# Credit Card Fraud Detection Using Machine Learning Algorithms

**Preprint** · December 2023

DOI: 10.13140/RG.2.2.14806.63044

**3 authors:**

Unam Sam
Bournemouth University
**1** PUBLICATION **1** CITATION

SEE PROFILE

Godfrey Moses
Bournemouth University
**1** PUBLICATION **1** CITATION

SEE PROFILE

Taiwo Olajide
Bournemouth University
**1** PUBLICATION **1** CITATION

SEE PROFILE

# *Credit Card Fraud Detection Using Machine Learning Algorithms*

Unam Sam
Computing and Informatics
Bournemouth University
Bournemouth, United Kingdom
s5557882@bournemouth.ac.uk
https://orcid.org/0009-0000-4148-7116

Godfrey Moses
Computing and Informatics
Bournemouth University
Bournemouth, United Kingdom
s5560990@bournemouth.ac.uk
https://orcid.org/0009-0004-2107-4786

Taiwo Olajide
Computing and Informatics
Bournemouth University
Bournemouth, United Kingdom
s5519082@bournemouth.ac.uk
https://orcid.org/0009-0009-5580-0439

***Abstract*— This comprehensive review paper investigates the current status of credit card fraud detection by using both traditional and advanced machine learning techniques. The text presents a range of methods, each with its own advantages and disadvantages. These methods include Decision Trees (DT), Logistic Regression (LR), K-Nearest Neighbours (KNN), Neural Networks (NN), Naive Bayes (NB), Genetic Algorithms (GA), Hidden Markov Models (HMM), Support Vector Machines (SVM), Fuzzy Logic-based Systems (FLBS), Hybrid Approaches, and Privacy-preserving Techniques. DT sacrifice generalisation capacity in exchange for interpretability, making them prone to overfitting. On the other hand, LR's performance is hindered by its susceptibility to outliers. Although NN excel at detecting intricate patterns, they may be relatively demanding in terms of computational resources. The accuracy of the speed-loving NB model is compromised by the lack of feature independence. These many techniques provide complex trade-offs in terms of accuracy, interpretability, scalability, and privacy concerns when strategically integrated into real-world financial security frameworks.**

***Keywords—credit card fraud, machine learning Support Vector Machines***

## I. Introduction

The use of credit cards and the advent of online purchasing have significantly facilitated the lives of both consumers and retailers [1]. Regrettably, the advent of the digital revolution has seen a significant surge in instances of credit card theft. Credit card fraud is a significant challenge for financial institutions and individuals worldwide, including unauthorised transactions, identity theft, and account hijacking [2]. Credit card fraud is a pressing issue for effective remedies, given the financial ramifications and the erosion of trust in digital payment mechanisms.

The detection of fraud has become inadequate with the rise of intricate fraudulent schemes, rendering rule-based systems and human evaluations insufficient [3]. Manual assessments are characterised by their time-consuming nature, high costs, and susceptibility to human error. Conversely, rule-based systems sometimes lack the necessary adaptability to effectively address emerging fraud tendencies. The emergence of machine learning (ML) algorithms has prompted the banking industry to explore more sophisticated and automated approaches to fraud detection. This review paper assesses the current state of credit card fraud detection (CCFD) with ML algorithms, by critically analysing the effectiveness, advantages, and limitations of various strategies.

## II. LITERATURE REVIEW

The issue of credit card fraud remains prevalent, hence requiring the development of novel and enhanced techniques for its detection [4]. This section will provide an overview of the key findings and observations derived from extensive study on the detection of credit card fraud, including both traditional and ML methodologies.

### *A. Credit Card Fraud Detection Before Machine Learning*

In the past, the area of CCFD was mostly influenced by rule-based and heuristic-driven methods, until the emergence of ML techniques [5]. Although these methodologies proved to be beneficial, they were not devoid of limitations and often exhibited inconsistent performance in identifying fraudulent transactions.

#### *1) Traditional Approaches*

1. Rule-Based Systems: The development of rule-based systems was an early technique for identifying fraudulent behaviour. These programmes used thresholds and specified criteria to identify potentially fraudulent financial dealings. A rule may, for instance, issue a warning if an international transaction followed a domestic one by less than a certain amount of time. While rule-based systems were easy to understand and deploy, they were not flexible enough to accommodate changing fraud schemes [6].
2. Heuristic-Based Approaches: Heuristic-based strategies were used, drawing upon expert knowledge, to formulate rules and tactics with the purpose of detecting instances of fraud [7]. According to Adebayo et al. [8], the efficacy of these methodologies was notable in detecting prevalent fraudulent patterns, although their effectiveness was limited when faced with intricate and innovative forms of fraud. In addition, it is worth noting that manual reviews, which have traditionally been used for fraud detection, have proven to be both time-consuming and labor-intensive. This is primarily due to the fact that these evaluations need ongoing modifications to the heuristics, relying heavily on human expertise [9], [10].
3. Threshold-Based Alerts: The threshold system imposes limitations on the magnitude or frequency of transactions. Notifications are sent when financial transactions exceeded certain thresholds. Nevertheless, a significant drawback of these approaches was the considerable quantity of false positives they generated. For example, reducing the barrier has the potential to expose a greater number of fraudulent activities; yet, it also has the potential to uncover legitimate transactions of considerable value [11].

### 2) *Limitations of Pre- Machine Learning Approaches*

Despite their historical importance in fraud detection, these traditional methods had notable limitations as mentioned by Al Smadi et al. [12]:

1. Lack of Adaptability: Conventional methods had difficulties in effectively adjusting to the dynamic nature of evolving fraud practises. Once those engaging in fraudulent activities were familiar with the established regulations and limitations, they could readily adapt their strategies in order to circumvent them.
2. High False Positives: Threshold-based systems often generate a substantial quantity of false positives, resulting in operational inefficiencies due to the need of manually reviewing several genuine transactions.
3. Resource-Intensive: Heuristic-based methodologies heavily depended on human knowledge and incurred significant costs for maintenance, necessitating regular revisions to ensure continued efficacy.
4. Inability to Detect Complex Patterns: Conventional approaches demonstrated efficacy in detecting uncomplicated instances of fraud, although encountered challenges in recognising intricate and dynamic patterns.

### 3) *Fraud Prevention and Mitigation*

Prior to the emergence of ML, endeavours to address credit card fraud mostly centred on a blend of conventional methodologies, such as fraud investigation units, and educational initiatives aimed at cardholders [13]. Financial institutions have developed specialised teams to conduct investigations into transactions and patterns that are deemed suspicious. These teams often collaborate closely with law enforcement organisations in their efforts [14]. Furthermore, the dissemination of information to cardholders on secure practises and the active monitoring of their own transactions played a pivotal part in the prevention of fraudulent activities. Cardholders are advised to instantly report any suspicious activity, and several banking institutions offered online transaction monitoring tools [15].

## B. *Machine Learning Algorithms in Credit Card Fraud Detection*

The use of ML algorithms for the purpose of CCFD has seen a surge in popularity in recent years, as shown by the work of Dornadula and Geetha [16]. This approach is favoured for its ability to provide reliable and adaptable solutions. These algorithms use the cognitive abilities of data-driven decision making in order to detect intricate patterns of fraudulent activities [17]. LR, DT, RF, SVM, neural networks, and deep learning models represent a subset of the methodologies used in the field.

### 1) *Logistic Regression (LR)*

As stated by Mohammed and Maram [18], LR is a popular approach for performing binary classification problems. Concurrently, there has been an upsurge in fraudulent activities due to the modern trend of using credit cards as the principal payment method, which has highlighted the need for strong fraud detection algorithms. LR ease of use and comprehension make it a popular choice. Credit card fraud may be detected using transaction characteristics such as amount, location, and merchant type, and LR has been used successfully to do so. Despite its benefits, LR may fail to detect non-linear, intricate fraud schemes. In light of recent technical developments, Tanouz et al. [19] investigate the critical necessity of a reliable fraud detection system. The paper suggests using LR, random forest, and Naive Bayes, among other ML-based classification methods, to address the dataset imbalance that is common in credit card fraud situations. To tackle the growing problem of credit card fraud, it is crucial to use ML techniques. Important evaluation metrics for these algorithms include recall, accuracy, precision, F1 score, confusion matrix, and Roc-auc score [20]. The need of applying sophisticated ML approaches to reduce fraudulent credit card activity is highlighted by the study, which highlights the need for an all-encompassing assessment methodology that focuses on algorithmic performance using key criteria. Similarly, Rathore et al. [20] investigates the use of Data Science and ML to combat the widespread problem of credit card fraud, which is becoming more common. Within the setting of highly unbalanced datasets common in credit card transaction data, the study examines four well-known ML algorithms: Decision Tree, Random Forest, K-nearest neighbours, and LR. Time, place, purchase type, value, seller, and consumer preferences are some of the transactional factors covered by this study. These varied data variables are input into models that assess the possibility of fraud in transactions by use of statistical methods [21]. The importance of using well-established ML methodologies to fight credit card fraud is highlighted in this review. It emphasises that thorough model comparisons and extensive transactional data are essential for successful research in this field. The importance of ML approaches, especially LR and other methodologies, in solving the difficult problems of CCFD has been emphasised in these works. Building reliable and efficient fraud detection systems relies heavily on these approaches, which place a focus on assessing performance indicators and using varied datasets.

