, noise is utilized to the information samples. In the empirical assessment, a vary of widespread methods has been used, which includes NB, SVM, and DL. A publicly accessible deposit card records set has been used for assessing by using AdaBoost and majority voting combination approaches with the help of individual models, and hybrid models. The experimental outcomes exhibit positively that the gadget of majority vote casting achieves robust accuracy costs in the recognition of fraud cases in credit card [28].

The Hybrid based method integrates three techniques: the technique of RFE to less the count of features, the approach of HPO to estimate optimized algorithms for RFC-based mannequin, and the SMOTE to clear up the imbalanced statistics problem. Research is carried out mannequin on three massive datasets, used with the aid of the Machine Learning community as referenced datasets, to validate findings, and it demonstrates its capability to acquire excessive precision output all through the CCFD phase. Moreover, the model ensures a very suitable output regardless of the used datasets in phrases of evaluation with recent works [29].

Research suggests an HM in this work by using both DM and statistical activities, cost-sensitive learning, resampling, feature selection, and for the CCFD. By using GA, useful facets are identified in the first step. Next, depending on the nature of testing and feedback outward appearance methodologies, the ultimate resampling approach is calculated. Finally, in the Adaboost algorithm, the cost-sensitive C4.5 algorithm is utilized as a beginner. This proposed method's accuracy and sensitivity have been 96.59 percent and 67.52 percent respectively. This methodology suggests that in contrast to other DM algorithms, the hybrid recommended approach has the right performance for detecting fraud transactions [30].

The study presents an HM that joins unsupervised and supervised methods to enhance the precision in fraud detection. Compared and checked on an actual, annotated, CCFD dataset are unsupervised outlier ratings, computed at quite a several stages of granularity. Experimental findings indicate that the combination is profitable and that the precision of the detection is also improved [31].

3.10 LIMITATIONS

The performance of hybrid models is strong, but still exits some drawbacks, as execution time is high compared to other techniques, the real-world credit card dataset is a major challenge that is difficult to find and only works with balanced datasets.

TABLE 5. HYBRID METHODS

Reference	Dataset	Pre-processing	Method	Result	Limitations or F/W
[24]	-The dataset from UCI Repository -Instances: 30000 -Attributes: 24	-Cross-Validation Technique is used to separate data	-Hybrid Classification Algorithms. KNN & Naïve Byes	Accuracy -99.95%	Execution Time can be increased
[25]	Original Banking Dataset from an institution in Nigeria -Records: 1356243	-Data Cleaning by removing less transaction data and inactive status data -DBSCAN for Clustering	Combined DBSCAN Algorithm with Rule-Based Algorithm (DBSCAN-Rule base)	F-Measure for datasets -A=0.800 -B=0.667 -C=1.000 -D=0.696 -E=0.444 -F=0.500 -G=0.609	A functioning knowledge base system can be more useful. More evidence is needed for the efficacy of the multi-algorithm.
[26]	Credit Card Dataset. Genuine and Fraud Transaction	-Communal Detection -Spike Detection -Cross-Validation	HSVM	Average Best Performance 0.2	Updating data in a database is compulsory for scalable results.
[27]	Credit Card Dataset Taken from UCI Website. Features: 20 Tuples: 1000	-Pre-processing using Rough Set -Split Data 60:40 Higher Dependency Feature Selected	Hybrid Approach by using DT and Rough Set	Performance 84.25%	Besides, organizations' resources should be concentrated on more fraudulent transactions to reduce fraud levels.
[28]	Actual credit card dataset from a Malaysian institution	-PCA for Transformation -Under Sampling	DS+GBT DT+DS DT+GBT DT+NB NB+GBT NN+NB RF+GBT	Accuracy 100.000% 100.000% 100.000% 99.999% 99.999% 99.999% 99.999%	Online learning models for future work will be applied to the methods studied in this paper. Besides, other models of online learning will be studied.
[29]	-DB1(European Data) -DB2(Pyism Data) -DB3(10 Million Credit Card Transaction)	-SVM-Recursive Feature Elimination (RFE) for Feature Selection -SMOTE is for Over-Sampling -GridCV for Hyper-Parameter Optimization	ML Methods -C5.0 -LR -BBN -ANN -RF -SVM -NB - HM	Accuracy: HM = 99% C5.0=96% SVM=96% ANN=96% LR=96% NB=93% BBN=94% KNN=95%	Writers' plan is to investigate expanded model on an extremely complicated dataset with carried high level concept and skewed data by constructing an adaptive CCFD system.
[30]	Original CCFD dataset	-Feature selection by using GA and Chi Statistic	Hybrid Approach -C4.5 -AdaBoost	Our proposed method's precision and	Comparing this approach with other proposed approaches

