Credit Card Fraud Detection Methods: A Review

Sumedh N. pundkar[1]* and Dr Mohd Zubei[1]
[1]Departement of Computer Science & Engineering., *Madhyanchal Professional University Ratibad, Bhopal, M.P., INDIA*

*Corresponding author : chetanhpatil@gmail.com

Abstract. In today's context, the term "fraud" has become closely intertwined with credit card-related deceit. Recent years have witnessed a notable surge in both credit card utilization and fraudulent activities. Detecting and thwarting fraud necessitates a meticulous analysis of customers' spending patterns. The ubiquity of credit card use for both online and in-store transactions has un-fortunately led to a parallel rise in recognition valentine scam occurrences. While the primary area of deception discovery is the documentation of sham incidents, the urgency of promptly flagging such events cannot be overstated. The con-temporary landscape heavily favors credit card usage, a trend that inadvert-ently contributes to the annual expansion of fraudulent gains. This unlawful practice exerts a pernicious influence on the global economy at large, exac-erbating its impact year after year. Numerous cutting-edge methods, including as data mining, machine learning, fuzzy logic, genetic programming, sequence alignment, artificial intelligence, and fuzzy logic, have become indispensable in the fight against this threat when it comes to identifying credit card fraud. This study delves into the intricate integration of data mining methodologies, showcas-ing their robust potential to provide comprehensive coverage against fraudu-lent activities while maintaining a controlled balance between false alarms and detection accuracy. Within the financial sector, the challenge of credit card fraud detection remains both persistent and pressing. This paper intro-duces an innovative paradigm aimed at fortifying credit card fraud detection. This is achieved by synergistically harnessing the capabilities of the Artificial Underground Over-sampling Practice (SMOTE), the potency of Adaptive Increasing (ADABoost), and the privacy-enhancing attributes of Federated Learning. The incorporation of federated learning serves a dual purpose: not only does it address prevailing data privacy concerns, but it also significantly augments the precision of fraud detection across a diverse array of geograph-ically distributed data sources

Keywords : Adaptive Boosting, Machine Learning (ML), and Synthetic Minority Over-sampling Technique(SMOTE)(ADABoost),

1 Introduction

Fraud encompasses the illicit act of misappropriating monetary funds or tangi-ble goods. The criminal intentions of fraud are not readily apparent. Credit card fraud is a form of fraudulent activity that mostly focuses on credit cards, howev-er it may also extend to other areas. Credit card fraud refers to the illicit act of utilizing a credit card or any other kind of payment instrument with the intention of unlawfully obtaining funds from a financial transaction. The incidence of créance card fraud is increasing within the financial services sector. Credit card fraud detection models play a key role in both academic and commercial envi-ronments because to the inherent challenges associated with previous approach-es. The significance of fraud has undergone a transformation in light of techno-logical advancements. Credit card theft poses a significant risk to contemporary organizations. Credit card fraud refers to the occurrence of unauthorized transac-tions made using a credit card that has been either lost or stolen. Various systems, models, procedures, and preventative actions can effectively mitigate this prac-tice. Credit card issuers and banks own significant quantities of client credit card transaction data. A significant number of individuals are issued credit cards. Cardholders make payments based on their level of dedication. The utilization of credit cards in China is seeing an upward trend; nonetheless, a limited number of cardholders employ them for routine online or in-store purchases. The situation of confiance cards is seldom due to the prevailing lack of faith in the payment sys-tem. The implementation of fraud detection techniques utilizing cardholder pur-chase data, namely by analyzing current purchasing behavior, is a promising ap-proach to mitigate credit card fraud. This method holds significant importance for both banking institutions offering credit services and the advancement of E-commerce. Fraud detection systems are triggered when perpetrators successfully bypass fraud prevention measures and engage in fraudulent activities. The prolif-eration of technology and improved communication channels has led to the dis-semination of information as well as fraudulent activities, resulting in substantial global financial losses. Anderson (2007) identified and provided descriptions for many forms of fraud. Credit card fraud encompasses several illicit activities, such as the unauthorized acquisition of credit card information, the production and use of counterfeit cards, the act of forging signatures, the submission of de-ceitful credit card applications, as well as the perpetration of online or electronic frauds in instances where the cardholder is not physically present. Detecting credit card fraud poses significant challenges and is a widespread issue. Confi-dential information includes just the revealed details of the amount, MCC (Mer-chant Category Code), acquirer number, date/time, and merchant address. Deci-sion trees, neural networks, and case-based reasoning are commonly employed in several fraud detection systems and models. In order to acquire knowledge about fraudulent tendencies, a substantial dataset comprising both genuine and fraudu-lent transactions is necessary for the implementation of these algorithms. In-stances of fraudulent transactions are few within a singular firm. In this dis-course, it is essential to consider the academic nature of the user's text. It is im-perative to rewrite the user's Figure 1 depicts the worldwide distribution of issued Master credit cards, highlighting their widespread use across many areas. The graph presumably illustrates the global growth and distribution patterns of these cards, indicating their widespread acceptance and utilization within the financial system. Figure 2 presents the global issuance of Visa credit cards, emphasizing their enormous international circulation. The graph presumably represents the patterns of issuance, reflecting the prevalence and broad use of Visa cards in the global financial domain.

