INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, e-ISSN 1841-9844, 14(6), 670-690, December 2019.

An Empirical Study of AML Approach for Credit Card Fraud Detection–Financial Transactions

A. Singh, A. Jain

Ajeet Singh
University School of Information, Communication & Technology
Guru Gobind Singh Indraprastha University, Delhi, India
ajeet.usit.025164@ipu.ac.in

Anurag Jain*
University School of Information, Communication & Technology
Guru Gobind Singh Indraprastha University, Delhi, India
*Corresponding author: anurag@ipu.ac.in

Abstract: Credit card fraud is one of the flip sides of the digital world, where transactions are made without the knowledge of the genuine user. Based on the study of various papers published between 1994 and 2018 on credit card fraud, the following objectives are achieved: the various types of credit card frauds has identified and to detect automatically these frauds, an adaptive machine learning techniques (AMLTs) has studied and also their pros and cons has summarized. The various dataset are used in the literature has studied and categorized into the real and synthesized datasets.The performance matrices and evaluation criteria have summarized which has used to evaluate the fraud detection system.This study has also covered the deep analysis and comparison of the performance (i.e sensitivity, specificity, and accuracy) of existing machine learning techniques in the credit card fraud detection area.The findings of this study clearly show that supervised learning, card-not-present fraud, skimming fraud, and website cloning method has been used more frequently.This Study helps to new researchers by discussing the limitation of existing fraud detection techniques and providing helpful directions of research in the credit card fraud detection field.

Keywords: Credit card fraud, cashless transaction, data mining technique, fraud detection.

1 Introduction

Nowadays, the use of credit cards has significantly increased on both online and offline purchases because of the fast growth of the e-commerce and online banking system. When someone uses other persons credit card for personal benefit without the knowledge of the owner of the credit card is known as credit card fraud. The Association of Certified Fraud Examiners defines a fraud as *"the use of one's occupation for personal enrichment through the deliberate misuse or application of the employing organization's resources or assets"* [63]. Individuals and government suffer large financial losses across the world every year due to the lack of sophisticated fraud detection system.

In recent years, a million dollars losses due to the card frauds and many countries have affected. According to the fraud facts report-2017, the UK payment credit cards losses has been increased 9% in 2016 from Euro 567.5 million in 2015 to Euro 618.0 million. This point out that payment card fraud is increasing yearly. The total spending on cards has reached Euro 904 billion in 2016, with 19.1 billion transactions happened during the 2015 [66].

In India, the State Bank of India (SBI) has blocked 0.6 million (6 lakh) debit cards due to a cyber attack (malware-related breach) to a YES Bank ATM network on October 19, 2016. It was one of the latest and largest security breaches where in Indian banking history, millions of the debit cards were compromised. The biggest financial data breaches reached almost 3.2 million (32lakhs) debit cards across 19 private banks (i.e. HDFC Bank, etc.).

According to the National Payments Corporation of India report, 90 ATMs suffered losses and 641 customers lost 13 million (1.3 Crores) Indian Rupee due to fraudulent transactions.

According to the Nilson Report, the card fraud losses reached $21.84 billion in 2015, $24.71 billion in 2016 and $27.69 billion in 2017. This is also informed that the actual amount of losses could increase in 2020 [64].

According to the annual report of Internet crime complain 2018 (IC3), online crime has increased from 2008 to 2017 due to many types of frauds such as credit card fraud, identity fraud, etc.

The total loss has increased from 239.1 million in 2008 to $1,418.7 million in 2017 due to frauds. Table1 presents the thousands of complaints between 2017 and 2008 received by the IC3 and happened financial loss in millions of USA dollar [65].

Table 1: Financial loss due to cyber crime

Year	**Complaint**	**Money Loss**	**Year**	**Complaint**	**Money Loss**
2017	301,580	$1,418.7	2012	289,874	$ 581.4
2016	298,728	$1,450.7	2011	314,246	$ 485.25
2015	288,012	$1,070.7	2010	303809	$121.71
2014	269,422	$ 800.5	2009	336,655	$559.7
2013	262,813	$ 781.8	2008	275,284	$265.0

It is clearly observed that financial loss has been increased due to card frauds while the number of complaints in some year has also decreased. These enormous numbers of losses define the importance of fraud prevention and detection system [51]. This study focuses on the banking domain in particular for credit card frauds. The common credit card frauds are application fraud, counterfeiting, identity theft fraud, phishing, and skimming fraud.

Frauds have to be explored and identified as quickly as possible in order to stop fraudulent activities [10, 34]. To address these frauds, various fraud prevention and detection mechanisms are deployed to detect and prevent credit card fraud. The purpose of credit card fraud prevention mechanism is to prevent and stop a fraudulent transaction before it happens in the initial phase, and also prevent the occurrence of different cyber attacks on your computer system, networks and your data by using numbers of prevention methods namely communal detection spike detection, PIN, Internet Security System, firewall and cryptography algorithms [1, 25].

Fraud detection is second stage mechanism to detect a fraud by applying various data mining, machine learning and bio-inspired techniques namely Decision Tree, Artificial Neural Network, Artificial Immune Systems (AIS), K Nearest Neighbor(KNN), Support Vector Machine(SVM), Genetic Algorithm (GA) and Hidden Markov Model (HMM) [12, 46, 55, 60]. These techniques are usually suitable for both types of transactions as normal and fraudulent in order to learn fraud patterns and customer patterns. The goal of the credit card fraud detection system is to maximize true positive and minimize false positive predictions of legitimate transactions.

The main contribution of this research is to identify the frequently occurred credit card frauds and methods committed to obtain credit card information illegally. A systematic review of the existing machine learning techniques (MLTs) for credit card fraud detection are also discussed, and their advantages and limitations are reported in this paper. Based on the outcome of this paper, the research gap is identified among the existing fraud detection techniques. Various credit card fraud datasets are also compared on the basis of their features and used these datasets to implementation various existing machine learning techniques.

The research process of this paper is explained in figure 1 based on adaptive existing machine learning techniques to credit card fraud detection.

The rest of the paper is designed as follows. Section 2 explores the different category of credit card frauds and methods. Section 3 discusses credit card fraud detection issues and challenges and Section 4 presents the various credit card fraud detection techniques. The various dataset, primary and derived attributes are summarized in Section 5. Section 6 presents performance matrices and evaluation criteria to evaluate the fraud detection system. Research conclusion and discussion are presented in Section 7.