		-Resampling by D-optimal Design -K-Mean Clustering		reliability were 96.59% and 67.52%, respectively. This indicates that to detect fraud transactions, our hybrid suggested approach has good efficiency.	and documenting the findings is recommended.
[31]	Credit card fraud detection dataset	-Outlier Detection -K-mean Clustering -Under-sampling	Hybrid approach by combining Supervised and unsupervised technique	Good accuracy obtained through using cluster-based GM-2 outlier scores as extra features from 54 days on a test dataset.	Future research with the clustering metric could shed more light on the importance of this method.

3.11 MISCELLANEOUS

An algorithm Lightgbm is proposed for detecting frauds. After that, a comparison is made with other methods like Logistic Regression, SVM, and Xgboost. The accuracy of Lightgbm is 0.982 as opposed to Logistic Regression 0.926, SVM 0.952, and Xgboost 0.971 but Lightgbm is performed very properly to others comparatively [32].

REDBSCAN algorithm is used to decrease the number of samples and it helps to remain the form of data. The comparison made with the SVM technique and AUC of SVDD is 0.9775 and SVM 0.9460. When SVDD is applied except REDBSCAN, it takes 194 seconds and when utilized with REDBSCAN, it takes 1.69 seconds which is an awful lot faster. REDCSCAN algorithm provides faster and preferred results [1].

The Reduction function can't dispose of the unproductive records so a two-stage function reduction and Random under-sampling are proposed, threshold sampling and instance hardness are also applied. Outlier deduction, under-sampling, characteristic reduction, and classification manner are used as the lookup methodology. Five classification algorithms are used NB, KNN, SVM, LR, and RF. Before making use of two-stage feature reduction and under-sampling accuracies of classification algorithms are KNN 0.99954, SVM 0.99949, LR 0.99933, NB 0.99933, and RF 0.99959. After making use of the proposed method, accuracies of classification algorithms are KNN 1, SVM 0.60131, LR 0.85621, 0.98693, and RF 0.98093. The overall performance of two-stage feature reduction seems not to be good [33].

Nine classifiers' performance is compared which consist of Pipelining and Ensemble Learning is the essential classifier and others are RF, Quadrant Discriminant Analysis, Ada Boost, MultiLayer Perception, K-Nearest Neighbour, Naïve Byes, and Logistic Regression. 70% of information from the dataset is used to train and 30% is used for testing the model. The accuracy of Pipelining is 99.9999% and Ensemble Learning accuracy is 99.99% which confirmed that these two-algorithms are carried out very well alternatively of other classifiers. KNN classifier carried out worst with an accuracy of 94.4%. Dataset is to be used to preserve balance in the ADASYN method [34].

Two unsupervised algorithms are proposed SIMPLEKMEANS and Principal Component Analysis to develop an anti-fraud application. PCA is a very dynamic device that is used for some calculations and tests the relationship between the transactions. It can easily apply to massive datasets. SKM is used for rapid and simpler recognition of prison and fraud transactions. Five bank facts are amassed to put in force this model. The model identifies which financial institution has fraud and felony transactions [35].

Four ML classifiers are used to scan that are RUSBoost, AdaBoost, Naïve Bayes, and RUSMRN. The comparison of these techniques is observed by the RUSMRN method presents the very best accuracy of 79.73% as antagonistic to different techniques. Other methods produce a common accuracy like AdaBoost 57.73%, Naïve Bayes 70.13%, and RUSBoost gave 57.73% accuracy. So, the RUSMRN classifier performs outclass in a period of sensitivity and accuracy [36].

For identification of fraud Improved Adaboost algorithm is proposed with some different measures like a danger control algorithm this will assist for built-in learning. This approach increases the accuracy of the identification of credit card fraud that reaches 96.50%. After getting the accuracy of the suggested method comparison is once made with other detection techniques like SVM, C4_5, and AutoEncoder and observed that enhancement in Adaboost algorithms provides proper results [37].

Data mining-based strategies are used to find out frauds on online transactions. Clustering, Research association, and classification are applied to data. Mix method is used for the classification of BDC. Then, the system signature swindler is utilized to gather all frauds and in the last gadget is applied on the web for fraud detection on the banking cart. Talking about performance minimization of FN skill that cases of fraud are no longer passing, and maximization of relevance shows that frauds are now not producing alerts after the detection of real frauds for doing that finished purpose to maximization of the coverture [38].