### 2) *Decision Trees (DT) and Random Forests (RF)*

The handling of non-linear data poses no significant challenges for DT and RF, as shown by Zhang [22]. As an ensemble learning technique that uses several DT, RF is well-known for its ability to withstand data noise and overfitting. Due to its independent tree creation as shown in figure 1.0, RF is computationally efficient, which highlights its practical value in different sectors, particularly in CCFD. Because it is an ensemble approach, RF uses the combined wisdom of many DT. The combination of many distinct trees, each trained on a different sample of the data, gives it resilience against overfitting. In the context of identifying fraudulent credit card transactions in the face of noise and changing trends, this property gives RF the intrinsic potential to generalise successfully to unknown data cases. The ensemble nature and multiplicity of component trees make RF interpretability a difficulty, despite its usefulness. While its interpretative capability is limited compared to simpler models, its efficacy in reliably recognising fraudulent transactions is not diminished by its intrinsic complexity. Multiple high-quality research back up the claim that RF is effective in detecting credit card fraud. As an example, Bhattacharyya et al. [23] used actual data from a global credit card operation to do a study. Their results demonstrated that RF outperformed other methods in detecting fraud with a better success rate and fewer false positives. To address the problem of idea drift in credit card transaction data, Dal Pozzolo et al. [24] investigated RF-based models. Their experiments using real-world datasets showed significant improvements in warning accuracy, offering a solid method for adapting to changing consumer habits. Further, research has compared RF to other methods, including neural networks, SVMs, LRs, and KNNs; for example, Dal Pozzolo et al. [25], Christopher et al. [26], and V´eronique et al. [27]. All things considered, these studies prove that RF is the best option when it comes to accuracy, AUC, and predictive power. And that's across a wide range of datasets and classification circumstances.

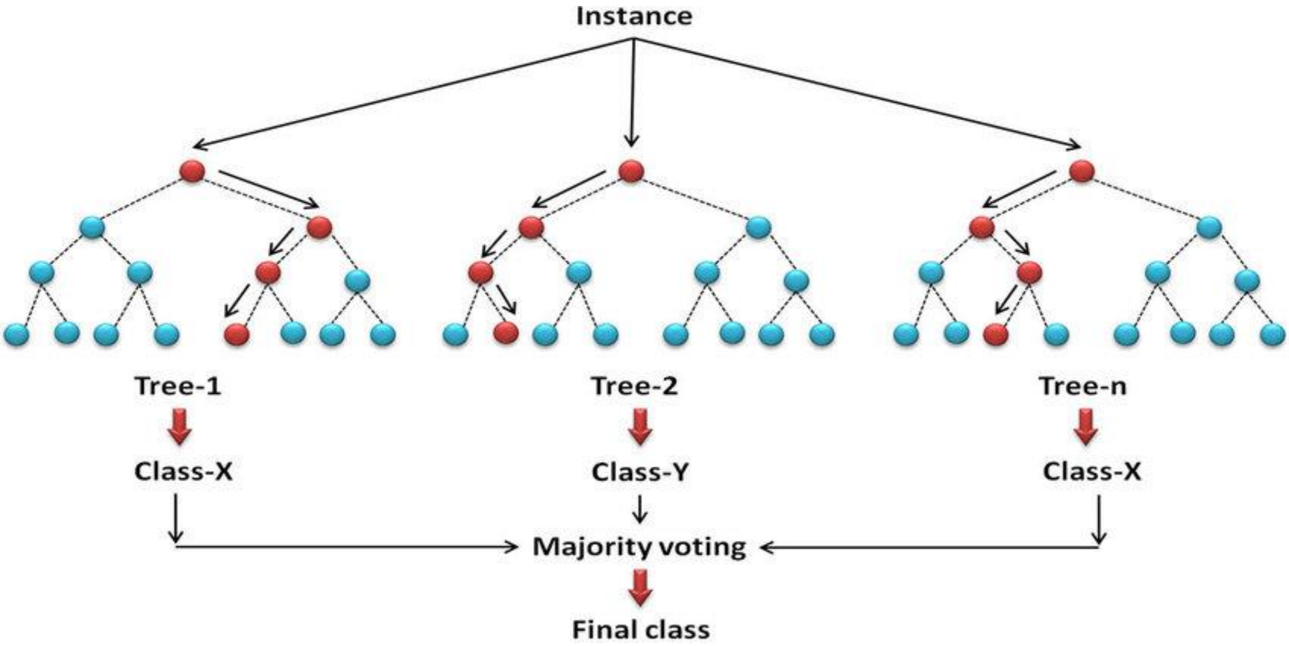

Figure 1.0: Framework of RF [28]

While, the introduction of decision tree algorithms like C4.5 and ID3 marked a significant milestone in the development of learning systems by enabling the handling of continuous data. To effectively analyse issues and develop accurate DT, these methods arrange nodes into branches and leaves that lack offspring. This research emphasises the significance of DT by demonstrating their basic function in mitigating complex problems. It recognises the groundbreaking accomplishments of C4.5 and ID3. In their study, Gedela et al. (2022) used the AdaBoost classifier to address the issue of credit card theft often faced by banks and other financial institutions. The research shown impressive levels of accuracy, with AdaBoost achieving a rate of 99.43% and DT achieving a rate of 94.8%. AdaBoost also revealed a noteworthy f-score of 99.48% when compared to LR, artificial neural networks, DT, and Naive Bayes using a dataset consisting of 284,807 transactions [29]. According to qualitative study, AdaBoost outperforms other methods in terms of accuracy when it comes to identifying fraud. According to Mathaw et al. [30], there has been an increase in credit card theft due to the significant growth of online commercial transactions. In their study on fraud detection, they emphasise the exceptional precision and satisfactory F-scores achieved by ML methods such as DT and RF. The significance of advanced ML techniques in addressing online credit card theft is emphasised by the fact that feature selection is a crucial step in enhancing model accuracy [31]. In light of the extensive use of credit cards for both online and offline transactions, Ganesh et al. [29] undertakes a study on the persistent rise in credit card fraud. The study investigates both traditional ML techniques and those specifically designed to address imbalanced data. Various supervised ML models are evaluated using simulated transaction data. The models encompass Support Vector Machine, Random Forest, Decision Tree, and LR. This research emphasises the need of robust ML techniques in detecting credit card fraud in situations when transaction data is imbalanced. DT are adaptable and may be utilised in many circumstances since they are not reliant on any one characteristic. Kokinaki [32] demonstrated the creation of similarity trees, which consist of nodes representing shared attribute values and linked by edges. However, manual verification is one of the limitations. Significant progress has been made in the field of intrusion detection systems, namely in the use of DT, especially inductive DT. This synthesis focuses on the advantageous characteristics of DT in addressing issues within intrusion detection systems and similarity trees, as shown by prior research [33]. This investigation emphasises the vital importance of DT in fraud detection, showcasing their adaptability and advantages in handling complex data structures. Furthermore, it tackles significant constraints in other domains, including intrusion detection and fraud detection. Researchers have shown that these algorithms exhibit a higher level of proficiency in detecting intricate fraudulent schemes. However, it is possible for these models to overfit the data, hence limiting their capacity to generalise.

### 3) *K Nearest Neighbour (KNN)*

KNN is a well-regarded supervised ML technique that is very effective in identifying instances of fraudulent activity in credit card transactions. It is particularly recognised for its application in regression and classification research [21]. An effective approach for addressing fraudulent behaviour during transactions, it enhances detection capabilities while minimising false-positive incidents. However, to use the KNN approach for fraud detection, it is essential to have two key estimates: calculating the distance between each transaction in the dataset and determining the correlations between transactions. The CCFD system is well-suited for this strategy due to its capability to identify fraudulent activity while also considering limitations on memory use. Data separation and over-sampling techniques facilitate the detection of anomalies in target variables, hence enhancing the efficiency of processing datasets of varying sizes. While the KNN technique has several advantages, it also has some limitations. Challenges occur, especially when handling large data sets, because of their significant memory requirements and tendency to amplify non-essential data characteristics [19]. This limitation is inherent to the CCFD process and has the potential to impact accuracy and recall metrics. The decline in the method's effectiveness becomes more apparent when more data is introduced. The high accuracy rate of KNN in CCFD, as shown by its reported 97.69% success rate in identifying fraudulent transactions [20], positions it as a valuable tool when combined with other notable ML techniques such as Naive Bayes and LR. However, doing a more in-depth analysis of the Cross-Correlation Feature Detection (CCFD) technique with the implementation of KNN)resulted in a 72% accuracy rate. This emphasises the significance of context and the constraints associated with this specific approach [19]. Although KNN is often efficient and reliable, it is important to carefully evaluate its limitations when working with huge datasets and its tendency to prioritise non-essential attributes. To effectively include KNN into the CCFD process, it is crucial to recognise and address the underlying limitations and challenges. Addressing these limitations has the capacity to enhance the utilisation of KNN in CCFD, making it more precise and adaptable to a broader spectrum of data scenarios.