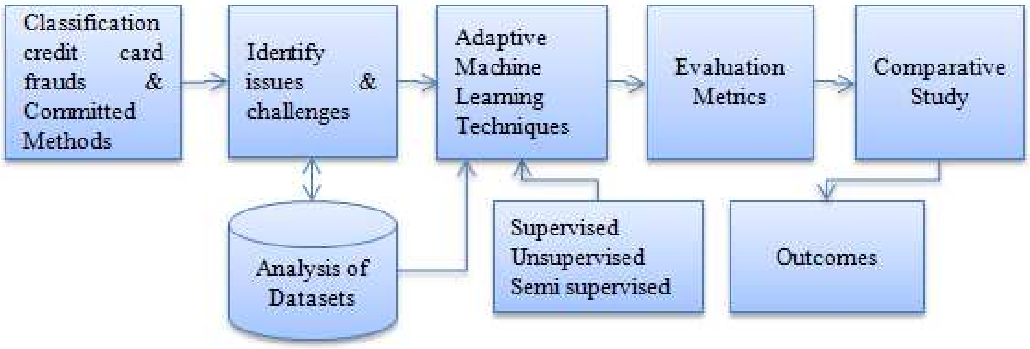
Classification credit card frauds & Committed Methods
Identify issues & challenges
Adaptive Machine Learning Techniques
Evaluation Metrics
Comparative Study
Analysis of Datasets
Supervised
Unsupervised
Semi supervised
Outcomes

Figure 1: Flow diagram of the machine learning-based credit card fraud detection system

2 Classification of credit card frauds

Criminals use a variety of methods and tools for obtaining card information needed for online transactions and gets access to the cardholders account.

The various credit card frauds are observed based on literature. The common types of credit card frauds include application fraud, transaction fraud, bankruptcy fraud ([16]), and counterfeit fraud ([12, 50]) are summarized in table 2.

3 Credit card fraud detection issues and challenges

Researchers are facing several issues and challenges to detect fraud on time efficient manner.Credit card issuing banks required efficient and well-organized fraud detection system to run their business in a healthy way.The numbers of issues and challenges are summarized in table 3 based on literature.

4 Fraud detection techniques based on Machine Learning

The credit card fraud detection techniques (CCFDTs) are grouped into two general categories such as fraud analysis (misuse detection), and user behaviors analysis (anomaly detection). The CCFDTs are also further categorized as supervised learning which uses labeled training

Table 2: Various credit card frauds

Fraud Name	**Definition**	**Method**	**Location of Occurs**
Application [1, 12, 50]	If a fraudsters fill the application form with fake information,or maybe even real, but stolen identity information.	Steal identity namely bank statements, telephone bill, and electricity bill, etc.	Financial organization.
Transaction [1, 12, 47]	Used stolen card details: When someone obtains others card details to make a fraudulent transaction without legitimate cardholder knowledge	Keylogger, Sniffers, Site cloning, Physical stolen card.	Anywhere.
Account Takeover	When fraudsters pose as a genuine cardholder to gain access and control of an account then withdraw funds, makes illegal transactions, and may change account details.	Brute force botnet attacks, phishing,& malware.	Anywhere.
Counterfeit [12, 50]	It is the practice of illegally copying a legitimate credit card details	Fake websites, Site cloning, ATM skimming device.	ATM/ Place where CC uses.
Card-Not-Present [1, 12, 50]	When fraudsters obtain card information for personal use	Internet, Hacking, Email phishing campaigns.	Anyplace.
Skimming	Keeps the credit card into a skimming machine to make a copy and save the important card information.	Skimming Machine	Shopping mall, Restaurants, etc.
Shoulder Surfing	When fraudsters looking over the cardholders shoulder to steal credit card details while cardholder enter card details an electronic device and a Web.	Digital devices like camera	ATM, POS.

data, and unsupervised learning uses unlabeled training data, and a hybrid technique called semi-supervised learning.

4.1 Supervised learning method

Supervised learning is the machine learning technique. The supervised learning and classification technique are used for fraud analysis (misuses detection) at the transaction level and transactions categorized as a fraud or normal transaction based on the historical data. The supervised learning techniques are ruled induction, decision trees, case-based reasoning, support vector machines (SVMs), neural networks techniques, linear regression, logistic regression, naive Bayes, linear discriminate analysis,and k-nearest neighbor algorithm which used to credit fraud detection.

Table 3: Various issue and challenge for credit card fraud detection

Issues	Description
No standard credit card dataset [25, 60]	It is the biggest issues in fraud detection domain. There is no standard, real, and benchmark dataset available to evaluate proposed fraud detection methods. In most of the cases, researchers have been used their own dataset (Synthetic dataset) for doing research .
Skewed class distribution [1,46,50, 60]	The skewed distribution (or imbalanced class) is one of the most serious problems. Very small percentages of all card transitions are fraudulent as compare to normal transition. In a supervised learning technique, imbalance problem occurs when the minority class percentage is not very high.
Cost-sensitive classification problem [45, 50]	Credit card Transactions are misclassification (fraudulent transaction as a legitimate transaction and legitimate transaction as a fraudulent transaction) due to this financial impact ranging from a few to thousands of money.
Presently no suitable evaluation criterion [50, 60]	Standard evaluation criterion is not available for assessing and comparing the results of a fraud detection system.The accuracy is not suitable metrics CCFD because data set is imbalanced.
Lack of proper algorithm [50]	A powerful algorithm is required to detecting a new type of normal and fraudulent pattern.The Data mining and Machine Learning algorithm has its own advantages and disadvantages (table4).
Fraudsters behaviour dynamic [60]	Fraudsters are changed their behaviour from time to time for getting card details and bypass detection system or modify fraud styles.
Cardholder Behavior(Concept Drift) [1].	The cardholder is always changing their behaviour may be specific situation/ occasions (e.g New Year), the buying power of users will be enhanced. If the CCFD system does not consider this as normal changes, will be considered as fraud behaviour
Pattern Recognition Algorithm [1]	Pattern recognition algorithm is used to recognize fraudster pattern and customer pattern to minimize fraud cases.

4.2 Unsupervised learning method

The user behavior analysis can be performed by using unsupervised learning and clustering technique that identifies a transaction based on the account behavior of users. In unsupervised learning, all transactions are unlabeled and the algorithms learn to the inherent structure from the input transactions. Unsupervised learning is a more powerful technique to identify fraud and non-fraud transaction. The most common unsupervised techniques are k-means, Self-organizing Map (SOM) technique, and Hidden Markov Model (HMM) technique.

4.3 Semi-supervised learning method

The semi-supervised learning method falls between supervised and unsupervised learning method. Semi-supervised learning is a machine learning task that makes use a very small number of labeled data (previously known) and a large number of unlabeled data (unknown data) to detect a card fraud. Semi-supervised learning could decrease the effort needed by supervisors to classify training data. The supervised, unsupervised and semi supervised techniques are introduced following which was used to detect credit card fraud (CCF).