A transaction blockading rule is proposed that confirms the transaction security. The rank alert is additionally generated by way of a studying algorithm that draws on scores certain to each alert. In the end, the machine is instructed and upgraded with the dataset which helps in the well-timed investigation and additionally blocks the credit card if variants are found on pattern [39].

Raw datasets and new datasets are created with the assist of transformation and discount of data. Supervised based classification is applied by making use of Network Bayesian classifiers like TAN, K2, NB, J48, and logistic. All classifiers firstly run-on raw facts these classifiers do not provide good results. But after the transformation when these classifiers run on data, they all execute above with 95% accuracy. So, consequences in pre-processed records are exceptional [40].

SOM method is used to revalidate the clusters for this association rules and it is utilized for every cluster. Here is trouble that association rules practice on categorical facts so, it needs to convert numeric information into categorical data. Association rules are simply if-then statementing that decides the connection between attributes. That gives thinking about which item is regarded in data. This can help us to display the transactions and fraud detection with the aid of the usage of unsupervised learning [41].

The Hidden Markov Model is used to identify a transaction that is legal or illegal and does not produce a low false alarm. Cardholder's spending addiction is beneficial in HMM to discover the frauds except for the use of fraud signatures. HMM preserve the database updated with patterns and transaction behaviours and it looks if any uncertain transaction goes through which is diverse from the past conduct of that purchaser and it generates the alarm and blocked the client on time. This model is making certain that the detection processing is very convenient and it helps to cast off the complexity [42].

Four classification strategies are used due to imbalanced records inexact alarm rate, equilibrium classification rate, fraud detection rate, and Matthews's correlation coefficient. The output of the algorithm listed is also compared with other classifiers such as SV, KNN, NB, and RF. The proposed algorithm gives the highest detection rate and a decrease in the false alarm rate [43].

Pre-processing is finished with the aid of checking missing values, sampling, and fact splitting. After pre-processing some ML algorithms are proposed to check the AUC rate like KNN, MLP, DT, and Cost-Sensitive Ensemble methods. The AUC fee of these methods is 0.87, 0.95, 0.90, 0.99. Proposed Cost-Sensitive Ensemble technique performed properly as in contrast to others. The limitation of these findings is that the power of cost-sensitive classifiers in the handling of extraordinarily negatively unbalanced records is also underlined [44].

BiLSTM-MaxPooling-BiGRU-MaxPooling stands totally on Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU) model. Pre-processing is carried out by way of Random Under Sampling, Random Oversampling, and SMOTE. It has been additionally carried out six classifiers for computing ML: NB, Voting, Ada boosting, RF, DT, and LR. Set side by side the outcomes of ML methods and model, the effects show that mannequin undertakes good as experiments get a score of 91.37 percent [45].

Pre-processing is accomplished with the aid of Sampling, Outlier Detection, K-Mean Clustering, and Normalization. An unconventional technique that can discover outliers in massive datasets and is resilient to evolving traits is a key contribution. AUPRC on Proposed Algorithm: 0.2916 and AUROC on Proposed Algorithm: 0.9311. Dissimilar clustering algorithms and formalization of the suggested strategy can be examined in further studies [46].

To optimize deposit card fraud detection techniques, the study counselled a model for applying classification, regression, and feature selection for pre-processing. There are six key components of the mannequin, setting, meaning, method, compare, determine, and act. The decision-making mechanism that suggests the interaction between a range of elements and the float of expertise from one element to the subsequent is integral for these mannequin elements. The mannequin can be checked to determine the degree to which it lets in the techniques of detecting deposit card fraud to make a more informed choice that will enhance accuracy and effectiveness [47].

The effectiveness of the proposed solution is measured by way of the ability for assessment. Comparison with different lookup findings and ultra-modern algorithms for machine learning, together with RF, logistic regression, vector computer for linear support, KNN, DT, and NB. In phrases of AUC, accuracy, F1-score, and precision for pre-processing, the experimental outcomes show that the

suggested technique outperformed the different desktop getting to know algorithms and carried out the absolute best efficiency. The effects to reveal that different classifiers are most excellent to the proposed algorithm. The findings also highlight the importance and gain of implementing a fantastic approach for optimizing parameters to increase the proposed approach's predictive efficiency [48].