### *4) Support Vector Machines (SVM)*

SVM has shown to be an important tool for regression and classification research, particularly in the field of CCFD [34]. Studying intricate patterns in the use of customers' credit cards is a frequently explored area of study. SVM techniques aid in the classification of consumer behaviours as either fraudulent or legal by examining payment patterns extracted from datasets. SVM has exceptional performance and yields dependable classification outcomes when just a limited number of attributes are extracted from the dataset [35]. However, difficulties arise when dealing with datasets that are too large, often including over 100,000 entries. SVM's performance is diminished and it becomes inefficient when dealing with large datasets because to the computational burden it imposes. This is especially true when considering real-time applications. Rtayli et al. [36] proposed a hybrid technique that combines Random Forest Classifier (RFC) and SVM methods to overcome the limitations of SVM in handling imbalanced and high-dimensional datasets. The purpose of this hybrid technique is to enhance the efficiency of fraud detection by tackling the challenges of selecting relevant features in large, imbalanced datasets with few instances of fraudulent transactions. The evaluation of this hybrid model included the use of metrics such as area under the curve, recall, and accuracy. Through the integration of RFC with SVM, we successfully attained a remarkable accuracy rate of 95%. This integration notably reduced the occurrence of false-positive transactions and enhanced sensitivity to 87%. By integrating various approaches, there is the possibility of enhancing the effectiveness of fraud detection, especially when dealing with extensive and imbalanced datasets [36]. While this model has exceptional performance in fraud detection, it presents privacy concerns when evaluating metrics like as recall and accuracy, which directly impact the security of financial transactions. A potential way to address these privacy concerns is the implementation of a federated learning framework. This approach would allow for localised data training and the use of ANN to enhance the accuracy of categorization. It is important to mention that the RFC algorithm may exhibit good performance with small datasets, but it has difficulties in scaling and achieving satisfactory results with larger datasets. Consequently, the continued challenge is in identifying a resolution that achieves a favourable equilibrium between precision and operational effectiveness when managing substantial volumes of credit card transaction data. In summary, SVM perform well in credit card transaction classifications when the number of features is limited, but their performance deteriorates when applied to larger datasets. Further investigation is crucial to explore privacy-preserving models like federated learning and other strategies to enhance the accuracy and data privacy in CCFD. The integration of the RFC with SVM is a potential solution for addressing unbalanced datasets. Researchers have made good use of its capacity to locate optimum decision boundaries in high-dimensional regions [16], [37]. The effectiveness of SVM may be tweaked with the use of a customised kernel and hyperparameter settings.

### *5) Naive Bayes (NB)*

NB classifiers use probabilistic classifiers that rely on Bayes conditional probability to categorise data into their most likely classes. This method is often used in the identification and prevention of fraudulent activities. The classifiers are appealing because to their efficacy and interpretability, especially when working with input data that has a high level of dimensionality. Their efficacy in intricate decision-making is heightened as they facilitate the integration of expert knowledge into ambiguous statements. The presumption of conditional independence among the features in the dataset, however, may significantly diminish their predictive efficacy. When confronted with duplicate attributes, this assumption often results in reduced precision. Mohammed et al. [38] conducted a study to evaluate the effectiveness of the NB classifier, among other ML algorithms, in detecting credit card fraud in large and imbalanced datasets. The research showcased the comparative efficiency of the NB approach in identifying fraud, as compared to the RF and Balanced Bagging Ensemble (BBE) classifier, using real-world datasets. However, a notable disadvantage arose due to an imprecise accuracy rate, leading to many false alarms and reducing its effectiveness in preventing fraudulent activities. In a similar manner, Mahmud et al. [39] examined the performance of many ML algorithms in identifying instances of credit card fraud. The research found that DT-based models had greater classification accuracy compared to NB techniques, as measured by metrics such as classification accuracy and fraud detection rate. Bahnsen et al. [40] developed a Bayes minimum risk technique to address the financial consequences of CCFD by include the real monetary expenditures associated with it. They demonstrated that the proposed framework has the potential to reduce costs compared to current approaches such as LR, DT C4.5, and RF, utilising a real-world transactional dataset.

Mahmoudi and Duman [41] conducted a study where they compared the performance of Linear Fisher discriminant analysis with NB, ANN, and DT. Their approach proved its practical advantages over conventional measures by surpassing the current cutting-edge in both classical performance metrics and conserving the overall available limit, using a real-world dataset obtained from an undisclosed Turkish bank. Collectively, these studies demonstrate the nuanced effectiveness of NB classifiers in identifying instances of credit card fraud. While their speed and interpretability are commendable, their ability to manage conditional independence of attributes is limited. Therefore, their applications need thorough consideration, especially in situations where the accurate and efficient detection of fraud is crucial.

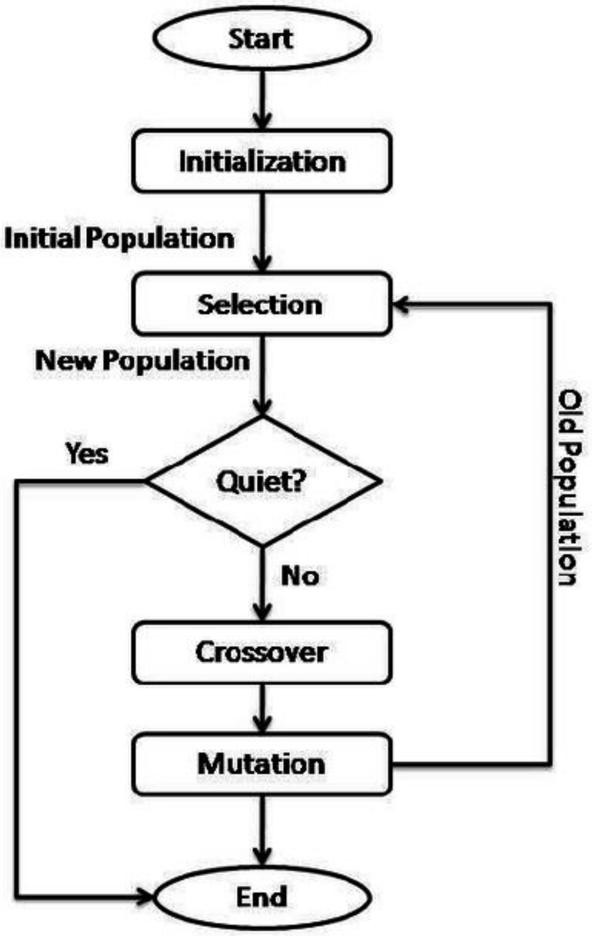

Figure 2.0: Flow chart of GA [42]

*6) GENETIC ALGORITHMS (GA)*

John Holland's [43] concept of GA was derived from natural evolutionary processes. The objective of these algorithms is to identify the optimal solution from a collection of alternatives, referred to as chromosomes, that are represented as binary strings. Its workflow is shown in figure 2.0. GA used for prediction may effectively detect potential cases of credit card fraud. By categorising credit card transactions as either suspicious or non-suspicious, these algorithms enhance security for both credit card issuers and their users. Bentley et al. [44] proposed a genetic programming solution that used a grading system to provide credit to trustworthy customers.

Their findings, derived from conducting tests using a dataset of 4,000 transactions, emphasised the significance of rules that exhibited the highest level of prediction accuracy. This approach shown superior performance compared to prior GA-based systems in terms of accuracy, particularly when compared to models that depended on user activity, such as Cha's algorithm.

Previous studies have mostly examined the accuracy of predictions (True Positive Rate - TPR) and the frequency of errors (False Negative Rate - FNR). However, Chan et al.

[45] introduced a method to anticipate suspicious activity by evaluating models using cost models. To enhance the predictive capability, Bei [46] proposed the integration of several algorithms. The articles include thorough explanations of several procedures, including diagnostic, resolution, best-match, density-selection, probabilistic-curve,

and negative-selection methods. Probabilistic and neighborhood-based algorithms have shown potential as categorization approaches, according to the results. The researchers also suggest using diagnostic algorithms to aid in decision-making when faced with uncertain situations, as well as to calculate relative risk and confidence metrics. GANNs, which integrate GAs with NN, are primarily propelled by natural selection. GANN integrates GA into the architecture of NN by incorporating NN into the DNA of the GA. This approach involves the evaluation of parameter strings by generating a pool of individuals picked at random, followed by the training of NN utilising genetic data. GANN approaches use either the GA in isolation or back-propagation training to evaluate performance and determine the optimal network topologies for improved accuracy in predicting fraud detection. In summary, GA are very effective optimisation and search tools since they emulate the mechanisms of evolution. These algorithms have been useful in enhancing precision, constructing prognostic models, and optimising decision-making procedures for the detection of credit card fraud. When used with NN in GANN strategies, they have significant potential for enhancing fraud detection techniques.

### 7) *HIDDEN MARKOV MODEL (HMM)*

The HMM governs the shifts between its restricted set of states by means of transition probabilities [47]. Every state is intricately connected to a probability distribution as shown in figure 3.0. In this framework, the probability distribution linked to certain states is used to generate probable outcomes or observations. The model, HMM, is named as such because while the visible outcomes of the data are known, the underlying states remain concealed. The HMM's inherent capacity to rapidly examine transactions and identify any fraudulent activities makes it a quick and efficient choice for CCFD. One of the key advantages of this system is its ability to effectively differentiate between legitimate and potentially fraudulent transactions, hence reducing the number of false positive transactions [45]. The HMM relies on the cardholders' spending behaviour as the main measure for identifying fraud, making it a crucial aspect of the model. By categorising these acts into several expenditure categories, often labelled as low, moderate, and high, it effectively organises individuals' spending patterns [49]. The fraud detection system, which is based on HMM, uses this classification method to detect potentially fraudulent behaviour in credit card transactions. Moreover, the HMM has a high level of proficiency in replicating the temporal patterns and dynamic sequences seen in financial data. The system successfully identifies potentially fraudulent activities by leveraging the sequential structure of the data, enabling it to identify slight variations or deviations in cardholder activity. In summary, the HMM is very efficient in identifying instances of credit card theft via the analysis of sequential data and behavioural patterns [47]. It is essential in addressing the challenges of detecting credit card fraud in real-life scenarios as it may minimise the occurrence of incorrect identifications and accurately categorise various sorts of expenses.