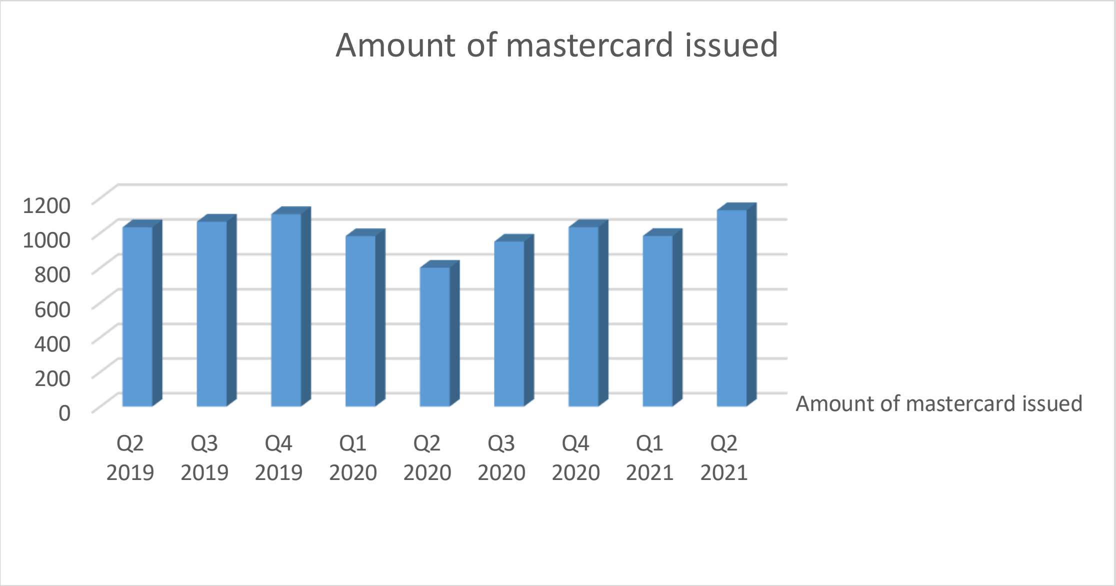

Fig. 1 Amount of Master credit card issued worldwide

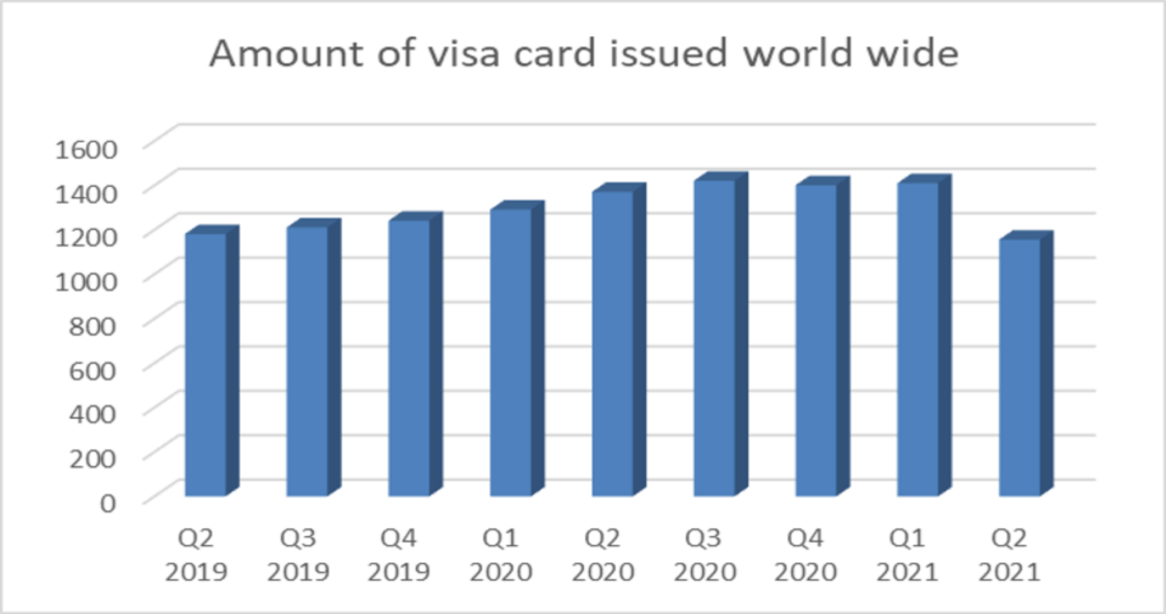

Fig. 2 Amount of visa credit card issued worldwide

1.1 Motivation

The development of an efficient detection system in the field of CCFD entails the utilization of intricate methods and approaches. The present analysis is based on a comprehensive examination of the literature pertaining to the issues prevalent in the CCFD. The identified difficulties have served as a catalyst for the devel-opment of a proposed solution that aims to be efficacious in addressing these concerns. The volume of credit card transactions is significant and exhibits het-erogeneity. Users employ credit cards for a multitude of purposes, dependent upon geographical areas and currencies. This observation underscores the broad range of fraudulent transactions. This issue has served as a source of motivation for me to develop a viable résultat that can effectively detect fraudulent transac-tions regardless of their geographical origin. The work of fraud detection is simi-larly characterized by many objectives. Banks and financial institutions are re-quired to consistently provide their users with a satisfactory experience and high-quality service. Hence, the utilization of customer datasets for experimental purposes is a significant challenge due to the need to maintain service availabil-ity and privacy. In order to address this particular difficulty, my drive prompts me to offer the concept of federated learning as a means of ensuring data privacy. The presence of fraudulent transactions and the passage of imbalanced datasets provide significant challenges in the field of credit card fraud detection (CCFD). Acquiring real-time datasets of credit card transactions poses significant chal-lenges. Due to adherence to the General Data Protection Regulation, banks and other financial institutions do not disclose the personal information of their clients (GDPR). Therefore, obtaining datasets to detect credit card fraud poses a serious challenge for academics. My passion drives me to assist research groups and data scientists operating within the financial industry in developing a system that ef-fectively addresses the issues associated with acquiring large volumes of data for the purpose of constructing robust appareil learning models. According to the source provided, [3] The user did not provide any text to rewrite.