4.4 Artificial Neural Networks

ANNs constitute a set of interconnected nodes designed to imitate the functioning of the human brain (Maes et al. [31]). A neural network based fraud detection system has trained with the earlier data of cardholder spending behavior and the tested software on synthetically generated data (Guo and Li [21]). ANNs has been configured by supervised, unsupervised, and semi-supervised learning methods for classification or clustering of transaction group. Chen et al. [14] employs SVM and ANN techniques to investigate the time-varying fraud problem. The performance of the ANN is compared with SVM, outcomes show that ANN has the highest training accuracy. ANN two techniques namely Back Propagation Neural Network (BPNN) and Self-organizing Maps Neural Network (SOMNN) are mostly used for credit card fraud detection.

Back Propagation Neural Network

BPNN is a supervised learning technique, a generalization of the delta rule, and suitable for the feed-forward network, that has no feedback. The feed forward network contains three layers namely input, hidden, and output layers to credit card frauds detection. Chen et al. [14] deployed SVM and BPNN (ANN) techniques to investigate the time-varying fraud problem. The performance of the BPNN (78%) is compared with SVM (67%); outcomes show that ANN has the highest training accuracy.Another research (Ghosh et al. [20]) deployed a multilayer feed forward neural network model to fraud detection from Mellon Bank and reduction total fraud losses from 20% to 40%.

Self-organizing map

Self-organizing Map (SOM) [27] is a neural network model based on unsupervised learning method for the analysis and visualization of high-dimensional data. SOM neural network used to detect credit card fraud based on customer behaviour. SOM consist of two layer namely input mapping layer. The input layer forward the incoming transactions to the mapping layer for performing the clustering technique. The mapping layer is to map all transactions with cardholder's behavior for detecting hidden patterns. Finally, the mapping layer produces results in the form of fraudulent and genuine transaction.

According to Quah and Sriganes [42], the SOM used to decipher,filter and analyze customer behavior for the detection of credit card fraud in the banking sector.

Olszewski [39] has suggested a fraud detection method based on the user accounts visualization. The proposed SOM visualization method was applied in three different areas such as telecommunication, computer intrusion, and credit card fraud detection. This method is more suitable for the credit card as compared to the other areas. The CC dataset contains 10,000 accounts details from 1 January 2005 to 1 March 2005 and identified 100 instances as fraudulent.

The SOM visualization method was determined based on the produced values of the Precision (1.0000 vs 0.8257, 0.7200, 0.5696) Recall (1.0000 vs0.9000, 0.9000, 0.9000) Accuracy (1.0000 vs 0.8550, 0.7750, 0.6100), F1 score (1.0000 vs 0.8612, 0.8000, 0.6977) and AUROC (1.0000 vs. 0.9415, 0.9285, and 0.8265) for a credit card. This is because of less unbalance dataset and small size of dataset. There may be chance of the accuracy and sensitivity will get dropped if dataset become large and more unbalance.

4.5 Bayesian network

Bayesian Network is a probabilistic graphical model denoting a set of random variables and their conditional dependencies through a directed acyclic graph introduced Cooper and Herskovits in 1992 (Panigrahi et al. [41]).

Bayesian Network used the concept of Bayes theorem to determine the probability of a given hypothesis to be true.

$$P(Fraudulent/Evidence) = \frac{(P(Evidence/Fraudulent) * P(Fraudulent))}{P(Evidence)} \tag{1}$$

Where P(Fraudulent/Evidence) is the posterior probability condition on Hypothesis. Moreover,P(Fraudulent) and P(Evidence) is the prior probability of Hypothesis.P(Evidence /Fraudulent) is called the likely hood. The Fraud probability represented by P(Fraudulent/Evidence) gives the observed user behavior. Bayesian network(BN) has suggested for the purpose of credit card fraud detection based on user behavior.Bayesian belief networks and artificial neural networks(ANN) techniques have used credit card fraud detection by (Maes et al. [31]). Result presented that Bayesian network has superior fraud detection capability than ANN.

According to Kirkos et al. [26], the Bayesian Belief Network model achieved the best performance to correctly classify 90.3% of the validation sample in a 10-fold cross validation procedure. The accuracy rates of the Neural Network and Decision Tree model were 80% and 73.6%, respectively.

4.6 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a supervised learning technique [53]. The Euclidean distance method is used to calculate the distance between two transactions (input data transaction and current transaction) for every data transaction in the dataset and distances are arranged in increasing order. The k items have the lowest distance to the input data transaction point are selected. KNN technique classifies any new incoming transaction by calculating the nearest point if the nearest neighbor is a fraud than transaction considers as fraud and if the nearest neighbor is not fraud than transaction consider as legal.

The Euclidean distance (D_{ij}) between two input vectors (X_i, X_j) is given by $\sum$

$$D_{ij} = \sqrt{\sum_{k=1}^{n} (X_{ik} - Y_{jk})^2}, (k = 1, 2, 3.....n.) \tag{2}$$

A simple matching coefficient is frequently used for categorical attributes. In this method, both legitimate and fraudulent transactions are to be fed in order to train the data sets. The K-Nearest Neighbor technique has optimized for better distance metrics. The various papers (Seeja and Zareapoor [46]),(Zareapoor et al. [60]) have performed a KNN technique to detect a credit card fraud and compare their result with other data mining techniques and This method is fast with minimum false alerts.

4.7 Artificial immune system

The artificial immune system (AIS) is part of artificial intelligence based on the biological symbol of the human immune system or natural immune system developed by Neal in 1998 [38]. It is a user-based model to discriminate the incoming transaction as genuine or fraudulent. According to Tuo et al. [54], artificial immune systems have four algorithms negative selection, clonal selection, immune network, and dendritic cells to identify fraudulent transaction detection. These algorithms have been applied to real data to detect fraud for achieving high accuracy.

Worng et al. [57] has developed a system called artificial immune system for fraud detection of credit card (AISCCFD). AISCCFD has a high level system model and consisted of the database, two subsystems (system interface) and AIS engine. AISCCFD System has identified input transactions as non-fraudulent or fraudulent.

Halvaiee and Akbari [22] introduced a new credit card fraud detection model using AIS called "AIS-based Fraud Detection Model". The model has added some enhancements to the Artificial Immune Recognition System algorithm which helped to raise the precision, reduces cost, and system training time. AIS-based Fraud Detection Model enhanced fraud detection rate of 23%, reduced cost 85%, and training time 40%. This research has implemented a parallel model in a test environment for fair minimization in training time.