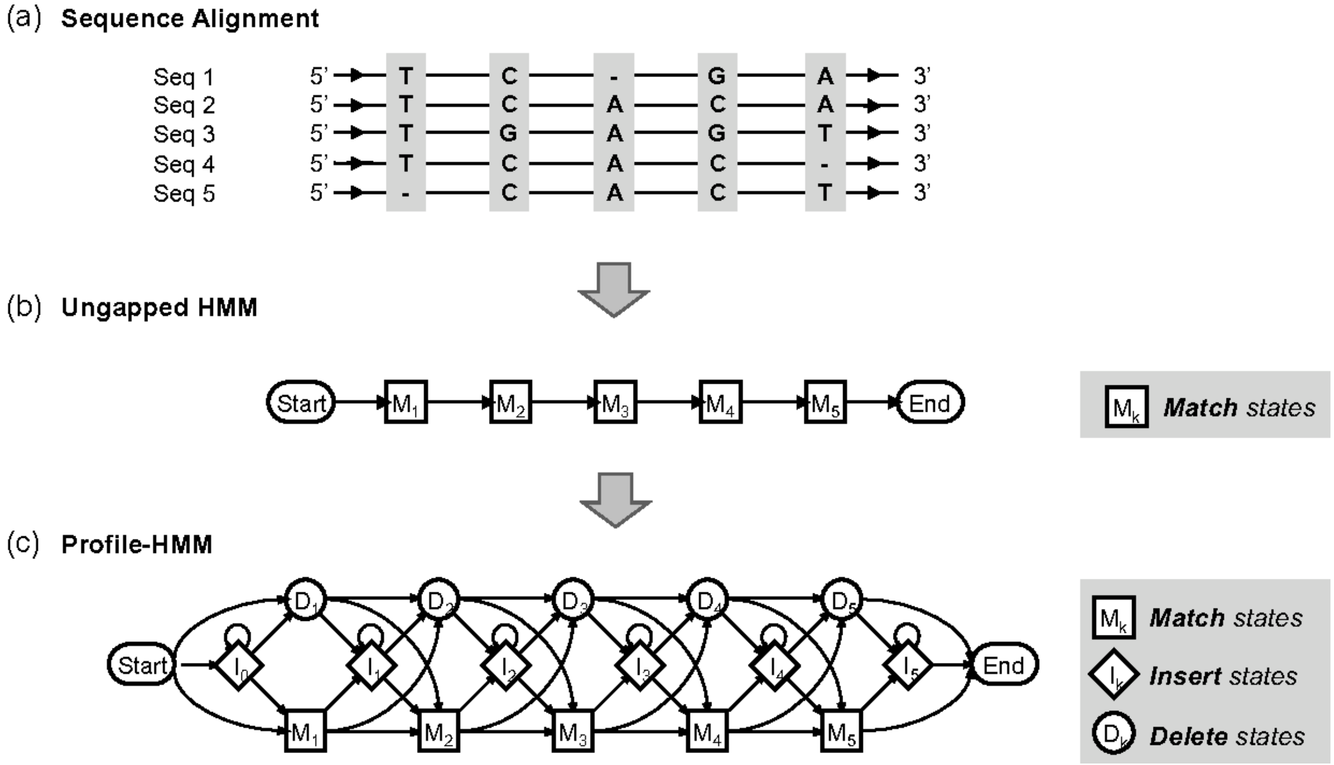

Figure 3.0: Framework of HMM [48]

### 8) *FUZZY LOGIC BASED SYSTEM (FLBS)*

Zadeh [50] first proposed the concept of FLBS in 1988. These systems provide a reliable approach to dealing with input and output variable uncertainties by making use of sets and fuzzy numbers to describe values in a way that is easy to understand. Credit card transactions may be effectively analysed for signs of fraud using this framework's detailed classification system, which places operations and transactions into low-, moderate-, or high-risk buckets. Fuzzy logic plays a crucial role in CCFD by strengthening security measures, allowing transactions to be categorised according to risk categories, and reducing the likelihood of fraudulent activity [51]. Fuzzy Neural Networks (FNN) and Fuzzy Darwinian Systems (FDS) are the two main variants of FLBS that were developed recently. These systems address different areas of fraud detection and categorization. When dealing with large datasets that include inaccurate or incomplete information, as is common in real-world applications, FNN are a potent tool to use. A system based on FNN-using knowledge discovery (FNNKD) was highlighted in an effort to speed fraud detection inside credit card transactions using Granular Neural Networks (GNN) [47], [52]. At the same time that it identifies possible fraudsters, this method greatly speeds up network training. The efficiency of FNN is shown by its ability to quickly identify fraud without sacrificing decent accuracy.

Contrarily, Fuzzy Darwinian Detection [52], which falls under the umbrella of Evolutionary-Fuzzy Systems, uses genetic programming to produce complex fuzzy logic rules. This method is great at protecting credit card information since it can distinguish between legitimate and fraudulent purchases. When a fuzzy expert system is combined with a genetic programming unit, it achieves excellent accuracy with a lower false-positive rate than other systems. Despite the impressive capabilities of FLBS in improving credit card security via detailed classification and quick fraud detection, there are a few things to keep in mind. It is necessary to do further research and optimisation on FDS in order to improve their efficiency due to the high computational intensity they may have. The use of FLBS, which include FNN and FDS, shows promise in detecting and classifying credit card fraud. They make a huge leap towards improving credit card security by quickly classifying transactions according to risk categories and handling uncertainty. For improved efficiency and broad use in financial fraud detection, however, more research into optimising these systems and reducing computing complexity is necessary.

*9) Neural Networks (NN) and Deep Learning (DL)*

The results attained by deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been notably remarkable [34]. Capturing complex patterns requires a substantial investment of effort and data. Its framework is shown in figure 4.0. In light of the increasing number of valentine scams and the dramatic increase in the use of debit cards for both in-store and online purchases, Sumanth et al. [54] set out to develop a method for accurately identifying cases of credit card fraud. The need for a large credit card dataset for testing and training has been validated by the meteoric rise of e-commerce. This study shows a comprehensive strategy that uses a deep neural network in conjunction with SVM, Deep Neural Network (DNN), and NB algorithms. The result is a system that is very good at accurately identifying credit card fraud. In a similar vein, Alarfaj et al. [55] tackle the widespread problem of credit card fraud in online purchases, which is fueled by the popularity and ease of these types of transactions. The research examines a wide variety of ML algorithms, including XG Boost, DT, RF, SVM, LR, and Extreme Learning Method, in an effort to enhance detection accuracy and alleviate fraud losses. This study thoroughly examines several CNN topologies, layering effects, and model configurations to find the best ones, highlighting the increasing importance of deep learning techniques. The suggested model does better than current ML and DL algorithms in detecting credit card fraud, with actual data showing an accuracy of 99.9%, a f1-score of 85.71%, a precision of 93%, and an AUC of 98%. In addition, the research explores ways to improve real-world efficacy and eliminate false negatives, showcasing the potential of these tactics to successfully combat credit card fraud. Credit card fraud causes enormous yearly financial losses for banks; Esenogho et al. [56] addresses this by addressing the boom in both conventional and online transactions made possible by recognising valentines. By suggesting a novel strategy, this work successfully negotiates the intricacies of biassed credit card datasets and changing purchasing behaviours. They use a hybrid data resampling approach that integrates a NN ensemble classifier with adaptive boosting (AdaBoost) with a long short-term memory (LSTM) neural network.

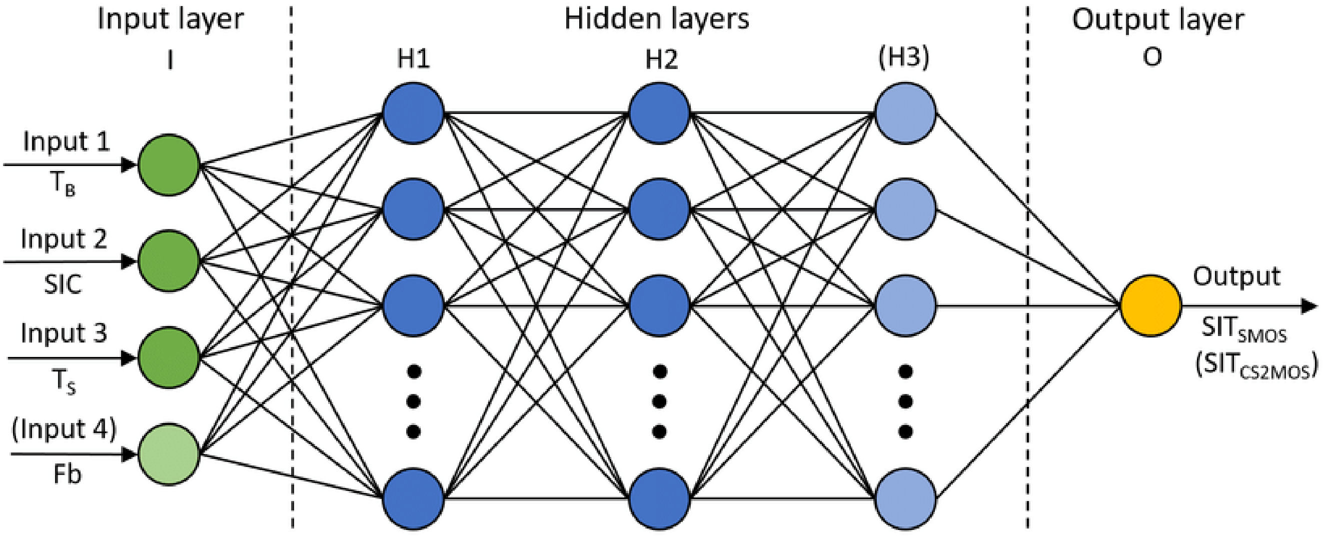

Figure 4.0: NN architecture [53]

CCFD benefits greatly from this innovative method, which is described in [30]. It uses cutting-edge strategies to deal with changing buying habits and dataset biases. Taken as a whole, these studies highlight how deep learning algorithms, ensemble classifiers, and innovative data strategies can fight against financial transaction fraud, and how important advanced neural network techniques are for detecting credit card fraud. ANNs are well recognised and often used technology in the field of supervised CCFD. The ability to address intricate patterns inherent in fraudulent conduct is a notable advantage of ANNs, which are well-known for their adaptability to complex data structures. However, the use of these methods is accompanied by challenges such as encountering local minima, overfitting, and noise, and their training requires a substantial amount of computational resources. A study conducted by Patidar and Sharma [57] uses NN to detect instances of credit card fraud. Their approach employs a GA to construct the network, which decides crucial characteristics such as the topology, number of hidden layers, and nodes. The objective of this strategy is to enhance the precision of fraud detection by improving the architecture of the network. Kim and Kim [58] propose a unique approach that integrates NN with a fraud density map inside an aggregated framework. In this case, the fraud density map result is combined with the fraud ratio generated by a neural classifier based on the feature vector. When tested using transaction data from a Korean credit business, our model outperforms a traditional neural network classifier. This discovery underscores the need of using other data sources to augment the precision of fraud detection.