Table 1 : List of Abrevation

Abréviation	**Full form**
ML	Machine Learning
CCFD	credit card fraud detection.
GDPR	General Data Protection Regulation
SMOTE	Synthetic Minority Over-sampling Technique
ADABoost,	Adaptive Boosting
POS	point-of-sale (POS)
True Positive Rate	True Positive Rate
DNN	Deep Neural Network

2 Types of Fraud

Credit card fraud, telephone scams, and hacking are all discussed in this paper, along with bankruptcy fraud, counter feiting, application fraud, and behavioural fraud. Offline credit card fraud

2.1.1 Lost or Stolen Card:

When a credit card is lost or stolen, fraudsters can use it to make unauthorized purchases. They may use the card at physical stores, restaurants, or other estab-lishments before the cardholder realizes the card is missing and reports it.

2.1.2 Counterfeit Cards:

Fraudsters create counterfeit credit cards by duplicating the information from legitimate cards. This can be done through skimming devices or by obtaining the cardholder's information from other sources. Counterfeit cards are then used for in-person transactions, often at retail stores or ATMs.

2.1.3 Card Impersonation:

In card impersonation fraud, a fraudster pretends to be the legitimate card-holder by presenting a physical card with the cardholder's information. This can happen when a thief acquires a physical card, or they may combine stolen card information with a fake identification document.

2.1.4 Card Alteration :

Fraudsters alter legitimate credit cards by modifying the information on the card, such as the cardholder's name, expiration date, or security code. This allows them to use the altered card for fraudulent transactions without raising suspicion.

2.1.5 Mail Theft:

Fraudsters may target mailboxes or intercept mail to obtain credit cards that are being sent through postal services. They can use the stolen cards for in-person transactions or sell them to other criminals.

2.1.6 Card Skimming:
Card skimming can occur offline as well, where fraudsters install devices called skimmers on physical card readers, such as ATMs or point-of-sale (POS) terminals. These skimmers capture the cardholder's information when they swipe or insert their card. The stolen information is then used to create counterfeit cards or make fraudulent purchases.

2.1.7 Card Trapping:
Card trapping involves the use of physical devices or mechanisms to trap a card inside an ATM or other card-accepting device. When the cardholder leaves, the fraudster retrieves the trapped card and uses it for fraudulent purposes.

2.1.8 Manual Card Imprinting:
In some instances, particularly in situations where electronic payment systems are unavailable or unreliable, merchants may use manual card imprinters to cap-ture credit card information. Fraudsters can abuse this process by capturing the card imprint and using the information for fraudulent transactions.

3 Literature Review

3.1.Neural networks:

In the study by C. H. Sumanth et al. (2022), the central objective is to detect instances of credit card fraud effectively. The surge in valentine scams corre-sponds to the rise in debit card usage for both in-person and online expenses. This rise is directly related to e-commerce's explosive expansion. Creating an extensive credit card dataset for testing and training is the research's main task. After training, a deep neural network is used to classify user-provided data together with techniques like Support Vector Machine (SVM), Deep Neural Network (DNN), and Naive Bayes. As the literature highlights, this all-encompassing method produces a very effective system for the precise identification of credit card fraud.

The research conducted by F. K. Alarfaj et al. (2022) addresses the prevalent issue of credit card fraud in online transactions, fueled by the ease and populari-ty of such transactions. With a focus on enhancing detection accuracy and mini-mizing fraud losses, the study explores an array of machine learning techniques, XG Boost, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, and Extreme Learning Method. The research emphasizes the growing significance of deep learning algorithms, aiming to improve accura-cy. Employing various convolutional neural network architectures, the study systematically analyzes layering effects and model configurations to achieve optimal results. With accuracy, f1-score, precision, and AUC tuned at 99.9%, 85.71%, 93%, and 98%, respectively, the suggested model outperforms current machine learning and deep learning algorithms in credit card fraud detection, according to empirical research. Additionally, meth-ods for addressing false negatives and achieving real-world effectiveness are explored. This study underscores the potential of the proposed methods in effec-tively countering credit card fraud, supported by empirical evidence [10].

In the research by E. Esenogho et al. (2022), the surge in both traditional and online purchases facilitated by recognition valentines has been driven by the expanding realm of electronic commerce and communication systems. However, this trend has also led to an increase in credit card fraud, resulting in substantial financial losses for banks annually. Addressing this challenge, the study focuses on developing accurate fraud detection algorithms, a task complicated by biases inherent in credit card datasets and the dynamic nature of user purchase behav-iors. To overcome these issues, the authors propose an innovative approach that combines a neural network ensemble classifier with a hybrid data resampling strategy. The proposed method involves utilizing adaptive boosting (AdaBoost) in conjunction with a long short-term memory (LSTM) neural network as the foundation for the ensemble classifier. This study, as outlined in [10], presents a valuable contribution to credit card fraud detection, leveraging advanced tech-niques to tackle the complexities associated with evolving purchase patterns and biased datasets

3.2.Decision Tree:

In the evolution of learning systems, notable contributions were made with the introduction of C4.5 (Quinlan, 1993) and ID3 (Quinlan, 1986), representing con-tinuous data decision tree methodologies. These decision trees entail branches leading to various options, classifying nodes as either branch nodes or leaf nodes devoid of offspring. Leveraging data mining, these approaches extract diverse forms of classifying information through training. Subsequently, this knowledge is harnessed to construct decision trees that effectively deconstruct intricate problems into more manageable components, embodying the principle of devel-oping accurate, resource-efficient decision trees. This review highlights the foundational role of decision trees in simplifying complex problems based on precise classification, influenced by the pioneering works of C4.5 and ID3.