This paper introduces a fraud detection mannequin by using the K-Star ML algorithm, making the use of German Credit and Australian Credit datasets to take a look at the findings. With a resulting classification precision of one hundred percent, a very low FP rate (0.00), and a very high TP rate of 1.00, the algorithm is proved fantastically successful and efficient. Both experiments are carried out in the simulation putting of WEKA DM and ML [49].

The proposed answer relies on the instruction of the autoencoder for normal information reconstruction. By specifying a recreation error threshold and recognizing cases with the most suitable threshold as anomalies are identified. The overall performance of the algorithm is once capable of to observe, fraud transactions linking 64 percent at the threshold = 5, 79 percentage at the threshold = 3, and ninety-one percentage at the threshold = 0.7, fifty-seven percent in the unbalanced dataset is higher than overall performance in contrast to logistic regression [50].

3.12 LIMITATIONS

The above review has potential limitations. Acquiring samples of data is a hard task, some outlier detection algorithms not able to differentiate the outliers, some approaches are too costly, and imbalanced data is still a challenge.

TABLE 6. MISCELLANEOUS

Reference	Dataset	Pre-processing	Methods	Results	Limitations or F/W
[32]	-IEEE-CIS Dataset. -1 million Samples. -400 features variable for each sample	-Data Cleaning with binary classification -Feature Engineering with Correlation Coefficient -Nan Values Filling with -999	-LR -SVM -Xgboost -Lightgbm	Accuracy: -0.926 -0.952 -0.971 -0.982	NA
[1]	-Payism synthetic dataset. -1 million data. -1142 labeled fraud	-Normalization (0-1) -Sample Reduction by using DBSCAN and REDBSCAN	-SVM -SVDD	-AUC:0.9460 -AUC:0.9775	Hard to acquire samples. Need more tuning for SVDD
[33]	-ULB Credit Card Detection Dataset. -Input Features 30. -248807 Sales	-MCD Algorithm for Outlier Detection -IHT Algorithm for Sampling -Correlation-based and PCA for Feature Reduction	-KNN -SVM -LR -NB -RF	Accuracy: -1 -0.60131 -0.85621 -0.98693 -0.98039	-Accuracy Reduced. -Not able to differentiate the outliers
[34]	-ULB Machine Learning Group. -Transaction 248807	-PCA for Transformation -Data distribution with zero mean and unit variance -Data Sampling	-LR -NB -KNN -MultiLayer	Accuracy -98.2% -99.6% -94.4% -98.4%	NA

		using ADASYN Method	Perception -Ada Boost -QDA -RF -Ensemble Learning -Pipelining	-98.5% -97.3% -99.7% -99.99%. -99.9999%	
[35]	Forged and Randomized data is generated	-PCA and SKM for Transformation -K-mean Clustering	Unsupervised Algorithms: Principal Component Analysis and SIMPLEKMEANS	A good model is proposed. Legal and fraud transaction detected -Error Rate 2%	By repeating the 'k means' method many times with different initial clusters, this risk could be minimized at the cost of increasing the execution time.
[36]	Payment data of Credit Card holder's Dataset taken from the bank. Taiwan. 25000 Observation. 5529 are from UCI Machine learning database	Data Sampling using RUS	-RUSMRN Algorithm -RUSBoost -AdaBoost -NB	Accuracy -79.73% -77.8% -57.73% -70.13%	NA
[37]	-Commercial Bank Dataset --n Samples -I Attributes	-PCA to Determines Initial Weights -Lower-Sampling -SMOTE for Oversampling	-PCA AdaBoost -AutoEncoder -SVM -C4_5	F-Measure -0.9730 -0.9674 -0.9533 -0.7568	NA
[38]	Banking Database	-Segmentation	-Classification -Research Association	Maximization if coverture Couv=TP/F = TP/TP+FN	Hierarchal Classification method can generate better result but it's too costly
[39]	Banking Credit Card Dataset	-Handle Imbalanced Data	Learning Algorithm	Expected Results are good	Algorithm Enhancement needed to make it simple and user friendly
[40]	-Credit Card Raw Dummy Dataset -Credit Card Transformed Dataset	-10-fold cross-validation -PCA for Anomaly transaction detection	Bayesian Network Classifiers -K2 -TAN -Naïve Bayes -Logistics	Accuracy -K2: 95.8% -TAN: 99.7% -Naïve Bayes: 96.7% -Logistics: 100.0% -J48: 100.0%	Real-Time data should be used.