In their pursuit of privacy-preserving techniques, Wang et al. [59] developed a distributed DNN methodology. Through this privacy-centric approach, financial institutions and other enterprises may share data without revealing any personally identifying information. Their study yields performance metrics, namely the AUC, that are comparable to the non-private reference point. This is achieved by using a real-world dataset for fraud detection, which comprises millions of transactions. This new technology ensures both data privacy and robust fraud detection capabilities. This study demonstrates the many applications of ANNs in detecting cases of credit card fraud. Demonstrating the adaptability and capacity of ANN to enhance the accuracy of fraud detection, while also considering crucial issues like privacy and data sharing, they tackle the intricacies of fraud detection by investigating different approaches, such as GA for optimising the network, combining with external density maps, and employing privacy-preserving distributed algorithms.

**Table 1.0: Advantages and Disadvantages of Different ML Approaches**

| Algorithm | Advantages | Disadvantages |
|---|---|---|
| **DT** | ▪ Easy to interpret and understand decisions.<br>▪ Handle both numerical and categorical data well.<br>▪ Automatically handle missing values and feature selection. | ▪ Prone to overfitting, especially with complex trees.<br>▪ Can create biased trees if some classes dominate the data. |
| **LR** | ▪ Simple and efficient for binary classification.<br>▪ Outputs probabilities for outcomes. | ▪ Assumes a linear relationship between features and outcomes.<br>▪ Sensitive to outliers and multicollinearity. |
| **KNN** | ▪ Intuitive and simple to implement.<br>▪ Robust to noisy data and irrelevant features. | ▪ Computationally expensive for large datasets or many features.<br>▪ Sensitive to the choice of distance metric and k value. |
| **NN** | ▪ Ability to capture complex patterns in data.<br>▪ Adaptability to diverse data types. | ▪ Prone to overfitting, especially with insufficient data.<br>▪ Computationally expensive and requires substantial computing power and time for training. |
| **ANN** | ▪ Suitable for complex pattern recognition.<br>▪ Can model nonlinear relationships in data | ▪ Prone to overfitting with large networks and training data.<br>▪ Requires a large amount of data for effective training.<br>▪ Complex architectures can be hard to interpret |
| **NB** | ▪ Simple and fast for classification tasks.<br>▪ Efficient with high-dimensional data. | ▪ Assumes independence between features, which can limit performance in cases of strong dependencies.<br>▪ Sensitive to irrelevant features and the presence of rare combinations of features in the training data. |
| **GA** | ▪ Effective in exploring large solution spaces and finding near-optimal solutions.<br>▪ Suitable for optimization problems with many parameters. | ▪ Computationally expensive for complex problems and large datasets.<br>▪ Initialization and parameter tuning can significantly impact performance. |
| **HMM** | ▪ Ability to model sequential data and hidden states.<br>▪ Efficient for time-series analysis. | ▪ Sensitive to the choice of the number of hidden states.<br>▪ Assumes Markov property (independence of future states given the present) which might not always hold true in real-world scenario. |
| **SVM** | ▪ Effective in high-dimensional spaces.<br>▪ Versatile due to different kernel functions. | ▪ Computationally intensive for large datasets.<br>▪ Difficult to interpret complex models.<br>▪ Sensitive to the choice of kernel and regularization parameters. |
| **FLBS** | ▪ Suitable for handling imprecise and uncertain data.<br>▪ Can represent linguistic terms and expert knowledge in a structured manner. | ▪ Interpretability can be challenging with complex fuzzy systems.<br>▪ Requires careful design of fuzzy rules which might need domain expertise.<br>▪ Complex systems might suffer from scalability issues and increased computational complexity. |
| **Hybrid Approach** | ▪ Combines strengths of different methods for better performance.<br>▪ Offers versatility and adaptability. | ▪ Complexity in parameter tuning and integration of multiple models.<br>▪ May suffer from interpretability issues when combining diverse methods.<br>▪ Potential scalability and computational resource issues when incorporating multiple techniques |

*10) Strengths and Weaknesses of ML Approach*

Strengths

1. High Accuracy: According to Bin Sulaiman et al. [60], the use of well trained and optimised ML algorithms can provide notable levels of precision in the detection of fraudulent transactions.
2. Real-time Detection: Ileber et al. [17] claim that the algorithms can examine transactions in real time, allowing for instantaneous intervention in the event that fraudulent conduct is uncovered.
3. Scalability: Since ML works well with huge datasets, it is a good fit for banks that process millions of transactions every day [34].

Weaknesses

1. Data Imbalance: A substantial proportion of credit card transactions, according to Pombal et al. [61], are legitimate, resulting in a disparity between social classes. Within the realm of fraud detection, this phenomenon has the potential to result in models that exhibit bias and exhibit subpar performance.
2. Overfitting: The phenomenon of overfitting may manifest in some ML algorithms when they lack regularisation techniques, as discussed by Mehta et al. [62].
3. Complexity and Interpretability: Deep learning models are often seen as "black boxes," despite their remarkable capabilities, which poses challenges in comprehending the rationale behind their decision-making [63].

### *C. Hybrid Approach*

The use of ML methods has revolutionised the CCFD environment, greatly enhancing the efficiency and effectiveness of fraud detection. A groundbreaking study developed a new approach by combining loan fraud detection methods with ML used for credit card transactions [64]. By combining data mining techniques with the Extreme Gradient Boosting (XGBoost) algorithm, this strategy achieved excellent results in CCFD. The researchers used a mixed approach that included trained and unsupervised ML algorithms to prioritise the storage of important data without direct access. Two important models, PK-XGBoost and XGBoost, were used in this hybrid framework. They displayed different performance, with PK-XGBoost significantly outperforming the regular XGBoost in terms of fraud detection efficiency [65], [66]. While guaranteeing strong user privacy, this higher performance measure greatly improved fraud detection. The privacy guarantees included into this method, however, created a bottleneck as the number of credit card transactions skyrocketed. It should be noted that in rare occasions, XGBoost overfits datasets, which means that several parameters need to be adjusted and synchronised in order to get the best possible accuracy. To further investigate the hybrid approach to CCFD, researchers used methods like RF and Isolation Forest to identify suspicious transactions [67]. The first part of this comprehensive strategy employed supervised learning to decipher abnormalities, while the second part used unsupervised learning to identify them. The methodology's effectiveness with real-world datasets proved its mastery of high-speed data [21], [68]. The detection of user geolocation was a part of this system's evaluation, which allowed for the differentiation of fraudulent transactions according to geolocation trends. But this method brought up privacy and confidentiality issues, which shows how important it is to have strict assessment standards to protect sensitive data.

Given these results, it is clear that models guaranteeing data secrecy and enhanced accuracy in CCFD, especially with bigger datasets, are urgently needed. Maintaining user privacy while implementing strong fraud detection techniques is an extremely difficult task. For credit card transaction security to be trustworthy and reliable, future research should focus on building models that can identify fraud more accurately while still protecting sensitive data.

### *D. Privacy preserving Technique*

The effectiveness of training datasets is of utmost importance in the field of ML. A large dataset is required for effective training, hence several research have used credit card data while maintaining strict privacy standards. One research analysed 300,000 user accounts using a combination of supervised ML and blockchain technologies, with a focus on the Ethereum network. The results demonstrated that changing the parameters had a substantial effect on the recall and accuracy metrics. While blockchain technology's decentralised structure effectively protects data privacy, it also adds potential risks, such as issues with scalability, higher processing resources and energy consumption, and the security of data stored in wallets [69]. These constraints make its widespread use in actual CCFD difficult and expensive for banks.

Blockchain technology's decentralised nature offers significant benefits when it comes to safeguarding data privacy. However, there are a number of limitations to its use in CCFD, therefore banks and other financial companies should proceed with care when adopting it. To guarantee adherence to stringent privacy standards, it is crucial to comply with the General Data Protection Regulation (GDPR) while using data for experimental purposes. Learning via gossip and federated learning methods have also been the subject of research. Federated learning, which is semi-decentralized, has shown better performance in protecting privacy and improving fraud detection efficiency than gossip learning, which is ineffective because it lacks a centralised control framework [70], [71].