Gadi et al. [19] has employed the Artificial Immune Systems on credit card dataset for fraud detection and outcomes are compared with Neural Network, Bayesian Network, Naive Bayes, and Decision Trees based on three strategies such as default 35.66 (3.21%), optimized 24.97 (5.43%), and robust 23.30 (2.29%). The AIS produced the best classifiers and give high accuracy compared with other fraud detection methods respectively.

Huang et al. [23] have applied the AIS model for fraud detection. The AIS based model combines two AIS algorithms with behavior-based intrusion detection using Classification and Regression Trees (CART).This hybrid method applied on same data with combination CART algorithm (TP1%,TN85%, FP15%,FN19%), CSPRA algorithm (3% undetected: TP81%, TN 89% , FP11%, FN 16%), DCA algorithms (TP78%, TN90%, FP10%,FN22%) and Stacking Bagging algorithm (TP89%, TN95%, FP5%, FN 11%) to calculate each algorithm performance by using standard evaluation criteria. It was observed that the performance of AIS is better than other techniques.

4.8 Decision tree

The Decision tree (DT) is a supervised learning and statistical data mining technique. The research (Koikkinaki [28]) implemented decision trees by using various machines learning-based algorithms such as the Iterative Dichotomiser3 (ID3), Successor of ID3 (C4.5), Random forest, Classification and Regression Tree (CART). These techniques have been applied to a credit card database for fraud detection based on detection time and accuracy. DT technique divides the complex problems into smaller problems and resolves through repeatedly using the procedure (Bai et al. [5]).

According to Sachin and Duman, 2011 [44], decision rules have applied to determine the class of an incoming transaction as a genuine or fraudulent. Gadi et al. [19] used DT technique to computing result based on three strategies such as default 32.76 (4.83%), optimized 27.84 (4.16%), and robust 27.87 (4.21%).

Minegishi and Niim [35] has developed a decision tree learning based method called Very Fast Decision Tree learner (VFDT) to solve the imbalanced data stream problem.

Sahin et al. [45] has developed and implemented a cost-sensitive decision tree induction algorithm to identify fraudulent credit card transactions on a real world credit card data set. The performance of this approach was compared with the traditional classification models on a real data set and result showed that cost-sensitive decision tree algorithm outperforms the existing traditional classification methods.

4.9 Support Vector Machine

The SVM is a statistical and supervised machine learning (ML) technique used for both regression and classification tests [15]. SVM has found to be effective and popular in a diversity of classification tasks because it is mostly used for solving classification problems. SVMs contain two important properties that are the margin optimization and kernel called a radial basis function that can be used to learn complex regions.

The kernel function is used to transform the dataset. The role of this function is a mapping of transactions between the input space and a higher dimensional space [13].

The kernel function is described by:

$$K(X1, X2) = \Phi\{(X1), (X1)\} \tag{3}$$

where $\Phi : X \to H$ map transactions in input space X to higher dimensional space H.

The hyperplane is used to separate the transactions, after applying the kernel function to the dataset.

Hyperplane is described by:

$$\{W, X\} + B = 0 \tag{4}$$

The main role of the hyperplane is to maximize the separation between both transactions, which helps to reduce potential errors caused by over training.

Therefore, the classification for a support vector machine defined as:

$$\sum_i \alpha_i Y_i K(X_i, X) + B = 0 \tag{5}$$

The Gaussian radial basis function and polynomial function are the most common kernel functions used which is dependent on the dataset and classification requirements [30]. An instance of transaction class is that separated by a hyperplane. Transaction classes of two distinct cards are represented in blue and black bullets (data samples). Support vectors work as a boundary to each class of credit card transaction. If transactions lie outside support vector boundary, considered as an Outlier or fraud activity. New samples are classified based on measuring its distance from the hyperplane (Nguwi and Cho [37]).

SVM has been classified into four categories namely Binary support vector system (Chen et al. [12]), classification model based on decision tree and SVM (Sachin and Duman [44]). Novel questionnaire responder transaction approach with SVM (Chen et al. [13], and Class Weighted SVM model (Lu and Ju [30]). It has been developed on Class Weighted Support Vector Machine for credit card fraud detection.

The model is more appropriate for solving fraud detection problem with high.The Various research (Zareapoor et al. [60], Seeja and Zareapoor [46], Bhattacharyya et al. [6] have used SVM Technique for credit card detect fraudulent transactions.

4.10 Hidden Markov model

Hidden Markov Model (HMM) is a probability and statistical Markov model which can be used to model sequential data. It is a double embedded random process with two states, one is unobserved ("hidden") and other is open to all. It is a finite set of states, each of which is associated with a probability distribution. HMM are effectively applied to many areas such as speech recognition, robotics, bio-informatics, data mining etc. This paper focuses on credit card fraud detection using HMM. The cardholder spending profile is divided into a low, medium, and high. The cardholder profile consists of a set of spending information namely money spent on each transaction, purchase time of last good, type of purchasing goods, the name of the merchant and last place purchase goods, etc. HMM is trained based on the cardholders normal behavior and if incoming credit card transaction is not accepted by the model with high probability that transaction is considered as a fraudulent transaction. HMM is also maintaining a log of the previous transaction. Bhusari and Patil [7] discussed how HMM used to detect credit card transaction fraud with a low false alarm 8% and found out more than 88% transactions are genuine. Another researches([2,18,24]) are also used HMM to detect a credit card fraud. These machine learning techniques are also contained some pros and cons.

Table 4 presents a comparison of these techniques based on learning, classification and clustering approach to detect credit card fraud, Learning Approach/Categories (LA/C).

5 Dataset and its attributes

The performance of any technique is purely dependent on the quality of the dataset. The datasets used by the various researcher to evaluate their proposed techniques and implemented methodologies are studied and summarized in table 5. The credit card datasets have been categorized as real world datasets (like Mellon Bank) and synthetic datasets (like UCSD-2009), and their primary and derived attributes discussed in details which play an important role in enhanced fraud detection rate and accuracy.

The credit card dataset contains two types of attributes, namely primary and derived attributes. These attributes were used in various studies [6, 25, 32, 36, 61] to detect credit card transaction frauds.

Primary attributes:
When credit card users are made transactions, a set of primary transnational attributes are generated. Based on the research papers mentioned in table 5, following attributes in table 6 are diagnosed as primary features of the credit card.

These primary attributes (features) help to design a good fraud detection system.

Beside primary attributes, **derived attributes** are also important but it is a very difficult task to extract the appropriate derived attribute from primary attributes.