			-J48		
[41]	Credit Card Transaction	-SOM Technique for Clustering. Revalidate Clusters by Association Rules -Data Conversion numeric to categorical	Unsupervised Learning Clustering	It allows transaction management rules to be generated automatically in a learning process and allows them to be continually improved in an environment of constantly changing information in an automated system.	The need for a safe testing framework is growing with the increasing ratio of cases of bank fraud and cybercrime. And to this dilemma, this is a direct solution. It may refer not only to a customer (debit-ant) but also to the duplex verification of the seller (credit-ant).
[42]	Cardholder's Database	NA	Hidden Markov Model	Well Secured Model. High Accuracy. Low false Alarms.	Credit Card Blocking chances increased due to mistakenly entering the wrong data.
[43]	Credit Card Holder Dataset. Transaction: 1098101. Customers: 5601	-Data Refining by removing the transactions who have only 1 transaction in the dataset -Segregate Transaction into legal and fraud	LINGO Algorithm	False Alarm rates are decreased. Quality Improved	The future work proposed using LINGO3G as an improved LINGO algorithm that has many enhancements, such as achieving very fast clustering over large records of snippets.
[44]	Credit Card Dataset from Kaggle. Observation 284,807	-PCA for Transformation - 10-fold Cross-Validation -Feature importance is assessed via Person Correlation	-KNN -MLP -DT -Cost-Sensitive Ensemble	AUC -0.87 -0.95 -0.90 -0.99	NA
[45]	Dataset for IEEE-CIS Fraud Detection provided by Kaggle	-Random Under Sampling. -Random Over Sampling. -Synthetic Minority Oversampling Technique (SMOTE).	-BiLSTM -BiGRU	AUC with RUS: 90% AUC with ROS: 91.37% AUC with SMOTE: 90%	Different more Sampling techniques can be applied for better results.
[46]	Original world credit card fraud detection dataset from Kaggle	-PCA for Transformation -Outlier Detection	Consistency Estimation Method	AUPRC on Proposed Algorithm: 0.2916. AUROC on Proposed Algorithm: 0.9311	Different clustering algorithms and formalization of the proposed approach can be part of

		-K-Mean Clustering -Normalization b/w 0 to1			future work.
[47]	Transaction dataset	Feature selection according to User previous transactions of same nature	Non-Linear Regression Method	In making a more educated decision, the results of this study assess the extent to which it supports the methods of detecting credit card fraud.	The model can be evaluated to assess the degree to which it facilitates methods of detecting credit card fraud to make a well organized decision that will enhance precision and performance.
[48]	-European Dataset -UCSD-FICO Dataset	-Feature selection by Using PCA and IG -Cross-Validation -Resampling	OLightGBM	Accuracy 98.40%.	NA
[49]	-German credit dataset -Australian credit dataset Obtained UCI	-10-Fold Cross-Validation test -Mathiew Correlation Coefficient Matric for Features	-SVM -Naïve Bayes -LWL -HMM -MLP -K-Star	Accuracy SVM=78.4% Naïve Bayes=77.2% HMM=70% LWL=70% MLP=99.3% K-Star=100%	NA
[50]	Credit card fraud detection dataset Normal transaction=284,315 Fraud transaction= 492	-Data Splitting into 80% and 20% -F1-Score	-LR -Autoencoder	-The accuracy of the Linear regression algorithms is Balance data=97.23 Unbalanced data=99.91 -Accuracy of Autoencoder (Thr=5) =98.70 (Thr=3) =97.70 (Thr=1) =90.02 (Thr=0.7) =80.00	To insure by giving the clear concept of applying fraudulent work and its classification, fraudsters can escape to mislead by comparing their previous knowledge. Saudi companies also follow to apply this methodology.

4 DISCUSSION

As discussed in this comparative study, some issues have been addressed whenever a transaction occurs while others are remained to be located that shows the direction of the future for highlighting and focusing of attention in the area of CCFD. After exploring and examining the limitations some extensions can be useful to improve the accuracy and other measures to detects frauds in credit cards. Some effective methods of data pre-processing may be useful, such as sampling methods, clustering algorithms, and some advanced methods of selection of features. For the Selection of Features, the majority voting method could be used, the Genetic Algorithm could be used to decrease the dataset, the optimized parameter selection kernel function can be used, and DL methods could be used for pre-processing.

To get better results, some algorithms need to be changed, such as RF, SVDD, learning algorithm, and LINGO to LINGGO3G. With wide and small datasets, DL and DL Hybrid models for fraud detection can be used to boost fraud detection. In time ahead, writers are ready to develop online fraud detection system instead of offline.