Finding ways to keep personal information private while using CCFD is a difficult balancing act between data security and efficiency in operations. More research into blockchain technology is required to fully understand its potential for protecting data privacy, while also taking into account its limits and the expenses that come with it. As banks and other financial organisations seek for privacy-preserving methods, federated learning stands out as a strong option. It shows potential in reducing privacy issues and improving fraud detection efficiency. In order to improve operational efficiency in CCFD and bring these systems into compliance with strict privacy requirements, more research is needed.

**Table 2.0: Previous works on CCFD using ML approach**

| Authors | Title | Techniques |
|---|---|---|
| Omar et al.,[72] | Predicting fraudulent financial reporting using artificial neural network | AAN, fraud triangle theory |
| Adewumi et al.,[73] | A hybrid firefly and support vector machine classifier for phishing email detection | FFA with SVM |
| Pozzolo et al., [74] | Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy | RF |
| Gadi et al., [75] | Comparison with Parametric Optimization in Credit Card Fraud Detection | DT, BN, NB, NN, and AIS |
| Qiu, and He [76] | Machine Learning- and Evidence Theory-Based Fraud Risk Assessment of China's Box Office | Regression models. |
| Melo-Acosta et al., [77] | Fraud Detection in Big Data using Supervised and Semi-supervised Learning Techniques | Balanced RF |
| Wei et al., [78] | Effective detection of sophisticated online banking fraud on extremely imbalanced data | NN, decision forest |
| Ade [79] | Logistic Regression Learning Model for Handling Concept Drift with Unbalanced Data in Credit Card Fraud Detection System | MLPNN, back propagation algorithm, LR |
| Gómez et al., [80] | End-to-end neural network architecture for fraud scoring in card payments | ANN |
| Carneiro et al., [81] | A data mining-based system for credit-card fraud detection in e-tail. | RF, SVM, LR |
| Zareapoor et al., [82] | Application of credit card fraud detection: Based on Bagging ensemble classifier | DT algorithms |
| Ramaki et al., [83] | A systematic review on intrusion detection based on the Hidden Markov Model | HMM |
| Wang et al., [84] | Imbalanced SVM-Based Anomaly Detection Algorithm for Imbalanced Training Datasets | SVM |

## III. Research Findings of Machine Learning Applications for Credit Card Fraud Detection

In this study, Chen and Lai [85] examine the use of deep learning models, such as CNNs and RNNs, to address the issue of identifying fraudulent transactions in credit card systems. It is said that deep learning models provide significant use in combating fraudulent activities due to their capacity to detect intricate and constantly evolving fraudulent schemes. Afriyie et al. [86] conducted a study to explore the applicability of DT and RF in the detection of fraudulent credit card payments. The findings demonstrate that these algorithms have exceptional performance in detecting intricate fraud patterns and effectively handle non-linearities present in the data. Overfitting is a phenomenon that is cautioned against in academic literature. Consequently, the emphasis is placed on the use of regularisation procedures. The viability of SVM in identifying credit card fraud is investigated by Xia [87]. The author highlight the capability of SVM to effectively identify optimal decision boundaries in feature spaces with a large number of dimensions. This ability facilitates accurate identification of fraudulent activities.

Furthermore, Aftab et al. [88] conducted a study to examine an enhanced random forest approach for detecting instances of credit card fraud. The need of continuous algorithmic innovation is emphasised in order to effectively combat the evolving tactics used by fraudsters.

Hybrid methodologies, which integrate traditional methodologies with ML algorithms, have been suggested as an intermediate solution by many researchers [89], [90].

Sohony et al. [91] offer an ensemble methodology that combines rule-based systems with ML methods. It is said that the integration of both approaches has the potential to provide a fraud detection system that is both adaptable and robust. Other research works on CCFD are included in table 2.0 above.

## IV. Challenges and Future Directions

ML algorithms have considerable promise, while more efforts are required to fully exploit their capabilities [92]. The authors highlight the issue of class bias in the detection of credit card fraud. The use of resampling and cost-sensitive learning algorithms is emphasised by researchers as a necessary measure to address this issue [93], [94].

Mishra et al. [95] engage in a discussion on the challenges associated with elucidating the outcomes of deep learning models. These models have the potential for achieving high levels of accuracy; yet, they are often characterised by their opaque nature, which poses challenges in effectively communicating their decision-making processes and outcomes to relevant stakeholders.

## V. Conclusion

The use of both traditional and modern ML techniques in CCFD has a diverse array of advantages and disadvantages. When confronted with intricate and constantly evolving fraud patterns, conventional methods, although providing clarity and ease of use, might sometimes be inadequate. However, ML approaches like DT provide interpretability but are prone to overfitting. LR exhibits efficiency but is sensitive to outliers. NN excel in pattern detection but need significant computational resources. NB, despite its seeming speed, is based on assumptions that undermine its accuracy. GA, HMM, SVM, FLBS, and Hybrid Approaches demonstrate dynamic choices in terms of accuracy, interpretability, scalability, and privacy. In order to navigate through this labyrinth, it is necessary to carefully choose algorithms, taking into consideration the advantages and disadvantages of accuracy, interpretability, and computational requirements. The objectives of future research should focus on enhancing models to resist emerging kinds of fraud and narrowing socioeconomic disparities. The development of more precise fraud detection technology, along with more feasible and reliable means of avoiding fraud, will continue to rely on the collaborative endeavours of academic institutions, data scientists, and financial businesses.

## Acknowledgment

## References

[1] L. Einav, P. Klenow, J. Levin, and R. Murciano-Goroff, “Customers and Retail Growth.,” *J. Monet. Econ.*, 2021, [Online]. Available: https://doi.org/10.1016/j.jmoneco.2021.09.004.

[2] M. Dhone and G. Regulwar, “Learning For Anomaly Detection.,” *J. Emerg. Technol. Innov. Res.*, 2020, [Online]. Available: https://doi.org/10.1201/b10867-5.

[3] S. Daliri, “Using Harmony Search Algorithm in Neural Networks to Improve Fraud Detection in Banking System.,” *Comput. Intell. Neurosci.*, 2020, [Online]. Available: https://doi.org/10.1155/2020/6503459.

[4] A. Al-Faqeh, A. Zerguine, M. Al-Bulayhi, A. Al-Sleem, and A. Al-Rabiah, “Credit Card Fraud Detection via Integrated Account and Transaction Submodules,” *Arab. J. Sci. Eng.*, vol. 46, pp. 10023–10031, 2021, [Online]. Available: https://doi.org/10.1007/s13369-021-05856-5.

[5] L. Moumeni, M. Saber, I. Slimani, I. Elfarissi, and Z. Bougroun, “Machine Learning for Credit Card Fraud Detection.,” *Lect. Notes Electr. Eng.*, 2021, [Online]. Available: https://doi.org/10.1007/978-981-33-6893-4_20.

[6] I. Sadgali, N. Sael, and F. Benabbou, “Human behavior scoring in credit card fraud detection.,” *IAES Int. J. Artif. Intell.*, vol. 10, pp. 698–706, 2021, [Online]. Available: https://doi.org/10.11591/IJAI.V10.I3.PP698-706.

[7] C. M. R. da Silva, E. L. Feitosa, and V. C. Garcia, “Heuristic-based strategy for Phishing prediction: A survey of URL-based approach,” *Comput. Secur.*, vol. 88, p. 101613, 2020, [Online]. Available: https://doi.org/10.1016/j.cose.2019.101613.

[8] O. Adebayo, T. Favour-Bethy, O. Owolafe, and O. Adebola, “Comparative Review of Credit Card Fraud Detection using Machine Learning and Concept Drift Techniques.,” *Int. J. Comput. Sci. Mob. Comput.*, vol. 12, pp. 24–48, 2023, doi: 10.47760/ijcsmc.2023.v12i07.004.

[9] L. Delamaire, A. Hussein, and P. John, “Credit card fraud and detection techniques: A review.,” *Banks Bank Syst.*, vol. 4, pp. 57–68, 2009.

[10] E. Marazqah Btoush, X. Zhou, R. Gururajan, K. Chan, R. Genrich, and P. Sankaran, “A systematic review of literature on credit card cyber fraud detection using machine and deep learning.,” *Peer J Comput Sci.*, vol. 9, p. e1278, 2023, doi: 10.7717/peerj-cs.1278.

[11] W. Hilal, S. A. Gadsden, and J. Yawney, “Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances,” *Expert Syst. Appl.*, vol. 193, p. 116429, 2022, [Online]. Available: https://doi.org/10.1016/j.eswa.2021.116429.

[12] B. Al Smadi and M. Min, “A Critical review of Credit Card Fraud Detection Techniques.,” pp. 0732–0736, 2020, doi: 10.1109/UEMCON51285.2020.9298075.

[13] E. Balagolla, W. Fernando, R. Rathnayake, M. Wijesekera, A. Senarathne, and K. Abeywardhana, “Credit Card Fraud Prevention Using Blockchain.,” in *2021 6th International Conference for Convergence in Technology (I2CT)*, 2021, pp. 1–8, [Online]. Available: https://doi.org/10.1109/I2CT51068.2021.9418192.

[14] J. Pan, “Deep Set Classifier for Financial Forensics: An application to detect money laundering,” 2022, [Online]. Available: https://doi.org/10.48550/arXiv.2207.07863.

[15] T. Vairam, S. Sarathambekai, S. Bhavadharani, A. Dharshini, N. Sri, and T. Sen, “Evaluation of Naïve Bayes and Voting Classifier Algorithm for Credit Card Fraud Detection.,” in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2022, vol. 1, pp. 602–608, [Online]. Available: https://doi.org/10.1109/ICACCS54159.2022.9784968.