Withrow et al. [56] have suggested transaction aggregation strategy approach to extract derived attributed based on the following three steps from the primary attributes available in the credit card transaction datasets:

(a) Grouping transactions by card identification number.
(b) Selecting transactions made in a previous period.
(c) Grouping selected transactions on the basis of one of the primary feature.

The derived attributes are used to accurate analysis of cardholders buying behaviors.The derived attributes are summarized in table 7.

6 Performance matrices and evaluation criteria

The standard evaluation criteria have been followed by many researchers [4,30,36,46,49,61] for evaluating their proposed model of credit card fraud detection.

The main objectives of these evaluation parameters (criteria) are to minimize false detection of fraud/non-fraud transactions and maximize the actual detection of fraud /non-fraud transactions.The formulae used by the many researchers for evaluation of their model have been presented in table 8 (P= Positive, N = Negative, FP = False Positive,FN = False Negative,TP = True Positive, TN =True Negative).

The performance metrics are discussed in table 8 that used by the various researcher to evaluate their method and determine the performance of fraud detection techniques.

Table 4: Comparison of various fraud detection techniques

(LA/C)	Techniques	Pros	Cons
Supervised/ Classification	Back Propagation Neural Network	To determine the best performance of the system, we need a retrain of parameters such as the number of hidden neurons, learning rate, and momentum	It requires log training time and extensive testing.
	Bayesian Network	System processing, accuracy, and detection speed are very high.	Required excessive training and expensive.
	K-nearest neighbor (KNN)	There is no requirement of the predictive model before transaction classification in a dataset.	The fraud detection accuracy depends on the measure of distance between two neighbors. This technique cannot detect the fraud at the time of transaction.
	Support Vector Machine (SVM)	It is capable of detecting the fraudulent activity at the time of the transaction and solving nonlinear classification problems. SVM technique provides a unique solution for the optimality problem is bulging.	This technique is hard for auditors to process outcomes due to the transformation of input set and sometimes it fails to detect fraud cases.It is expensive, medium accuracy and has a low speed of detection. Lack of results transparency.
	Decision Tree (DT)	A decision tree is simple to use, implement, display and easy to understand. It can handle well non-linear data. It requires low computational power for training, high flexibility, good explainable.	This technique involves the complex algorithm and even a small change in the data can distract the structure of the tree. Need to check each transaction one by one. Choosing splitting criteria are also complex.
Unsupervised/ Clustering	Self-organizing Map	This technique is simple to implement and very easy for auditors are given visual nature of outcomes.	Visualization requires auditor observation.This technique is not being fully automated.
	Hidden Markov Model (HMM)	HMM technique is fast in fraud detection and capable of detecting the fraudulent activity at the time of the transaction. HMM-based models reduce the False Positive (FP) transactions predict as fraud, though they are genuine User.	It cannot detect the fraud in the initial few transactions. It is highly expensive and low accuracy. It cannot scalable to large size datasets.
Semisupervised classification	Artificial immune system	This is highly suitable for classification problems with imbalanced data.Pattern recognition capability is also a high and powerful technique for learning. It is diversity, dynamically changing coverage, multilayered, and Inexpensive.	Need a very high computational power for operation to make it unsuitable for real-time functions.The handling missing data are poor and need higher training time.

Table 5: Analysis of credit card dataset used between 1994 and 2019

Year	**Dataset sample size**	**Source**
1994	1,100,000 transactions authorized in a two month period.	Mellon Bank [20]
1999	500,000 records of both banks. Chase Bank (20% fraud and 80% non-fraud distribution) and First Union Bank (15% fraud,non-fraud 85%).	Chase Bank and First Union Bank [9]
2002	Data Come from SQL server database containing sample Visa Card transactions.	Synthetically created Data [52].
2005	6000 credit card data, normal accounts (84%) and fraudulent account(16%).	Major US bank [29]
2008	About 50 million transactions (49,858,600 transactions), one million (1,167,757 credit cards) credit cards from a single country.	International CC operators transactions [40]
2009	25,000 payment observations,fraud (5,529) and non-fraud (19,471).	Taiwan bank [59]
2010	462279 unique customers data have 4 million transactions & 5417 fraud transaction	Financial institute in Ireland(WebBiz)^3 [8]
2011	250,000 transactions, fraud (1050), & remaining none fraud.	Turkish bank [17]
2012	640361 transactions contain by 21746 credit cards.	A Large Australian bank [57]
2013	22 million CC transactions, fraud (978) and remaining non-fraud.	Bank's CC data warehouses [45]
2014	Details of 41647 transactions and 3.74% is fraudulent.	Brazilian bank transaction [39]
2014	100000 transaction.But 21000 transaction for training and 19918 for testing out of 40,918 unspecified transactions.	UCSD-FICO [46]
2015	9,387 transactions, 939 fraud and 8,448 non fraud	Real life dataset from an anonymous bank in Turkey [32].
2016	10000 records, numeric data with 334 input attributes.	UCSD-2009 (Synthesized) [55]
2017	10000 transactions, 20 attributes of German Credit data	UCI respiratory [3]
2018	30,000,000 individual transactions, 82,000 transactions fraud	E-commerce company of China [58]
2019	German Credit card1000 transaction &21 attributes.	UCI ML respiratory (Real data) [48, 49]
2019	Ten million credit card transactions with 8 variables	https://packages. revolution analytics.com (Real data) [33]

Table 6: Primary attributes of credit card

Attribute Name	**Description**
Account Number	Identification number of the Credit Cardholders
Card Number	Is a Credit Card number
Cardholders Age	Present an age of Credit Cardholders
Cardholders Gender	Gender of the Credit card holder (male, female, other)
Type of Card	Visa debit, Master Card, American Express, etc.
Transaction ID	Transaction identification number
Transaction Place	Place where transaction made (Determine by System IP Address, if online)
Entry Mode	Chip and pin, magnetic stripe
Transaction Time/-Date	Time and Date of the transaction
Amount	Amount of the transaction n USD, EURO, INR, etc.
Terminal Type	Transactions by the Internet, ATM, POS,Mobile, etc.
Country Transaction	Where transaction made by cardholders
Country Home	Residence country of cardholders
Merchant Name	Name of merchant who is provides service to cardholders
Merchant Code	Is a merchant identification number
Merchant Group	Merchant group identification
Merchant City, Country	Name of merchant city & country