In the study conducted by B. Gedela et al. (2022), the substantial challenge faced by financial institutions in identifying fraudulent activities related to credit cards is addressed. The research employs the AdaBoost classifier to detect potentially suspicious financial transactions. The suggested algorithm's efficacy is evaluated using a variety of techniques, such as logistic regression, artificial neural networks, decision trees, and Naive Bayes. The dataset com-prises a total of 284,807 transactions, divided into training [n=227,845 (80%)] and test [n=56,962 (20%)] subsets. Among these transactions, 492 are classified as fraudulent. The performance of the AdaBoost algorithm is evaluated and com-pared against various other machine learning techniques, using metrics like accu-racy, sensitivity, specificity, precision, and f-score. Interestingly, the high detection accuracies of 99.43%, 90.9%, 95.3%, 94.8%, and 94.8% for AdaBoost, Na-ive Bayes, logistic regression, artificial neural network, and decision tree algorithms, respectively, are all displayed. AdaBoost's f-score, at a significance level of 0.05, reaches an impressive 99.48%. The suggested AdaBoost algorithm beats alternative techniques including Naive Bayes, logistic regression, artificial neural networks, and decision tree algorithms in the identification of credit card fraud, according to qualitative analysis [11]. The advantages of the AdaBoost technique in improving the precision of fraud detection in credit card transactions is highlighted by this study.

In the study by J. C. Mathaw et al. (2022), the surge in online business transac-tions owing to digitization's growing popularity is evident. The convenience of electronic payments and virtual shopping, driven by time and cost savings, has attracted consumers. Concurrently, the rise of recognition valentine scams paral-lels the increase in connected shopping. Due to the increasing use of credit cards in online transactions, credit card fraud has increased. Illegally obtaining credit card and user data for unauthorized teleshopping operations is a common fraudulent avenue. Amid the challenge of identifying fraud within a vast pool of legitimate transactions, more effective detection techniques are imperative. The research proposes a method for accurately categorizing potentially fraudulent financial transactions. Leveraging Machine Learning (ML) algorithms like Deci-sion Trees, Random Forests, Logistic Regressions, and K-Nearest Neighbours, the study focuses on detecting credit card scams. Both Random Forest and Decision Tree algorithms demonstrate high accuracy and satisfactory F-scores. Future steps involve employing a feature selection process to enhance classifier model accuracy by isolating crucial data characteristics [16]. This work underscores the significance of advanced ML techniques in the accurate detection of credit card fraud within the context of increasing online business conventions.

According to J. B. et al.'s investigation from 2022, credit card fraud is still on the rise even though credit cards are a common form of payment for both online and offline transactions. In order to protect customers from inadvertent payments for unauthorized transactions, banks and credit card firms, among other financial institutions, must promptly detect fraudulent credit card activity. The challenge arises from the higher frequency of legitimate transactions, rendering the development of a model ca-pable of distinguishing them from fraudulent ones more intricate. This study compares the traditional method of regular machine learning algorithms for such data with algorithms designed expressly to handle extremely imbalanced data that is prejudiced against fraud transactions. This study compares and assesses the efficacy of many supervised machine learning models, including Support Vector Machine, Random Forest, Decision Tree, and Logistic Regression, in identifying fraudulent mobile money transactions. These models are rigorously tested using data generated by a simulator based on actual business transactions [12]. This review emphasizes the significance of robust machine learning techniques in effectively addressing the challenge of detecting credit card fraud within the realm of imbalanced transaction data.

Decision trees offer inherent advantages due to their adaptability without the need for parameters or specific data distributions, enhancing their versatility. Their nearly explicit logic contributes to various applications, notably the creation of similarity trees as demonstrated by Kokinaki (1997). In similarity trees, nodes sharing attribute values are linked by edges, forming leaves associated with an intensity factor reflecting the ratio of attributes to total matching conditional labeling requirements (s). This method, known as the similarity tree ap-proach, stands out for its user-friendly nature, visual representation, and com-prehensibility. However, its limitations encompass manual verification of transactions. In the realm of intrusion detection systems, decision trees, particularly inductive decision trees, have been successfully employed to establish compara-ble outcomes [Fan et al., 2001]. This synthesis underscores the advantageous attributes of decision trees in addressing various challenges within the context of similarity trees and intrusion detection systems, drawing from past research find-ings.

3.3.Logistic Regression:

In the investigation by D. Tanouz et al. (2021), the prevalence of credit card usage has overtaken cash as the preferred payment method in recent times. In response to the escalating occurrence of fraud, propelled by technological advancements, the need for an effective fraud detection algorithm has become im-perative. The study proposes the adoption of diverse machine learning-based classification algorithms, including logistic regression, random forest, and Naive Bayes, to handle the inherent imbalance in the dataset. The research concludes by evaluating critical metrics such as accuracy, precision, recall, f1 score, confu-sion matrix, and Roc-auc score [13]. This review underscores the significance of employing machine learning techniques to combat rising credit card fraud, with a focus on a comprehensive assessment of algorithm performance through essen-tial evaluation metrics.