[16] V. N. Dornadula and S. Geetha, “Credit Card Fraud Detection using Machine Learning Algorithms,” *Procedia Comput. Sci.*, vol. 165, pp. 631–641, 2019, [Online]. Available: https://doi.org/10.1016/j.procs.2020.01.057.

[17] E. Ileberi, Y. Sun, and Z. Wang, “A machine learning based credit card fraud detection using the GA algorithm for feature selection.,” *J Big Data*, vol. 9, p. 24, 2022, [Online]. Available: https://doi.org/10.1186/s40537-022-00573-8.

[18] N. H. Mohammed and S. C. R. Maram, *Fraud Detection of Credit Card Using Logistic Regression*. 2022.

[19] 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 Int. Conf. Intell. Comput. Control Syst. (ICICCS), Madurai, India*, pp. 967–972, 2021, doi: 10.1109/ICICCS51141.2021.9432308.

[20] A. S. Rathore, A. Kumar, D. Tomar, V. Goyal, K. Sarda, and D. Vij, “Credit Card Fraud Detection using Machine Learning,” in *2021 10th International*

*Conference on System Modeling & Advancement in Research Trends (SMART), MORADABAD, India*, 2021, pp. 167–171, [Online]. Available: 10.1109/SMART52563.2021.9676262.

[21] O. Vynokurova, D. Peleshko, O. Bondarenko, V. Ilyasov, V. Serzhantov, and M. Peleshko, "Hybrid machine learning system for solving fraud detection tasks.," in *2020 IEEE third international conference on data stream mining & processing (DSMP), IEEE*, 2020, pp. 1–5.

[22] Q. Zhang, "Financial Data Anomaly Detection Method Based on Decision Tree and Random Forest Algorithm," *J. Math.*, vol. 2022, no. 9135117, p. 10, 2022, [Online]. Available: https://doi.org/10.1155/2022/9135117.

[23] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study.," *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011.

[24] A. D. Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection and concept-drift adaptation with delayed supervised information.," in *In Neural Networks (IJCNN), 2015 International Joint Conference on IEEE*, 2015, pp. 1–8.

[25] A. D. Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective.," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4915–4928, 2014.

[26] W. Christopher, D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams, "Transaction aggregation as a strategy for credit card fraud detection.," *Data Min. Knowl. Discov.*, vol. 18, no. 1, pp. 30–55, 2009.

[27] V. V. V´eronique *et al.*, "Apate: A novel approach for automated credit card transaction fraud detection using network-based extensions.," *Decis. Support Syst.*, vol. 75, pp. 38–48, 2015.

[28] O. Ghorbanzadeh *et al.*, "Gully erosion susceptibility mapping (GESM) using machine learning methods optimized by the multi-collinearity analysis and K-fold cross- validation.," *Geomatics, Nat. Hazards Risk.*, vol. 11, pp. 1653-1678., 2020, doi: 10.1080/19475705.2020.1810138.

[29] Ganesh et al., "Credit Card Fraud Detection with Unbalanced Real and Synthetic dataset using Machine Learning models," in *2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC), Chennai, India.*, 2022, pp. 73-78., doi: 10.1109/ICESIC53714.2022.9783529.

[30] 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," in *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India*, 2022, pp. 265–272, doi: 10.1109/ICAIS53314.2022.9742830.

[31] S. Dubey, K. Mundhe, and A. Kadam, "Credit card fraud detection using artificial neural network and back propagation.," in *2020 4th international conference on intelligent computing and control systems (ICICCS).*, 2020, pp. 268–273.

[32] A. I. Kokkinaki, "On atypical database transactions: Identification of probable frauds using machine learning for user profiling," in *Proceedings of the IEEE Knowledge & Data Engineering Exchange Workshop, KDEX*, 1997, pp. 107–113.

[33] A. Guezzaz, S. Benkirane, M. Azrour, and S. Khurram, "A Reliable Network Intrusion Detection Approach Using Decision Tree with Enhanced Data Quality.," *Commun. Networks Secur.*, pp. 1–1230593, 2021, [Online]. Available: https://doi.org/10.1155/2021/1230593.

[34] 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," *IEEE Access*, vol. 10, pp. 39700–39715, 2022, doi: 10.1109/ACCESS.2022.3166891.

[35] S. Sasank, G. Sahith, K. Abhinav, and M. Belwal, "Credit card fraud detection using various classification and sampling techniques: a comparative study," pp. 1713–1718, 2019.

[36] N. Rtayli and N. Enneya, "selection features and support vector machine for credit card risk identification.," *Procedia Manuf.*, vol. 46, pp. 941–8, 2020.

[37] D. Prajapati, A. Tripathi, J. Mehta, K. Jhaveri, and V. Kelkar, "Credit Card Fraud Detection Using Machine Learning," *2021 7th IEEE Int. Conf. Adv. Comput. Commun. Control. ICAC3 2021*, 2021, doi: 10.1109/ICAC353642.2021.9697227.

[38] R. A. Mohammed, K.-W. Wong, M. F. Shiratuddin, and X. Wang, "Scalable machine learning techniques for highly imbalanced credit card fraud detection: A comparative study.," *Pacific Rim Int. Conf. Artif. Intell.*, pp. 237–246, 2018.

[39] M. S. Mahmud, P. Meesad, and S. Sodsee, "An evaluation of computational intelligence in credit card fraud detection.," in *In Computer Science and Engineering Conference (ICSEC), 2016 International*, 2016, pp. 1–6.

[40] A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. ¨orn Ottersten, "Cost sensitive credit card fraud detection using bayes minimum risk.," in *In Proceedings-2013 12th International Conference on Machine Learning and Applications*, 2013, pp. 333–338.

[41] N. Mahmoudi and E. Duman, "Detecting credit card fraud by modified fisher discriminant analysis.," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2510–2516, 2015.

[42] A. Dastanpour, S. Ibrahim, and M. Mashinchi, "Used Genetic Algorithm for Support Artificial Neural Network in Intrusion Detection System.," 2014, [Online]. Available: https://www.researchgate.net/publication/261028418_Used_Genetic_Algorithm_for_Support_Artificial_Neural_Network_in_Intrusion_Detection_System.

[43] J. H. Holland, *Adaptation in natural and artificial systems*. university of michigan press, 1975.

[44] et al. Bentley, "Fuzzy darwinian detection of credit card fraud."," in *The 14th Annual Fall Symposium of the Korean Information Processing Society.*, 2000,

vol. 14.

[45] et al. Chan, Philip K., “Distributed data mining in credit card fraud detection.,” in *IEEE Intelligent Systems and Their Applications*, 1999, pp. 67–74.

[46] Y. Bei, “Detection and Resolution of Data Confliction in the Integration of Heterogeneous Information Sources.,” *J. Beijing Univ. Technol.*, 2008.

[47] M. Syeda, Y.-Q. Zhang, and Y. Pan, “Parallel granular neural networks for fast credit card fraud detection,” in *2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE’02. Proceedings*, 2002, vol. 1, p. 02CH37291.

[48] B.-J. Yoon, “Hidden Markov Models and their Applications in Biological Sequence Analysis,” *Curr. Genomics*, vol. 10, pp. 402-415., 2009, [Online]. Available: https://www.semanticscholar.org/reader/f7181dd96101b60fd397ef5d53c152b56fbff056.

[49] D. L. Talekar and K. P. Adhiya, “Credit Card Fraud Detection System: A Survey.,” *Int. J. Mod. Eng. Res.*, vol. 4, p. 9, 2014.

[50] L. A. Zadeh, “Fuzzy logic,” *Computer (Long. Beach. Calif).*, pp. 83–93, 1988.

[51] M. Zareapoor, K. R. Seeja, and M. A. Alam, “Analysis on credit card fraud detection techniques: based on certain design criteria,” *Int. J. Comput. Appl.*, vol. 52, p. 3, 2012.

[52] Z. Zojaji, R. E. Atani, and A. H. Monadjemi, “A survey of credit card fraud detection techniques: data and technique oriented perspective.,” 2016.

[53] C. Herbert, J. Munoz-Martin, D. Llaveria, M. Pablos, and A. Camps, “Sea Ice Thickness Estimation Based on Regression Neural Networks Using L-Band Microwave Radiometry Data from the FSSCat Mission.,” *Remote Sensing.*, vol. 13, p. 1366, 2021, doi: 10.3390/rs13071366.

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

[55] 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,” vol. 10, pp. 39700–39715, 2022, doi: 10.1109/ACCESS.2022.3166891.

[56] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, “A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection.,” *IEEE Access*, vol. 10, pp. 16400–16407, 2022.

[57] R. Patidar and L. Sharma, “Credit card fraud detection using neural network.,” *IJSCE ISSN 2231-2307*, vol. 1, no. NCAI2011, 2011.

[58] M.-J. Kim and T.-S. Kim, “A neural classifier with fraud density map for effective credit card fraud detection.,” in *In International Conference on Intelligent Data Engineering and Automated Learning*, 2002, pp. 378–383.

[59] Y. Wang *et al.*, “Privacy preserving distributed deep learning and its application in credit card fraud detection.,” in *In 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, 2018, pp. 1070–1078.

[60] R. Bin Sulaiman, V. Schetinin, and P. Sant, “Review of Machine Learning Approach on Credit Card Fraud Detection.,” *Hum-Cent Intell Syst*, vol. 2, pp. 55–68, 2022, [Online]. Available: https://doi.org/10.1007/s44230-022-00004-0.

[61] J. Pombal, A. F. Cruz, J. Bravo, P. Saleiro, M. A. T. Figueiredo, and P. Bizarro, “Understanding Unfairness in Fraud Detection through Model and Data Bias Interactions,” 2022, [Online]. Available: http://arxiv.org/abs/2207.06273.