Table 7: Derived attributes of credit card

Attribute Name	**Description**
Amount-Same-Day	Total amount spent on the same day up to this transaction
Number-Same-Day	Total number of transactions on the same day up to this transaction
Amount-Same-Month	Total amount in the same month up to this transaction.
Average-Same-Month	Average amount spent over a month up to this transaction.
Amount-same-Merchant	All transactions amount spent in the same merchant up to present transaction.
Number-Same-Merchant	Total number of transactions with the same merchant during 6 month
Number-Same-Hour	Total number of transactions in the same hour up to present transaction.
Amount-Same-Hour	Total amount spent in the same hour up to present transaction.
Transaction amount over a month	Average amount spent per transaction over a month on all transactions up to present transaction.
Previous-Amount	Previous amount of transaction.
Average amount over 3 months	Average amount spent over the course of 1 week during the past 3 months

Table 8: Evaluation criteria for Credit card fraud detection

Measure	**Formula**	**Description**
Accuracy (Detection rate)	TN+TP/TP+TN+FP+FN	The percentage of correctly predicated transactions.
Precision (Hit rate)	TP/TP+FP	It is the number of classified fraud transactions that actually are fraud transactions, and gives the accuracy of the fraud transaction.
False Positive Rate (False alarm rate)	FP/FP+TN	The ratio of credit card fraud detected incorrectly.
True Positive Rate (Sensitivity or recall)	TP/TP+FN	It is the number of correctly classified fraud transaction and gives the accuracy of the fraud transaction.
True negative rate (Specificity)	TN/TN+FP	It is the amount of correctly classified non-fraud and gives the accuracy of the non fraud transaction.
False Negative Rate	FN/P=1-TP rate	It is the number of credit card transaction classified as a non-fraud as a percentage of all fraudulent.
ROC	True positive rate plotted against false positive rate	Relative Operating Characteristic curve, a comparison of TPR and FPR as the criterion changes.
Cost	100 * FN + 10 *(FP +TP)	Present a cost of FDS.
F-measure	2*(Precision*Recall)/ (Precision + Recall)	The Weighted average of the precision and recall, Gives the harmonic mean of precision and recall.
Balanced Classification Rate	1*(TP/P+TN/N)	It is the average sensitivity and specificity.
Mathews Correlation Coefficient	(TP*TN)-(FP*FN)/ {(TP+FP)(TP+FN) (TN+FP)(TN+FN)}	Used as a measure of the quality of binary classification, correlation coefficient between the observed and predicate binary classification.
Geometric Mean	(Sensitivity *Specificity).05	Gives the geometric of fraud and non-fraud accuracies.
Weighted-Accuracy	W*Sensitivity+(1-w)*Specificity	Provide performance summary that indicators of each tradeoff.

Table 9: The Performance comparison of various DM & ML Techniques

Reference	**Method**	**Sensitivity**	**Specificity**	**Accuracy**
[22]	AIS	33.6-52.6%	97.8-98.1%	94.6-96.4%
[6]	LR	24.6-74.0%	96.7-99.8%	96.6-99.4%
	SVM	43.0-68.7%	95.7- 99.8 %	95.5-99.4%
	RF	42.3-81.2%	97.9-99.8%	97.8-99.6%
[36]	KNN	81.82%	97.61%	96.52%
	KNN+DRF	81.97 %	98.62%	97.47%
[4]	NB	82.10-95.15%	97.54-98.96%	97.52-97.69%
	KNN	82.85-93.75%	100-100 %	97.15-97.92%
	LR	71.55-58.33%	29.39-53.13%	36.39-54.86 %
[43]	SVM	55.43-73.60%	70.41- 73.41%	70.41-73.41%
	GP	85.64-95.09%	89.27-94.14%	89.27-94.14%
	NN (FF)	67.24-80.21%	75.32-78.77%	75.32-78.77%
	GMDH	87.44-93.46%	88.34-95.18%	88.14-93.00%
	LR	62.91-65.23%	70.66-78.88%	66.86-70.86%
	NN(Prob)	87.53 -98.09%	94.07-98.09%	95.64-98.09%
[11]	RUSMRN	53.36%	88.13%	79.73%
	RUSBoost	50.3%	86.16%	77.8%
	AdaBoost	31.4%	83.1%	57.73%
	Naive Bayes	40.2%	78.8%	70.13%

6.1 Comparative study of learning technique

In this study, the performance of a variety of Data mining & machine learning technique in terms of sensitivity, specificity, and accuracy are presented in table 9. This study has been classified credit card fraud detection researches based on these performance measures. The result of various researches showed in table 9 and identify that specificity and accuracy were slightly better of $KNN{+}DRF$ technique [36] with compare to other methods. The sensitivity, specificity, and accuracy have been better of NB and KNN technique [4]. It clearly identified that some results are vary based on credit card transactions dataset size and detection techniques. If credit card dataset is too large than accuracy could be decreased. This study has been classified credit card fraud detection researches based on these performance measures. The result of various researches showed in table 9,sensitivity and accuracy was slightly better for Self-organization map [8] with compare to other methods.The sensitivity, specificity, and accuracy have been better of NB and KNN technique [4]. It clearly identified that some results are vary based on credit card transactions dataset size and detection techniques. If credit card dataset is too large than accuracy could decrease.

7 Conclusion and future work

In this paper, various research papers based on credit card fraud are studied and discussed based on the finding of these papers, the various credit card frauds are classified and among them, card-not-present fraud and skimming frauds are more frequently occurred.The fraudsters mostly used website cloning, the false merchant website, and phishing methodology to steal credit card detail.The challenges and issues to prevent and detect the fraud have also been discussed and identified that one of the biggest issues is benchmark dataset which has unskewed, balanced and free from the anomaly and real-world dataset.

The pros and cons of numerous fraud detection techniques are discussed and concluded that the major researchers were performed under a supervised learning method based on available datasets.Different convention dataset was used to compare the performance of various methodologies on the base of various performance measures using the sensitivity,specificity,and accuracy.This paper also presented a comparative analysis of different fraud detection techniques and identified that NB (95.15%, 98.96%, and 97.69%),KNN (93.75%, 100%,and 97.92%) & KNN-DRF (81.97%, 98.62%,and 97.47%) givens better results among the studied techniques. The primary and derived attributes of credit card dataset have been playing an important role in enhanced fraud detection rate and accuracy.

In this paper, techniques based on machine learning are discussed. The study can be extended for bio-inspired algorithms which have not explored in the paper and the result can be compared with traditional MLT. More effects can be made to get the real dataset from the credit card issuing and managing organization, so that the exact techniques can be compared on the real dataset.

In future, this work can further be extended on the enhancement of fraud detection techniques based on the detection accuracy, precision, MCC and improvement of fraud detection evaluation criteria. Security guidelines are needed for credit card users to make them aware of how to use and secure card details.