In the study by A. S. Rathore et al. (2021), the prevalence and increasing fre-quency of credit card fraud prompts the application of Data Science and its coun-terpart, Machine Learning, to address this challenge. The research focuses on comparing the efficacy of four widely recognized machine learning algorithms against the backdrop of heavily imbalanced datasets. The algorithms under scru-tiny include Decision Tree, Random Forest, K-nearest neighbors, and Logistic regression. This comparison is conducted within the context of transactional da-ta, encompassing variables such as time, location, purchase type, value, vendor, and customer preferences. These data attributes are fed into diverse models em-ploying statistical techniques to assess the potential for fraud within a given transaction [14]. This review underscores the significance of employing well-established machine learning methods to tackle credit card fraud, emphasizing the use of comprehensive transactional data and robust model comparisons as integral components of the research.

3.4.Genetic algorithms:

Fraud prediction algorithms have garnered significant attention in the field. Bentley et al. (2000) propose a notable approach utilizing genetic programming to formulate logical rules for the classification of credit card transactions into suspicious or legitimate categories. Despite its conceptual strengths, this strategy presents certain limitations. Their empirical study draws from a database encom-passing 4,000 transactions across 62 distinct fields. The evaluation involved a comparative analysis of datasets using a "tree" structure, along with segregated training and evaluation samples. To achieve this, a comprehensive exploration of rules was conducted across diverse domains, with predictive performance emerg-ing as the most influential criterion.

The potential of this approach extends beyond credit card fraud detection, as suggested by its applicability to real-world home insurance data. Another notable contribution, the algorithm proposed by Chan et al. (1999), focuses on predicting behaviors indicative of suspicious activities. This study introduces an alternative perspective by evaluating models based on cost considerations, diverging from the traditional reliance on True Positive Rate (TPR) and False Negative Rate (FNR) metrics.

Further strides in predictive accuracy are proposed by Wheeler & Aitken (2000), who advocate for a novel amalgamation of algorithms to enhance predictive ca-pabilities. Their comprehensive framework encompasses techniques ranging from diagnosis and resolution to optimal matching, density selection, probabilis-tic curves, and negative selection. The research underscores the effectiveness of probabilistic algorithms and neighborhood-based approaches as suitable classifi-cation techniques. However, it also underscores the need for additional diagnos-tic algorithms to address nuanced decision-making scenarios and the incorpora-tion of confidence and relative risk measures for improved performance.

An intriguing convergence of genetic algorithms and neural networks emerges through the Genetic Algorithms with Neural Networks (GANN) paradigm. Draw-ing inspiration from natural evolution, GANN optimizes neural network configu-rations through genetic algorithms. The crux of this integration lies in addressing the challenges of effectively combining these methodologies. GANN generates a diverse array of network architectures through randomization, evaluated via ge-nomic data that examines parameter combinations within neural network struc-tures. Notably, the subsequent evaluation process is facilitated by back-propagation training.

While some GANN techniques focus solely on genetic algorithms to identify optimal neural network architectures, a noteworthy approach involves parameter exploration and ranking in the absence of a predefined training dataset. Genetic algorithms, a subset of Evolutionary Algorithms (EA), offer a robust solution to optimization and search tasks, leveraging mechanisms like modification, inheritance, selection, and crossover

3.5.Clustering techniques:

Bolton and Hand present two clustering algorithms for identifying behavioral fraud (2002). Peer group analysis can identify accounts that have diverged from their peers. Such accounts are flagged. Fraud analysts found them. Peer group analysis suggests that if a group of accounts has been behaving similarly for some time, and one account starts to behave markedly differently, that account should be highlighted. Breakpoint analysis is a separate approach that tests the hypothesis that an account should be reviewed if a card usage notification is re-ceived individually. Break-point analysis can also identify irregular card transac-tions. Unexpectedly large transactions and frequent use without the cardholder's knowledge are red flags (s).

3.6 KNN (K nearest Neighbour)

One popular supervised machine learning method for classification and regression analysis applications is the K-Nearest Neighbors (KNN) algorithm. Supervised learning is a methodology that has proven to be efficacious. The utilization of this technique contributes to the enhancement of detection capabilities and the reduction of false-positive occurrences. In order to identify instances of fraudulent activity in credit card transactions, a supervised technique is used [14]. Two estimations are required when using the K-nearest neighbors (KNN) fraud detection approach: the distance between transaction occurrences in the dataset and the correlation of transactions. The K-Nearest Neighbors (KNN) approach is well-suited for the identification and detection of fraudulent behav-ior that occurs throughout transactional periods. The utilization of over-sampling techniques and data separation has the potential to facilitate the identification of abnormalities within the target variables. Hence, it may be deemed suitable for consideration in the context of CCFD with regards to memory constraints. It has the potential to facilitate CCFD by using limited memory and computational re-sources. The proposed method offers a more expedient strategy for processing datasets of varying sizes. When compared to other anomaly-based approaches, KNN demonstrates superior accuracy and efficiency [12]. The utilization of this method is prevalent in the identification of like patterns seen in past transactions conducted by the individual possessing the card. A few machine learning techniques that are frequently used are K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression (LR). The K-Nearest Neighbors (KNN) method exhibits a high accuracy rate of 97.69% in accurately identifying fraudulent transactions in the context of credit card detection, according the referenced source [13]. The system has achieved optimal performance. The efficiency of KNN has been demonstrated in terms of all metrics employed, as it exhibited a lack of false-positive classifications. A further investigation was conducted utilizing the K-Nearest Neighbors (KNN) algorithm, resulting in a 72% accuracy rate for the Cross-Correlation Feature Detection (CCFD) technique [12]. While the authors performed iterative experiments using the K-Nearest Neighbors (KNN) method, it is important to acknowledge the inherent limitations of this approach. The K-nearest neighbors (KNN) technique is known for its high memory requirements, since it tends to magnify non-essential attributes of the data. Similarly, it exhibits deficiencies in the aforementioned trials. The performance of the K-nearest neighbors (KNN) method deteriorates when it is provided with a substantial vol-ume of data. Consequently, these constraints exert an influence on the accuracy and recall metrics inside the CCFD (Constraint-based Collaborative Filtering with Diversity) process.