5 CONCLUSION

This review offers an overview of different methods of data mining for the CCFD. With time, frauds in credit cards have risen. This study presents to find out more ways effectively in the field of fraud detection. One of these methods, or a combination of them, can be used to detect fraud. Most researchers face some challenges, such as unavailability of real datasets, unbalanced datasets, and size of datasets, which is a difficult subject of research to detect credit card fraud.

REFERENCES

[1] M. Khedmati, M. Erfani, and M. GhasemiGol, "Applying support vector data description for fraud detection," *arXiv*, pp. 1–6, 2020.

[2] K. R. Seeja and M. Zareapoor, "FraudMiner: A novel credit card fraud detection model based on frequent itemset mining," *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/252797.

[3] L. S. V S S and S. Deepthi Kavila, "Machine Learning For Credit Card Fraud Detection System," *Int. J. Appl. Eng. Res.*, vol. 13, no. 24, pp. 16819–16824, 2018.

[4] Y. Lucas *et al.*, "Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs," *Futur. Gener. Comput. Syst.*, vol. 102, pp. 393–402, 2020, doi: 10.1016/j.future.2019.08.029.

[5] N. Carneiro, G. Figueira, and M. Costa, "A data mining based system for credit-card fraud detection in e-tail," *Decis. Support Syst.*, vol. 95, pp. 91–101, 2017, doi: 10.1016/j.dss.2017.01.002.

[6] V. N. Dornadula and S. Geetha, "Credit Card Fraud Detection using Machine Learning Algorithms," *Procedia Comput. Sci.*, vol. 165, no. 20, pp. 631–641, 2019, doi: 10.1016/j.procs.2020.01.057.

[7] Y. Lucas *et al.*, "Multiple perspectives HMM-based feature engineering for credit card fraud detection," *Proc. ACM Symp. Appl. Comput.*, vol. Part F147772, pp. 1359–1361, 2019, doi: 10.1145/3297280.3297586.

[8] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random forest for credit card fraud detection," *ICNSC 2018 - 15th IEEE Int. Conf. Networking, Sens. Control*, pp. 1–6, 2018, doi: 10.1109/ICNSC.2018.8361343.

[9] S. Xuan, G. Liu, and Z. Li, *Refined weighted random forest and its application to credit card fraud detection*, vol. 11280 LNCS. Springer International Publishing, 2018.

[10] F. F. Noghani and M.-H. Moattar, "Ensemble Classification and Extended Feature Selection for Credit Card Fraud Detection," *J. AI Data Min.*, vol. 5, no. 2, pp. 235–243, 2017.

[11] J. R. D. Kho and L. A. Vea, "Credit card fraud detection based on transaction behavior," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2017-Decem, no. 2, pp. 1880–1884, 2017, doi: 10.1109/TENCON.2017.8228165.

[12] M. Hammed and J. Soyemi, "An implementation of decision tree algorithm augmented with regression analysis for fraud detection in credit card," *Int. J. Comput. Sci. Inf. Secur.*, vol. 18, no. 2, pp. 79–88, 2020, [Online]. Available: https://sites.google.com/site/ijcsis/.

[13] A. Husejinović, "Credit card fraud detection using naive Bayesian and c4.5 decision tree classifiers," *Period. Eng. Nat. Sci.*, vol. 8, no. 1, pp. 1–5, 2020, doi: 10.21533/pen.v.

[14] Geetika and D. G. Gupta, "Original Research Paper Computer Engineering MACHINE LEARNING APPROACH FOR CREDIT CARD FRAUD Geetika," no. 9, pp. 2017–2019, 2019.

[15] M. Phil, C. Science, V. College, N. Geetha, and T. Kavipriya, "an Identification and Detection of Fraudulence in Credit Card Fraud Transaction System Using Data Mining Techniques," pp. 1238–1241, 2018.

[16] D. V. and D. R., "Behavior Based Credit Card Fraud Detection Using Support Vector Machines," *ICTACT J. Soft Comput.*, vol. 02, no. 04, pp. 391–397, 2012, doi: 10.21917/ijsc.2012.0061.

[17] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, "Real-time credit card fraud detection using machine learning," *Proc. 9th Int. Conf. Cloud Comput. Data Sci. Eng. Conflu. 2019*, pp. 488–493, 2019, doi: 10.1109/CONFLUENCE.2019.8776942.