Acknowledgment

We gratefully acknowledge the funding agency, the University Grant Commission (UGC), Delhi, Government of India for providing financial support to complete this work. And also, USICT, Guru Gobind Singh Indraprastha University, Delhi, India to provided the facilities to complete research and Editors-in-Chief of IJCCC.

Bibliography

[1] Abdallah, A.; Maarof, M. A.; Zainal, A.(2016). Fraud detection system: A survey, *Journal of Network and Computer Applications*, 68, 90-113, 2016.

[2] Agrawal, A.; Jain,A.; Kumar, B. S.(2019). Deep Learning Based Classification for Assessment of Emotion Recognition in Speech, *Available at SSRN 3356238.*

[3] Alam, F.; Pachauri, S. (2017). Detection using weka, *Advances in Computational Sciences and Technology*, 10(6), 1731-1743, 2017.

[4] Awoyemi, J. O.; Adetunmbi, A. O.; Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis, *2017 International Conference on Computing Networking and Informatics (ICCNI)*, IEEE, 1-9, 2017.

[5] Bai, B.; Yen, J.; Yang, X. (2008). False financial statements: characteristics of China's listed companies and cart detecting approach, *International journal of information technology* & *decision making*, 7(2), 339-359, 2008.

[6] Bhattacharyya, S.; Jha, S.; Tharakunnel, K. Westland, J. C. (2011). Data mining for credit card fraud: A comparative study, *Decision Support Systems*, 50(3), 602-613, 2011.

[7] Bhusari, V.; Patil, S.(2011). Application of hidden markov model in credit card fraud detection, *International Journal of Distributed and Parallel systems* 2(6), 203, 2011.

[8] Brabazon, A.; Cahill, J.; Keenan,P.; Walsh,D.(2010).Identifying online credit card fraud using artificial immune systems, *Evolutionary Computation (CEC), 2010 IEEE Congress on* IEEE, 1-7, 2010.

[9] Chan, P. K.; Fan, W.; Prodromidis, A. L.; Stolfo, S. J.(1999). Distributed data mining in credit card fraud detection, *IEEE Intelligent Systems and Their Applications*, 67-74, 1999.

[10] Chandola, V.; Banerjee, A.; Kumar, V. (2009). Anomaly detection: A survey, *ACM computing surveys (CSUR)*, 41(3), 1-72, 2009.

[11] Charleonnan, A. (2016). Credit card fraud detection using RUS and MRN algorithms, *2016 Management and Innovation Technology International Conference (MITicon)*, IEEE, MIT73-MIT76, 2016.

[12] Chaudhary, K.; Yadav, J.; Mallick, B. (2012). A review of fraud detection techniques: Credit card, International, *Journal of Computer Applications* , 45 (1), 39-44, 2012.

[13] Chen, R.C.; Chiu, M.L.; Huang, Y.L.; Chen, L.T. (2004). Detecting credit card fraud by using questionnaire responded transaction model based on support vector machines, *nternational Conference on Intelligent Data Engineering and Automated Learning*, Springer, 800-806, 2004.

[14] Chen, R.C.; Luo, S.T.; Liang, X.; Lee, V.(2005). Personalized approach based on svm and ann for detecting credit card fraud, *International Conference on Neural Networks and Brain*, IEEE, 810-815, 2005.

[15] Cortes, C.; Vapnik, V.(1995). Support-vector networks, *Machine learning*, 20(3), 273-297.

[16] Craciun, M.; Ratiu, C.; Bucerzan, D.; Manolescu, A. (2013). Actuality of Bankruptcy Prediction Models used in Decision Support System, *International Journal of Computers Communications & Control*, 8(3), 375-383, 2013.

[17] Duman, E.; Ozcelik, M.H.(2011). Detecting credit card fraud by genetic algorithm and scatter search, *Expert Systems with Applications*, 38(10), 13057-13063, 2011.

[18] Falaki, S.; Alese, B.; Adewale, O.; Ayeni, J.; Aderounmu, G.; Ismaila, W.(2012). Probabilistic credit card fraud detection system in online transactions, *Int. J. Softw. Eng. Appl*, 6 , 69-78, 2012.

[19] Gadi, M. F. A.; Wang, X.; do Lago, A. P. (2008). Credit card fraud detection with artificial immune system, *in:International Conference on Artificial Immune Systems*, Springer, 119-131, 2008.

[20] Ghosh, S.; Reilly, D. L.(1994). Credit card fraud detection with a neural-network,In System Sciences, *Proceedings of the Twenty-Seventh Hawaii International Conference on*, IEEE, 3, 621-630, 1994.

[21] Guo, T.; Li, G.Y. (2008). Neural data mining for credit card fraud detection, *2008 International Conference on Machine Learning and Cybernetics*, IEEE, 7, 3630-3634, 2008.

[22] Halvaiee, N. S.; Akbari, M. K.(2014). A novel model for credit card fraud detection using artificial immune systems, *Applied Soft Computing*, 24, 40-49, 2014.

[23] Huang, R.; Tawfik, H.; Nagar, A.K. (2010). A novel hybrid artificial immune inspired approach for online break-in fraud detection, *Procedia Computer Science*, Elsevier, , 1(1), 2733-2742, 2010.

[24] Iyer, D.; Mohanpurkar, A.; Janardhan, S.; Rathod, D.; Sardeshmukh, A. (2011). Credit card fraud detection using hidden markov model, *Information and Communication Technologies (WICT), World Congress on*, IEEE, 1062-1066, 2011.

[25] Jha, S.; Guillen, M.; Westland, J. C. (2012). Employing transaction aggregation strategy to detect credit card fraud, *Expert systems with applications*, 39(16), 12650-12657, 2012.

[26] Kirkos, E.; Spathis, C.; Manolopoulos, Y.(2007). Data mining techniques for the detection of fraudulent financial statements, *expert systems with applications* , 32(4), 995-1003, 2007.

[27] Kohonen, T. (1990). The self-organizing map, *Proceedings of the IEEE*, 78(9), 1464-1480, 1990.

[28] Kokkinaki, A. I. (1997). On atypical database transactions: identification of probable frauds using machine learning for user Profiling, *Knowledge and Data Engineering Exchange Workshop, IEEE proceedings*, 107-113, 1997.

[29] Kou, G.; Peng, Y.; Shi, Y.; Wise, M.; Xu,W. (2005). Discovering credit cardholders behavior by multiple criteria linear programming, *Annals of Operations Research*, 135, 261-274, 2005.

[30] Lu, Q.; Ju, C. (2011). Research on credit card fraud detection model based on class weighted support vector machine, *Journal of Convergence Information Technology*, 6, 2011.