3.7 Privacy preserving Technique

In the context of machine learning, the process of training a dataset is of ut-most importance. In order to achieve effective training, it is necessary to supply machine learning algorithms with a substantial amount of data. Numerous studies have been conducted with credit card data in a manner that ensures privacy preservation. One of the tests employed the supervised machine learning tech-nique in conjunction with blockchain technology. The technology was employed within the Ethereum network, and the operation was conducted across a total of 300,000 individual accounts. The findings indicate that modifying parameters has an impact on the accuracy and recall values. Furthermore, it has been recog-nized that the utilization of blockchain technology may pose a potential hazard due to its decentralized nature [53]. Nevertheless, the decentralised structure of blockchain technology renders it an efficient means of safeguarding data privacy. However, the utilization of decentralized technology in real-world applications for CCFD is accompanied by several restrictions. These limits encompass scala-bility concerns and the challenge of securely storing data within wallets. Addi-tionally, this process requires a significant amount of computational resources, resulting in increased energy consumption. Consequently, it is a costly approach, and its use is not universally standardized. Therefore, it might not be the best idea for banks and other financial institutions to use blockchain technology for CCFD. The General Data Protection Regulation's (GDPR) rules must be followed while using data for experimental purposes. The research was conducted using the gossip learning and federated learning approaches. The ineffectiveness of gossip learning approaches has been at-tributed to the absence of a central control structure. In contrast, F.L. has demon-strated superior performance due to its semi-decentralized character (52, 54).

3.8 Block Chain Technology

Several applications utilizing blockchain technology have garnered significant public interest. This phenomenon is predicated upon the premise that it trans-cends the confines of centralized servers, such as those employed by financial institutions and other organizations. Instead, it offers a decentralized model in which user behavior is contingent upon the characteristics of Blockchain tech-nology. Malicious software has the potential to instigate fraudulent activities within blockchain transactions. The authors Michal et al. (2019) have introduced a supervised appareil learning methodology within the context of blockchain technology (56). The approach has been employed by the authors on the Ethere-um network.

A sample size of 300,000 accounts was used for the experiment. The results were compared to the outcomes of random forest, support vector machines (SVM), and XGBoost algorithms. [57]. The ex-periment's findings indicate that the accuracy and recall values are influenced by different transaction settings. Additionally, it has been proposed that Blockchain is an autonomous technology. The dependence on this particular factor may provide a significant risk, particularly within the realm of the financial industry. Hence, our study is grounded in a pragmatic methodology utilizing federated learning, a semi-decentralized strategy that concurrently guarantees efficacy and privacy.

Table 2 : Comparision of Previous work

References	Methods	Data Set	Advantages	Disadvantage	Benefit
[4]	Deep learning _logistic regression	Nigiria Bank	It improve real life entries of transaction	Its does not work if the transaction not based on the real time	95% for detecting fraudulent transaction
[17]	ANN	Uci websites	Effective result once ANN	Time - consuming when trained with aneelaing simulation	92% for detecting fraudulent transaction
[23]	KNN NB LR DT CF LAN	European card holder	CFL-ANN Redue mean square error	KNN is time consuming	97.56% fraudlent detection
[15]	RF -1 RTRF, RF-2 CRF	Chinese E commerce Firm	R-F performes well other DT algorithm	Data imbalance is lacking	97.77 Accuracy 89.46 precision
[18]	Support vector machine, neural network	Chinese financial institutions	Overall performance of CCFD is enhanced	Time consuming process	99.21% accuracy Recall of 95.20%
[19]	Random forest, logistics regression, SVM, DT.KNN	Credit card from European Data set	Classification method produced higher accuracy in prediction	It requires classification of anomalies earlier	Overall accuracy is achieved higher in contrast to random forest

4. Model Designing

Detecting credit card fraud is still a major problem for the financial sector. This study presents a novel method that combines the power of Adaptive Boosting (ADABoost), Federated Learning, and Synthetic Minority Over-sampling Technique (SMOTE) to improve credit card fraud detection. By leveraging federated learning, this model addresses data privacy concerns and improves fraud detection accuracy across multiple distributed data sources.

4.1 Introduction:

Credit card fraud continues to evolve in complexity, necessitating advanced detection mechanisms. Traditional methods may struggle with imbalanced data and privacy concerns related to centralized data processing. To address these challenges, this paper proposes a Federated Learning-based approach that incorporates SMOTE and ADABoost.