[18] W. Xu and Y. Liu, "An optimized SVM model for detection of fraudulent online credit card transactions," *Proc. - 2012 Int. Conf. Manag. e-Commerce e-Government, ICMeCG 2012*, pp. 14–17, 2012, doi: 10.1109/ICMeCG.2012.39.

[19] N. B. Muppalaneni, M. Ma, and S. Gurumoorthy, *Soft Computing and Medical Bioinformatics*. Springer Singapore, 2019.

[20] J. Akhilomen, "Data Mining Application for Cyber Credit-Card Fraud Detection System What Is Cyber Credit-Card Fraud or No Card Present," *13th Ind. Conf. ICDM 2013, Lect. Notes Comput. Sci.*, pp. 218–228, 2013.

[21] D. Cheng, S. Xiang, C. Shang, Y. Zhang, F. Yang, and L. Zhang, "Spatio-Temporal Attention-Based Neural Network for Credit Card Fraud Detection," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 01, pp. 362–369, 2020, doi: 10.1609/aaai.v34i01.5371.

[22] F. Z. El Hlouli, J. Riffi, M. A. Mahraz, A. El Yahyaouy, and H. Tairi, "Credit Card Fraud Detection Based on Multilayer Perceptron and Extreme Learning Machine Architectures," *2020 Int. Conf. Intell. Syst. Comput. Vision, ISCV 2020*, pp. 3–7, 2020, doi: 10.1109/ISCV49265.2020.9204185.

[23] X. Yu, X. Li, Y. Dong, and R. Zheng, "A Deep Neural Network Algorithm for Detecting Credit Card Fraud," *Proc. - 2020 Int. Conf. Big Data, Artif. Intell. Internet Things Eng. ICBAIE 2020*, pp. 181–183, 2020, doi: 10.1109/ICBAIE49996.2020.00045.

[24] D. Kaur, "Machine Learning Approach for Credit Card Fraud Detection (KNN & Naïve Bayes)," *Mach. Learn. Approach Credit Card Fraud Detect. (KNN Naïve Bayes)(March 30, 2020)*, 2020.

[25] A. Oluwafolake and O. A. Solomon, "A multi-algorithm data mining classification approach for bank fraudulent transactions," *African J. Math. Comput. Sci. Res.*, vol. 10, no. 1, pp. 5–13, 2017, doi: 10.5897/ajmcsr2017.0686.

[26] V. Mareeswari and G. Gunasekaran, "Prevention of credit card fraud detection based on HSVM," *2016 Int. Conf. Inf. Commun. Embed. Syst. ICICES 2016*, no. Icices, pp. 1–4, 2016, doi: 10.1109/ICICES.2016.7518889.

[27] R. Jain, B. Gour, and S. Dubey, "A Hybrid Approach for Credit Card Fraud Detection using Rough Set and Decision Tree Technique," *Int. J. Comput. Appl.*, vol. 139, no. 10, pp. 1–6, 2016, doi: 10.5120/ijca2016909325.

[28] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018, doi: 10.1109/ACCESS.2018.2806420.

[29] N. Rtayli and N. Enneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization," *J. Inf. Secur. Appl.*, vol. 55, no. September, p. 102596, 2020, doi: 10.1016/j.jisa.2020.102596.

[30] S. Beigi and M.-R. Amin-Naseri, "Credit Card Fraud Detection using Data mining and Statistical Methods," vol. 8, no. 2, pp. 149–160, 2019, doi: 10.22044/JADM.2019.7506.1894.

[31] F. Carcillo, Y. A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Inf. Sci. (Ny).*, no. xxxx, 2019, doi: 10.1016/j.ins.2019.05.042.

[32] D. Ge, J. Gu, S. Chang, and J. H. Cai, "Credit card fraud detection using lightgbm model," *Proc. - 2020 Int. Conf. E-Commerce Internet Technol. ECIT 2020*, pp. 232–236, 2020, doi: 10.1109/ECIT50008.2020.00060.

[33] D. Trisanto, N. Rismawati, M. Mulya, and F. Kurniadi, "Effectiveness Undersampling Method and Feature Reduction in Credit Card Fraud Detection," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 2, pp. 173–181, 2020, doi: 10.22266/ijies2020.0430.17.

[34] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, "Credit Card Fraud Detection using Pipeling and Ensemble Learning," *Procedia Comput. Sci.*, vol. 173, no. 2019, pp. 104–112, 2020, doi: 10.1016/j.procs.2020.06.014.