[62] P. Mehta *et al.*, “A high-bias, low-variance introduction to Machine Learning for physicists,” *Phys. Rep.*, vol. 810, pp. 1–124, 2019, [Online]. Available: https://doi.org/10.1016/j.physrep.2019.03.001.

[63] V. Hassija, V. Chamola, A. Mahapatra, and E. Al., “Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence.,” *Cogn Comput*, 2023, [Online]. Available: https://doi.org/10.1007/s12559-023-10179-8.

[64] L. Jakaite, V. Schetinin, J. Hladuvka, S. Minaev, A. Ambia, and W. Krzanowski, “Deep learning for early detection of pathological changes in X-ray bone microstructures: case of osteoarthritis.,” *Sci Rep.*, 2021, [Online]. Available: https://doi.org/10.1038/s41598-021-81786-4.

[65] H. Wen and F. Huang, “Personal loan fraud detection based on hybrid supervised and unsupervised learning.,” in *2020 5th IEEE international conference on big data analytics (ICBDA)*, 2020, pp. 339–343.

[66] W. Li, S. Lin, and X. et al. Qian, “An evidence theory-based validation method for models with multivariate outputs and uncertainty.,” *Simulation*, vol. 97, pp. 821–34, 2021, [Online]. Available: https://doi.org/10.1177/00375497211022814.

[67] M. Zięba, S. Tomczak, and J. Tomczak, “Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction.,” *Expert Syst Appl.*, vol. 58, pp. 93–101, 2016, [Online]. Available: https://doi.org/10.1016/j.eswa.2016.04.001.

[68] A. Rai and R. Dwivedi, “Fraud detection in credit card data using unsupervised machine learning based scheme.,” pp. 421–426., 2020.

[69] M. Ostapowicz and K. Żbikowski, “Detecting fraudulent accounts on blockchain: a supervised approach.,” *Cham Springer*, 2019, [Online]. Available: http://scholar.google.com/scholar_lookup?&title=Detecting fraudulent accounts on blockchain%3A a supervised approach&publication_year=2019&author=Ostapowicz%2CM&author=Żbikowski%2CK.

[70] G. Danner, Á. Berta, I. Hegedűs, and M. Jelasity,

"Robust fully distributed minibatch gradient descent with privacy preservation.," *Secur Commun Netw.*, pp. 1–15, 2018, [Online]. Available: https://doi.org/10.1155/2018/6728020.

[71] W. Yang, Y. Zhang, and et al. Ye K, "FFD: a federated learning based method for credit card fraud detection.," *Cham Springer*, 2019, [Online]. Available: http://scholar.google.com/scholar_lookup?&title=FFD%3A a federated learning based method for credit card fraud detection&publication_year=2019&author=Yang%2CW&author=Zhang%2CY&author=Ye%2CK.

[72] N. Omar, Z. A. Johari, and M. Smith, "Predicting fraudulent financial reporting using artificial neural network," *J. Financ. Crime*, vol. 24, pp. 362–387, 2017, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10.1108/JFC-11-2015-0061.

[73] O. A. Adewumi and A. A. Akinyelu, "A hybrid firefly and support vector machine classifier for phishing email detection," *Kybernetes.*, vol. 45, pp. 977–994, 2016, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10.1108/K-07-2014-0129.

[74] A. D. Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy IEEE Trans.," *Neural Networks Learn. Syst.*, vol. 29, p. 8, 2018, [Online]. Available: https://scholar.google.com/scholar_lookup?title=Credit Card Fraud Detection%3A A Realistic Modeling and a Novel Learning Strategy&publication_year=2018&author=A. Dal Pozzolo&author=G. Boracchi&author=O. Caelen&author=C. Alippi&author=G. Bontempi.

[75] M. F. A. Gadi, X. Wang, and A. P. do Lago, "Comparison with Parametric Optimizationin Credit Card Fraud Detection.," in *Proceedings of the Seventh International Conference on Machine Learning and Applications San Diego, CA, USA*, 2008, pp. 11–13.

[76] S. Qiu and H.-Q. He, "Machine Learning- and Evidence Theory-Based Fraud Risk Assessment of China's Box Office," *IEEE Access*, vol. 6, pp. 75619–75628, 2018, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10.1109/ACCESS.2018.2883487.

[77] G. E. Melo-Acosta, F. Duitama-Muñoz, and J. D. Arias-Londoño, "Fraud Detection in Big Data using Supervised and Semi-supervised Learning Techniques," in *Proceedings of the IEEE Colombian Conference on Communications and Computing (COLCOM) Cartagena, Colombia*, 2017, pp. 16–18, [Online]. Available: https://scholar.google.com/scholar?q=G.E. Melo-Acosta F. Duitama-Muñoz J.D. Arias-Londoño Fraud Detection in Big Data using Supervised and Semi-supervised Learning Techniques Proceedings of the IEEE Colombian Conference on Communications and Computing (COLCOM) Cartagena%2C Colombia 16–18 August 2017.

[78] W. Wei, J. Li, L. Cao, Y. Ou, and J. Chen, "Effective detection of sophisticated online banking fraud on extremely imbalanced data World Wide Web," vol. 16, pp. 449–475, 2013, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10.1007/s11280-012-0178-0.

[79] P. K. Ade, "Logistic Regression Learning Model for Handling Concept Drift with Unbalanced Data in Credit Card Fraud Detection System," *Proc. Second Int. Conf. Comput. Commun. Technol. Hyderabad*, pp. 24–26, 2015, [Online]. Available: https://scholar.google.com/scholar?q=P.K. Ade Logistic Regression Learning Model for Handling Concept Drift with Unbalanced Data in Credit Card Fraud Detection System Proceedings of the Second International Conference on Computer and Communication Technologies Hyderabad%2C India 24–26 July 2015.

[80] J. A. Gómez, J. Arévalo, R. Paredes, and J. Nin, "End-to-end neural network architecture for fraud scoring in card payments," *Pattern Recognit. Lett.*, vol. 105, pp. 175–181, 2018, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10.1016/j.patrec.2017.08.024.

[81] N. Carneiro, G. Figueira, and M. Costa, "A data mining-based system for credit-card fraud detection in e-tail Decis.," *Support Syst.*, vol. 95, pp. 91–101, 2017, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10.1016/j.dss.2017.01.002.

[82] M. Zareapoor and P. Shamsolmoali, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier," *Procedia Comput. Sci.*, vol. 48, pp. 679–685, 2015, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10.1016/j.procs.2015.04.201.

[83] A. A. Ramaki, A. Rasoolzadegan, and A. J. Jafari, "A systematic review on intrusion detection based on the Hidden Markov Model," *Stat. Anal. Data Min. ASA Data Sci. J.*, vol. 11, pp. 111–134, 2018, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10.1002/sam.11377.

[84] G. P. Wang, J. X. Yang, and R. Li, "Imbalanced SVM-Based Anomaly Detection Algorithm for Imbalanced Training Datasets," *ETRI J.*, vol. 39, pp. 621–631, 2017, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10.4218/etrij.17.0116.0879.

[85] J. Chen and K.-L. Lai, "Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert.," *J. Artif. Intell. Capsul. Networks*, vol. 3, pp. 101–112, 2021, doi: 10.36548/jaicn.2021.2.003.

[86] J. K. Afriyie *et al.*, "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," *Decis. Anal. J.*, vol. 6, p. 100163, 2023, [Online]. Available: https://doi.org/10.1016/j.dajour.2023.100163.

[87] J. Xia, "Credit Card Fraud Detection Based on Support Vector Machine.," *Highlights Sci. Eng. Technol.*, vol. 23, pp. 93–97, 2022, doi: 10.54097/hset.v23i.3202.

[88] A. Aftab, I. Shahzad, A. Sajid, M. Anwar, and N. Anwar, “Fraud Detection of Credit Cards Using Supervised Machine Learning Techniques.,” *Pakistan J. Emerg. Sci. Technol.*, p. 4, 2023, doi: 10.58619/pjest.v4i3.114.

[89] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions.,” *SN Comput.*, vol. SCI. 2, p. 160, 2021.

[90] J. Huang, Q. Chang, and J. Arinez, “Product Completion Time Prediction Using A Hybrid Approach Combining Deep Learning and System Model,” *J. Manuf. Syst.*, vol. 57, pp. 311–322, 2020, [Online]. Available: https://doi.org/10.1016/j.jmsy.2020.10.006.

[91] I. Sohony, R. Pratap, and U. Nambiar, “Ensemble learning for credit card fraud detection.,” pp. 289–294, 2018, doi: 10.1145/3152494.3156815.

[92] G. Dheepak and D. Vaishali, “A Comprehensive Overview of Machine Learning Algorithms and their Applications.,” *Int. J. Adv. Res. Sci. Commun. Technol.*, 2021, [Online]. Available: https://doi.org/10.48175/ijarsct-2301.

[93] M. Alamri and M. Ykhlef, “Survey of Credit Card Anomaly and Fraud Detection Using Sampling Techniques.,” *Electronics*, vol. 11, no. 23, p. 4003, 2022, [Online]. Available: https://doi.org/10.3390/electronics11234003.

[94] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, “Credit card fraud detection in the era of disruptive technologies: A systematic review,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 1, pp. 145–174, 2023, [Online]. Available: https://doi.org/10.1016/j.jksuci.2022.11.008.

[95] R. Mishra, G. Reddy, and H. Pathak, “The Understanding of Deep Learning: A Comprehensive Review.,” *Math. Probl. Eng.*, pp. 1–15, 2021.