4.2 Data Federation and Privacy:

Federated Learning maintains sensitive data local while facilitating model training across decentralized data sources. Participating institutions maintain control over their data, contributing to a global model without sharing raw data. This approach ensures privacy compliance and promotes collaborative fraud detection.

4.3 Data Preprocessing:

The credit card transaction data at each participating institution is often imbalanced, posing challenges for fraud detection. SMOTE is employed locally on each institution's dataset to oversample the minority class, addressing the class imbalance issue.

4.4 Feature Engineering and Selection:

Feature engineering and selection are essential for accurate fraud detection. Relevant features are extracted from local data, and collaborative feature selection techniques are applied to construct a comprehensive feature set for the federated model.

4.5 Proposed Model Architecture:

The proposed model leverages federated learning to create a collaborative fraud detection model. The architecture includes the following steps:Fig. 3 presents the envisioned architecture for a proposed model, illustrating the structural framework. This visual likely outlines the components, connections, and flow of information within the model, aiding in understanding its design and functionality at a glance. The diagram assists in conveying the conceptual organization of the proposed model's key elements.

a. Data Localization: Each participating institution applies SMOTE and prepares its data for training.

b. Secure Aggregation: Encrypted model updates are aggregated securely across institutions, preserving privacy.

c. Global Model: The aggregated updates form a global model, benefiting from insights across all data sources.

d. ADABoost Integration: The global model is incorporated into the ADABoost algorithm for boosted classification performance.

e. Model Distribution: The enhanced global model is securely distributed back to participating institutions.

The results of proposed Federated Learning-based model will , combined with SMOTE and ADA Boost, effectively enhances fraud detection accuracy while ensuring data privacy. The collaborative nature of the model promotes better generalization to diverse fraud patterns.