[35] M. R. Lepoivre, C. O. Avanzini, G. Bignon, L. Legendre, and A. K. Piwele, "Credit Card Fraud Detection with Unsupervised Algorithms," *J. Adv. Inf. Technol.*, vol. 7, no. 1, pp. 34–38, 2016, doi: 10.12720/jait.7.1.34-38.

[36] A. Charleonnan, "Credit card fraud detection using RUS and MRN algorithms," *2016 Manag. Innov. Technol. Int. Conf. MITiCON 2016*, pp. MIT73–MIT76, 2017, doi: 10.1109/MITICON.2016.8025244.

[37] H. Zhou, L. Wei, G. Chen, P. Lin, and Y. Lin, "Credit card fraud identification based on principal component analysis and improved adaboost algorithm," *Proc. - 2019 Int. Conf. Intell. Comput. Autom. Syst. ICICAS 2019*, pp. 507–510, 2019, doi: 10.1109/ICICAS48597.2019.00111.

[38] H. El-Kaime, M. Hanoune, and A. Eddaoui, *The Data Mining: A Solution for Credit Card Fraud Detection in Banking*, vol. 756. Springer International Publishing, 2019.

[39] N. Kalaiselvi, S. Rajalakshmi, J. Padmavathi, and J. B. Karthiga, "Credit Card Fraud Detection Using Learning to Rank Approach," *7th IEEE Int. Conf. Comput. Power, Energy, Inf. Commun. ICCPEIC 2018*, pp. 191–196, 2018, doi: 10.1109/ICCPEIC.2018.8525171.

[40] O. S. Yee, S. Sagadevan, and N. H. A. H. Malim, "Credit card fraud detection using machine learning as data mining technique," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 1–4, pp. 23–27, 2018.

[41] P. Vikrant Agaskar, M. Babariya, S. Chandran, and N. Giri, "Unsupervised Learning for Credit Card fraud detection," *Int. Res. J. Eng. Technol.*, vol. 4, no. 3, pp. 2395–56, 2017, [Online]. Available: https://www.irjet.net/archives/V4/i3/IRJET-V4I3608.pdf.

[42] P. Yadav, P. Wangade, M. Thakur, M. Fakih, and G. Hegde, "Proposed Distributed Data Mining in Credit Card Fraud Detection," *Int. Res. J. Eng. Technol.*, pp. 460–463, 2016.

[43] M. Hegazy, A. Madian, and M. Ragaie, "Enhanced Fraud Miner: Credit Card Fraud Detection using Clustering Data Mining Techniques," *Egypt. Comput. Sci. J.*, vol. 40, no. 03, pp. 1110–2586, 2016.

[44] T. A. Olowookere and O. S. Adewale, "A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach," *Sci. African*, vol. 8, p. e00464, 2020, doi: 10.1016/j.sciaf.2020.e00464.

[45] H. Najadat, O. Altiti, A. A. Aqouleh, and M. Younes, "Credit Card Fraud Detection Based on Machine and Deep Learning," *2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020*, no. Section IX, pp. 204–208, 2020, doi: 10.1109/ICICS49469.2020.239524.

[46] U. Porwal and S. Mukund, "Credit card fraud detection in E-commerce," *Proc. - 2019 18th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. IEEE Int. Conf. Big Data Sci. Eng. Trust. 2019*, pp. 280–287, 2019, doi: 10.1109/TrustCom/BigDataSE.2019.00045.

[47] B. K. Nkomo and T. Breetzke, "A conceptual model for the use of artificial intelligence for credit card fraud detection in banks," *2020 Conf. Inf. Commun. Technol. Soc. ICTAS 2020 - Proc.*, 2020, doi: 10.1109/ICTAS47918.2020.233980.

[48] A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light

Gradient Boosting Machine," *IEEE Access*, vol. 8, pp. 25579–25587, 2020, doi: 10.1109/ACCESS.2020.2971354.

[49] S. Africa, "Credit Card Fraud Detection using k-star Machine Learning Algorithm DADA Emmanuel Gbenga MAPAYI Temitope , OLAIFA Olowasogo Moses ," pp. 2–15, 2019.

[50] M. A. Al-Shabi, "Credit Card Fraud Detection Using Autoencoder Model in Unbalanced Datasets," *J. Adv. Math. Comput. Sci.*, vol. 33, no. 5, pp. 1–16, 2019, doi: 10.9734/jamcs/2019/v33i530192.