[31] Maes, S.; Tuyls, K.; Vanschoenwinkel, B.; Manderick,B.(2002) . Credit card fraud detection using bayesian and neural networks, *Proceedings of the 1st international naiso congress on neuro fuzzy technologies*, 261-270, 2002.

[32] Mahmoudi, N.; Duman, E.(2015). Detecting credit card fraud by modified Fisher discriminant analysis, *Expert Systems with Applications*, 42(5), 2510-2516.

[33] Makki, S.; Assaghir, Z.; Taher, Y.; Haque, R.; Hacid, M. S.; Zeineddine, H. (2019). An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection, *IEEE Access*, 7, 93010-93022, 2019.

[34] Meyer, A.; Zimmermann, H.-J. (2011). Applications of Fuzzy Technology in Business Intelligence, *International Journal of Computers Communications & Control*, 6(3), 428-441, 2011.

[35] Minegishi, T.; Niimi, A. (2011). Detection of fraud use of credit card by extended VFDT, *Internet Security (WorldCIS), 2011 World Congress on*, IEEE, 152-159, 2011.

[36] Nami, S.; Shajari, M. (2018). Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors, *Expert Systems with Applications*, 110, 381-392, 2018.

[37] Nguwi, Y.Y.; Cho, S.Y. (2010). An unsupervised self-organizing learning with support vector ranking for imbalanced datasets, *Expert Systems with Applications* , 37 (12), 8303-8312, 2010.

[38] Nunes, C. L.; De Castro, L. N.; Timmis, J. (2002). *Artificial immune systems: a new computational intelligence approach*, Springer Science and Business Media, 2002.

[39] Olszewski, D.(2014). Fraud detection using self-organizing map visualizing the user profiles, *Knowledge-Based Systems*, 70, 324-334, 2014.

[40] Paasch, C. A. (2008). *Credit card fraud detection using artificial neural networks tuned by genetic algorithms*, Hong Kong University of Science and Technology, Hong Kong, 2008.

[41] Panigrahi, S.; Kundu, A.; Sural, S.; Majumdar, A. K. (2009). Credit card fraud detection: A fusion approach using Dempster-Shafer theory and bayesian learning,*Information Fusion*, 10(4), 354-363, 2009.

[42] Quah, J. T.; Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence, *Expert systems with applications*, 35 (4),1721-1732.

[43] Ravisankar, P.; Ravi, V.; Rao, G. R.; Bose, I.(2011). Detection of financial statement fraud and feature selection using data mining techniques, *Decision Support Systems*, 50(2), 491-500.

[44] Sahin, Y. G.; Duman, E. (2011). Detecting credit card fraud by decision trees and support vector machines, *Processing of the international multi conference Engineering and computer Scientist*, 1, 2011.

[45] Sahin, Y.; Bulkan, S.; Duman, E.(2013). A cost-sensitive decision tree approach for fraud detection, *Expert Systems with Applications*, 40(15), 5916-5923, 2013.

[46] Seeja, K.; Zareapoor, M.(2014). Fraudminer: a novel credit card fraud detection model based on frequent item set mining, *The Scientific World Journal*, ID 252797, 2014.

[47] Singh, A.; Jain, A. (2018l). Study of cyber attacks on cyber-physical system, *Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT)*, 26-27, 2018.

[48] Singh, A.; Jain, A. (2019). Adaptive Credit Card Fraud Detection Techniques Based on Feature Selection Method, *Advances in Computer Communication and Computational Sciences*, Springer, Singapore, 167-178, 2019.

[49] Singh, A.; Jain, A. (2019). A Novel Framework for Credit Card Fraud Prevention and Detection (CCFPD) Based on Three Layer Verification Strategy, *Proceedings of ICETIT 2019*, Springer, Cham, 935-948, 2019.

[50] Sorournejad, S.; Zojaji, Z.; Atani, R. E.; A. H. Monadjemi (2016). A survey of credit card fraud detection techniques:Data and technique oriented perspective, *arXiv preprint arXiv*:1611.06439, 2016.

[51] Stolfo, S.; Fan, D. W.; Lee, W.; Prodromidis, A.; Chan, P. (1997). Credit card fraud detection using meta-learning: Issues and initial results, *In AAAI-97 Workshop on Fraud Detection and Risk Management*, 83-90, 1997.

[52] Syeda, M. ; Zhang, Y.-Q.; Pan,Y. (2002). Parallel granular neural networks for fast credit card fraud detection, *Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference*, IEEE, 1, 572-577, 2002.

[53] Tripathi, K. K.; Pavaskar, M. A.(2012). Survey on credit card fraud detection methods,*International Journal of Emerging Technology and Advanced Engineering*, 2(11), 721-726, 2012.

[54] Tuo, J.; Ren, S.; Liu, W.; Li,X.; Li,B.; Lei, L. (2004). Artificial immune system for fraud detection, *IEEE International Conference on Systems, Man and Cybernetics*, IEEE, 2, 1407-1411, 2004.

[55] West, J.; Bhattacharya, M.(2016). Intelligent financial fraud detection: a comprehensive review,*Computers & security*, 57, 47-66, 2016.

[56] Whitrow, C.; Hand, D.; Juszczak, P.; Weston, D.; Adams, N. (2009). Transaction aggregation as a strategy for credit card fraud detection, *Data Mining and Knowledge Discovery*, 18 (1), 30-55, 2009.

[57] Wong, N.; Ray, P.; Stephens,G.; Lewis,L.(2012). Artificial immune systems for the detection of credit card fraud: an architecture, prototype and preliminary results, *Information Systems Journal*, 22 (1), 53-76, 2012.

[58] Xuan, S.; Liu, G.; Li, Z.; Zheng, L.; Wang, S.; Jiang, C. (2018). Random forest for credit card fraud detection, *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, IEEE, 1-6, 2018.

[59] Yeh, I.C.; Lien,C.h.(2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients, *Expert Systems with Applications*, 36(2), 2473-2480, 2009.

[60] Zareapoor, M.; Seeja, K.; Alam, M. A.(2012). Analysis on credit card fraud detection techniques: Based on certain design criteria, *International Journal of Computer Applications*, 52 (3), 35-42, 2012.

[61] Zhang, X.; Han, Y.; Xu, W.; Wang, Q.(2019). HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture, *Information Sciences*, In Press, 2019.

[62] http://www.economictimes.indiatimes.com, October 31, 2016. R. Dave, *Credit, debit card frauds and how you can avoid them*, Accessed on Nov 10, 2018.

[63] https://www.acfe.com

[64] https://www.nilsonreport.com/upload

[65] https://www.ic3gov.in

[66] https://www.financialfraudaction.org.uk