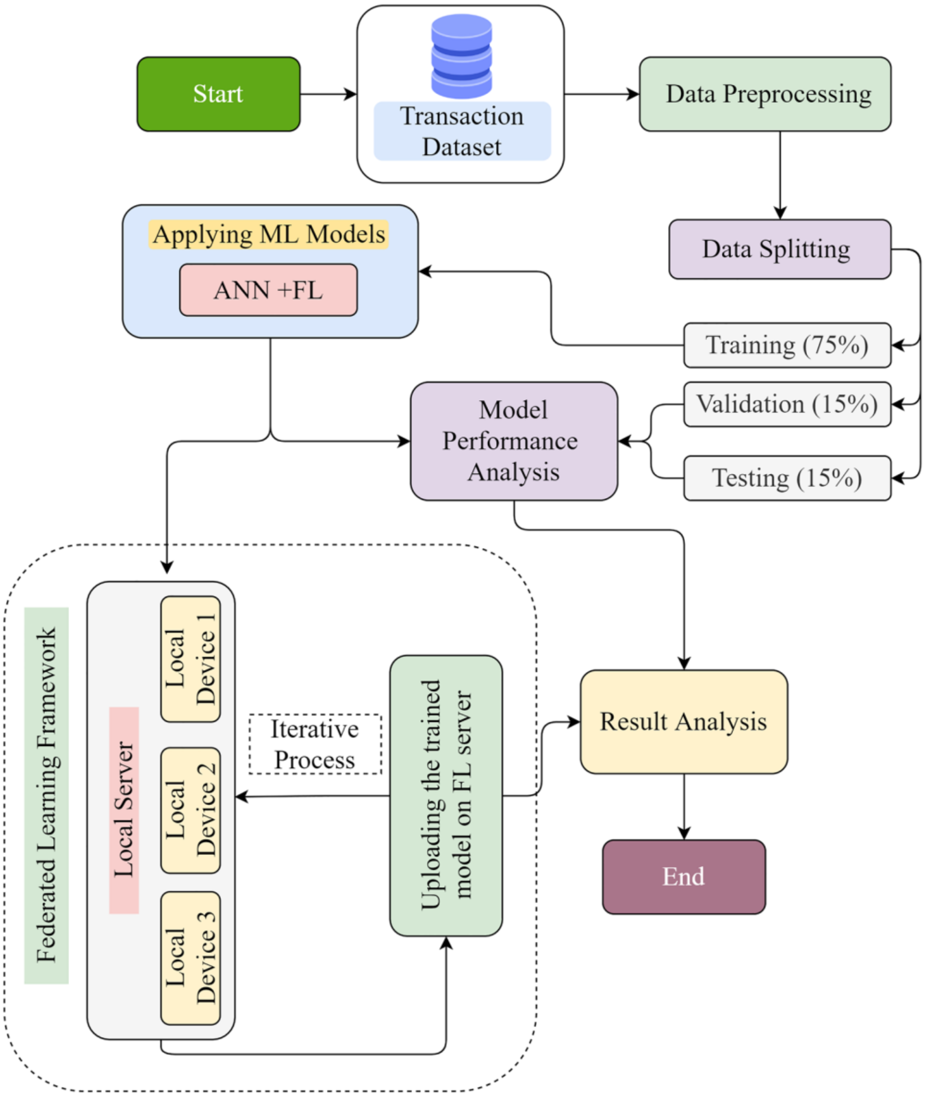

Fig 3. Proposed model Architecture

5 Conclusion

This review study investigates the many CCFD strategies that have been employed. It is clear from analysis that using ML approaches to improve CCFD accuracy is an excellent idea. Large datasets are necessary for model training, nevertheless, in order to prevent data imbalance. We can obtain a wider range of data by utilizing real-time information, while privacy concerns still need to be addressed. Our suggested approach allows us to train the model in a privacy-preserving way using the real-time datasets. The ability of the machine learning model to identify fraudulent transactions can be improved by using an ANN in a federated learning framework. With the help of real-world statistics, the hybrid strategy that has been suggested can effectively change the way CCFD is done while opening up new possibilities for the banking and finance sectors. By working together, the suggested approach can assist banks and financial institutions in making use of real-time datasets, which will benefit everyone in the process of creating a successful CCFD system. While the suggested approach uses the real-time datasets in a privacy-preserving manner and is successful in terms of CCFD, it has drawbacks when it comes to practical implementation. Every bank and other financial institution has its own set of guidelines, and they are very stringent about following them. It will be difficult to modify the suggested approach because banks and other financial institutions have their own constraints and depend more on internal resources than on a centralized strategy. Despite the lack of central data sharing, even the trained model will eventually pick up patterns that hackers might be able to decipher. Therefore, work still needs to be done to win over banks and other financial institutions to the adoption of this technology, even with the limits in place.

References

[1] Lucas Y, Portier P-E, Laporte L, et al. Multiple perspectives HMM-based feature engineering for credit card fraud detection. In: ACM, 2019. p. 1359–1361.

[2] Duman E, Elikucuk I. Solving credit card fraud detection problem by the new metaheuristics migrating birds optimization. Berlin: Springer; 2013.

[3] Botchey FE, Qin Z, Hughes-Lartey K. Mobile money fraud prediction— a cross-case analysis on the efficiency of support vector machines, gradient boosted decision trees, and Naïve Bayes algorithms. Information. 2020;11:383. https:// doi. org/ 10. 3390/ info1 10803 83.

[4] Ogwueleka FN. Data mining application in credit card fraud detection system. J Eng Sci Technol. 2011;6:311–22.

Sriram Sasank JVV, Sahith GR, Abhinav K, Belwal M. Credit Card fraud detection using various classification and sampling techniques: a comparative study. In: IEEE, 2019. p. 1713–1718.

[5] Ojugo AA, Nwankwo O. Spectral-cluster solution for credit-card fraud detection using a genetic algorithm trained modular deep learning neural network. JINAV J Inf Vis. 2021;2:15–24. https:// doi. org/ 10. 35877/ 454RI. jinav 274.

[6] Majhi SK, Bhatachharya S, Pradhan R, Biswal S. Fuzzy clustering using SALP swarm algorithm for automobile insurance fraud detection. J Intell Fuzzy Syst. 2019;36:2333–44. https:// doi. org/ 10. 3233/ JIFS- 169944.

[7] Darwish SM. An intelligent credit card fraud detection approach based on semantic fusion of two classifiers. Soft Comput. 2019;24:1243–53. https:// doi. org/ 10. 1007/ s00500- 019- 03958-9.

[8] C. H. Sumanth, P. P. Kalyan, B. Ravi and S. Balasubramani., "Analysis of Credit Card Fraud Detection using Machine Learning Techniques," *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2022,pp.1140-1144.doi:10.1109/ICCES54183.2022.9835751

[9] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," in *IEEE Access*, vol. 10, pp. 39700-39715, 2022. doi:10.1109/ACCESS.2022.3166891

[10] J. C. Mathew, B. Nithya, C. R. Vishwanatha, P. Shetty, H. Priya and G. Kavya, "An Analysis on Fraud Detection in Credit Card Transactions using Machine Learning Techniques," *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, 2022, pp. 265-272. doi:10.1109/ICAIS53314.2022.9742830

[11] J. B, J. A. K. R and D. P. S. Ganesh, "Credit Card Fraud Detection with Unbalanced Real and Synthetic dataset using Machine Learning models," *2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC)*, Chennai, India, 2022, pp. 73-78. doi:10.1109/ICESIC53714.2022.9783529

[12] D. Tanouz, R. R. Subramanian, D. Eswar, G. V. P. Reddy, A. R. Kumar and C. V. N. M. Praneeth, "Credit Card Fraud Detection Using Machine Learning," *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2021, pp. 967-972. doi: 10.1109/ICICCS51141.2021.9432308

[13] A. S. Rathore, A. Kumar, D. Tomar, V. Goyal, K. Sarda and D. Vij, "Credit Card Fraud Detection using Machine Learning," *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, MORADABAD, India, 2021, pp. 167-171. doi: 10.1109/SMART52563.2021.9676262

[14]. Vynokurova O, Peleshko D, Bondarenko O, Ilyasov V, Serzhantov V, Peleshko M. Hybrid machine learning system for solving fraud detection tasks. In: 2020 IEEE third international conference on data stream mining & processing (DSMP), IEEE; 2020. p. 1–5.

[15] Rai AK, Dwivedi RK. Fraud detection in credit card data usingunsupervised machine learning based scheme. In: IEEE, 2020. p. 421–426.

[16] Dubey SC, Mundhe KS, Kadam AA. Credit card fraud detection using artificial neural network and back propagation. In: 2020 4th international conference on intelligent computing and control systems (ICICCS). IEEE; 2020. p. 268–273.

[17] Patidar R, Sharma L. Credit card fraud detection using neuranetwork. Int J Soft Comput Eng (IJSCE), 2011;1(32–38).

[18] Dhankhad S, Mohammed E, Far B. Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In: IEEE, 2018. p. 122–125.